

Master thesis submitted in order to be awarded the Master Degree of Master of
Science in Photonics Engineering

OPTICAL SPECTROSCOPY AND MACHINE LEARNING EMPOWERING MULTI-MYCOTOXIN DETECTION IN THE AGRIFOOD INDUSTRY

Wannes Luts de Martelaere

June 2025

Promotor: Prof. Dr. Lien Smeesters, Supervisor: Ir. Indy Magnus

Faculty of Engineering

Abstract

Mycotoxins continue to pose a serious threat to the global agrifood supply chain, yet conventional detection methods—chemical techniques, like liquid chromatography tandem mass spectroscopy—are inherently destructive and poorly suited for real-time, batch-wide, or in-field monitoring. In response to the widespread challenges of mycotoxin contamination and the limitations of existing analytical tools, this thesis proposes the development and rigorous validation of an optical spectroscopic detection platform, integrated with machine learning models, for the reliable identification of the most common mycotoxins on staple crops such as maize and wheat.

Hitherto, existing approaches have largely been limited to single-mycotoxin, single-matrix detection, relying on full-spectral data to achieve model performance. Advancing beyond the current state-of-the-art, this work aims to enable the simultaneous, multiplexed detection of mycotoxins across diverse food matrices by integrating complementary spectroscopic modalities—including fluorescence spectroscopy at multiple excitation wavelengths and ultraviolet, visible, and near-infrared (UV-VIS-NIR) reflectance spectroscopy.

To reduce system complexity and data dimensionality without sacrificing performance, strategic wavelength selection is applied to transition from full-spectrum models to models operating on only a discrete subset of the most informative features, thus paving the way for integration into real-time, in-line, and in-field monitoring systems, allowing for batch-wide contamination assessment at scale. Both classification models (for qualitative and semi-quantitative analysis) and regression models (for quantitative estimation) are developed within this framework.

In a first step, multi-mycotoxin-contaminated samples were prepared using ground wheat and corn standards with certified toxin concentrations. The targeted mycotoxins include aflatoxins (AFLA), deoxynivalenol (DON), zearalenone (ZEA), fumonisins (FUM), ochratoxin A (OTA), and T-2/HT-2 toxins. Spectral data was acquired via fluorescence spectroscopy at excitation wavelengths of 395, 385, 375, 365, and 256nm, along with upconversion fluorescence measurements at 750nm, and complementary UV-VIS-NIR reflectance spectra.

Next, using this spectral data, classification models were trained to categorize samples as safe, risky, or unsafe with respect to various mycotoxins. Models were developed for individual matrices (corn and wheat) and for joint classification across both. An extensive model architecture search revealed that Linear Discriminant Analysis (LDA), Linear Support Vector Machines (LSVM), and Random Forests (RF) were consistently the best-performing classifiers. A Leave-One-Sample-Out Cross-Validation (LOSO-CV) scheme was employed to ensure models were trained on all available data while being validated on unseen samples.

Further hyperparameter tuning showed that LDA, paired with area-normalized spectral data at 8nm resolution, delivered the most robust and generalizable performance for both wheat and corn. For corn, the highest classification accuracy for AFLA, ZEA, DON, and FUM was achieved using fluorescence data at 395nm. For wheat, NIR reflectance measurements yielded

the best performance for detecting DON, HT-2/T-2, and OTA simultaneously. When detecting DON across both matrices, K-Nearest Neighbor (KNN) classifiers using Euclidean distance with 3–10 neighbors outperformed other methods, achieving 82% accuracy in binary classification at a 500 $\mu\text{g}/\text{kg}$ threshold using the fluorescence data.

To reduce model complexity and enhance interpretability, a subsequent phase focused on spectral feature selection, for which Minimum Redundancy Maximum Relevance (MRMR) and Sequential Feature Selection (SFS) were applied and compared. While MRMR provided rapid dimensionality reduction, SFS consistently achieved superior model performance, albeit at increased computational cost. Through SFS, the spectral inputs could be reduced to just the 8–10 most informative wavelengths, while retaining 95% of the performance achieved using the full spectral dataset.

Further performance improvements were realized by aggregating features across all spectroscopic modalities and selecting the top 10 most informative features overall. For corn, this approach led to final binary classification accuracies of 88.8% for AFLA (using a threshold of 2.88 $\mu\text{g}/\text{kg}$), 94.7% for DON (1250 $\mu\text{g}/\text{kg}$), 88.7% for ZEA (50 $\mu\text{g}/\text{kg}$), and 91.3% for FUM (900 $\mu\text{g}/\text{kg}$)—results comparable to existing literature benchmarks, but now in a multiplexed setting—for the first time to our knowledge.

In addition to classification, regression models were developed to estimate toxin concentrations directly, offering quantitative insight beyond class-based decisions. A comprehensive model search and preprocessing optimization identified Partial Least Squares Regression (PLSR) with 9 PLS-components, applied to area-normalized data at 10nm resolution, as the most effective regression approach.

The best performance was achieved using NIR-reflectance data for quantifying the combined presence of DON, HT-2/T-2, and OTA in wheat samples. This model reached an R^2 -value of 0.94 with a Root Mean Squared Error (RMSE) of 72 $\mu\text{g}/\text{kg}$. For DON quantification on corn, the model achieved an R^2 -value of 0.83 and RMSE of 296 $\mu\text{g}/\text{kg}$. Joint regression across both corn and wheat matrices for DON yielded an R^2 -value of 0.82 with an RMSE of 270.5 $\mu\text{g}/\text{kg}$, indicating accurate detection, even below legal limits.

In conclusion, extending beyond the state-of-the-art, this thesis has demonstrated that a multimodal spectroscopic platform, augmented by machine learning, can effectively enable both qualitative and quantitative detection of multiplexed mycotoxins on staple grains. Through tailored feature selection and model optimization, the proposed approach reduces system complexity while maintaining high performance—presenting a scalable solution for real-time, in-line, and in-field mycotoxin monitoring. These findings position the proposed method as a competitive and industrially viable alternative to conventional chemical analyses—capable of meeting stringent regulatory requirements while significantly reducing time, cost, and operational complexity. Moreover, by enabling individualized sample-based screening, the system has the potential to improve food safety and reduce waste across the supply chain.

Keywords: *optical spectroscopy, machine learning, multi-mycotoxin detection, aflatoxin, deoxynivalenol, zearalenone, fumonisins, ochratoxin, T-2/HT-2, in-line and in-field monitoring, agrifood, corn and wheat, non-invasive, Linear Discriminant Analyses (LDA), Partial Least Squares Regression (PLSR), Sequential Feature Selection (SFS).*

Contents

1	Introducing Mycotoxins In Agrifood	1
1.1	A Brief Background	1
1.1.1	Some History On Mycotoxins	2
1.1.2	Risks Posed By Mycotoxins	2
1.1.3	Mitigation Of Mycotoxins	3
1.1.3.1	Pre-harvest	3
1.1.3.2	Post-harvest	4
1.1.3.3	Mycotoxin Monitoring	4
1.2	Optical Spectroscopy For Agrifood	5
1.2.1	Challenges In Mycotoxin Detection	6
1.2.1.1	Sample Complexities	6
1.2.1.2	Model Development Difficulties	7
1.2.1.3	Industrial Viability	8
1.3	Objectives	8
1.4	Overview Of The Thesis	10
2	Mycotoxin Monitoring: A Literature Review	13
2.1	Mycotoxins On Maize & Wheat	13
2.1.1	Aflatoxin (AFLA)	14
2.1.2	Ochratoxin A (OTA)	15
2.1.3	Fusarium Toxins	15
2.1.3.1	Deoxynivalenol (DON)	16
2.1.3.2	T-2 And HT-2 Toxins	16
2.1.3.3	Zearalenone (ZEA)	17
2.1.3.4	Fumonisin (FUM)	17
2.1.4	Multi-mycotoxin: Co-contamination	18
2.2	Regulation Of Mycotoxins On Maize And Wheat	19
2.3	Chemical Platforms For Mycotoxin Detection	20
2.3.1	Chromatographic Techniques	21
2.3.1.1	State-of-the-art	22
2.3.2	Immunological Techniques	25
2.3.2.1	State-of-the-art	25
2.3.3	Biosensors	29
2.3.3.1	State-of-the-art	29
2.4	Need For Non-Invasive Alternatives	31
2.5	Optical Spectroscopy For Mycotoxin Detection; State-of-the-art	32
2.5.1	Comparing Spectroscopic Platforms	38

2.5.2	Fundamentals And State-of-the-art Of Fluorescence And Reflectance Spectroscopy	39
2.5.2.1	Reflectance Spectroscopy	40
2.5.2.1.1	State-of-the-art	40
2.5.2.2	Fluorescence Spectroscopy	41
2.5.2.2.1	Notes On The Use Of Fluorescence	41
2.5.2.2.2	State-of-the-art	42
2.5.2.3	Upconversion Fluorescence For Mycotoxin Detection	43
2.5.3	Chemometrics And ML For Mycotoxin Detection	43
2.5.3.1	Model Development Process	44
2.5.3.2	Review Of ML-Models For Spectroscopic Mycotoxin Detection	44
2.5.3.2.1	State-of-the-art; Classification Models	45
2.5.3.2.2	State-of-the-art; Regression Models	46
2.5.3.3	A Review Of Feature Selection In Literature	46
2.6	Summary And Research Gaps	47
3	Materials And Methods	49
3.1	Multi-mycotoxin Samples	49
3.1.1	The Standards	50
3.1.2	Sample Synthesis	51
3.1.2.1	The Corn Samples	52
3.1.2.2	The Wheat Samples	54
3.2	Measurement Setup	55
3.2.1	Reflection Spectroscopy Setup	55
3.2.1.1	Small Reflection Sphere; Setup	55
3.2.1.2	Small Reflection Sphere; Measurement Protocol	56
3.2.1.3	Large Reflection Sphere; Setup	58
3.2.1.4	Large Reflection Sphere; Measurement Protocol	58
3.2.2	Fluorescence; Setup	59
3.2.3	Fluorescence; Measurement Protocol	60
3.3	Spectroscopic Data Acquisition And Results	61
3.3.1	The Small Sphere Reflectance Measurements	61
3.3.1.1	Comparing The Bottom And Side Measurements	62
3.3.1.2	Small Sphere Measurements For Corn	63
3.3.1.3	Small Sphere Measurements For Wheat	64
3.3.2	Large Sphere Reflectance Measurements	66
3.3.3	Fluorescence Measurements	67
3.3.3.1	One-Photon Fluorescence	67
3.3.3.2	Two Photon Fluorescence	73
3.4	Summary Of The Chapter	75
4	Classification Model Development	77
4.1	From Proof-of-Concept to Robust Validation	77
4.1.1	Reproducing The Initial Results	78
4.1.2	Extending To Reflectance Data	79
4.1.3	Reproducibility	80
4.1.4	Generalization Challenges And Need For LOSO-CV	81
4.2	Classifier Development	83
4.2.1	Classification Classes	84

4.2.2	The Model Development Strategy	86
4.2.3	Notes On The Code Used For Model Development	87
4.3	Classification Of Mycotoxins On Corn	88
4.3.1	Model-Search: Fluorescence	89
4.3.1.1	Hyperparameter Sensitive Models	96
4.3.1.2	Evaluation Of PLS-DA For DON Classification	96
4.3.2	Model-Search: Reflectance	98
4.3.3	Computational Cost Comparison	103
4.3.4	Model Optimization	104
4.3.5	Feature Selection	111
4.3.5.1	10 Feature Models	113
4.3.5.2	Complexity For AFLA Detection	113
4.3.6	Combining Complementary Data	115
4.3.6.1	The Best 3-Class Classifiers	117
4.3.6.1.1	3-Class DON-Classifiers	117
4.3.6.1.2	3-Class ZEA & FUM-Classifiers	120
4.3.6.2	The Best 2-Class Classifiers	121
4.3.7	Conclusion	124
4.4	Classification Of Mycotoxins On Wheat	124
4.4.1	Model-Search	125
4.4.2	Model Optimization	129
4.4.3	Feature Selection	131
4.4.4	Combining Complementary Data	133
4.4.5	Experiment: Restricting The Spectral Range	134
4.4.6	Conclusion	136
4.5	Classification Of Mycotoxins On Corn And Wheat	137
4.5.1	Model-Search	137
4.5.2	Model Optimization	141
4.5.3	Feature Selection	142
4.5.4	Best Model & Conclusion	144
4.6	Chapter Conclusion	145
5	Regression Model Development	147
5.1	Regression Model For Mycotoxin Detection On Wheat	148
5.1.1	Model-Search	149
5.1.2	Computational Cost Comparison Of Regressors	151
5.1.3	Model Optimization	152
5.1.4	The Best Model: Mycotoxins On Wheat	154
5.2	Regression Model For Mycotoxin Detection On Corn	155
5.2.1	The Best Model: DON On Corn	157
5.3	Regression Model For Mycotoxin Detection On Corn And Wheat	158
5.3.1	The Best Model: DON On Wheat And Corn	160
5.4	Feature Selection	160
5.5	Conclusion: DON Detection Using Regression Models	164
6	Validating Model Performance	167
6.1	Validation Samples	167
6.2	Spectral Data Of The Validation Samples	168
6.3	Validating Full Spectrum Classifiers	170

6.4	Feature Based Classifiers	172
6.5	Conclusion On Model Generalizability	174
7	Concluding Remarks & Future Perspectives	175
7.1	Summary Of Results	175
7.2	Limitations And Future Perspectives	178
A	Theoretical Background On ML For Mycotoxin Detection	181
A.1	Spectral Data And Preprocessing Steps	181
A.1.1	Trimming	182
A.1.2	Smoothing	183
A.1.3	Normalization	183
A.1.4	Ratios	184
A.2	Machine Learning Models	184
A.2.1	Classification Models	185
A.2.1.1	Linear Discriminant Analysis (LDA)	186
A.2.1.2	Quadratic Discriminant Analysis (QDA)	187
A.2.1.3	Support Vector Machines (SVM)	187
A.2.1.4	K-Nearest Neighbors (KNN)	188
A.2.1.5	Decision Trees (DT) And Random Forests (RF)	189
A.2.1.6	Artificial Neural Networks (ANN)	190
A.2.1.7	Partial Least Squared Discriminant Analyses (PLS-DA)	191
A.2.2	Regression Models	191
A.2.2.1	Support Vector Regression (SVR)	192
A.2.2.2	Random Forest Regressors (RFR)	193
A.2.2.3	Artificial Neural Network Regressor (ANN-MLPR)	193
A.2.2.4	Lasso And Elastic Net	194
A.2.2.5	Partial Least Squares Regression (PLSR)	195
A.3	Feature Selection Methods	195
A.3.1	Principal Component Analysis (PCA)	195
A.3.2	Sequential Feature Search (SFS)	196
A.3.3	Minimum Redundancy Maximum Relevance (MRMR)	197
A.4	Model Validation & Performance Metrics	197
A.4.1	Validation Strategies	198
A.4.2	Key Performance Metrics	199
A.4.2.1	Classification Models	199
A.4.2.2	Regression Models	200

Chapter 1

Introducing Mycotoxins In Agrifood

1.1 A Brief Background

Ensuring food safety in the agrifood industry is an enduring, complex and ever evolving challenge, particularly in the context of mycotoxin detection. These toxic secondary metabolites, produced by certain filamentous fungi, pose a significant health risks to both humans and animals even at ultra-trace concentrations [1, 2].

The term "*mycotoxin*" derives from the Greek words *mykes* (fungus) and *toxicum* (poison), signifying its literal meaning: "*fungal poison*" [1]. These compounds, although not essential for fungal growth or primary metabolic function [3, 4], are often synthesized as by-products under specific environmental stressors, such as nutrient depletion or oxidative stress [5]. Their production may confer ecological advantages, including protection against microorganisms, nematodes, insects, animals, and humans, enhancing the fungi's ability to compete in their environment [3, 6, 7].

Functionally, mycotoxins constitute a diverse class of non-enzymatic, abiotic fungal metabolites that present substantial risks to both human and animal health [4, 8, 9]. These toxic compounds can infiltrate the food supply chain at multiple stages, ranging from crop growth to harvesting, transport, storage, and post-processing [9]. Contamination occurs either directly, through fungal proliferation on crops, or indirectly, via the incorporation of contaminated raw materials into processed foods. Additionally, mycotoxins can bio-accumulate in animal products such as milk, meat, and eggs when livestock consume contaminated feed, effectively making animal-products secondary exposure vectors [1, 10, 11].

The primary biological effects of mycotoxins in higher vertebrates arise through ingestion, inhalation, or dermal exposure, eliciting toxicological responses collectively termed mycotoxicoses [4, 12]. These conditions are notably non-transmissible, unresponsive to conventional drug or antibiotic therapies, and often exhibit a seasonal incidence pattern linked to environmental factors [4]. Mycotoxicoses can result in a broad spectrum of acute and chronic health effects, contingent upon the nature and extent of exposure—including carcinogenic, mutagenic, and immunosuppressive outcomes [9]. A more detailed analysis of the specific health threats posed by the most hazardous mycotoxins commonly found in maize and wheat is provided in Section 2.1.

Beyond the severe health risks associated with mycotoxin poisoning, the economic impact of contamination is equally profound. It is estimated that 20–25% of harvested fruits and vegetables are lost annually due to post-harvest fungal contamination, even in highly developed agricultural economies [3]. In the United States alone, mycotoxin-related losses surpassed 1 billion USD per year, as early as the beginning of the 2000s [3]. Staple crops such as maize, wheat, barley, and groundnuts are particularly susceptible, with up to 25% of global cereal production affected by mycotoxin contamination [4]. According to Mesterházy et al. [13], mycotoxin contamination contributes to the loss of approximately 1.3 billion metric tons of food annually. Beyond direct crop losses, contamination in animal feed presents additional challenges. When livestock consume contaminated feed, mycotoxins can accumulate in animal-derived products, creating secondary exposure risks for humans and moreover, compromising animal health, impairing growth, and reduce overall productivity, leading to substantial financial losses for the livestock industry [9].

1.1.1 Some History On Mycotoxins

Throughout recorded history, several major outbreaks of human and animal diseases now recognized as mycotoxicoses have been documented. One of the earliest and most infamous was *ergotism* or "*St. Anthony's Fire*", reported in France as early as 943 AD. Caused by alkaloid compounds produced by *Claviceps purpurea* (ergot) infecting rye and other cereals, ergotism led to the deaths of thousands during the Middle Ages. Its cause, however, was not conclusively identified until the 19th century [1].

During World War II, Russia's Orenburg Province witnessed an outbreak of *Alimentary Toxic Aleukia* (ATA) linked to the consumption of overwintered, moldy grain. The disease, characterized by dermal necrosis, hemorrhaging, leucopenia, and bone marrow degeneration, reached mortality rates of up to 60%. Later studies identified *Fusarium poae*, *Fusarium sporotrichioides*, and *Cladosporium* species as the primary causative molds [1].

Around the same time, Japan reported *yellowed-rice syndrome*, involving severe liver damage in animals fed mold-infected rice. Toxic compounds, later known as "*yellow rice toxins*", were isolated and linked to *Penicillium* species, though broader scientific interest in mycotoxins remained limited [1].

Although *Aspergillus flavus* was known to cause ear mold in corn by 1920, its ability to produce highly toxic metabolites was not recognized until the 1960s [9]. The turning point came with England's infamous "*Turkey-X Disease*", in which tens of thousands of turkey poults, ducklings, and other young farm animals died after consuming peanut meal contaminated by *A. flavus* [1, 9, 10]. Investigations uncovered a series of fluorescent compounds—later identified as aflatoxins (AFLA)—responsible for the fatalities. This period marked a pivotal shift, establishing mycotoxin research as a scientific priority and prompting the development of strategies to regulate and mitigate fungal contamination in the global food supply.

1.1.2 Risks Posed By Mycotoxins

Following the discovery of aflatoxins in poultry feed and the growing recognition of their severe toxicological impact, research into mycotoxins—their biosynthesis, toxicology, and the environmental conditions that promote their occurrence—emerged as a critical scientific priority. To date, over 400 distinct mycotoxins have been identified, although only approximately 30 are considered to pose significant risks to human and animal health [14]. Notably, the mere pres-

ence of mycotoxin-producing fungi in food does not necessarily indicate contamination, as toxin biosynthesis is highly contingent upon specific environmental and physiological triggers [1]. Once produced, however, mycotoxins are remarkably persistent, often remaining within the food matrix even after the fungal colonies have been eliminated [1].

The fungal genera most commonly associated with mycotoxin contamination in the human food chain include *Aspergillus*, *Fusarium*, *Penicillium*, *Alternaria*, *Trichothecium*, *Cladosporium*, *Byssoschlamys*, and *Sclerotinia* [1, 12]. These fungi thrive on a wide variety of substrates and under diverse environmental conditions, moisture levels, pH-levels, and temperatures [1]. Among these, *Aspergillus*, *Fusarium*, and *Penicillium* species are the principal contributors to mycotoxin contamination in the food supply chain [12], particularly affecting staple commodities such as cereal grains, fruits, nuts, and dairy products [1, 2].

Each of these fungal species can create a multitude of different mycotoxins. Among the most hazardous toxins are *Aflatoxin* (AFLA), *Patulin*, *Citrinin*, *Ochratoxin*, *Zearalenone* (ZEA), *Fumonisin* (FUM), and the *Trichothecene* group—including *Deoxynivalenol* (DON), T-2, and HT-2 toxins [1]. These toxic compounds exhibit a remarkable diversity of molecular architectures, ranging from simple heterocyclic structures with molecular weights as low as 50Da (g/mol) to highly complex, multi-ring macromolecules exceeding 500Da. Their small size and structural variability not only enable them to evade immune detection—contributing to their persistence in food and feed—but also hinder effective recognition and removal, thereby complicating mitigation efforts [3]. Consequently, mycotoxin contamination often remains undetected until clinical symptoms of mycotoxicoses emerge in affected populations, underscoring the stealthy and persistent threat posed by these fungal metabolites [3].

1.1.3 Mitigation Of Mycotoxins

Effective mycotoxin mitigation necessitates a comprehensive, multi-stage strategy that not only targets contamination but also ensures continuous monitoring across the entire food production process. This encompasses both pre- and post-harvest interventions, alongside integrated detection and monitoring systems capable of identifying mycotoxin presence at critical control points within the supply chain.

In addition to preventive measures, emerging decontamination strategies—spanning chemical, physical, and biological approaches—aim to neutralize or remove mycotoxins from contaminated commodities. However, these latter methods remain largely experimental and currently fall short of delivering scalable, safe, and cost-effective solutions suitable for industrial deployment.

The following sections provide a concise overview of key mitigation techniques, outlining their practical applications, technological maturity, and inherent limitations.

1.1.3.1 Pre-harvest

Pre-harvest interventions are critical in minimizing fungal colonization and subsequent mycotoxin production. These include timely harvesting to reduce prolonged exposure to environmental stressors, careful handling to prevent mechanical damage, and thorough drying to lower moisture content—an essential factor in suppressing fungal proliferation during storage [7, 14]. Additionally, sorting techniques, whether manual or automated via optical systems, facilitate the removal of damaged or defective produce, which is inherently more vulnerable to fungal contamination [3].

Complementing these immediate actions, agronomic best practices—such as optimized soil management, strategic crop rotation, and rigorous hygiene protocols to prevent cross-contamination

via equipment—serve as proactive defenses against fungal infestation and mycotoxin synthesis [14]. Nonetheless, exclusive reliance on such measures is often inadequate, particularly in resource-constrained environments or under increasingly variable climatic conditions, highlighting the need for adaptive mitigation frameworks and robust monitoring strategies.

1.1.3.2 Post-harvest

Post-harvest strategies constitute a vital second line of defense in mitigating mycotoxin contamination. However, conventional processing methods—such as roasting—are largely ineffective, as many mycotoxins possess exceptional thermal stability, allowing them to endure standard food processing conditions without significant degradation [4, 14]. In response to these limitations, a range of advanced physical decontamination techniques have been explored, including ozonization, irradiation, and filtration. While these methods have demonstrated potential in reducing mycotoxin levels, they frequently compromise the sensory qualities and nutritional value of food products, thereby constraining their broader industrial applicability [14].

Chemical detoxification offers a more direct approach to mycotoxin degradation. Agents such as calcium hydroxide, ammonium hydroxide, and hydrogen peroxide have proven effective in breaking down even the most stable compounds, such as aflatoxins, by converting them into less toxic or inert derivatives. Despite their demonstrated efficacy, the large-scale implementation of chemical treatments remains limited by practical concerns—including high operational costs, the risk of residual toxicity, and compliance with regulatory standards [15]. Thus, while chemical detoxification holds promise, its industrial scalability and safety profile remain key barriers to adoption.

An alternative, and increasingly promising, route is biological detoxification—particularly microbial degradation. Certain microorganisms, such as *Lactobacillus* and *Bifidobacterium*, have shown the ability to bind aflatoxins, thereby reducing their bioavailability and toxicity [14]. Moreover, specific yeast strains exhibit antifungal activity capable of directly inhibiting fungal growth and suppressing aflatoxin biosynthesis at its source [14]. Although biological strategies present an attractive, eco-friendly alternative, they remain largely experimental. Further research is needed to optimize these methods for stability, efficacy, and cost-efficiency before they can be viably scaled for widespread application across the food and feed supply chains.

While post-harvest interventions—including physical, chemical, and biological detoxification strategies—can significantly reduce mycotoxin concentrations, none offer a fully effective or scalable solution for complete decontamination at the industrial level. Owing to the chemical resilience of many mycotoxins and the complex, variable conditions under which they are produced and persist, mitigation efforts must be reinforced by reliable monitoring. Continuous surveillance is crucial for preventing contaminated products from entering the food and feed supply chains, and ensuring regulatory compliance.

The following section introduces mycotoxin monitoring as the final—and arguably most critical—component of a holistic mitigation strategy, examining current detection methodologies, their inherent limitations, and emerging innovations that aim to deliver accurate, real-time contamination assessment at scale.

1.1.3.3 Mycotoxin Monitoring

Ensuring compliance with regulatory thresholds requires comprehensive, continuous monitoring throughout all stages of food production and processing [16]. As of 2025, Liquid Chromatography-tandem Mass Spectrometry (LC-MS/MS) and immunoassays such as Enzyme-Linked Immunosorbent Assays (ELISA) remain the gold standards for mycotoxin detection and surveillance

[2, 17, 18]. Besides these, a wide range of chemical methods, immunological assays, and biosensor technologies has been developed to identify and quantify mycotoxins in various food matrices. Among these, chemical techniques offer the highest sensitivity for detecting multiple mycotoxins at ultra-trace levels [17]. However, these methods are inherently labor-intensive, costly, and demand complex sample preparation, solvent-based extractions, and extensive laboratory infrastructure [2, 19]. In contrast, immunological approaches provide faster, more accessible alternatives but are limited by multiplexing challenges and cross-reactivity issues in complex food matrices [20, 21, 22, 23, 24, 25]. Although biosensors represent a promising emerging technology, they still face significant hurdles in reproducibility and cost-effectiveness that hinder large-scale implementation [2].

Moreover, these conventional detection methods are inherently destructive, rendering tested samples unusable after analysis [26]. Coupled with low throughput and high operational costs, such limitations confine their use to small, representative subsamples rather than enabling comprehensive, batch-wide screening [27]. As a result, chemical analyses typically yield an average contamination estimate for the sampled subset—mitigating the influence of outliers but often failing to capture localized contamination spikes [26]. Given the heterogeneous distribution of mycotoxins within food products [28], this sampling strategy can lead to a misrepresentation of actual contamination levels across the broader supply chain. This not only contributes to unnecessary food waste and economic losses but also fails to ensure the complete exclusion of contaminated products, thereby perpetuating ongoing food safety risks [16].

The limitations of current mycotoxin monitoring methods—namely their destructive nature, low throughput, and dependence on costly, labor-intensive procedures—underscore the urgent need for alternative technologies that enable accurate, real-time contamination assessment. To meet this need, non-destructive, high-throughput approaches have gained increasing attention, with optical spectroscopy emerging as a particularly promising candidate.

1.2 Optical Spectroscopy For Agrifood

Optical spectroscopy—encompassing ultraviolet (UV), visible (VIS), and near-infrared (NIR) reflectance and transmittance spectroscopy, as well as fluorescence spectroscopy, among others—offers a powerful, non-invasive platform for material identification and characterization. These techniques exploit the interaction between light and matter to capture distinct spectral signatures that reflect a sample’s chemical composition and physical properties [29]. The core principle involves wavelength-selective absorption, scattering, and emission of light, corresponding to electronic, vibronic, or Ro-vibronic transitions within the molecular structure of the material.

Identification can be achieved either through the analysis of light emitted intrinsically by the sample—as in fluorescence spectroscopy—or through the interrogation of externally sourced light that is reflected, refracted, scattered, or absorbed by the sample matrix. Each modality yields a unique spectroscopic fingerprint, enabling accurate detection and classification of target compounds. A more detailed discussion of these techniques and their specific applications in mycotoxin detection is presented in Section 2.5.2.

Optical spectroscopy is particularly well-suited for seamless integration into both food sorting systems and handheld devices, enabling real-time, non-destructive monitoring in in-line industrial and in-field agricultural settings [16]. In food sorting systems, items are transported over a conveyor belt where they are illuminated by lasers or LEDs at discrete wavelengths. The re-

sulting reflected, scattered, or fluorescent light is collected via lenses and mirrors and directed toward photodetectors or imaging sensors. To ensure industrial feasibility while enabling in-line processing, only a limited set of wavelengths—typically 8 to 10—is monitored [30]. A central processing unit analyzes the spectral data in real time to determine whether each item meets predefined quality criteria. Contaminated items are promptly removed using targeted air jets, allowing for item-level precision without reliance on bulk-sample testing. Handheld devices, typically miniaturized grating-based or filter-based spectrometer modules, enable portable and on-site analysis [16]. Although compact hyperspectral analyzers exist, their high cost often makes it more economical to restrict analysis to a limited set of wavelengths.

Spectroscopic food sorting systems are already widely deployed in industrial production lines for quality control and foreign object detection [31]. They are particularly prevalent in the processing of staple crops such as maize, rice, soybeans, and potatoes, where they are used to detect fungal infections, adulterants, ripeness, and internal defects [31, 32]. Beyond structural assessment, spectroscopic platforms have demonstrated high sensitivity for detecting specific chemical compounds at trace levels [33].

1.2.1 Challenges In Mycotoxin Detection

Returning to the primary focus of this thesis—the in-situ monitoring of mycotoxins on corn and wheat—it is pertinent to ask whether spectroscopic techniques can provide a viable solution. A substantial body of research has explored this very question, demonstrating the feasibility of detecting mycotoxins directly or indirectly on food matrices without the need for analyte extraction or chemical preprocessing. Various spectroscopic modalities have shown promise for non-invasive detection, and a detailed overview of the current state-of-the-art is presented in Section 2.5.

Despite considerable progress, several critical challenges persist. These can be broadly grouped into three categories: (1.2.1.1) challenges related to sample complexities, (1.2.1.2) challenges in model development, and (1.2.1.3) challenges concerning industrial viability.

1.2.1.1 Sample Complexities

One of the primary obstacles arises from the complex nature of the sample matrix and the desire to implement in-situ measurements. Mycotoxins typically occur in ultra-trace concentrations and generate weak spectroscopic signals that are often masked by the dominant background response of the food-matrix itself. In corn and wheat, this background is shaped by major biochemical constituents such as starch, chlorophyll, and naturally fluorescent compounds—including tryptophan, riboflavin, tyrosine, and phenylalanine [34]. Consequently, the subtle spectral signatures of mycotoxins are frequently obscured, making direct detection highly challenging.

Moreover, not all mycotoxins exhibit strong intrinsic optical responses. For example, fluorescence spectroscopy is well-suited for detecting aflatoxins due to their inherent fluorescence. In contrast, compounds such as deoxynivalenol lack significant native fluorescence, making this modality unsuitable for their direct detection.

In such cases, identification instead relies on indirect spectral markers—matrix-level changes associated with the biosynthetic activity of fungi during toxin production [31]. These secondary alterations, including protein degradation and shifts in lipid and carbohydrate content, can serve as useful proxies for contamination. However, they introduce additional analytical complexity and necessitate the use of naturally contaminated samples. Artificially spiked samples do not reproduce these biochemical alterations, and are therefore insufficient for building models for the

in-situ monitoring of mycotoxins. This significantly restricts the range of usable samples, as only naturally occurring concentrations and toxin compositions can be utilized; it is not possible to generate controlled samples at arbitrary concentrations or compositions.

1.2.1.2 Model Development Difficulties

Given these obstacles, one might reasonably conclude that in-situ spectroscopic detection of mycotoxins isn't feasible. However, the application of chemometrics—a suite of multivariate data analysis techniques—has proven instrumental in extracting meaningful insights from high-dimensional and noisy spectral data. Through mathematical and statistical modeling, chemometric methods uncover hidden patterns and relationships, transforming raw spectra into interpretable and actionable information [35].

Building upon this foundation, advanced machine learning techniques offer an even more powerful and flexible framework for decoding subtle spectral variations linked to mycotoxin contamination [30]. By training classification and regression models on labeled datasets, ML algorithms can effectively differentiate between contaminated and uncontaminated samples—even when spectral differences are indirect, subtle, and context-dependent. In doing so, ML significantly extends the capabilities of traditional chemometric approaches.

Despite their potential, ML-based spectroscopic systems for mycotoxin detection face significant hurdles—chief among them, the limited availability of labeled training data. Building robust and generalizable models requires large and representative datasets paired with accurate chemical reference measurements to establish ground-truth contamination levels. However, collecting such datasets is both resource-intensive and time-consuming, often limiting the number of viable training samples to only a few dozen.

Compounding the challenge is the inherently heterogeneous distribution of mycotoxins. While chemical analysis techniques require tens of grams of material and thus provide only an averaged contamination level, spectroscopic measurements probe localized regions of the sample. This mismatch introduces ambiguity, as the locally measured concentrations may differ significantly from the batch-level reference values used as ground truth during model training.

As a result, most successful applications to date have focused on classification models, particularly those designed for binary sorting based on predefined thresholds of total mycotoxin content. However, a more selective subdivision of individual mycotoxins is required for food safety monitoring compliant with regulatory frameworks.

Furthermore, in contrast to classification, regression models—which aim to provide quantitative estimates of contamination—are considerably more challenging to develop due to their higher data requirements and the need for fine-grained ground truth labels [31]. Consequently, research in this area remains comparatively limited.

Another challenge lies in the selection of an appropriate learning algorithm, which involves balancing competing demands such as model complexity, processing speed, accuracy, and selectivity. Highly complex models may offer improved performance but often at the cost of computational efficiency—factors that are crucial for real-time applications. Conversely, simpler models may generalize better on small datasets and enable faster deployment, but they risk reduced sensitivity to subtle spectral variations. As such, algorithm selection remains a critical design decision that must be tailored to both the application context and the practical constraints of the detection system.

1.2.1.3 Industrial Viability

Another critical factor is the feasibility of deploying these techniques in industrial settings. While laboratory-grade spectrometers can record full-spectral data across hundreds of wavelengths, practical implementation in food sorting systems or handheld devices is constrained by hardware limitations. As noted earlier, commercial systems often operate using only a limited number of discrete wavelengths—typically between 8 and 10.

This constraint highlights the importance of feature selection: the process of identifying the most informative wavelengths for model training. By selecting only the key spectral features, it becomes possible to maintain detection performance comparable to full-spectrum models while significantly reducing system complexity, data acquisition time, and computational overhead. This is crucial for industrial integration and improves both scalability and interpretability [16]. A final challenge is ensuring that the model’s detection accuracy meets regulatory thresholds for mycotoxin contamination—without such compliance, adoption by the food industry is unlikely.

1.3 Objectives

Mycotoxin contamination poses a persistent threat to public health and results in substantial economic losses within the agrifood sector. While conventional chemical and immunological methods offer high analytical accuracy, they are typically slow, destructive, and unsuitable for large-scale, real-time monitoring. This has created a pressing need for innovative detection approaches that are non-invasive, rapid, and compatible with high-throughput workflows. Although spectroscopic techniques face challenges such as weak analyte signals, matrix interference, limited availability of annotated data, and hardware constraints, recent advances in instrumentation and data-driven modeling have begun to overcome these obstacles, offering a viable path toward practical implementation.

Chemometric methods enable the extraction of relevant information from complex spectral profiles, while machine learning enhances detection capabilities by identifying subtle, indirect spectral signatures. Additionally, feature selection helps reconcile the trade-off between detection accuracy and hardware feasibility, supporting the development of scalable, high-throughput, and industry-ready solutions. Together, these innovations are poised to transform spectroscopic mycotoxin detection from a laboratory-bound technique into a practical tool for real-world food safety monitoring.

This thesis aims to address these challenges by exploring the potential of optical spectroscopy platforms enhanced with machine learning techniques to detect multiple mycotoxins simultaneously on both corn and wheat. The primary objectives of this research are to:

1. **Assess The State-of-the-Art:** Perform a comprehensive evaluation of current detection methods—including chemical, immunological, biosensor, and spectroscopic approaches—by examining their effectiveness, benefits, and drawbacks. Particular focus will be given to techniques enabling multiplexed detection, which allow for the simultaneous identification of multiple mycotoxins.
2. **Optical Spectroscopic Measurements:** Compose samples from corn and wheat reference standards, contaminated with multiple mycotoxins and develop an extensive spectroscopic dataset, encompassing a variety of measurement techniques, including fluorescence and upconversion fluorescence acquired under multiple excitation wavelengths and powers, as well as reflection spectroscopy spanning the UV-VIS and NIR spectral regions. This

comprehensive dataset will serve as the foundational basis for the subsequent development, optimization and evaluation of advanced detection models.

3. **Evaluate Classification And Regression (Models):** Develop robust classification and regression models by integrating chemometric analysis with advanced machine learning algorithms. These models will be designed to simultaneously detect multiple mycotoxins in corn, wheat, or combined corn and wheat samples. The performance of the various spectral datasets will be compared to identify the most informative spectroscopic platforms, and complementary data will be merged to determine the optimal detection framework. The final models will be rigorously evaluated by benchmarking against current state-of-the-art detection methods.
4. **Implement Wavelength Selection:** Investigate and apply advanced feature selection techniques to reduce spectral complexity and data dimensionality. This approach seeks to balance detection sensitivity with system simplicity, ultimately tailoring the platform for industrial integration.

In summary, this thesis aims to bridge the gap between analytical precision and industrial feasibility by developing an advanced spectroscopic detection framework for multi-mycotoxin identification on both corn and wheat. By leveraging multiple spectroscopic modalities—fluorescence, upconversion fluorescence, UV-VIS, and NIR reflectance spectroscopy—this work seeks to establish a robust, high-throughput detection platform. Central to this effort is the collection of a comprehensive spectroscopic dataset, enabling the rigorous evaluation of various machine learning models to identify the most effective approach for both single and multiplexed mycotoxin detection. A key challenge addressed in this work is the reduction of spectral complexity, with a focus on optimizing feature selection to limit the number of analyzed wavelengths to align with industrial standards for real-time food sorting applications, while merging information from various complementary modalities. Finally, the performance of the developed detection framework will be critically benchmarked against existing state-of-the-art methods, ensuring compliance with regulatory standards and assessing its practical implementation potential for large-scale food safety monitoring.

1.4 Overview Of The Thesis

This thesis is organized into several interlinked chapters, each contributing a key component to the overall research framework. Figure 1.1 presents a high-level schematic of the workflow, illustrating the progression from sample synthesis through spectral data acquisition and preprocessing, to the development of classification and regression models, each integrated with feature selection. A detailed discussion on each of these aspects will be given in the following chapters.

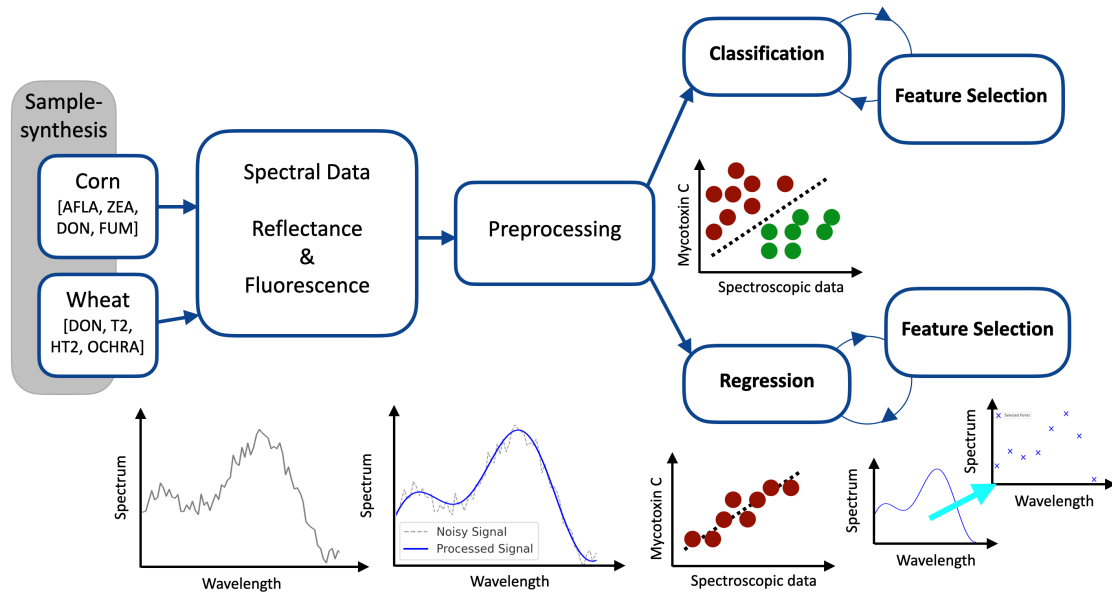


Figure 1.1: Schematic overview of the thesis workflow. From sample synthesis through reflectance and fluorescence spectral acquisition, the workflow branches into classification and regression pathways, each incorporating preprocessing and feature selection steps.

A breakdown of the content of each chapter is given below:

- Chapter 1: Introduction;** The introductory chapter contextualizes the mycotoxin contamination issue within the agrifood industry, emphasizing the health risks, economic implications, and the need for improved detection methods. It provides an overview of traditional and emerging mycotoxin detection approaches and introduces the motivation for exploring optical spectroscopy as a viable alternative. The objectives of the research are outlined, along with the key challenges that must be addressed.
- Chapter 2: Literature Review;** This chapter provides an in-depth analysis of the current body of knowledge surrounding mycotoxin contamination in corn and wheat. It begins by examining the diverse range of mycotoxins, their growth conditions, and the resultant health implications, while also addressing the complex interactions arising from co-contamination. Next, it reviews the regulatory settings for mycotoxins in both the EU and US, providing the baseline performance benchmarks. Thereafter the chapter critically evaluates the state-of-the-art in mycotoxin detection by discussing both established methodologies (such as liquid chromatography and immunoassays) and emerging technologies, including bio-chip solutions and optical spectroscopy.

Finally, it explores the theoretical foundations of various spectroscopic platforms, machine learning classification and regression models, and advanced feature selection techniques aimed at reducing spectral complexity.

- **Chapter 3: Materials and Methods;** This chapter begins by outlining the overall workflow of the thesis, presenting the experimental roadmap. In the second part, a detailed discussion of sample preparation—using certified reference standards—is provided. This is followed by a description of the experimental setups, including both small and large integration spheres for broadband diffuse reflection spectroscopy and a tunable laser configuration for fluorescence measurements, along with the corresponding measurement methodologies used across both modalities. The chapter then presents a comprehensive visualization and analysis of the acquired spectral data, highlighting the impact of various measurement parameters and establishing a solid foundation for the subsequent development and evaluation of predictive models.
- **Chapter 4: Classification Models;** This chapter focuses on the development and optimization of machine learning–based classifiers for multi-mycotoxin detection using spectroscopic data. It begins by outlining the initial state of the project prior to the thesis work and explains how this context informed the decision to adopt a Leave-One-Sample-Out Cross-Validation strategy. The chapter then presents the model architecture search and identifies the most effective classifiers, followed by a critical benchmarking of their performance relative to the current state-of-the-art. Special emphasis is placed on feature selection techniques, which are used to reduce model complexity by isolating the most informative wavelengths. The resulting trade-off between model performance and complexity is analyzed in detail.
- **Chapter 5: Regression Models;** This chapter follows a parallel approach to the prior chapter, now looking into regression model development. It begins by presenting the preprocessing and model-search process, followed by a detailed discussion of the best-performing regressors, with performance critically compared to the current state-of-the-art. In the final stage, feature selection is revisited as a strategy to reduce input dimensionality while preserving predictive accuracy.
- **Chapter 6: Validating Model Performance;** In this penultimate chapter, the generalizability of the developed classification models is evaluated using a newly composed set of validation samples. These samples were prepared from independently sourced reference materials, entirely distinct from those used during model training. The chapter begins by detailing the composition and spectral properties of these novel samples, followed by an assessment of model performance on this external dataset. This final validation experiment provides critical insight into the model’s ability to generalize beyond the narrowly defined training domain, and assesses whether toxin-specific spectral features were effectively learned in the presence of broader biological variability.
- **Chapter 7: Concluding Remarks & Future Perspectives;** This final chapter reflects on the key outcomes of the thesis, examining both classification and regression models alongside the trade-offs introduced by spectral feature selection. It underscores the strengths and limitations of the developed models and proposes avenues for future research to further enhance performance, scalability, and real-world applicability.

Chapter 2

Mycotoxin Monitoring: A Literature Review

Building on the introduction’s emphasis on the critical need for improved mycotoxin detection platforms, this chapter delivers a comprehensive review of the principal mycotoxins of concern, the regulatory thresholds that define minimal performance requirements, and the current state-of-the-art in available detection technologies. We critically assess both conventional methods—such as chemical chromatography and immunoassays—and emerging approaches like biosensors and optical sensors, systematically highlighting their respective strengths and intrinsic limitations. Through this analysis, we establish the motivation for using optical spectroscopy as a promising non-invasive alternative, and a detailed state-of-the-art is presented. The fundamental operating principles of key spectroscopic modalities are thereafter outlined, alongside an in-depth discussion of the advanced chemometric and machine learning models used for mycotoxin detection from the spectral data. This integrated review forms the conceptual and technical foundation for the experimental and methodological advancements presented in the subsequent chapters.

2.1 Mycotoxins On Maize & Wheat

Cereal crops such as *Zea mays* (maize, also referred to as corn) and *Triticum aestivum* (wheat) are highly susceptible to mycotoxin contamination, posing a significant challenge to food safety. The most common mycotoxins affecting these crops include *aflatoxins* (AFLA), *fumonisin* (FUM), *deoxynivalenol* (DON), and *zearalenone* (ZEA), as well as *ochratoxin A* (OTA) and T-2 and HT-2 toxins [1]. Contamination can arise at multiple critical stages of production: pre-harvest, during harvest and drying, and throughout post-harvest storage [1, 9]. Environmental factors such as high humidity, drought stress, and physical damage to kernels—resulting from insect infestation or mechanical handling—significantly amplify the risk of fungal colonization and subsequent mycotoxin production [9]. Moreover, ongoing climatic shifts are expected to further exacerbate these stress conditions, potentially increasing the global occurrence and severity of mycotoxin contamination [10].

The global prevalence of mycotoxin contamination in maize is well-documented. A survey of 11,237 samples collected from 75 countries between 2014 and 2017 revealed that 92%¹ of the

¹This includes not only samples with contamination levels exceeding legal limits, but also those containing trace amounts.

samples contained at least one mycotoxin, with many exhibiting co-contamination by multiple toxins [9]. Among these FUM was the most frequently detected, present in 82% of samples at a median level of 1.4mg/kg, with some samples reaching as high as 171.9mg/kg. Aflatoxins were detected in 24% of the samples, with concentrations as high as 1.4g/kg in some cases. Additionally, ZEA and DON were found in 49% and 72% of samples, respectively, with median levels of 76 μ g/kg and 581 μ g/kg respectively. Such widespread contamination underscores the challenge of ensuring food safety in both domestic and international markets.

The following subsections introduce the key mycotoxins relevant to this thesis—AFLA, DON, ZEA, FUM, OTA, and T-2/HT-2—detailing the producing fungal species, the toxin structures, toxicological effects, and spectroscopic properties. Additionally, critical environmental and biological factors influencing mycotoxin biosynthesis will be discussed. Finally, the phenomenon of co-contamination—where multiple mycotoxins coexist within the same food matrix—will be examined, with particular emphasis on potential synergistic effects on health.

2.1.1 Aflatoxin (AFLA)

Genus *Aspergillus* comprises over 300 identified species [4], although only a limited subset produces mycotoxins. Among these, aflatoxins (AFLA) are the most significant, primarily produced by *Aspergillus flavus* and *Aspergillus parasiticus*, with *Aspergillus nomius* contributing to a lesser extent [1]. Aflatoxins are among the most potent natural hepatocarcinogens, exhibiting both acute toxicity and significant carcinogenic potential [9]. Epidemiological evidence also indicates a synergistic effect between aflatoxin exposure and hepatitis-B virus infection, significantly elevating the risk of hepatocellular carcinoma [3, 36]. Beyond carcinogenicity, aflatoxins exert immunosuppressive, mutagenic, and teratogenic effects, further underscoring their severe impact on health [9].

Over 20 distinct types of aflatoxins have been identified, among these, four primary forms—B₁, B₂, G₁, and G₂—are the most frequently detected in contaminated foods and animal feed [1, 12]. The labels "B" and "G" correspond to the characteristic blue and green fluorescence emitted under ultraviolet (UV) light, while the subscript numbers distinguish major and minor variants [1, 12, 22]. Aflatoxin B₁ (AFB₁) is notably the most toxic and carcinogenic of the group, classified as a group one carcinogen by the International Agency for Research on Cancer (IARC) [9].

The chemical structure of AFB₁ is 2,3,6a,9a-tetrahydro-4-methoxycyclopenta[c]furo[2,3:4,5]furo[2,3-h]chromene-1,11-dione, with formula C₁₇H₁₂O₆ [37]. The molecular structures of AFB₁, AFB₂, AFG₁ and AFG₂ are shown in Figure 2.1.

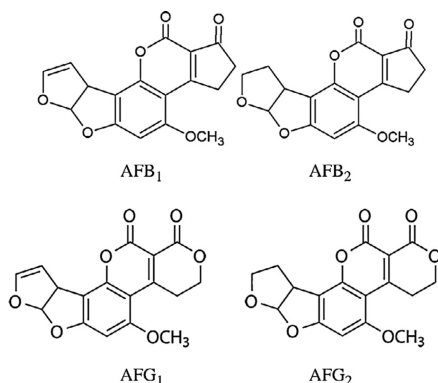


Figure 2.1: The molecular structure of the 4 main aflatoxin forms: AFB₁, AFB₂, AFG₁, and AFG₂ [4].

Spectroscopically aflatoxins strongly absorb light in the 200-250nm range and between 300 and 400nm, with a prominent absorption peak around 365nm. Their fluorescence emission typically spans the 400-500nm range, peaking near 440nm, although the exact emission characteristics depend on the excitation wavelength used and environment of the toxins [30, 38, 39].

2.1.2 Ochratoxin A (OTA)

Penicillium species are commonly observed as blue-green, powdery molds on corn kernels, with *Penicillium Oxalicum* particularly associated with ear rot in the field [9]. Unlike *Aspergillus*, *Penicillium* species are more prevalent in cooler climates and can proliferate at grain moisture levels as low as 16% to 17%, making them prominent contaminants in stored grain [9]. The genus *Penicillium* produces a wide range of mycotoxins, notably patulin and ochratoxins [12, 40]. Ochratoxins (OTs) are synthesized primarily by *Penicillium verrucosum*, as well as by several *Aspergillus* species, including *Aspergillus ochraceus* and *Aspergillus carbonarius* [41].

Among the various forms synthesized, ochratoxin A (OTA) is the most prevalent and toxicologically significant. OTA is frequently detected in cereals, coffee, dried fruits, and red wine, and it is of particular concern due to its potent nephrotoxic effects [42]. It has been implicated as a possible human carcinogen and is associated with kidney damage; notably, exposure to OTA has been linked to Balkan endemic nephropathy—a fatal chronic renal disease observed in parts of Eastern Europe.

OTA has a molecular formula $C_{20}H_{18}ClNO_6$ and its chemical structure is depicted in Figure 2.2. Its IUPAC name is N-[(3R)-5-chloro-8-hydroxy-3-methyl-1-oxo-3,4-dihydro-1H-isochromen-7-yl] carbonyl-L-phenylalanine [43]. This complex structure features an isochromenone core linked via a carbonyl group to L-phenylalanine, combining both lipophilic and polar elements. The presence of a chlorine atom, a hydroxyl group, and other functional moieties contributes to OTA's stability and biological reactivity [43]. Spectroscopically OTA shows intrinsic fluorescence with absorption in the 301-395nm range, and broad emission in the range 400-475nm [44, 45].

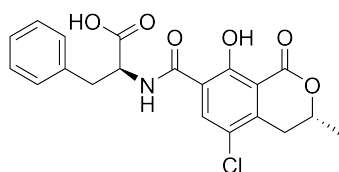


Figure 2.2: The molecular structure of ochratoxin A (OTA) [43].

2.1.3 Fusarium Toxins

The genus *Fusarium* comprises a diverse array of toxigenic fungi that significantly impact cereal crops such as maize and wheat [46]. Globally, at least 22 toxigenic *Fusarium* species have been identified on these crops, with their distribution varying by geographic latitude [9]. One particularly prominent species, *Fusarium graminearum*, is the primary causative agent of Gibberella ear rot—or "red ear rot"—in corn, a disease that typically initiates at the ear tip and spreads to produce extensive red or pink mold throughout the kernel, thereby threatening crop quality and yield [9, 47].

In addition to field infections, post-harvest conditions play a critical role in the proliferation of *Fusarium* species. Suboptimal storage conditions, especially when grain moisture exceeds 19%, facilitate fungal growth and mycotoxin production, thereby amplifying food safety risks [9, 46].

Furthermore, *Fusarium* species synthesize a wide array of mycotoxins, over 200 distinct compounds, including trichothecenes—such as deoxynivalenol (DON), T-2, and HT-2 toxins—as well as zearalenone (ZEA) and fumonisins (FUM) [4, 9].

2.1.3.1 Deoxynivalenol (DON)

Among the various trichothecenes, deoxynivalenol (DON), commonly known as *Vomitoxin*, is one of the most frequently detected mycotoxins in maize and wheat. Despite being one of the less toxic trichothecenes, it still poses significant health risks, particularly to livestock. It disrupts protein and DNA synthesis, leading to reduced feed intake and impaired growth rates in animals [46]. Swine are especially sensitive to DON exposure, with feed refusal commonly occurring at dietary concentrations of approximately 8mg/kg [3]. In humans, DON has been associated with *alimentary toxic aleukia* (ATA), a severe illness historically documented in the former USSR (Union of Soviet Socialist Republics) [9].

Chemically DON is classified as 12,13-epoxy-3,4,15-trihydroxytrichothec-9-en-9-one [48], with its molecular structure depicted in Figure 2.3. Spectroscopically, it exhibits no intrinsic fluorescence; consequently, detection via fluorescence modalities relies exclusively on indirect signatures arising from matrix alterations associated with mycotoxin metabolism [49].

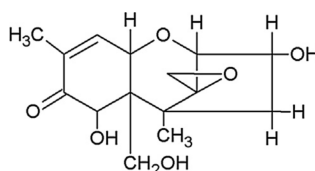


Figure 2.3: The molecular structure of Deoxynivalenol (DON) [4].

2.1.3.2 T-2 And HT-2 Toxins

T-2 and HT-2 toxins are also trichothecenes produced primarily by various *Fusarium* species, including *Fusarium sporotrichioides*, *Fusarium langsethiae*, *Fusarium acuminatum*, and *Fusarium poae* [42]. These toxins are commonly detected in cereals such as oats, barley, and wheat. Both pose significant health risks due to their high cytotoxic and their disrupting nature to DNA and RNA synthesis, affecting rapidly dividing cells in the gastrointestinal tract, skin, and lymphoid tissues [4]. Moreover, T-2 toxin is capable of being absorbed through the skin as well as by ingestion of contaminated grains, thereby increasing exposure risks via both dietary and dermal routes. In addition to DON, T-2 toxin was one of the primary agents linked to the 1940 outbreak of ATA in Russia [1].

Chemically T-2 is known as (2 α , 3 α , 4 β , 8 α)-4,15-bis(acetyloxy)-3-hydroxy-12,13-epoxytrichothec-9-en-8-yl 3-methylbutanoate, the chemical structure of the toxin is shown in Figure 2.4. HT-2 is structurally defined as 2-(acetyloxymethyl)-10,11-dihydroxy-1,5-dimethylspiro[8-oxatricyclo-dodec-5-ene-12,2'-oxirane]-4-yl 3-methylbutanoate. The trichothecene core shared by both toxins is primarily responsible for the potent biological activity of the mycotoxins [4]. Just like was the case for DON, T-2 and HT-2 toxins exhibit little to no intrinsic fluorescence, requiring indirect detection through matrix alterations [49].

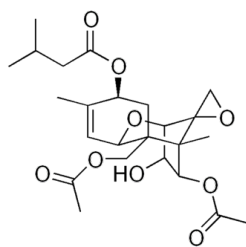


Figure 2.4: The molecular structure of T-2 [4].

2.1.3.3 Zearalenone (ZEA)

Zearalenone (ZEA) is another mycotoxin produced by *Fusarium* species, with contamination typically occurring under cool, moist conditions either pre-harvest or during inadequate storage [4]. The primary toxicity of ZEA stems from its strong estrogenic activity, due to its structural similarity to endogenous estrogens [9].

Chemically, ZEA is known as 6-(10-hydroxy-6-oxo-trans-1-undecenyl)- β -resorcylic acid lactone, with the molecular formula $C_{18}H_{22}O_5$ [9]. The molecular structure of ZEA is shown in Figure 2.5. ZEA exhibits significant heat stability, withstanding temperatures exceeding 150°C . Degradation is only observed under extreme conditions, such as very high temperatures or exposure to alkaline environments [4].

Spectroscopically it exhibits blue-green fluorescence when excited by long wavelength UV light (360nm) and a more intense green fluorescence when excited with short wavelength UV light (260nm) [22]. In methanol, ZEA exhibits absorption maxima at 236, 274, and 316nm, with maximum fluorescence emission at 450nm when excited at 314nm [50].

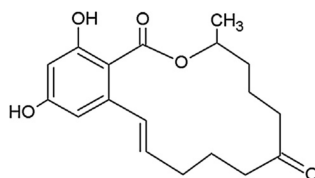


Figure 2.5: The molecular structure of Zearaleone (ZEA) [4].

2.1.3.4 Fumonisin (FUM)

Lastly, we present Fumonisin (FUM). These toxins are produced by at least 16 fungal species, most of which belong to the *Fusarium* genus [51]. FUM is commonly found in maize [4, 9], where contamination typically occurs before harvest or shortly afterwards, especially during the early stages of storage [4]. In addition to maize, FUM has also been detected in lesser amounts in other crops such as wheat, barley, rice, and coffee beans [10, 52].

Fumonisin are primarily recognized for their toxicity in animals [9, 53]. Regarding human health, early research established a strong correlation between fumonisin contamination and esophageal cancer [9]. Today, fumonisin B₁ (FB₁) is classified as a Group 1 carcinogen by the International Agency for Research on Cancer [53].

From a chemical perspective, fumonisins are a family comprising at least 28 polyketide-derived, nonfluorescent mycotoxins, with a structure similar to sphinganine [4, 10]. Among these, Fumonisin B₁ (FB₁), B₂ (FB₂), and B₃ (FB₃) are the most prevalent and thoroughly studied [9, 10]. FB₁ is chemically defined as 1,2,3-propane-tricarboxylic-acid,1,1-[1-(12-amino-4,9,11-trihydroxy-2-methyl-tridecyl)-2-(1-methylpentyl)-1,2-ethane-diyl]ester [4, 9]. Its complex chemical structure is shown in Figure 2.6. These toxins are also heat stable, and only significantly reduce in abundance at temperatures exceeding 220°C [51], and there is little degradation during fermentation processes [4]. Just like was the case for the other trichothecenes, FUM does not exhibit significant intrinsic fluorescence.

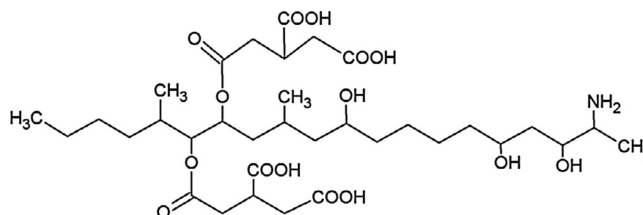


Figure 2.6: The molecular structure of Fumonisin B₁ (FB₁) [4].

2.1.4 Multi-mycotoxin: Co-contamination

Until now, our discussion has focused on the individual effects of mycotoxins—namely AFLA, DON, ZEA, FUM, OTA, and T-2/HT-2. However, when multiple mycotoxins co-occur, they can interact in additive, synergistic, or antagonistic ways, often amplifying their toxic effects beyond the impact of individual toxins [7, 53]. This phenomenon is particularly common in maize and wheat, where *Aspergillus*, *Penicillium*, and *Fusarium* species flourish, frequently resulting in the simultaneous detection of for example AFB₁ and FB₁ [53].

Such co-contamination not only complicates the detection process but also raises serious concerns about the heightened health risks associated with combined mycotoxin exposure [7]. These interactions further intensify the challenges of ensuring the safety of food and feed products, and highlights the need for methods capable of multiplexed mycotoxin detection and monitoring.

For example, studies have demonstrated that while DON and FB₁ individually caused only marginal reductions in weight gain in pigs, their combined presence resulted in a marked synergistic effect, significantly impairing growth [7, 54]. In contrast, the co-occurrence of DON and FB₁ in chickens showed an additive effect on growth reduction, highlighting species-specific differences in toxin interactions [7, 55]. The interaction between DON and ZEA has been observed to vary depending on concentration and biological systems, ranging from additive to synergistic toxicity [7, 56, 57]. Additionally, AFB₁ and FB₁ have shown antagonistic cytotoxic effects in certain cellular systems [7], conversely, AFB₁ in combination with ZEA or DON has been reported to enhance cytotoxicity [53], further illustrating the complexity and unpredictability of these interactions.

Our understanding of the combined effects of co-occurring mycotoxins remains limited, underscoring the need for further research. Advancing analytical methods capable of simultaneously detecting multiple mycotoxins is crucial for improving risk assessment and ensuring food safety. However, existing regulations primarily focus on individual mycotoxins, overlooking their po-

tential interactions. As research progresses, it is likely that regulatory frameworks will evolve to impose stricter limits not only on individual mycotoxins but also on multi-mycotoxin contamination. Therefore, any platform developed for industrial applications should be designed with future compliance in mind, offering robust multi-mycotoxin monitoring capabilities at ever decreasing concentration thresholds.

The following section outlines the current regulatory limits for mycotoxin contamination relevant to this thesis, focusing on specific toxin/matrix combinations. For maize, thresholds for AFLA, DON, ZEA, and FUM are reviewed, while for wheat, the established limits for DON, T-2, HT-2, and OTA are examined across both European and American markets.

2.2 Regulation Of Mycotoxins On Maize And Wheat

Regulatory limits for mycotoxins in corn and wheat vary considerably across regions, reflecting differences in food safety priorities, legislative frameworks, and monitoring capabilities [7]. Both the regulatory commission of the European Union (EU [58]) and the United States Food and Drug Administration (US-FDA [59]) have set action levels to account for the unavoidable presence of these toxins in the food supply chain. Table 2.1 summarizes the limits for human consumption established by the EU for aflatoxins (AFLA), deoxynivalenol (DON), fumonisins (FUM), and zearalenone (ZEA) in unprocessed and ground corn, as well as the limits for DON, ochratoxin A (OTA), T-2, and HT-2 in unprocessed and ground wheat. Table 2.2 presents the corresponding thresholds set by the US-FDA, which indicate levels above which legal action may be taken to prevent contaminated products from reaching the market. Although the FDA has updated guidelines to warn for toxins such as T-2, HT-2, OTA, and ZEA, no official action levels have as of yet been established for these toxins.

Current regulations in both the EU and US focus on individual mycotoxins, overlooking the increased risks associated with co-occurring toxins. Research has demonstrated that synergistic or additive effects from multiple mycotoxins can pose a greater threat to human and animal health than individual contaminants alone [7, 53]. Although the FDA has recently incorporated multi-mycotoxin monitoring into its compliance program, it has not yet established specific measures to prevent co-contamination from exceeding safety limits.

Since our research is conducted within the European Union, the EU Commission's maximum permissible levels serve as our critical benchmark. Any mycotoxin detection technique must achieve, at a minimum, the sensitivity required to comply with these regulatory thresholds. For maize intended for human consumption, the EU allows a maximum of $4\mu\text{g}/\text{kg}$ (ppb) for total aflatoxins, with a stricter limit of $2\mu\text{g}/\text{kg}$ specifically for AFB₁. The allowable level for deoxynivalenol (DON) in unprocessed maize is $1250\mu\text{g}/\text{kg}$, while fumonisins (FUM) and zearalenone (ZEA) are limited to $4000\mu\text{g}/\text{kg}$ and $350\mu\text{g}/\text{kg}$, respectively. For ground wheat the limit for DON is set at $1000\mu\text{g}/\text{kg}$, ochratoxin A (OTA) is restricted to $5\mu\text{g}/\text{kg}$, and the combined limit for T-2 and HT-2 toxins is limited to $20\mu\text{g}/\text{kg}$.

Regulatory limits for Aflatoxins (AFLA) in maize under EU legislation				
Mycotoxin	Commodity	B ₁ Limit [μg/kg]	Sum of B ₁ , B ₂ , G ₁ , G ₂ [μg/kg]	References
Aflatoxins (AFLA)	Unprocessed cereal grains	2	4	2023/915
	Maize flour, bran, or germ	2	4	2023/915
Regulatory limits for other mycotoxins in maize under EU legislation				
Mycotoxin	Commodity	–	Regulatory Limit [μg/kg]	Reference
Deoxynivalenol (DON)	Unprocessed maize grains	–	1250	2023/915
	Processed maize products	–	750	2023/915
	Maize for popping and popcorn	–	750	2024/1022
Fumonisin (FUM)	Unprocessed maize grains	–	4000	2023/915
	Processed maize-based foods	–	1000	2023/915
Zearalenone (ZEA)	Unprocessed maize grains	–	350	2023/915
	Maize snacks and cereals	–	100	2023/915
Regulatory limits for mycotoxins in wheat under EU legislation				
Deoxynivalenol (DON)	Unprocessed cereal grains	–	1000	2024/1022
Ochratoxin A (OTA)	Unprocessed cereals	–	5	2022/1370
Sum of T-2 and HT-2	Unprocessed cereal grains	–	50	2024/1038
	Milling products of cereals	–	20	2024/1038

Table 2.1: Regulatory limits for mycotoxins in corn and wheat for human consumption as established by the European Union’s regulatory commission [58]. The table is organized into three sections: one for aflatoxins in corn, one for other mycotoxins in corn (DON, ZEA, FUM), and one for toxins in wheat (DON, OTA, T-2, and HT-2).

Regulatory limits for mycotoxins in corn for the US-FDA		
Mycotoxin	Commodity	Regulatory Limit [μg/kg]
Aflatoxins (AFLA)	Foods	20
	Milk (aflatoxin M ₁)	0.5
Deoxynivalenol (DON)	Finished corn products (e.g., flour)	1000
Fumonisin (FUM) (FB ₁ +FB ₂ +FB ₃)	Degermed dry milled corn products	2000
	Whole/partially degermed dry milled corn products	4000
	Corn bran	4000
	Corn for mass production	4000
	Corn for popcorn	3000
Regulatory limits for mycotoxins in wheat for the US-FDA		
Deoxynivalenol (DON)	Finished wheat products (e.g., flour, bran)	1000

Table 2.2: Regulatory limits for human consumption, set by the US-FDA [59], for mycotoxins (AFLA, DON, FUM) in corn and wheat.

2.3 Chemical Platforms For Mycotoxin Detection

The enforcement of the maximum permissible levels—ranging from parts-per-billion to parts-per-million range—requires the deployment of highly reliable and sensitive detection platforms for monitoring mycotoxins in food and feed supply chains [2]. Moreover, as we learned contamination is frequently multiplexed, with multiple mycotoxins present simultaneously on the same produce, which necessitates detection methods capable of accurately identifying a spectrum of toxins at once [19].

Traditionally, two primary approaches have been employed for mycotoxin detection. The first relies on chemical chromatographic techniques, renowned for their high precision and ability to quantitatively analyze a wide range of toxins. In contrast, rapid screening methods—such as immunoassays and biosensors—provide faster, qualitative or semi-quantitative results. Despite

their respective advantages, both approaches depend on chemical extractions, making them inherently destructive and unsuitable for fast, in-line, batch-wide, or in-field scanning applications.

In the following sections (Sections 2.3.1 to 2.3.3), we present an in-depth review of the state-of-the-art chemical methodologies for mycotoxin detection. These sections provide background context that will later serve as a reference point for comparing the performance achieved using spectroscopic methods, whose state-of-the-art is discussed in Section 2.5. While informative, these background sections are not essential to the core contributions of this work and may be skimmed, unless a deeper understanding of conventional detection methods is of interest to the reader.

2.3.1 Chromatographic Techniques

The advent of thin-layer chromatography (TLC) revolutionized mycotoxin detection. First introduced in the 1960s in response to the *Turkey X* disease outbreak for identifying aflatoxins, TLC quickly became the leading chromatographic method for the study and analysis of mycotoxins [60]. Today TLC remains a valuable tool in certain applications, offering detection and quantification limits on the order of a few $\mu\text{g/kg}$ (ppb) [61, 62, 63].

Chromatographic techniques have however vastly improved, and new techniques like Liquid Chromatography with tandem Mass Spectrometry (LC-MS/MS) or high-performance Liquid Chromatography with tandem Mass Spectrometry (HPLC-MS/MS), enable the simultaneous detection of a vast number of mycotoxins, and detection limits as low as $0.01\mu\text{g/kg}$ [17, 18]. These newer chromatographic methods offer highly sensitive and specific detection platforms and are the industry standard for reference and regulatory measurements [64].

The first step in implementing chromatographic methods is the extraction of the analyte from the matrix. This step must be meticulously designed to efficiently separate the target analytes from the complex sample matrix without co-extracting excessive amounts of interfering matrix elements. This step is sample and matrix specific, adding complexity and demanding expertise to ensure complete analyte extraction² [2].

Traditional extraction techniques typically rely on solvent extraction, using organic solvents like methanol or acetonitrile mixed with water in various ratios to maximize extraction efficiency while mitigating matrix interference [2]. To enhance efficiency and reduce solvent consumption, more advanced methods such as Supercritical Fluid Extraction (SFE), Accelerated Solvent Extraction (ASE), and Microwave-Assisted Extraction (MAE) have been developed [64]. These advanced techniques not only offer improved reproducibility but are also considered more environmentally friendly compared to conventional approaches.

Another widely used method is the Quick, Easy, Cheap, Effective, Rugged, and Safe (QuEChERS) technique, which simplifies sample preparation and significantly reduces costs, making it a popular choice for many mycotoxin analyses [64]. However, while QuEChERS is effective for single mycotoxin detection, it is less suitable for multi-mycotoxin analysis because it does not extract all mycotoxins equally well, posing challenges for comprehensive multiplexed detection [64].

A second preparatory step required for the chromatographic methods is sample clean up. Techniques like TLC and LC-MS/MS or HPLC-MS/MS are super sensitive to matrix interference [64], so the clean up steps are required to efficiently separate the mycotoxins of interest from any

²Also, note that multi-toxin detection is especially complex, since different mycotoxins may require different optimized conditions to achieve complete extraction due to their varying physicochemical properties like their varying polarities [19].

interfering compounds. Common cleanup methods include Immuno-Affinity Columns (IAC), Solid-Phase Extraction (SPE), liquid-liquid partitioning, ion-exchange columns, and column chromatography [19, 64]. These sample clean up steps can take several hours to complete, and once again require trained professionals.

2.3.1.1 State-of-the-art

The state-of-the-art performance of TLC, HPLC-MS/MS, and LC-MS/MS for multi-mycotoxin detection in corn, wheat and related matrices is summarized in Table 2.4. This table presents key details, including the sample type (matrix), target mycotoxin(s), analytical method employed, and various performance metrics, including: the recovery [%], which indicates the proportion of a spiked mycotoxin successfully extracted and detected; the Limit Of Detection (LOD), representing the lowest detectable toxin concentration; the Limit Of Quantification (LOQ), specifying the lowest concentration that can be reliably quantified; and the Linear Range (LR), defining the concentration range over which the method is effective. Additionally, the table includes the year of publication and corresponding reference for each study, providing insights into the evolution of chromatographic techniques and their improving performance over time.

Many of the works cited in this table, as well as in Tables 2.6 and 2.7, were identified through two pivotal review studies: *"Mycotoxins in Food: Recent Developments in Food Analysis and Future Challenges; A Review"* by Shahzad Z. Iqbal [64] and *"Rapid and Sensitive Detection of Mycotoxins by Advanced and Emerging Analytical Methods: A Review"* by Jyoti Singh and Alka Mehta [22]. Additional relevant studies were also sourced through an independent Google Scholar search using key terms such as multi-mycotoxin detection, corn/maize/wheat, and the relevant detection technologies.

The collected studies demonstrate the evolution of chromatographic techniques for mycotoxin detection, highlighting their increasing sensitivity and improving ability to detect multiple toxins simultaneously. TLC, once the standard method, has been largely surpassed by LC-MS/MS and HPLC-MS/MS due to their vastly superior performance. Both techniques provide highly sensitive, specific, and multi-toxin detection platforms, with LOD as low as $0.01\mu\text{g/kg}$ in multi-toxin settings [17, 18], and only $0.004\mu\text{g/kg}$ for single toxin detection [2]. These methods consistently exhibit high recovery rates³ (typically 70–130%), ensuring accurate quantification across diverse food matrices. The ability to simultaneously detect multiple mycotoxins, including aflatoxins, deoxynivalenol, fumonisins, zearalenone, and ochratoxin, makes these techniques the gold standard for regulatory and reference measurements [64].

Despite technological advancements, chromatographic techniques remain labor-intensive, requiring extensive sample preparation—such as extraction and cleanup—to eliminate matrix interferences. They also necessitate highly skilled personnel and advanced laboratory infrastructure, making them less practical for routine monitoring. Furthermore, although methods like LC-MS/MS and HPLC-MS/MS deliver exceptional sensitivity, their long turnaround times (often ranging from several hours to days) and high operational costs hinder their accessibility, particularly in resource-constrained settings. This reliance on costly instrumentation and complex workflows underscores the challenges of deploying chromatographic techniques beyond specialized laboratories.

³Recovery rates above 100% may result from matrix effects that enhance the analytical signal, leading to a measured concentration greater than the spiked amount. This phenomenon, while seemingly anomalous, is acceptable if it is reproducible and properly corrected during method validation.

Chromatographic techniques								
Matrix	Target(s)	Method	Recovery [%]	LOD [$\mu g/kg$]	LOQ [$\mu g/kg$]	LR [$\mu g/kg$]	Year	Ref.
Almonds, cashew nuts, chestnuts, hazelnuts, pistachios	AFB ₁ , AFB ₂ , AFG ₁ , AFG ₂	TLC	-	-	5	-	1993	[62, 2]
Barley, maize and wheat flour	ZEA	TLC	91 – 111	100	-	-	2005	[65]
Corn	AFB ₁	TLC	82 – 84	2.00 – 5.00	-	-	2007	[66]
Corn-based food products	AFB ₁ , AFB ₂ , AFG ₁ , AFG ₂	TLC	-	-	2	-	2007	[61, 2]
Brazil nuts	AFB ₁ , AFB ₂ , AFG ₁ , AFG ₂	TLC	-	-	2	-	2013	[63]
Wheat, maize	NIV, DON, FUS-X, DAS, HT-2, T-2, NEO, ZON	LC MS-MS	-	0.2 – 3.3	-	HT-2: 10 – 250; DAS: 10 – 500	2009	[67]
Cereal-based samples	AFB ₁ , OTA, FB ₁ , DON, T2, HT-2, ZEA	LC/MS-MS	AFB ₁ : 61 – 88, OTA: 64 – 85, FB ₁ : 76 – 122, DON: 62 – 70, T-2: 61 – 75, HT-2: 62 – 79, ZEA: 64 – 78	AFB ₁ : 0.06, OTA: 0.4, FB ₁ : 8, DON: 20, T-2: 4, HT-2: 20, ZEA: 1.6	AFB ₁ : 0.13, OTA: 0.8, FB ₁ : 16, DON: 20, T-2: 8, HT-2: 20, ZEA: 3.2	AFB ₁ : 0.06 – 1.6, OTA: 0.4 – 10.0, FB ₁ : 8 – 200, DON: 20 – 500, T-2: 4 – 100, HT-2: 20 – 500, ZEA: 1.6 – 40	2017	[68]
Cereal grain	AFB ₁ , AFB ₂ , AFG ₁ , AFG ₂ , NIV, DON, FB ₁ , FB ₂ , T-2, HT-2, ZEA, OTA	LC/MS-MS	73.9 - 133.0	0.1 - 18.1	0.4 - 54.8	1.3 - 53	2017	[69]
Corn	AFB ₁ , OTA, ZEA, FB ₁	LC-MS/MS	79.2 – 113.4	[$\mu g/L$] AFB ₁ : 0.21, ZEA: 0.19, FB ₁ : 0.09, OTA: 0.24	-	[$\mu g/L$] AFB ₁ : 0.5 – 55.7, ZEA: 0.5 – 127.1, FB ₁ : 0.22 – 31.36, OTA: 0.56 – 92.57	2018	[70]

Matrix	Target(s)	Method	Recovery [%]	LOD [$\mu\text{g/kg}$]	LOQ [$\mu\text{g/kg}$]	LR [$\mu\text{g/kg}$]	Year	Ref.
Grain	AFB ₁ , AFB ₂ , AFG ₁ , AFG ₂ , OTA, DON, ZEA, T-2	LC-MS/MS	AFB ₁ : 85.3, AFB ₂ : 82.8, AFG ₁ : 80.9, AFG ₂ : 102, OTA: 75.7, DON: 80.3, ZEA: 73.7, T-2: 56.5	AFB ₁ : 0.1, AFB ₂ : 0.1, AFG ₁ : 0.1, AFG ₂ : 0.3, OTA: 0.5, DON: 6.4, ZEA: 1.5, T-2: 0.6	AFB ₁ : 0.3, AFB ₂ : 0.4, AFG ₁ : 0.4, AFG ₂ : 0.9, OTA: 1.6, DON: 21.3, ZEA: 4.8, T-2: 1.8	AFB ₁ : 0.5 – 13.1, AFB ₂ : 0.5 – 12.7, AFG ₁ : 0.5 – 13.1, AFG ₂ : 0.5 – 15.0, OTA: 0.8 – 21.9, DON: 105 – 2889, ZEA: 10.7 – 294, T-2: 5.6 – 154	2020	[71]
Walnut kernel	AFs	LC/MS-MS	-	0.004 – 0.013	0.012 – 0.042	-	2020	[2]
Maize	AFB ₁ , Beauvericin, Citrinin, Moniliformin, Sterigmatocystin, ZON	LC-MS/MS	52.1 - 100	240	-	-	2021	[72]
Peanut, maize, wheat	AFB ₁ , AFB ₂ , AFG ₁ , AFG ₂ , FB ₁ , T-2	HPLC-MS/MS	84.5 – 112.7	0.01 – 0.05	0.04 – 0.16	0.05 – 80	2022	[17]
Maize, wheat, watermelon	AFB ₁ , AFB ₂ , AFG ₁ , AFG ₂ , OTA, OTB, T-2	HPLC-MS/MS	83.5 – 108.5	0.02 – 0.06	0.08 – 0.20	0.2 – 100	2023	[18]

Table 2.4: The performance of chromatographic techniques for mycotoxin detection on corn and wheat over time, including Thin-Layer Chromatography (TLC), Liquid Chromatography with tandem Mass Spectrometry (LC-MS/MS), and high-performance Liquid Chromatography with tandem Mass Spectrometry (HPLC-MS/MS). The table lists the tested mycotoxins (Target(s)) and the substrate (matrix), along with performance metrics such as the recovery, Limit Of Detection (LOD), Limit Of Quantification (LOQ), and Linear Region (LR).

2.3.2 Immunological Techniques

In addition to chromatographic methods, rapid screening techniques—primarily based on immunological approaches—are widely employed for mycotoxin detection and monitoring. These platforms utilize highly specific recognition elements, such as antibodies, to selectively bind target mycotoxins, enabling accurate detection. This specificity allows immunological methods to bypass extensive sample cleanup procedures, thereby streamlining the detection process and reducing analysis costs [20, 21]. However, this same specificity also limits these techniques to single-mycotoxin detection unless multiple recognition elements are incorporated to enable multiplexing.

2.3.2.1 State-of-the-art

Adhuc 2025, a diverse array of immunological screening techniques has been developed for mycotoxin detection, including Lateral Flow Devices (LFD), Enzyme-Linked Immuno-Sorbent Assays (ELISA), and Lateral Flow Immuno-Assays (LFIA). In addition, advanced optical detection methods—such as Surface-Enhanced Raman Spectroscopy (SERS) and Fluorescence Resonance Energy Transfer (FRET)—have been integrated with highly specific recognition elements to target mycotoxins. This combined approach enables sensitive and selective detection across various sample matrices while eliminating the need for extensive cleanup procedures, allowing for direct analysis of extracted solutions and facilitating rapid detection.

However, as mentioned before, the high specificity of the recognition elements often hinders the simultaneous detection of multiple mycotoxins and residual matrix effects further limit performance at ultra-low concentrations [22]. Moreover, these methods typically yield only qualitative or semi-quantitative results—indicating whether a specific mycotoxin is present above a certain threshold rather than providing its precise concentration [73]. This restricts their utility as reference or regulatory tools, relegating them primarily to initial, cost-effective screening. When positive results are obtained, more accurate chromatographic techniques are required to generate conclusive data necessary for legal or regulatory actions [73].

To evaluate the current day performance and limitations of these immunoassay-based technologies, a state-of-the-art review was conducted, with a particular focus on studies examining corn and wheat matrices. The results, summarized in Table 2.6, highlight key performance metrics for multi-mycotoxin detection, including recovery rates, Limits Of Detection (LOD) and Quantification (LOQ), and Linear detection Ranges (LR). The table also includes the sample type, target mycotoxins, detection method, and publication year to illustrate technological advancements and trends over time.

Among the various immunoassay-based techniques, ELISA, established in the late 1970s, remains one of the most widely used screening methods due to its ease of use, cost-effectiveness, and rapid turnaround time [2]. ELISA operates based on antigen–antibody interactions, producing a colorimetric response measurable by spectrophotometry. Studies show that ELISA can achieve high recovery rates (typically 80–125%), indicating its reliability for mycotoxin detection. However, despite its broad applicability, ELISA methods are often limited in their ability to detect multiple mycotoxins simultaneously, as a consequence of the highly specific antigen–antibody interactions [20, 21].

The sensitivity of ELISA varies significantly depending on the target mycotoxin. Single-mycotoxin detection using ELISA has been reported with LODs as low as $0.02\mu\text{g}/\text{kg}$, making it suitable for regulatory applications in some cases [21]. However, for multi-mycotoxin detection, LODs are

generally higher, often in the range of 0.1–5 μ g/kg, limiting its use for trace-level detection [20]. Additionally, ELISA is prone to matrix interference, which can affect accuracy unless additional cleanup steps are performed.

LFIA has emerged as a fast alternative to ELISA, enabling on-site, real-time mycotoxin detection without the need for specialized laboratory equipment. LFIA offers LOD values as low as 0.04 μ g/kg for AFB₁ in some studies, making it highly competitive for rapid screening [74]. However, similar to ELISA, LFIA struggles with multi-toxin detection due to cross-reactivity limitations and often requires separate assays for different mycotoxins [75, 76]. While LFIA is particularly useful for field applications and preliminary screening, its semi-quantitative nature prevents it from being used as a regulatory tool.

More advanced integrated approaches combine optical detection techniques with mycotoxin-specific binding agents to enhance both sensitivity and specificity. For example, recent work has demonstrated that SERS-based detection can achieve limits of detection below 0.001 μ g/kg [24], far below legal limits. Moreover, these techniques can be scaled by incorporating various binding agents, enabling multiplexed detection of multiple mycotoxins [77], which makes this approach highly competitive with chromatographic techniques in terms of sensitivity, speed, and multiplexing capability. Similarly, FRET-based assays have achieved LODs as low as 0.014 μ g/kg for aflatoxins and 0.027 μ g/kg for OTA, demonstrating their potential for ultra-trace detection [25]. Despite these impressive detection capabilities, both SERS and FRET rely on expensive instrumentation and require highly trained personnel for mycotoxin extraction, limiting their practicality for in-line and in-field screening applications, but offering a pathway forward for replacing cumbersome chromatographic methods for in-lab settings.

Overall, immunological techniques offer a valuable alternative to chromatographic methods for mycotoxin screening, especially in low-resource settings and for rapid monitoring. Methods such as ELISA and LFIA are simple, affordable, and user-friendly; however, they often lack the sensitivity and specificity required for conclusive regulatory analysis. In contrast, advanced optical methods like SERS and FRET significantly lower detection limits and enable multiplexing. Nonetheless, their high cost, low throughput, and reliance on mycotoxin extraction prevent their application for industry-scale, batch-wide monitoring.

Immunological techniques								
Matrix	Target(s)	Method	Recovery [%]	LOD [$\mu\text{g/kg}$]	LOQ [$\mu\text{g/kg}$]	LR [$\mu\text{g/kg}$]	Year	Ref.
Corn	FB ₁ , FB ₂	ELISA	101.7 – 112.06	[$\mu\text{g/L}$] 5	-	[$\mu\text{g/L}$] 10 – 1000	2010	[78]
Maize	ZEA	ELISA	83 – 93	0.02	-	0 -	2014	[21]
Corn	AFB ₁ , OTA, ZEA	ELISA	83.36 – 125.28	[$\mu\text{g/L}$] AFB ₁ : 0.112, OTA: 0.319, ZEA: 0.377.	-	-	2020	[20]
Wheat, maize	AFB ₁ , ZEA, DON	LFIA	80 - 122	AFB ₁ : 0.05, ZEA: 1, DON: 3	-	-	2014	[76]
Maize	FB ₁ , DON	LFIA	FB ₁ : 89 – 110, DON: 83 – 105	[$\mu\text{g/L}$] FB ₁ : 2.0, DON: 40	FB ₁ : 20, DON: 400	[$\mu\text{g/L}$] FB ₁ : 0 – 10, DON: 0 – 160	2015	[75]
Cereal	FB ₁ , ZEA, T-2, DON, AFB ₁	LFIA	84.0 – 110.0	AFB ₁ : 2.5, ZEA: 0.5, DON: 0.5, T-2: 2.5, FB ₁ : 0.5	AFB ₁ : 0.27, ZEA: 0.24, DON: 0.9, T-2: 0.32, FB ₁ : 0.59	0.5 – 10 to 2.0 – 50	2020	[79]
Corn	FB ₁ , ZEA, OTA, AFB ₁	LFIA	82.36 - 116.23	[$\mu\text{g/L}$] FB ₁ : 3.27, ZEA: 0.70, OTA: 0.10, AFB ₁ : 0.06	-	[$\mu\text{g/L}$] FB ₁ : 4 - 80, ZEA: 0.8 – 40, OTA: 0.2 – 2, AFB ₁ : 0.1 – 1.25	2020	[80]
Cereal	AFB ₁ , T-2, ZEA	LFIA	82.1 – 107.5	AFB ₁ : 0.04, T-2: 0.40, ZEA: 1.21	AFB ₁ : 0.09, T-2: 0.99, ZEA: 1.56	AFB ₁ : 0 - 1, T-2: 0 - 15, ZEA: 0 - 40	2022	[74]
Corn/wheat	ZEA, OTA	LFIA/SERS	ZEA: 96.9 – 107.4, OTA: 96.0 – 110.7	[$\mu\text{g/L}$] ZEA: 0.054, OTA: 0.018	-	[$\mu\text{g/L}$] ZEA: 0.05 – 500, OTA: 0.01 – 100	2021	[81]

Matrix	Target(s)	Method	Recovery [%]	LOD [$\mu\text{g/kg}$]	LOQ [$\mu\text{g/kg}$]	LR [$\mu\text{g/kg}$]	Year	Ref.
Corn	AFB ₁ , ZEA, DON, FB ₁ , OTA, T-2	LFIA/SERS	78.9 - 106.2	[ng/L] AFB ₁ : 0.96, ZEA: 6.2, FB ₁ : 0.26, DON: 0.11, OTA: 15.7, T-2: 8.6	-	[$\mu\text{g/L}$] AFB ₁ : 0.0014 – 0.33, ZEA: 0.015 – 3.7, FB ₁ : 0.41 – 100, DON: 0.14 – 33.3, OTA: 0.027 – 6.7, T-2: 0.014 – 3.3	2020	[77]
Corn/rice/wheat	AFB ₁ , ZEA, OTA	SERS	83.8 - 108.1	AFB ₁ : 0.061/0.064/0.066, ZEA: 0.53/0.55/0.57, OTA: 0.27/0.26/0.29	-	[$\mu\text{g/L}$] AFB ₁ 0.001 – 1, ZEA: 0.01 – 10, OTA: 0.005 – 5	2022	[23]
Maize	ZEA, OTA, AFB ₁	SERS	ZEA: 93.12 – 102.3, AFB ₁ : 70.7 – 112.1, OTA: 85.4 – 92.6	[pg/L] ZEA: 0.159, AFB ₁ : 1.56, OTA: 2.015	[pg/L] AFB ₁ : 5.20, ZEA: 0.53, OTA: 6.72	1.5 pM - 10 nM	2022	[24]
Corn, red wine, wheat, peanut	AFB ₁ , OTA	FRET	AFB ₁ : 94.1 – 111.7, OTA: 98.7 – 112.2	[$\mu\text{g/L}$] AFB ₁ : 0.014, OTA: 0.027	-	[$\mu\text{g/L}$] AFB ₁ : 0.1 – 100, OTA: 0.1 – 100	2023	[25]

Table 2.6: Performance summary of immunological techniques for mycotoxin detection on corn over time—including LFD, ELISA, LFIA, SERS, and FRET. The table lists the tested mycotoxins (Target(s)) and substrate (matrix), along with key metrics like the recovery, Limit Of Detection (LOD), Limit Of Quantification (LOQ), and Linear Region (LR), as well as publication year and reference.

2.3.3 Biosensors

In addition to traditional chromatographic and immunological methods, biosensors have emerged as a promising alternative for mycotoxin detection. Similar to ELISA and lateral flow devices, biosensors offer rapid, user-friendly, and cost-effective analysis. Moreover, they deliver superior reproducibility, enhanced sensitivity, and improved accuracy, making them highly competitive for on-site testing and real-time monitoring [64].

Biosensors typically comprise three key components: a biological recognition element, a transducer, and a signal processing system [64]. The recognition element interacts selectively with the target mycotoxins and may include antibodies, enzymes, or nucleic acids, while advanced materials such as aptamers, molecularly imprinted polymers, and recombinant antibodies may further enhance specificity [64]. Upon binding to the target, this interaction generates a signal that is converted into a measurable output by the transducer. Different biosensor types utilize distinct transduction mechanisms: optical biosensors detect changes in light properties, such as surface plasmon resonance and fluorescence; piezoelectric sensors monitor mechanical variations, such as mass shifts in quartz crystal microbalance systems; and electrochemical sensors measure electrical changes, including impedance, potential, or current variations [64]. The signal processing system then amplifies and translates these responses into an interpretable format, enabling rapid and accurate detection of mycotoxins for real-time decision-making.

2.3.3.1 State-of-the-art

Biosensors represent a rapidly developing approach to mycotoxin detection, offering significant potential due to their high sensitivity, rapid response time, and suitability for on-site applications. Despite their relatively recent emergence in food safety analysis, some biosensor platforms have already demonstrated remarkable detection limits, reaching as low as 0.1 ng/L for ZEA [82, 83]. This level of sensitivity highlights their ability to detect mycotoxins even at ultra-trace concentrations, making them a powerful alternative to traditional chromatographic and immunological methods. The state-of-the-art performance of biosensors for mycotoxin detection in maize and related matrices is summarized in Table 2.7, which outlines key performance metrics, including recovery rates, LOD and LOQ, while also tracking the evolution of biosensor technologies over time.

A notable limitation of current biosensors is their restricted ability to detect multiple mycotoxins simultaneously [2]. Most biosensor platforms are developed for single-analyte detection or for detecting only a limited number of mycotoxins, which limits their application in multi-toxin screening scenarios. This contrasts with liquid chromatography-mass spectrometry, which can analyze a broad spectrum of mycotoxins within a single run. As a result, biosensors remain primarily focused on targeted detection rather than comprehensive screening.

Furthermore, while biosensors offer a simplified analytical workflow compared to chromatographic techniques, they still require sample preparation steps, such as extraction and sometimes cleanup, to minimize matrix interferences and enhance detection accuracy [2]. These additional processing steps increase complexity and demand skilled personnel, particularly in the case of electrochemical and optical biosensors, which rely on precise calibration and interpretation of signal outputs. Such constraints limit their widespread deployment in non-specialized settings.

Despite these challenges, biosensors continue to evolve, with ongoing advancements aimed at improving their multiplexing capabilities, automation, and ease of use.

Bio sensors							
Matrix	Target(s)	Method	Recovery [%]	LOQ	LR	Year	Ref.
Corn powder	AFB ₁	Electrochemical immunosensor based on single-walled carbon nanotubes/chitosan (SWNTs/CS)	94.5 - 103.7	[ng/L] 3.5	[ng/L] 10 - 100	2016	[84]
Corn	OTA, AFB ₁ , DON	Multiplexed capillary microfluidic immunoassay	-	[μg/L] OTA: 40, AFB ₁ : 0.1 - 0.2, DON: 10	[μg/L] OTA: 10 - 100, AFB ₁ : 0.1 - 1, DON: 10 - 100	2018	[85]
Corn	AFB ₁	Electrochemical impedance spectroscopy	94.5 - 103.7	[μg/L] 0.05	[μg/L] 0.1 - 100	2018	[82]
Maize	ZEA	Aptamer, transducer: amperometric (CV/DPV)	-	[ng/L] 0.17	[ng/L] 0.5 - 500	2019	[86, 2]
Maize	ZEA	Aptamer, transducer: potentiometric (CV/DPV)	91.6 - 104.4	[ng/L] 0.105	[ng/L] 0.5 to 5000	2019	[87]
Corn	FB ₁ , DON	Electrochemical immunosensor with gold nanoparticles (AuNPs) and polydimethylsiloxane (PDMS) microfluidics	FB ₁ : 88.6 - 95.8, DON: 82.4 - 95.8	[ng/L] FB ₁ : 97, DON: 35	[μg/kg] AFB ₁ : 0.3 - 140, OTA: 0.2 - 60	2019	[88]
Corn powder	AFB ₁ , ZEA	Two-plex non-faradaic electrochemical impedance biosensor with ZnO	>80%	[ng/mL] AFB ₁ : 0.005, ZEA: 0.05	[ng/mL] AFB ₁ : 0.01-9.151, ZEA: 0.1-25.6	2025	[89]

Table 2.7: The performance of biosensors for mycotoxin detection over time. This table lists the tested mycotoxins (Target(s)) and the substrate (matrix), along with performance metrics such as the recovery, Limit Of Quantification (LOQ), and Linear Region and (LR).

2.4 Need For Non-Invasive Alternatives

Thus far, we have reviewed a diverse array of detection techniques for mycotoxin analysis, each offering unique advantages and limitations. As of 2025, chromatographic methods remain the gold-standard, providing unmatched sensitivity and specificity for detecting multiple mycotoxins, even at ultra-trace levels. While essential for regulatory compliance, these techniques are hindered by their high costs, labor-intensive protocols, and complex sample preparation requirements. In contrast, immunological screening platforms—such as Enzyme-Linked Immuno-Sorbent Assays (ELISA)—offer rapid, cost-effective in-field monitoring. However, these methods are generally limited to qualitative results and are less effective at simultaneous multi-mycotoxin detection. Advanced immunological approaches, like biosensors, offer promising solutions for multiplexing, and strong improvements to sensitivity, though they still rely on mycotoxin extraction.

Furthermore, a fundamental limitation shared by all these methods is their inherently destructive nature. Every approach necessitates the physical removal of sample material, which is then chemically processed to extract mycotoxins prior to analysis, this prevents their use for in-line applications (e.g., conveyor belt integration), and confines testing to only a small, representative subset of a larger batch, making it impossible to assess contamination on an individual sample basis [27]. As a result, these methods only measure the average contamination within the sampled subgroup, assuming it representative of the entire batch. This averaging can mask the influence of localized hotspots or outliers [26]. Given that mycotoxins are often heterogeneously distributed—tending to cluster in localized portions of a batch rather than being uniformly dispersed—bulk sampling frequently fails to accurately represent true contamination levels across the food supply chain [28]. Such discrepancies pose significant risks: they can result in the unnecessary disposal of largely uncontaminated food, while simultaneously allowing contaminated products to enter the supply chain due to insufficient sampling coverage [16].

The critical importance of effective mycotoxin monitoring at all stages of production is further underscored by data from the Rapid Alert System for Food and Feed (RASFF)—a regulatory body of the European Union tasked with ensuring the safety of all food and feed products sold within the union. In 2022 alone, the most recent year for which data is available, a total of 4361 recalls and border rejections related to food safety were reported [90]. Mycotoxins—particularly aflatoxins (AFLA)—were the third most common cause, underscoring the substantial economic burden mycotoxins impose on the agrifood sector.

To reduce such occurrences, there is a clear need for improved mycotoxin screening and monitoring technologies—solutions that are **non-destructive**, **reagent-free**, **in-situ**, and capable of **real-time**, **high-throughput** screening. Such a method would allow for the direct assessment of every individual food item rather than relying on small, representative subsamples. In an ideal scenario, every maize/wheat kernel in a batch could be individually inspected, thereby improving food safety and reducing unnecessary waste by selectively removing contaminated samples from the supply chain without discarding entire shipments.

The only class of material identification techniques with the potential to achieve these objectives is optical spectroscopy. As discussed in the introduction, Section 1.2.1, spectroscopic methods unlike chemical or immunological approaches do not require analyte extraction or extensive sample preparation, making them inherently well-suited for in-line and in-field applications. By leveraging the interaction of light with matter, these techniques can provide rapid, in-situ, non-invasive compositional characterization of food products, enabling real-time decision-making without disrupting processing workflows.

2.5 Optical Spectroscopy For Mycotoxin Detection; State-of-the-art

Spectroscopy has played a pivotal role in microorganism detection for decades. Its origins date back to the 1950s, when near-infrared spectroscopy (NIRS) was first explored for this application [73]. In the 1980s, the technology was directly employed to detect fungal contamination in agricultural produce [91]. More recently, advanced techniques such as Raman spectroscopy have enabled the label-free, in-situ identification of fungal species, including *Fusarium* [92].

Beyond fungal detection, a variety of spectroscopic methods have proven effective in identifying mycotoxin contamination on agricultural products. It is important to note that these techniques do not always directly detect mycotoxins; rather, they often measure the impact of fungal metabolism on the food matrix during mycotoxin production, serving as an indirect proxy for contamination levels [31].

Reflection-based methods—including ultraviolet-visible-near infrared (UV-VIS-NIR) [26, 28, 93, 94, 95, 96, 97, 98], short-wave infrared (SWIR) [28, 93], mid-infrared (MIR) [27, 99], and Fourier Transform reflectance spectroscopy [51, 60, 100, 101]—along with Raman⁴ [27, 92, 93, 100], fluorescence spectroscopy [16, 28, 30, 93, 102], two-photon fluorescence spectroscopy [34], and Hyper-Spectral Imaging (HSI) [27, 28, 103, 104, 105, 106, 107], have all been explored for this purpose [108].

These approaches have been applied to a range of commodities, such as pistachio nuts [30], wheat [27], corn [26, 28, 93], rice [109], and peanuts [110]. Additionally, spectroscopic techniques have been investigated for detecting mycotoxins in processed foods, including apple juice [111], milk [27, 112], and cheese [83], albeit with varying degrees of success.

While optical spectroscopy presents a promising avenue for mycotoxin detection, it is not without challenges. As discussed in Section 1.2.1, the spectral signatures of mycotoxins are often weak and masked by the dominant spectral response of the surrounding food matrix. Additionally, inherent variability in agricultural products—arising from genetic differences, environmental growth conditions, and storage practices—further complicates the reliable identification of mycotoxin-specific features. Despite these obstacles, recent advancements in chemometric data analysis and machine learning have enabled the development of robust classification and regression models capable of extracting relevant information from complex spectral datasets. These innovations have significantly advanced the feasibility of real-time, non-invasive mycotoxin detection and monitoring.

To assess the current state-of-the-art in optical spectroscopy-based mycotoxin detection on both corn and wheat, a comprehensive literature review was conducted. The results are presented in Tables 3.09 and 3.10, which compile studies utilizing classification and regression models, respectively. Both tables provide a structured overview of the spectroscopic platforms employed, the wavelength ranges investigated, the types of models applied, and the performance metrics reported. Furthermore, they distinguish between studies conducted on whole versus ground samples, identify the specific target mycotoxins, and include the year of publication along with the corresponding references.

⁴Note that in this context, Raman spectroscopy is used without recognition elements for direct mycotoxin detection within the sample matrix, unlike the integrated SERS platform described in Section 2.3.3, which employs such elements.

Spectroscopic techniques for in-situ mycotoxin detection							
Technique	Wavelength [nm]	Sample matrix	Target	Classification model	Performance metrics	Year	Ref.
Fluorescence	Excitation: 375 Detection: 200-1000	Rice kernels and ground rice	AFLA, fungi	PCA	Accuracy: 97%, FPR: 2%	2022	[102]
Fluorescence	Excitation: 365 Detection: 440-720	Ground maize	AFLA	QSVM	Threshold: 10ppb Accuracy: 95.7%	2022	[93]
Fluorescence	Excitation: 365 Detection: 400-650	Maize kernels (naturally contaminated)	AFLA	-	Threshold: 0.6 μ g/kg	2023	[16]
Fluorescence	Excitation: 385 Detection: 400-850	Pistachio nuts	AFLA defects	Feature selection: SFS, Classifier: QDA	Accuracy: 99.2%, FPR: 0.8%	2023	[30]
Fluorescence	Excitation: 385 Detection: 440-720	Maize kernel	AFLA, FUM	SVM	Accuracy: 89.1%	2023	[28]
Fluorescence HSI	Detection: 400-600	Maize kernels (naturally contaminated)	AFLA	SAM (spectral angle mapper)	Threshold: 20ppb, Accuracy: 86%, FPR: 15%, Threshold: 100ppb, Accuracy: 88%, FPR: 16%	2010	[113]
Fluorescence HSI	Detection: 400-600	Maize kernels (artificially infected in the field)	AFLA	ANN	Threshold: 20ppb, Accuracy: 84%, Threshold: 100ppb, Accuracy: 0.92%	2010	[106]
Fluorescence HSI	Detection: 400-600	Maize kernels	AFLA	Feature selection: GA, Classifier: SVM	Threshold: 20ppb, Accuracy 87.7%, Threshold: 100ppb, Accuracy: 90.5%	2012	[105]
Fluorescence HSI	Detection: 400-600	Maize kernels	AFLA	ANN	Threshold: 20ppb, Accuracy: 87%, Threshold: 100ppb, Accuracy: 88%	2013	[114]
Fluorescence HSI	Detection: 400-700	Maize kernels	AFLA	LDA	Threshold: 20ppb, Accuracy: 78.9% - 77.2%, Threshold: 100ppb, Accuracy: 94.4% - 91.7%	2013	[103]
Fluorescence HSI	Detection: 398.77 - 700.82	Maize kernels	AFLA	LS-SVM, KNN	LS-SVM, accuracy: 86.67%, KNN, accuracy: 93.33%	2016	[104]
Fluorescence HSI	405, 435, 450, 470, 505, 525, 570, 590, 630, 645, 660, 700, 780, 850, 870, 890, 910, 940, 970	Maize kernels	ZEA	Feature selection: GA, Classifier: NN	Accuracy: 93.33%	2022	[107]

Technique	Wavelength [nm]	Sample matrix	Target	Classification model	Performance metrics	Year	Ref.
Raman	Excitation: 785 Detection: 400-1200cm ⁻¹	Maize kernels (naturally contaminated)	AFLA	LDA, KNN, PLS-DA	KNN, accuracy: 91.4%, LDA, accuracy: 91.4%, PLS-DA, accuracy: 53.4%	2015	[100]
Raman	Excitation: 785 Detection: 375-2814cm ⁻¹	Ground maize	AFLA	LSVM	Threshold: 10ppb, Accuracy: 87%	2022	[93]
Reflectance & Transmittance	Transmittance: 500 - 950, Reflectance: 550 - 1700	Maize kernels	AFLA	PLS-DA	High: 100ppb, Low: 10ppb, Accuracy: 95%	2001	[96]
Reflectance	500–1700	Maize kernels	AFLA	ANN	Threshold: 100ppb, Accuracy: 99%	2004	[97]
Reflectance	(VIS-NIR)	Maize kernels	DON	-	High: 18184ppb, Low: 654ppb	2016	[26]
Reflectance	4000–575cm ⁻¹	Ground maize	DON	DT	Threshold: 1750ppb, Accuracy: 79%, Threshold: 500ppb, Accuracy: 85%	2016	[99]
Reflectance	470, 527, 624, 850, 880, 910, 940, 970, 1040, 1070, 1200, 1300, 1450, 1550	Maize kernels	AFLA, FUM	LDA	threshold(s): AFLA: 10ppb, FUM: 1000ppb, Sensitivity: 77%, Specificity: 83%	2017	[98]
Reflectance	304 - 1086	Maize kernels	AFLA	RF	Sensitivity: 85.7%, Specificity: 97.3%	2019	[95]
Reflectance	304 - 1085	Maize kernels	AFLA	PLS-DA	threshold: 20ppb, Accuracy: 71%, Threshold: 100ppb, Accuracy: 82%, Threshold: 1000ppb, Accuracy: 92%	2022	[94]
Reflectance	420-1020	Ground maize	AFLA	QSVM	Threshold: 10ppb, Accuracy: 82.6%	2022	[93]
Reflectance	900-2500	Ground maize	AFLA	LSVM	Threshold: 10ppb, Accuracy: 82.6 – 95.7%	2022	[93]
Reflectance	420-1020	Maize kernel	AFLA, FUM	SVM	Accuracy: 71.7%	2023	[28]

Technique	Wavelength [nm]	Sample matrix	Target	Classification model	Performance metrics	Year	Ref.
Reflectance	900-2500	Maize kernel	AFLA, FUM	SVM	Accuracy: 95.7%	2023	[28]
FT-Reflectance	900-2500	Maize kernels (natu- rally contaminated)	AFLA	LDA, KNN, PLS-DA	KNN, accuracy: 92.0%, LDA, accuracy: 62.0%, PSL-DA, accuracy: 100.0%	2015	[100]
FT-Reflectance	1000-2500	Maize kernels (natu- rally contaminated)	AFLA	LDA, KNN, PLS-DA	KNN, accuracy: 96.0%, LDA, accuracy: 79.2%, PSL-DA, accuracy: 60.0%	2015	[100]

Table 2.9: Overview of studies employing spectroscopic techniques and classification algorithms for mycotoxin detection in corn and related samples. This table summarizes investigations using fluorescence spectroscopy, reflectance/transmittance spectroscopy, Fourier Transform reflection spectroscopy (FT), fluorescence Hyper-Spectral Imaging (HSI), and Raman spectroscopy. Key parameters listed include the optical platform used (technique), wavelength range, sample matrix, target mycotoxin(s), and performance metrics (thresholds, classification accuracy, and False Positive Rate (FPR)). The table also details the classification models applied, such as Quadratic and Linear Support Vector Machines (QSVM, LSVM), Quadratic and Linear Discriminant Analysis (QDA, LDA), Principal Component Analysis (PCA), Multivariate Linear Regression-Discriminant Analysis (MLR-DA), Partial Least Squares-Discriminant Analysis (PLS-DA), K-Nearest Neighbors (KNN), Neural Networks (NN), Decision Trees (DT), and Random Forests (RF).

Regression Models for Mycotoxin Quantification Using Spectroscopic Techniques							
Technique	Wavelength [nm]	Sample matrix	Target	Regression model	Performance metrics	Year	Ref.
Transmittance	670-1100	Wheat kernels	DON	PLSR (12 PCs)	Range: 400 – 13000ppb, R: 0.984, SECV: 381µg/kg	2003	[115]
Reflectance	350-2500	Wheat kernels	DON	PLSR	Range: 0.1-100.75mg/kg, R^2 : 0.84, Intercept: -0.89 ± 3.61mg/kg	2010	[116]
Reflectance	650-2500, 2nm steps	Ground maize (natu- rally contaminated)	FUM	PLSR (17 and 21 PCs)	Range: 0.357-11845mg/kg, mean: 3.353mg/kg Calibration R: 0.995/0.998, Validation R: 0.908/0.909, SECV: 0.232/0.144	2013	[51]
Reflectance	(NIR)	Wheat kernels	DON	PLSR	Range: 1-11mg/kg, R^2 : 0.46	2014	[31, 117]
Reflectance	400–2500	Wheat (kernels & milled)	DON	PLSR	Range: 1-11 mg/kg, Median: 2.66mg/kg, Kernels: R^2 : 0.89, SECV: 612µg/kg, Milled: R^2 : 0.91, SECV: 578µg/kg	2018	[31, 118]
Reflectance	400–2500	Maize kernels	FUM	PLSR	Range: 125-24200µg/kg, R^2 : 0.991, SEP: 0.984µg/kg	2021	[119]
Reflection	450–1050	Maize kernels	AFLA	PLSR	Levels: 0,3,5,10,20,30,50µg/kg, R^2 : 0.99, SECV: 1.70µg/kg, RPD: 9.90, LOD: 0.60µg/kg, LOQ: 1.81µg/kg	2025	[120]
FT-Reflectance	650–2500	Maize kernels	AFLA	PLSR	SECV: 0.8, R^2 : 0.2	2009	[60]
FT-Reflectance	650–2500	Ground maize	FUM	PLSR	SECV: 0.433, R^2 : 0.964, SEP: 0.839mg/kg	2012	[101]

Technique	Wavelength [nm]	Sample matrix	Target	Regression model	Performance metrics	Year	Ref.
FT-Reflectance	650-2500	Ground maize	FUM	PLSR	SECV: 0.232, R^2 : 0.995, SEP: 0.933mg/kg	2013	[51]
FT-Reflectance	10000-4000cm ⁻¹	Ground wheat	DON	PLSR	Range: 0-16000μg/kg, SEP: 1977μg/kg	2014	[31, 121]
FT-Reflectance	12000-4000cm ⁻¹	Ground barley	DON	PLSR	Range: 0.016-13.0mg/kg, R: 0.875, SECV: 268μg/kg, SEP: 399μg/kg	2018	[31, 122]

Table 3.10: Overview of studies employing spectroscopic techniques and regression algorithms for mycotoxin detection in wheat and related samples. This table summarizes investigations utilizing reflectance/transmittance spectroscopy and Fourier Transform reflection spectroscopy (FT). Key parameters include the optical platform used (technique), wavelength range, sample matrix, target mycotoxin(s), and performance metrics such as contamination range, mean contamination, correlation coefficient (R -value), coefficient of determination (R^2), Standard Error of Cross-Validation (SECV), Standard Error of Prediction (SEP), Ratio of Performance to Deviation (RPD), Limit Of Detection (LOD), and Limit Of Quantification (LOQ). The table also details the regression models applied, specifically Partial Least Squares Regression (PLSR), along with the number of Principal Components (PCs) used.

This state-of-the-art analysis, Tables 3.09 and 3.10, allows us to examine both the spectroscopic modalities employed as well as the ML-based chemometric models used for data processing—components that together constitute the complete mycotoxin detection platform.

In the following sections, we will first explore the spectroscopic techniques, with a focus on identifying those best suited for the intended application of realizing an in-line or in-field detection systems for industrial use. Subsequently, attention will shift to the computational domain, where we assess the performance of various classification and regression models to determine the most effective modeling approaches.

2.5.1 Comparing Spectroscopic Platforms

To enable a direct comparison of the spectroscopic techniques reviewed in the state-of-the-art, a consolidated summary is presented in Table 2.13. This comparative table evaluates the relative advantages and limitations of fluorescence, HSI, FT-reflectance, UV-VIS-NIR reflectance, and Raman spectroscopy based on key practical and analytical criteria relevant to the development of in-line and in-field mycotoxin detection systems.

The criteria considered include cost-effectiveness, measurement speed, sensitivity, specificity, portability, equipment requirements, data dimensionality, and computational demand. Each of these aspects is critical to the real-world deployment of mycotoxin detection systems.

Cost-effectiveness plays a pivotal role in scalability—particularly for industrial sorting systems that scan individual food items or for portable, farm-level devices. Measurement speed is essential for real-time assessment in high-throughput environments, such as processing lines. Sensitivity reflects the platform’s ability to detect low-level contamination, while specificity indicates its ability to differentiate target toxins from background noise within the food matrix. Portability is a key enabler for field-based screening, eliminating the need for centralized laboratory infrastructure. Equipment requirements refer to the volume and complexity of hardware needed to operate the system, favoring technologies that are simpler and more compact. Data dimensionality and computational demand capture the complexity of the generated data and the processing resources needed—both of which directly impact latency and power consumption.

Aspect	Fluo	UV-VIS-NIR	HSI	FT	Raman
Cost-effectiveness	+	+	–	–	–
Measurement Speed	+	+	+	–	–
Sensitivity	±	±	±	+	++
Specificity	±	±	±	+	++
Portability	+	+	±	–	±
Equipment Requirement	+	+	–	–	–
Data Dimensionality	Low	Low	High	Moderate	Low
Computational Demand	Low	Low	High	Moderate	Moderate

Table 2.13: Comparison of spectroscopic techniques (fluorescence, reflectance, Hyper-Spectral Imaging (HSI), Fourier Transform (FT), and Raman spectroscopy) for mycotoxin detection based on practical and analytical aspects as derived from the literature study presented in Tables 3.09 and 3.10. “+” indicates a relative advantage, “–” indicates a limitation, and “±” indicates a context-dependent or moderate performance.

Based on the comparative analysis, fluorescence and reflectance⁵ spectroscopy stand out as the most promising technologies for industrial in-field and in-line mycotoxin detection. These methods strike an effective balance between analytical performance, measurement speed, and operational simplicity, while maintaining minimal instrumentation requirements. Their capacity for rapid, high-throughput analysis makes them particularly suitable for integration into automated in-line sorting systems for real-time quality control [102]. Simultaneously, their compact design and low equipment demands position them as strong candidates for the development of cost-effective, handheld devices for on-site monitoring [16].

By contrast, techniques such as FT-spectroscopy and Raman spectroscopy, while offering superior sensitivity and specificity, face significant challenges to industrial implementation. Raman spectroscopy, for example, is known for its exceptional chemical specificity, enabled by the detection of molecular vibrational modes through narrow, well-defined spectral peaks. This allows for precise differentiation between chemically similar analytes—an area where NIR spectroscopy, with its broader and overlapping absorption bands, often falls short. However, this level of spectral resolution demands sophisticated instrumentation, including high-stability excitation lasers and highly sensitive optical detectors. These requirements substantially increase system complexity and cost, hindering the scalability of Raman spectroscopy for widespread industrial deployment. Similarly, FT-spectroscopy offers high spectral resolution and broad wavelength coverage but typically relies on a scanning interferometric mechanism. The presence of this mechanical component not only extends acquisition times but also complicates integration into high-speed, in-line production environments. As a result, both FT and Raman spectroscopy are limited by slower acquisition speeds and elevated instrumentation costs, relegating their use to laboratory-based confirmatory testing rather than large-scale, real-time screening or cost-effective in-field monitoring.

Although HSI shows potential for in-line integration, its industrial applicability is hindered by high implementation costs and the generation of large, high-dimensional datasets. Successful deployment demands substantial computational infrastructure and access to extensive annotated datasets for model development and training [123]. These requirements collectively undermine the cost-effectiveness of HSI for real-time, in-line monitoring.

Given these considerations, this master’s thesis focuses exclusively on UV-VIS-NIR reflectance and fluorescence spectroscopy—two techniques that offer an optimal balance between analytical performance and practical feasibility, for both in-line and in-field applications.

2.5.2 Fundamentals And State-of-the-art Of Fluorescence And Reflectance Spectroscopy

Before examining the state-of-the-art performance of fluorescence- and reflectance-based platforms in more detail, it is crucial to first establish a foundational understanding of the spectroscopic principles underpinning these techniques. The following sections begin with a concise overview of both methods, outlining the operating principles and spectral characteristics. This is followed by a focused discussion on their application to mycotoxin detection, highlighting both conventional and advanced implementations.

⁵It is also worth noting that different regions of the reflectance spectrum—UV, VIS, and NIR—entail distinct trade-offs. Sensor cost typically increases with wavelength, making NIR instrumentation more expensive than systems operating in the UV or visible range. However, the NIR region often contains rich and diagnostically relevant spectral information, particularly for organic compounds, which enhances its value for accurate and reliable mycotoxin detection.

2.5.2.1 Reflectance Spectroscopy

When light encounters a material interface, its behavior is determined by three main processes: transmission, reflection, and absorption. Both transmission and reflection are elastic processes, preserving the wavelength of the incident light. In contrast, absorption converts photon energy into other forms, such as heat or chemical energy, or it may lead to re-emission at different wavelengths, as seen in fluorescence.

Reflection can be categorized into two distinct types: **specular** and **diffuse**. Specular reflection directs light in a single, well-defined direction, with the angle of reflection matching the angle of incidence. In contrast, diffuse reflection results from scattering due to microstructural irregularities—such as grain boundaries, impurities, or surface inhomogeneities—which cause light to be redirected in multiple directions. Moreover, materials selectively absorb specific wavelengths due to their unique electronic, vibrational, and rotational electron transitions. This selective absorption imprints a distinct spectral fingerprint on the reflected or transmitted light, enabling precise material identification.

Reflectance spectroscopy capitalizes on these principles by analyzing the wavelength-dependent fraction of light reflected from a material's surface, making it especially effective for highly scattering or opaque materials. In comparison, transmittance spectroscopy measures the light that passes through a sample, which is more suitable for transparent or semi-transparent media. Both techniques are powerful tools for material identification, with their effectiveness largely determined by the optical properties of the sample. Given that this study focuses on opaque corn and wheat samples, reflectance spectroscopy is the preferred approach.

Selecting the wavelength range for detection depends on the analytes under investigation and their specific electronic or vibrational transitions. The relevant regions of the electromagnetic spectrum are conventionally divided as follows: UV: 100–380nm, VIS: 380–780nm, and IR: 780–25000nm. The IR domain itself is further subdivided into SWIR: 780–1100nm, NIR: 780–2500nm, and MIR: 2500–25000nm.

Among these, the UV-VIS range is particularly effective for probing electronic transitions, making it ideal for detecting compounds with strong chromophoric features. In contrast, the NIR region is primarily sensitive to vibrational transitions, capturing overtones and combination bands of hydrogen-containing functional groups such as $-\text{CH}$, $-\text{NH}$, $-\text{OH}$, and $-\text{SH}$. These interactions provide deeper insight into molecular structure and composition [35]. The deeper IR region extends into rotational transitions, but this part of the spectrum is rarely used in industrial settings due to the high complexity and cost of the required instrumentation, rendering it impractical for routine applications.

By spanning a broad spectral range—from electronic transitions in the UV-VIS region to vibrational modes in the NIR—reflectance spectroscopy captures a comprehensive molecular fingerprint of a material. This versatility makes it an invaluable, non-destructive tool for material characterization, quality control, and real-time monitoring in agro-industrial applications [35].

2.5.2.1.1 State-of-the-art

Reflectance spectroscopy's proven effectiveness in material characterization has spurred its broad adoption across industrial and scientific fields. Its origins date back to the 19th century, when William Herschel first identified the NIR region of the spectrum [35]. Building on this discovery, Karl Norris later pioneered the application of NIR spectroscopy in agricultural analysis, demonstrating its capacity to non-destructively assess the protein content of wheat [35]. Thanks to its rapid and non-invasive nature, reflectance spectroscopy quickly became a critical tool in food production—initially serving early-stage quality control—and has since evolved to include

contaminant detection, ripeness monitoring, and, more recently, mycotoxin detection [34, 35].

In the context of mycotoxin detection, reflectance spectroscopy has been extensively applied to agricultural products such as maize and wheat. Measurements in the UV-VIS-NIR range have demonstrated high sensitivity and specificity; for example, Random Forest (RF) classification models achieved 85.7% sensitivity and 97.3% specificity for AFLA detection [95]. Similarly, reflectance data in the VIS-NIR region has been used to classify AFLA contaminated maize kernels into high (> 100 ppb) and low (< 10 ppb) categories using Partial Least Squares (PLS) analysis, with accuracies reaching up to 95% [96]. The short-wave infrared (SWIR) region has also proven effective, where Linear Support Vector Machines (LSVM) achieved over 95% classification accuracy for detecting AFLA—and AFLA combined with FUM—at a detection limit of 10ppb [28].

Furthermore, regression models have been successfully implemented for detecting AFLA and other toxins on corn kernels using the VIS-NIR region, with Partial Least Squares Regression (PLSR) models achieving an R^2 value of 0.99 and a Limit Of Detection (LOD) of $0.60\mu\text{g/kg}$ [120].

2.5.2.2 Fluorescence Spectroscopy

Fluorescence is a form of luminescence—a phenomenon in which certain molecules absorb photons at specific wavelengths, causing a transition to an electronically excited state [29]. Once excited, these molecules rapidly relax to a lower vibrational level through nonradiative processes (e.g., heat dissipation) before returning to the ground state via the emission of light at a longer wavelength (i.e., lower energy) [124].

Fluorescence specifically occurs when the excited state is a singlet state, in which the electron in the excited orbital remains spin-paired with the ground-state electron, exhibiting opposite spin orientations. According to quantum mechanics, the transition from a singlet excited state back to the ground state is spin-allowed, which enables rapid decay with a typical excited-state lifetime of approximately 10ns [29, 124]. The excess energy is released as a photon of lower energy (longer wavelength), conserving the total energy of the system.

The spectral components observed in a fluorescence spectrum are determined by the available vibronic transitions of the analyte. The specific transitions that can occur are governed by quantum mechanical selection rules, which are derived from transition moment integrals [125]. However, a detailed discussion of these principles extends beyond the scope of this work. The important point is that the specific composition of the analyte dictates a unique set of spectral lines that are characteristic of its chemical structure. This spectral signature serves as a fingerprint, enabling precise identification and differentiation of the analyte based on its fluorescence properties.

2.5.2.2.1 Notes On The Use Of Fluorescence

Effective fluorescence detection depends critically on selecting the optimal excitation wavelength(s) to maximize analyte (mycotoxins) excitation—and consequently emission—while minimizing background fluorescence from matrix constituents unrelated to mycotoxin presence. Achieving this balance is essential for enhancing sensitivity, specificity, and overall measurement reliability.

In literature, for example, excitation at 365nm has been successfully employed for AFLA detection on maize [34]. However, identifying the ideal excitation wavelength can vary significantly depending on both the specific mycotoxin and the sample matrix. As such, a portion of this

thesis is dedicated to experimentally determining the optimal excitation wavelengths for each relevant toxin–matrix combination.

To support the search for optimal excitation wavelengths, it is useful to consider the fluorescence characteristics of the different mycotoxins, previously outlined in Section 2.1 and summarized again below.

AFLA strongly absorbs light in the 200–250nm range and between 300 and 400nm, with a prominent absorption peak around 365nm, emission typically spans 400–500nm, peaking near 440nm, although the exact emission profile depends on the excitation wavelength employed [30, 38, 39]. In contrast, ZEA displays blue-green fluorescence under long-wavelength UV light and intense green fluorescence under short-wavelength UV light. In methanol, ZEA shows absorption maxima at 236, 274, and 316nm, with maximum fluorescence emission at 450nm when excited at 314nm [50]. OTA exhibits intrinsic fluorescence with broad emission peaks centered around 460nm and a shoulder at shorter wavelengths when excited at 330nm in aqueous solutions [44]. By contrast, mycotoxins such as FUM, DON, and T-2/HT-2 exhibit little to no intrinsic fluorescence, necessitating indirect detection through changes induced in the surrounding matrix [49]. Importantly, the sample matrix itself contributes to background fluorescence, primarily due to naturally occurring proteins such as tryptophan, tyrosine, and phenylalanine in corn and wheat, which exhibit strong absorption in the 200–300nm range [34]. Additionally, chlorophyll present in food products—particularly in wheat—can produce significant fluorescence signals. While these endogenous emissions typically add to the background and should ideally be minimized to enhance analyte specificity, they may also contain diagnostically relevant information. In particular, matrix-related fluorescence alterations caused by the mycotoxin metabolism⁶ may serve as indirect indicators of mycotoxin contamination.

From these observations, it becomes evident that excitation in the 200–300nm range primarily stimulates matrix components, which may be advantageous for detecting matrix alterations associated with fungal metabolic activity. In contrast, operating in the 300–400nm range is more likely to selectively excite the mycotoxins themselves while suppressing background fluorescence from the matrix, thereby enhancing detection specificity for toxins specific markers.

Beyond conventional single-photon excitation, multi-photon excitation techniques—such as two-photon fluorescence—may also offer promising avenues for improving sensitivity and selectivity. These approaches are discussed in greater detail in Section 2.5.2.3.

2.5.2.2.2 State-of-the-art

Over the past century, fluorescence spectroscopy has evolved into an indispensable analytical tool across a wide range of scientific disciplines, including biology, chemistry, and materials science [124, 126]. In the last two decades, its relevance in food analysis has grown substantially, demonstrating high efficacy in characterizing chemical constituents, detecting contaminants, and verifying product authenticity [127]. Among its most recent applications, fluorescence spectroscopy has shown considerable promise for mycotoxin detection, enabling reliable in-situ mycotoxin analysis even in exceedingly complex food matrices [16, 28, 30, 93, 102, 105, 106, 113].

Fluorescence-based mycotoxin detection was first demonstrated in the 1960s when aflatoxins in peanut-based feed—implicated in the *Turkey X* Disease outbreak in England—were identified [1]. This breakthrough led to the discovery of other naturally fluorescent mycotoxins, such as

⁶Mycotoxin metabolism refers to the specific metabolic state of fungi during the production of distinct mycotoxins, and its effect on the matrix composition.

OTA, and established fluorescence as a widely adopted method for mycotoxin analysis.

Today, fluorescence spectroscopy is valued for its non-destructive, non-invasive, in-situ, rapid, and highly sensitive approach to food safety monitoring. However, detecting mycotoxins at ultra-low concentrations, as required by strict food safety regulations (Tables 2.1, and 2.2) remains a significant challenge, especially in a multi-toxin environment.

2.5.2.3 Upconversion Fluorescence For Mycotoxin Detection

While conventional fluorescence involves the absorption of a single quanta of light to excite an electron from the ground state to a higher energy level, upconversion fluorescence relies on the absorption of multiple lower-energy photons to achieve the same excitation [128]. This process, known as multi-photon excitation, is most commonly observed in Two-Photon Fluorescence (TPF) and Three-Photon Fluorescence (3PF).

In TPF, an electron in the ground state simultaneously absorbs two lower-energy photons, typically in the near-infrared region, to reach an excited state that would normally require a single, higher-energy photon in the visible or ultraviolet range. This process arises from nonlinear optical interactions, where the probability of excitation is proportional to the square of the excitation intensity [128]. As a result, high-intensity (pulsed) lasers are typically required to achieve efficient excitation. Once the molecule reaches the excited state, it undergoes the same fluorescence emission process as in conventional fluorescence, now emitting a shorter-wavelength photon than the excitation pump.

3PF extends this concept by requiring three simultaneous lower-energy photons to reach the excited state. The excitation probability in 3PF is even more dependent on laser irradiance, scaling proportionally to the cube of the excitation power [128]. As a result, 3PF requires even higher laser irradiance but provides even larger depth penetration and selectivity, making it valuable for advanced spectroscopic and imaging applications.

Despite requiring high excitation power densities and yielding lower overall fluorescence intensity, TPF and 3PF offer notable advantages. These techniques reduce background fluorescence [34, 129] and enable deeper sample penetration, making them highly suitable for biological imaging and deep-tissue analysis. In mycotoxin detection, the requirement for multiple photons for excitation enhances specificity and sensitivity by mitigating interference from background fluorescence [34]. Consequently, multi-photon excitation emerges as a promising approach for fluorescence-based mycotoxin detection in challenging agricultural and food safety applications, but does require the use of high power excitation lasers in combination with a high-efficient fluorescence collection and a high detecting sensitivity [30].

One aspect of this thesis will be to evaluate upconversion fluorescence for multi-mycotoxin detection, assessing its effectiveness compared to conventional fluorescence spectroscopy. Given the limited scope of existing literature on this topic, this study aims to provide a more comprehensive understanding of its usability for detecting ultra-trace mycotoxins directly on corn and wheat matrices.

2.5.3 Chemometrics And ML For Mycotoxin Detection

As established in the preceding sections, chemometric- and ML-techniques are essential for implementing optical spectroscopy-based mycotoxin detection platforms. While the models employed have already been catalogued in previous chapters—particularly in the literature study presented

in Tables 2.09 and 2.10—the emphasis so far has been on the spectroscopic modalities.

In this section, the focus shifts to fully integrated spectroscopic/ML detection platforms. We begin detailing the model development process, covering both classification and regression models, training procedures, validation strategies, and key performance metrics.

This background provides the necessary context for revisiting the state-of-the-art—now with an emphasis on complete detection systems that integrate both sensing and data processing components. The objective being to identify the most promising integrated platforms for effective mycotoxin detection as reported in literature.

2.5.3.1 Model Development Process

Developing a robust mycotoxin detection model involves a structured five-step process. (1.) One starts by collecting a representative set of spectral data from samples with known mycotoxin compositions to establish a labeled training dataset. (2.) Then one applies mathematical preprocessing—such as denoising and normalization—to enhance data quality and ensure consistency (Section A.1). (3.) Next one then trains a predictive model to correlate the spectral data with contamination of the sample (Section A.2). This may take the form of a classification model, which distinguishes between contamination classes based on threshold levels, or a regression model, which estimates continuous toxin concentrations from spectral features. (4.) This is followed by rigorous model validation to evaluate accuracy and generalizability. This is typically achieved using methods such as external hold-out validation, k-fold Cross-Validation (CV), or Leave-One-Sample-Out Cross-Validation (LOSO-CV), depending on data availability and structure. A more detailed discussion on these validation strategies, their strengths and limitations, is given in Section A.4 of the Appendix. (5.) Finally, feature selection algorithms may be applied to identify the most informative wavelengths, reducing data dimensionality, improving computational efficiency, and simplifying the required setup, ensuring ease of integration in industrial settings (Section A.3).

An in-depth theoretical description of these five steps for model development is given in Appendix A. Readers with limited prior knowledge of machine learning for classification and regression, validation structures, performance metrics, or feature selection methods are strongly encouraged to refer to the relevant parts of the appendix. Going forward, this work assumes the reader to be familiar with the theoretical foundations of machine learning model development.

We now turn our attention to the state-of-the-art in mycotoxin detection on wheat and corn, with a specific focus on the classification and regression models employed, as well as the use of feature selection techniques.

2.5.3.2 Review Of ML-Models For Spectroscopic Mycotoxin Detection

Revisiting the state-of-the-art documented in Tables 2.09 and 2.10, we now narrow our focus to the most promising spectroscopic modalities—reflectance and fluorescence. Within this framework, we examine the machine learning models utilized in the corresponding classification and regression studies.

However, directly comparing model performance across these studies proves challenging due to significant variability in model development workflows, training procedures, and validation strategies. Differences in mycotoxin types, food matrices, sample preparation methods (e.g., artificially spiked versus naturally contaminated), validation schemes, contamination ranges, classification thresholds, and even the physical state of samples (e.g., ground versus whole kernels) all contribute to this complexity.

Additionally, the performance metrics reported differ between studies. For example, some studies on classification report only accuracy, while others provide sensitivity, specificity, and the F1-score, with few presenting complete confusion matrices. Similarly, regression performance is typically characterized by either the R^2 or R -value along with various error metrics. However, critical values such as the Limit Of Detection (LOD) and Limit Of Quantification (LOQ) are often omitted, making it challenging to determine whether the models meet regulatory guidelines. Despite these challenges, general conclusions about model performance can be drawn by reviewing comparative analyses and identifying the most consistently successful methods.

Most studies focus on detecting a single mycotoxin—commonly AFLA [16, 30, 93, 102, 106, 113, 120], DON [26, 31, 99, 116, 117, 118], or FUM [28, 51, 119]—with only a few studies reporting simultaneous monitoring of multiple toxins [28, 98]. The work on classification models is extensive (see Table 2.09), covering a broad range of spectroscopic platforms and employing methods from Linear Discriminant Analyses (LDA) [30, 98, 103, 125] and Quadratic Discriminant Analyses (QDA) [30] to K-Nearest Neighbors (KNN) [100, 104], Linear Support Vector Machines (LSVM) [28, 93, 105], Quadratic Support Vector Machines (QSVM) [93], Decision Trees (DT) [99], Random Forests (RF) [95], and Artificial Neural Networks (ANN) [94, 96].

In contrast, regression models (see Table 2.10) have predominantly employed transmittance [115] and reflectance spectroscopy [31, 51, 116, 117, 118, 119, 120], with all identified studies finding Partial Least Squares Regression (PLSR) as the optimal model.

We now turn to a detailed examination of selected studies that have identified high-performing classification and regression models for mycotoxin detection, offering insights into the strengths and limitations of each approach.

2.5.3.2.1 State-of-the-art; Classification Models

Among the most insightful contributions to classification model development is the work of Indy Magnus et al. [30], who investigated multi-defect detection of aflatoxins and external contaminants in pistachios. This study showed that both KNN and LDA models achieved approximately 96% classification accuracy in distinguishing safe from unsafe samples, significantly outperforming QDA, which attained only 90.2% accuracy. Similarly, Kyung-Min Lee et al. [100] reported that KNN and LDA reached 91.4% accuracy for aflatoxin detection in maize.

SVM classifiers have also demonstrated promising results. For instance, Kim et al. [28] achieved 89.1% accuracy for the co-detection of aflatoxins and fumonisins on whole maize kernels using fluorescence spectroscopy, with performance varying significantly by spectral modality—95.7% accuracy using SWIR reflectance and 71.7% in the VIS-NIR region. Wei Liu et al. considered an ANN combined with feature reduction reached 93.3% accuracy for ZEA detection using fluorescence hyperspectral imaging [107]. Furthermore, decision tree-based methods have shown high specificity when applied on reflectance spectra; for example a sensitivity of 85.7% and specificity of 97.3% were obtained using Random Forests [95], while bagged decision trees attained 85% accuracy for DON detection [99].

It is important to note that these studies did not report their respective classification thresholds, complicating the assessment of how close to the legal limits these methods operated. Due to the wide range of outcomes reported in the literature, no single model or method can be considered universally optimal. As a result, this work adopts a systematic approach, evaluating a broad set of models to determine which performs best for our specific application.

2.5.3.2.2 State-of-the-art; Regression Models

Regression models have also been extensively studied for the detection of single mycotoxins on corn and wheat. For regression, PLSR consistently emerged as the optimal model. However, only limited studies, as detailed below, have reported results that are close to or complying to the legal limits. Most studies still suffer from insufficient LOD and LOQ values, indicating the need for further optimization.

Appaw et al. [120] demonstrated exceptional performance for aflatoxin detection in maize kernels using VIS-NIR reflectance spectroscopy. Their PLSR model showed an R^2 of 0.99, a Standard Error of Cross-Validation (SECV) of $1.70\mu\text{g/kg}$, and reported a LOD and LOQ of $0.60\mu\text{g/kg}$ and $1.81\mu\text{g/kg}$, respectively—values that are highly promising for regulatory compliance considering the legal limit for AFLA detection is $4\mu\text{g/kg}$.

Similarly, Levasseur et al. [31, 118] applied PLSR for DON quantification in wheat kernels, achieving R^2 values between 0.89 and 0.91 and SECV values ranging from 578 to $612\mu\text{g/kg}$, both well below the legal limit of $1250\mu\text{g/kg}$. Pettersson et al. achieved an R -value of 0.984 for samples extending from 400 to $13000\mu\text{g/kg}$ and a SECV of $381\mu\text{g/kg}$, indicating a resolution sufficient for regulatory monitoring.

For FUM on maize kernels, Tyska et al. [119] reported outstanding performance with a PLSR model yielding an R^2 of 0.991, though with a Standard Error of Prediction (SEP) of $984\mu\text{g/kg}$, which exceeds the legal limit of $350\mu\text{g/kg}$. Della et al. [51] considered FT-NIR reflectance on ground maize achieving calibration R -values between 0.995 and 0.998, with SECV values as low as 232 – $144\mu\text{g/kg}$, well below the legal threshold.

Collectively, these studies—targeting DON, FUM, and AFLA—demonstrate that PLSR-based regression models are capable of achieving prediction errors below the respective legal thresholds. However, while error metrics are commonly reported, essential analytical parameters such as the LOD and LOQ are most often overlooked. Nonetheless, the consistent performance of PLSR across different mycotoxins and matrices underscores its strong potential as a core regression method for the development of robust, industry-ready multi-mycotoxin detection platforms.

2.5.3.3 A Review Of Feature Selection In Literature

It is worth noting that all the models discussed above rely on full spectral data, using the entire range of available wavelengths to maximize predictive performance. However, for developing more streamlined and cost-effective platforms suitable for industrial use, it is often preferable to limit models to a smaller, carefully selected set of wavelengths. To address this, we introduced feature selection as a systematic approach to identify the most informative spectral features, allowing models to perform effectively with reduced input complexity.

Despite its promise, feature selection remains relatively underexplored in the context of mycotoxin detection. However, a handful of recent studies have begun to address this gap. For example, Magnus et al. [30] and Smeesters et al. [16] successfully applied techniques such as Sequential Feature Selection (SFS) to isolate key wavelengths, enhancing both model efficiency and suitability for industrial deployment.

Another notable precedent comes from the work of Stasiewicz et al. [98], who used a fixed set of just nine wavelengths, in the range of 470 – 1550nm , to develop a multispectral sorter for AFLA and FUM detection in Kenyan maize. Rather than using algorithmic feature selection, wavelengths were selected based on hardware constraints and industrial precedent, incorporating visible and near-infrared bands. Their classification models (LDA) achieved 77% sensitivity and

83% specificity in distinguishing contaminated kernels above and below regulatory thresholds. Importantly, their study demonstrated that even with a discrete and relatively small subset of features, meaningful toxin reduction—over 80% on average—could be achieved in sorted maize samples.

Here we can note that algorithmic feature selection methods will have the potential to improve upon these results by systematically identifying the most informative and least redundant wavelengths from a larger candidate pool. This data-driven approach could uncover non-obvious spectral combinations that maximize classification and regression performance, potentially enhancing sensitivity, specificity, and overall sorting efficiency beyond what is achievable through heuristic or hardware-constrained feature choices.

Taken together, the integration of algorithmic feature selection not only holds promise for refining the spectral components of detection systems but also exemplifies a paradigm shift toward precision-engineered models, wherein the confluence of statistical rigor, computational optimization, and domain-specific constraints coalesces to yield solutions that are not merely performant in silico, but robust, scalable, and translational across the diverse operational landscapes of real-world mycotoxin mitigation.

Building on this, our work aims to systematically compare a broad range of feature selection strategies, seeking to benchmark full-spectral-models against those based on selected wavelengths, with a focus on optimizing the balance between industrial feasibility and predictive performance.

2.6 Summary And Research Gaps

Optical spectroscopy, when integrated with machine learning, presents a powerful, non-invasive, and rapid alternative to conventional chemical analysis for mycotoxin detection. These optical approaches provide the capability for in-situ, real-time assessment of food and feed safety, aligning well with the demands of modern industrial settings. Despite this promise, their widespread adoption has been constrained by several critical limitations.

Most notably, current implementations predominantly target the detection of a single mycotoxin within a specific matrix under tightly controlled laboratory conditions. This narrow scope restricts scalability and limits practical applicability. Consequently, a clear gap remains: the development of a cost-effective, scalable, and industrially viable optical platform capable of simultaneous multi-mycotoxin detection across diverse agricultural matrices.

To address this challenge, the present study centers on reflectance and fluorescence spectroscopy, two modalities that strike an optimal balance between analytical performance and operational feasibility. On the computational front, this work employs machine learning for both classification and regression tasks. A wide spectrum of classification models—including LDA, SVM, KNN, DT, RF, and ANN—have shown promise in prior studies. To identify the most effective approach for our specific application, a systematic comparative analysis of these models will be undertaken. For regression PLSR has consistently demonstrated superior performance in the literature. Nonetheless, a dedicated benchmarking effort will be carried out to validate these findings within the context of our data.

Finally, to enhance system scalability and cost-effectiveness, this study explores algorithmic feature selection strategies aimed at reducing spectral complexity without compromising predictive

performance. Techniques such as Sequential Feature Selection (SFS) and Minimum Redundancy Maximum Relevance (MRMR) have emerged as viable tools for identifying compact, highly-informative wavelength subsets. By combining these strategies with multi-mycotoxin detection across multiple matrices (i.e., corn and wheat), this work seeks to overcome current limitations and facilitate the transition from laboratory research to industrial deployment.

Chapter 3

Materials And Methods

Building on the foundation laid in the previous chapter—which detailed the biochemical background of mycotoxins, the capabilities and limitations of spectroscopic methods, and the current state of machine learning for spectroscopic mycotoxin detection—we now transition from theoretical insight to experimental implementation. Guided by the overarching workflow introduced in Figure 1.1, this chapter starts with the preparation of a well-characterized series of corn- and wheat-based samples, synthesized by mixing certified, naturally contaminated reference materials in controlled ratios to yield known multi-mycotoxin compositions. Section 3.1 provides a comprehensive account of this preparation process.

The chapter then presents the measurement setups and the calibration and acquisition protocol (Section 3.2). Spectral data is acquired using both reflectance (UV-VIS and NIR) and fluorescence spectroscopy, with the latter employing multiple excitation wavelengths. A visual exploration of the collected spectral data is thereafter conducted to identify the most informative spectral regions (Section 3.3).

All this forms the experimental backbone for the model development and validation presented in Chapters 4 and 5, which address classification and regression modeling, respectively.

3.1 Multi-mycotoxin Samples

To develop accurate and robust predictive models, it is essential to construct a labeled dataset comprising spectral measurements from a diverse range of samples with known mycotoxin compositions. Ideally, this dataset should cover contamination levels ranging from non-detectable to concentrations well above established regulatory thresholds, thereby allowing both classification and regression models to learn across a wide dynamic range.

Notably, spectroscopic detection of mycotoxins typically relies on indirect indicators—namely, secondary effects such as compositional changes in the matrix induced by fungal metabolism during toxin production. This makes the use of naturally contaminated material critical. In contrast, spiked samples introduce only pure mycotoxins, without the associated matrix alterations, and therefore fail to capture the key spectral features that serve as reliable predictors. As a result, they offer limited utility for training models intended to operate under real-world conditions.

In light of these considerations, the ideal approach should involve sourcing naturally contaminated, whole corn and wheat kernels with various compositions of mycotoxins, collected from geographically and environmentally diverse regions. This would ensure that the dataset fully

reflects the natural variation in both matrix composition and toxin distribution—an essential requirement for developing models with real-world applicability.

However, several practical limitations make this strategy difficult to implement. First, reliably obtaining field-contaminated kernels with known contamination levels is both challenging and costly, as precise quantification requires labor-intensive and resource-demanding chemical analysis. Second, the heterogeneous and highly localized distribution of mycotoxins within grain batches poses a significant obstacle, since standard analytical methods typically require bulk quantities—often tens of grams—making it impossible to determine toxin concentrations at the single-kernel level, as only the averaged contamination of a larger batch can be determined. These constraints render the creation of a large, well-characterized set of individual, naturally contaminated kernels with known mycotoxin profiles exceedingly difficult.

Given these constraints—and recognizing that the present work is intended as an initial feasibility study into multi-toxin detection—we adopt a more controlled and reproducible approach to sample preparation. Rather than using whole kernels, we work with finely ground reference materials derived from naturally contaminated corn and wheat, sourced from certified providers such as Romer Labs [130]. These biopure standards are homogeneous in composition and come with known concentrations of specific mycotoxins, allowing for the precise formulation of new samples by mixing different ratios of these standards. This approach greatly simplifies sample preparation and reduces natural variability, creating a simplified scenario for initial model development. The specific mixing strategies used to generate the experimental datasets are detailed in the sections that follow, beginning with a brief overview of the reference standards used.

3.1.1 The Standards

The reference samples, provided by Romer Labs [130], were prepared through a meticulous process to ensure accuracy, homogeneity, and reliability. The process, as applied by Romer Labs, began by sourcing corn and wheat naturally infected with mycotoxins. The contaminated grains were then finely ground and sieved through a 30-mesh (0.595mm) to achieve a uniform particle size. This ground material was thoroughly homogenized to ensure an even distribution of mycotoxins, minimizing variability and enhancing reproducibility in analytical measurements.

To determine the mycotoxin concentrations, the samples underwent rigorous quality control. First, mycotoxins were extracted using a carefully formulated solvent, and then the extracts were analyzed using Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS). To ensure reliability, 30 different extracts were prepared and analyzed over at least five independent testing days. The results from these tests were averaged to obtain the most accurate and reproducible concentration estimate. Finally, the samples are stored in a fridge to minimize contamination changes over time.

For this study, six certified reference standards were used: four for corn and two for wheat. The corn standards included individual contaminations with aflatoxins (AFLA), deoxynivalenol (DON), and zearalenone (ZEA), along with a mixed standard containing AFLA¹, DON, ZEA, and fumonisins (FUM). For wheat, one mixed standard with multiple mycotoxins and one blank (mycotoxin-free) standard were used. A complete overview of the mycotoxins present in each standard and their certified concentrations is provided in Table 3.1.

¹Here AFLA represents the combined levels of AFB₁, AFB₂, AFG₁, and AFG₂.

Matrix	Contaminants	Certified Concentrations [$\mu\text{g/kg}$]
Corn	AFLA (AFB ₁)	AFB ₁ : (4.8 ± 1.9) ; AFB ₂ , AFG ₁ , AFG ₂ : < 1 [131]
Corn	DON	(2200 ± 600) [132]
Corn	ZEA	(62 ± 17) [133]
Corn	AFLA ¹ , DON, ZEA, FUM	AFLA: (6.0 ± 2.5) ; DON: (900 ± 200) ; ZEA: (103.4 ± 37.5) ; FUM: (2600 ± 600) [134]
Wheat	DON, T-2, HT-2, OTA	DON: (1000 ± 200) ; T-2: (49.1 ± 21.4) ; HT-2: (9.7 ± 4.2) ; OTA: (8.3 ± 3.9) [135]
Wheat	Blank	No detectable mycotoxins [136]

Table 3.1: Details on the certified reference standards used for corn and wheat sample preparation.

3.1.2 Sample Synthesis

With the reference standards detailed, we turn to sample preparation. Two independent sample sets were constructed—one for corn and one for wheat—to enable the development of crop-specific detection models.

For corn, the four reference standards—AFLA, DON, ZEA, and the mixed standard (MIX), which contains AFLA, DON, ZEA, and FUM—were used. Due to concentration constraints in the available materials, it was only possible to generate samples exceeding the legal limits for AFLA and DON. In contrast, the concentrations of ZEA and FUM in both the pure and mixed standards remained below their respective regulatory thresholds, limiting the achievable contamination range for these toxins. Nevertheless, methodologies that perform at sub-threshold concentrations are expected to scale well for higher contamination levels, as the spectral features associated with mycotoxins typically become more pronounced with increasing concentration.

An additional constraint for corn is the absence of an uncontaminated reference standard (blank). Consequently, all combined samples will inevitably contain some degree of mixed contamination. For example, a sample with a reduced DON concentration can only be obtained by diluting the pure DON standard (or the MIX) using either the AFLA or ZEA standards—thereby introducing additional mycotoxins by necessity. To tackle this sample complexity, high care needs to be taken during the sample preparation, since the considered models will merely learn to differentiate between different toxin levels with respect to the other toxins, and not with respect to uncontaminated samples.

Here, more specifically, care needs to be taken that the model learns from the individual toxins, which might not be the case if always the same type of dilution would be applied. If, for example, the DON standard is always diluted using AFLA, the resulting data may confound the effects of DON with those of AFLA, making it unclear whether the model is learning to detect DON or a correlated change in AFLA concentration. To avoid this, the sample series must be designed carefully to decouple the spectral effects of individual mycotoxins from one another.

A further complexity arises from the fact that each of the four reference standards originates from a different batch of corn. Although Romer Labs [130] ensures that all standards are derived from the same corn type and produced under comparable conditions, batch-to-batch variation

in matrix composition is inevitable.

While some matrix differences may result from metabolic changes induced by fungal activity during mycotoxin production, others may be unrelated to the mycotoxins. Consequently, when samples are composed by mixing standards, they may differ not only in mycotoxin-specific features but also in standard-specific characteristics. Given that spectroscopic mycotoxin detection often relies on the detection of these indirect markers, it is critical to ensure that the models learn to identify toxin-specific features, rather than confounding patterns tied to the standards themselves.

To address both challenges, the corn samples were deliberately composed to isolate toxin-specific features. The core strategy involved creating a series of samples in which the concentration of a target mycotoxin remained constant, while varying the combination of reference standards used to achieve the concentration. This approach helps decouple mycotoxin signals from matrix-specific effects, enabling more reliable identification of toxin-related spectral patterns.

For example, samples with a fixed concentration of aflatoxin (AFLA) were prepared by blending the pure AFLA standard with either the DON or ZEA standard, or by combining the mixed standard (which includes AFLA) with one of the single-toxin standards. This generated multiple samples with identical AFLA levels but differing concentrations of the other toxins, and variable matrix compositions. By systematically repeating this process across a range of concentrations for each target toxin, we created a diverse and balanced dataset. This approach ensures that any consistent spectral patterns learned by the models can be attributed to the target mycotoxin or its associated matrix effects, rather than to co-contaminants or confounding standard-specific signatures.

For wheat, only two reference standards were available: an uncontaminated blank and a sample contaminated with DON, OTA, and T-2/HT-2 toxins. Of these, only OTA exceeded the legal limit of $5\mu\text{g/kg}$, while DON, T-2, and HT-2 remained within regulatory thresholds. Due to the limited number of standards, it is not possible to isolate the individual spectral effects of each mycotoxin. Consequently, all models trained on the wheat dataset are constrained to learn the combined spectral signature of this specific toxin mixture. Furthermore, because only two unique reference materials were available for sample mixing, it is impossible to disentangle matrix-specific features from mycotoxin-related signals.

Despite these limitations, the inclusion of a blank standard made it possible to construct an equidistant concentration series—from zero (pure blank) up to the maximum contamination level of the standard. This approach is particularly advantageous for regression tasks, as it ensures uniform coverage across the full concentration range, thereby reducing bias that could arise from an unbalanced dataset.

3.1.2.1 The Corn Samples

Following the sample synthesis strategy, we prepared a series of 40 samples, each weighing 10 grams, by precisely combining different corn-reference standards using a high-accuracy scale ($\pm 0.01\text{g}$). After mixing, each sample was thoroughly homogenized to ensure an even distribution of mycotoxins. The resulting mixtures were stored in glass vials, providing a stable, inert environment and allowing for convenient handling throughout the study. To maintain sample integrity, all vials were refrigerated when not in use.

Each vial was labeled as Sxx , where “ xx ” denotes the sample number. Figure 3.1 presents the

prepared corn samples. A detailed overview of the samples² is provided in Table 3.1. The error on the concentration is calculated by propagating the standard deviation of the standard's concentration and the measurement uncertainty of the sample composition.

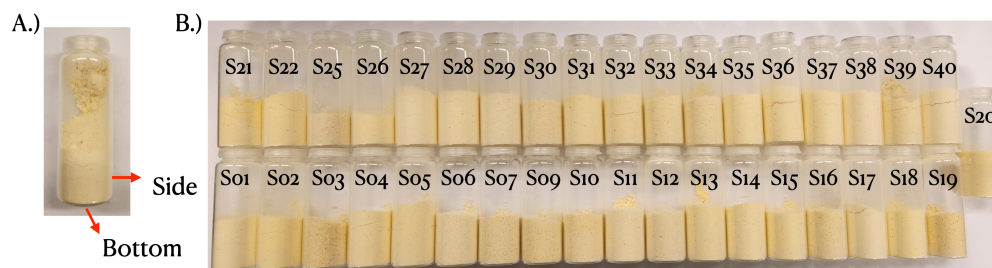


Figure 3.1: A.) A contaminated corn sample sealed in a glass vial, with measurement sites indicated; spectra were collected from multiple positions on both the bottom and side of the vial. B.) Most of the prepared corn samples. Visual distinction between samples is not possible, as all appear identical.

Mycotoxin On Corn Samples								
Sample	Mycotoxin Concentration				Sample Composition [± 0.01 g]			
	AFLA [μ g/kg]	DON [μ g/kg]	ZEA [μ g/kg]	FUM [μ g/kg]	AFLA [g]	DON [g]	ZEA [g]	Mix [g]
S01	4.80 $\pm$ 1.90	0	0.0	0	10.00	0.00	0.00	0.00
S02	0.00	2200 $\pm$ 600	0.0	0	0.00	10.00	0.00	0.00
S03	0.00	0	62.0 $\pm$ 17.0	0	0.00	0.00	10.00	0.00
S04	3.84 $\pm$ 1.52	440 $\pm$ 180	0.0	0	8.00	2.00	0.00	0.00
S05	1.92 $\pm$ 0.76	1320 $\pm$ 540	0.0	0	4.00	6.00	0.00	0.00
S06	3.84 $\pm$ 1.52	0	12.4 $\pm$ 3.5	0	8.00	0.00	2.00	0.00
S07	1.92 $\pm$ 0.76	0	37.2 $\pm$ 10.2	0	4.00	0.00	6.00	0.00
S08	0.00	440 $\pm$ 180	49.6 $\pm$ 13.6	0	0.00	2.00	8.00	0.00
S09	0.00	880 $\pm$ 360	37.2 $\pm$ 10.2	0	0.00	4.00	6.00	0.00
S10	0.00	1320 $\pm$ 540	24.8 $\pm$ 6.8	0	0.00	6.00	4.00	0.00
S11	3.84 $\pm$ 1.52	220 $\pm$ 90	6.2 $\pm$ 1.8	0	8.00	1.00	1.00	0.00
S12	1.92 $\pm$ 0.76	440 $\pm$ 180	24.8 $\pm$ 6.8	0	4.00	2.00	4.00	0.00
S13	6.00 $\pm$ 2.50	900 $\pm$ 200	103.4 $\pm$ 37.5	2600 $\pm$ 600	0.00	0.00	0.00	10.00
S14	5.40 $\pm$ 1.57	450 $\pm$ 100	51.7 $\pm$ 18.8	1300 $\pm$ 300	5.00	0.00	0.00	5.00
S15	3.00 $\pm$ 1.25	1550 $\pm$ 461	51.7 $\pm$ 18.8	1300 $\pm$ 300	0.00	5.00	0.00	5.00
S16	3.00 $\pm$ 1.25	450 $\pm$ 100	82.7 $\pm$ 20.6	1300 $\pm$ 300	0.00	0.00	5.00	5.00
S17	5.04 $\pm$ 1.60	180 $\pm$ 40	20.7 $\pm$ 7.5	520 $\pm$ 120	8.00	0.00	0.00	2.00
S18	1.20 $\pm$ 0.50	1940 $\pm$ 721	20.7 $\pm$ 7.5	520 $\pm$ 120	0.00	8.00	0.00	2.00
S19	1.20 $\pm$ 0.50	180 $\pm$ 40	70.3 $\pm$ 15.5	520 $\pm$ 120	0.00	0.00	8.00	2.00
S20	5.76 $\pm$ 2.04	720 $\pm$ 160	82.7 $\pm$ 30.0	2080 $\pm$ 480	2.00	0.00	0.00	8.00
S21	4.80 $\pm$ 2.00	1160 $\pm$ 249	82.7 $\pm$ 30.0	2080 $\pm$ 480	0.00	2.00	0.00	8.00
S22	4.80 $\pm$ 2.00	720 $\pm$ 160	95.1 $\pm$ 30.2	2080 $\pm$ 480	0.00	0.00	2.00	8.00
S23	4.80 $\pm$ 1.90	0	0.0	0	10.00	0.00	0.00	0.00
S24	0.00	2200 $\pm$ 900	0.0	0	0.00	10.00	0.00	0.00
S25	0.00	0	62.0 $\pm$ 17.0	0	0.00	0.00	10.00	0.00

²All samples were used for model development, with the exception of sample S08, which was lost prior to the start of this thesis and is therefore excluded from the remainder of the study.

S26	6.00 ± 2.50	900 ± 200	103.4 ± 37.5	2600 ± 600	0.00	0.00	0.00	10.00
S27	2.88 ± 1.14	880 ± 360	0.0	0	6.00	4.00	0.00	0.00
S28	0.96 ± 0.38	1760 ± 720	0.0	0	2.00	8.00	0.00	0.00
S29	2.88 ± 1.14	0	24.8 ± 6.8	0	6.00	0.00	4.00	0.00
S30	0.96 ± 0.38	0	49.6 ± 13.6	0	2.00	0.00	8.00	0.00
S31	0.00	1760 ± 720	12.4 ± 3.4	0	0.00	8.00	2.00	0.00
S32	3.84 ± 1.60	576 ± 128	88.5 ± 24.8	1664 ± 384	0.00	0.00	3.60	6.40
S33	2.88 ± 1.20	432 ± 96	81.9 ± 20.1	1248 ± 288	0.00	0.00	5.20	4.80
S34	1.92 ± 0.80	288 ± 640	75.2 ± 16.7	832 ± 192	0.00	0.00	6.80	3.20
S35	0.72 ± 0.30	2044 ± 792	12.4 ± 4.5	312 ± 72	0.00	8.80	0.00	1.20
S36	1.44 ± 0.60	1888 ± 686	24.8 ± 9.0	624 ± 144	0.00	7.60	0.00	2.40
S37	2.16 ± 0.90	1732 ± 580	37.2 ± 13.5	936 ± 216	0.00	6.40	0.00	3.60
S38	0.82 ± 0.34	288 ± 72	63.2 ± 14.4	356 ± 82	0.00	0.75	7.88	1.37
S39	3.84 ± 1.60	576 ± 128	88.5 ± 24.8	1664 ± 384	0.00	0.00	3.60	6.40
S40	5.18 ± 1.52	288 ± 64	33.1 ± 12.0	832 ± 192	6.80	0.00	0.00	3.20

Table 3.2: Overview of the prepared mycotoxin-contaminated corn samples. The first column lists the sample identifiers (*Sxx*). Columns 2 to 5 indicate the total concentration of each respective mycotoxin, while columns 6 to 9 detail the composition of each sample in terms of the specific reference standards used.

3.1.2.2 The Wheat Samples

For wheat, an approximately equidistant series of 18 contaminated samples was prepared by mixing the two available reference standards. Minor adjustments to the concentration intervals were made to align certain wheat samples with the DON-levels used in the corn dataset. This approach was chosen to facilitate the potential development of a unified model capable of detecting DON across both wheat and corn matrices.

The samples are visually shown in Figure 3.2, highlighting, just like for corn, their identical nature. The full composition of the synthesized wheat samples is presented in Table 3.2. The error on the concentration is once more calculated by propagating the standard deviation of the standard's concentration and the measurement uncertainty of the sample composition.



Figure 3.2: The full series of prepared wheat samples. Visual distinction between samples is not possible, as all appear identical.

Mycotoxin Content in Wheat Samples						
Sample	Mycotoxin Concentration				Sample Composition [± 0.01 g]	
	DON [μ g/kg]	T-2 [μ g/kg]	HT-2 [μ g/kg]	OTA [μ g/kg]	Contam. Sample [g]	Uncontam. Sample [g]
SoA	1000 ± 200	49.0 ± 21.4	8.7 ± 4.2	8.3 ± 3.9	10.00	0.00
SoB	900 ± 180	44.1 ± 19.3	7.8 ± 3.8	7.5 ± 3.5	9.00	1.00
SoC	850 ± 170	41.7 ± 18.2	7.4 ± 3.6	7.1 ± 3.3	8.50	1.50

S0D	800 ± 160	39.2 ± 17.1	6.9 ± 3.3	6.6 ± 3.1	8.00	2.00
S0E	720 ± 144	35.3 ± 15.4	6.3 ± 3.0	6.0 ± 2.8	7.20	2.80
S0F	650 ± 130	31.9 ± 13.9	5.7 ± 2.7	5.4 ± 2.5	6.50	3.50
S0G	610 ± 122	29.9 ± 13.0	5.3 ± 2.6	5.1 ± 2.4	6.10	3.90
S0H	576 ± 116	28.2 ± 12.4	5.0 ± 2.4	4.8 ± 2.3	5.80	4.20
S0I	500 ± 100	24.5 ± 10.7	4.4 ± 2.1	4.2 ± 2.0	5.00	5.00
S0J	450 ± 90	22.1 ± 9.6	3.9 ± 1.9	3.7 ± 1.8	4.50	5.50
S0K	400 ± 80	19.6 ± 8.6	3.5 ± 1.7	3.3 ± 1.6	4.00	6.00
S0L	350 ± 70	17.2 ± 7.5	3.0 ± 1.5	2.9 ± 1.4	3.50	6.50
S0M	288 ± 58	14.1 ± 6.2	2.5 ± 1.2	2.4 ± 1.1	2.90	7.10
S0N	180 ± 36	8.8 ± 3.9	1.6 ± 0.8	1.5 ± 0.7	1.80	8.20
S0O	130 ± 26	6.4 ± 2.8	1.1 ± 0.6	1.1 ± 0.5	1.30	8.70
S0P	100 ± 20	4.9 ± 2.1	0.9 ± 0.4	0.8 ± 0.4	1.00	9.00
S0Q	50 ± 10	2.5 ± 1.1	0.4 ± 0.2	0.4 ± 0.2	0.50	9.50
S0R	0.0	0.0	0.0	0.0	0.00	10.00

Table 3.3: Overview of the wheat samples contaminated with varying levels of DON, T-2, HT-2, and OTA. Columns 2 to 5 indicate the total concentration of each respective mycotoxin, while columns 6 and 7 detail the composition of each sample in terms of the specific quantity of each standard mixed.

3.2 Measurement Setup

With the wheat and corn samples fully prepared, the next phase of the study focuses on the reflection and fluorescence spectroscopic measurements. These complementary techniques were chosen for their proven effectiveness in prior studies and their practical suitability for integration into industrial systems, as discussed in Section 1.2.1.

The following sections outline the experimental setups employed, as well as the corresponding calibration and measurement protocols.

3.2.1 Reflection Spectroscopy Setup

For reflectance spectroscopy, both local and global measurement approaches were explored, utilizing either a small or a large integrating sphere. These setups were designed to capture the full spectrum of diffusely reflected light, independent of the reflection angle, across the UV–VIS–NIR wavelength range. The following sections provide a detailed description of both configurations, along with the associated calibration and measurement protocols.

3.2.1.1 Small Reflection Sphere; Setup

The local diffuse reflectance along the UV-VIS and NIR spectral regions was studied using a compact integrating sphere setup. A schematic of this setup is presented in Figure 3.3. This configuration used a fiber-coupled, broadband white light source, from Instrument Systems, which integrates deuterium and halogen lamps to produce an illumination beam covering the 200–2500nm spectral range.

To direct light toward the integrating sphere, a secondary fiber (Avantes FC-UVIR600-2, a multimode fiber with a core diameter of 600 μ m, enabling high transmittance) was employed. A compact SMA-terminated collimating lens (COL-UV/VIS, Avantes) was used to couple both fibers. The secondary multimode fiber was then connected to the illumination port of a small integrating sphere (AvaSphere-30, Avantes), which features a 98% diffusely reflective Spectralon[®]

coating. The sample is illuminated under an 8° incident angle to reduce specular reflections, by illuminating through a 6mm aperture, yielding a localized measurement area of 28.3mm^2 .

All reflected light, regardless of its emission angle, is then collected by the integrating sphere and transmitted via a secondary fiber (Avantes FCB-UVIR600-2, a bifurcated optical fiber) to a dual-channel spectrometer system. This system consists of two modules: the Avantes AvaSpec-3684, which covers the 200–1100nm range with a spectral resolution of 1.4nm, and the Avantes AvaSpec-256, covering 1000–1700nm with a resolution of 4nm. This dual-channel configuration enabled the acquisition of the entire UV–VIS–NIR spectrum, ensuring comprehensive spectral coverage in a single measurement.

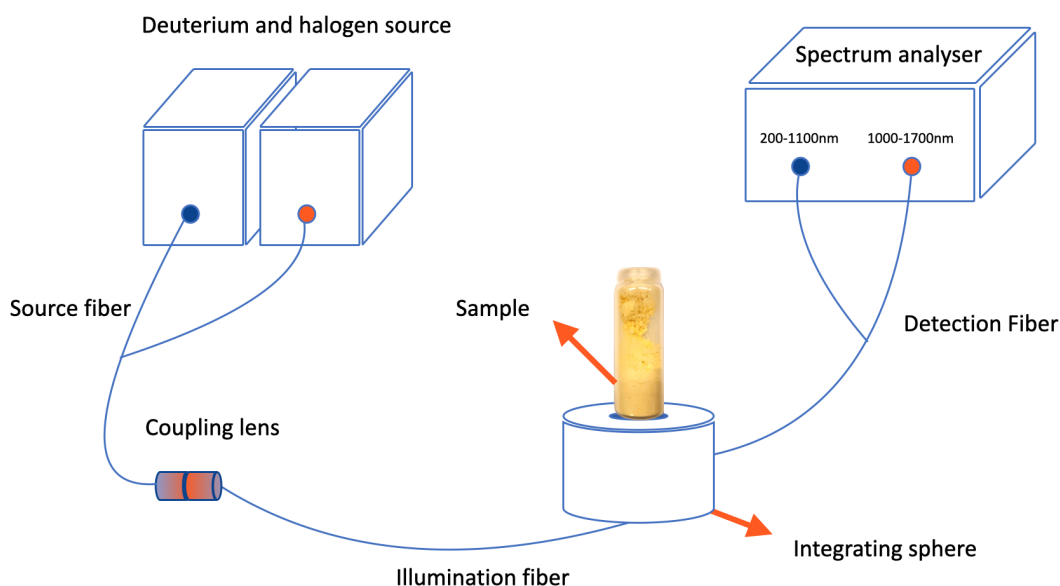


Figure 3.3: Schematic of the small integrating sphere setup used for localized reflectance measurements. This configuration captures diffuse reflectance spectra across the full UV–VIS–NIR range.

3.2.1.2 Small Reflection Sphere; Measurement Protocol

Before starting each measurement session, the light sources were allowed to warm up and stabilize for 10–15 minutes. Once stabilized, 15 measurements were collected per sample. Both spectrometers were connected to a computer for real-time data acquisition using Avantes AvaSoft 8 software. Within the software, pixel smoothing was configured to enhance signal clarity and reduce high-frequency noise: smoothing was set to 36 pixels for the UV–VIS range and 12 pixels for the NIR range, as specified within the Avantes manual and matching the spectrometer resolution. Automatic saturation detection was also enabled to prevent signal clipping and ensure reliable intensity readings.

Before commencing the actual measurements, the integration time was automatically configured. This was done by covering the measurement aperture with a certified reference tile featuring a 99.9% reflective coating (Spectralon[®] diffuse reflectance standard SRS-99-10) and using the

auto-configure function. Once the integration time was set (2-4 seconds, depending on exact optical alignment), a reference measurement was taken by capturing the spectrum of the reference tile. Subsequently, a dark measurement was recorded to calibrate the setup. The dark spectrum was obtained using an empty aperture while working in a dark room, capturing the effects of any stray-light reaching the spectrometer and any residual ambient light interference. Calibration accuracy was verified by ensuring that the measured reflectance of the reference tile remained consistently at 100% within the stable regions: 350–1000nm for the UV–VIS channel and 1000–1700nm for the NIR channel. Once proper calibration was confirmed, sample measurements commenced.

Prior to measurement, each sample was thoroughly homogenized by shaking and mixing to ensure an even distribution of mycotoxins. The prepared vial was carefully placed over the measurement aperture, ensuring it remained stationary throughout data acquisition. Measurements were conducted in two orientations—bottom and side—as illustrated in Figure 3.4. For bottom measurements, the vial was positioned vertically over the aperture; while for side measurements, a custom holder was used to secure the vial in a horizontal position. Once a measurement completed the reflectance of the samples was calculated automatically by the software as the ratio of the reflected light intensity and the reference spectrum, both corrected for dark, as described by equation 3.1,

$$reflectance(\lambda) = \frac{I_{Sample}(\lambda) - I_{Dark}(\lambda)}{I_{Reference}(\lambda) - I_{Dark}(\lambda)}. \quad (3.1)$$

This spectral data was saved as a ASCII file (.txt), and later processed in Python.

After each measurement, the sample was slightly repositioned to expose a new area to the aperture, and capture localized variations. This process was repeated 15 times for each sample and each orientation. To ensure homogeneity, the sample was thoroughly mixed again after every 5 measurements. Once all 15 measurements were completed for a sample, new dark and reference measurements were taken to correct for any gradual shifts in background signal or light source output, ensuring consistent accuracy. This entire procedure was then repeated for all samples. The actual measurements conducted and the obtained spectral data are detailed in Sections 3.3.1.2, and 3.3.1.3, for both corn and wheat samples, respectively.

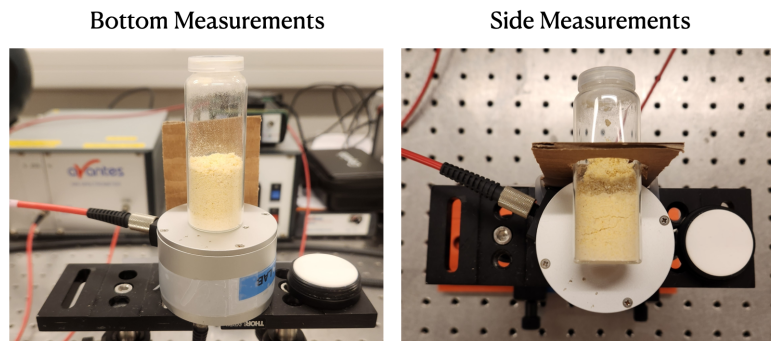


Figure 3.4: Sample placement for localized reflectance measurements using the small integrating sphere setup, schematically shown in Figure 3.3.

3.2.1.3 Large Reflection Sphere; Setup

In addition to localized reflectance measurements, the reflectance of the full sample was assessed using a large integrating sphere (LSM100) with a diameter of 25cm. This setup allowed measurements of the overall reflectance of a sample by placing the entire vial inside the sphere. A visual of this setup is provided in Figure 3.5.

Given the significantly larger size of the sphere, a high-intensity light source was required. For this purpose, we employed the Fianium-SC400 supercontinuum source, which generates a high-power broadband spectrum by pumping a photonic crystal fiber with a high-power pulsed laser (1064nm, 1.5W). This setup provided a spectral range extending from 415 to 1880nm, ensuring sufficient illumination across the UV-VIS-NIR spectrum. The same spectrum analyzer used in the previous setup, with same settings, was also used here.



Figure 3.5: Image of the large integration sphere, coated with a Spectralon[®] coating.

3.2.1.4 Large Reflection Sphere; Measurement Protocol

The used measurement procedure was largely similar to the protocol used for local measurements, as detailed in Section 3.2.1.2. Similarly, before starting measurements the light source was allowed to stabilize, now by operating for over an hour to reach stable output and minimize spectral fluctuations. To initiate the measurement process, the integration time was automatically set using an empty, closed sphere. For calibration, a reference measurement and a dark measurement were captured. The reference was obtained by scanning an empty sphere, while the dark signal was approximated by disconnecting the fiber from the sphere and recording measurements in a dark room.

To measure a sample, the vial was placed upright on the sample platform within the sphere, ensuring consistent positioning across all samples (see Figure 3.5). To prevent interference from labeling, written on top of each vial, the top was covered with a small white piece of paper, as open-vial measurements were not feasible due to the toxic nature of the mycotoxin-contaminated samples. Since we observed that the light source, the Fianium-SC400 supercontinuum, exhibited

instabilities, we increased the calibration frequency to one calibration before every measurement to compensate.

This measurement procedure was repeated three times for each sample, mixing the sample before every measurement. The measurements conducted and obtained spectral data are detailed in Section 3.3.2.

3.2.2 Fluorescence; Setup

For both conventional fluorescence and upconversion fluorescence spectroscopy measurements the same highly adaptable illumination and detection system was employed. A schematic of the experimental setup is presented in Figure 3.6.

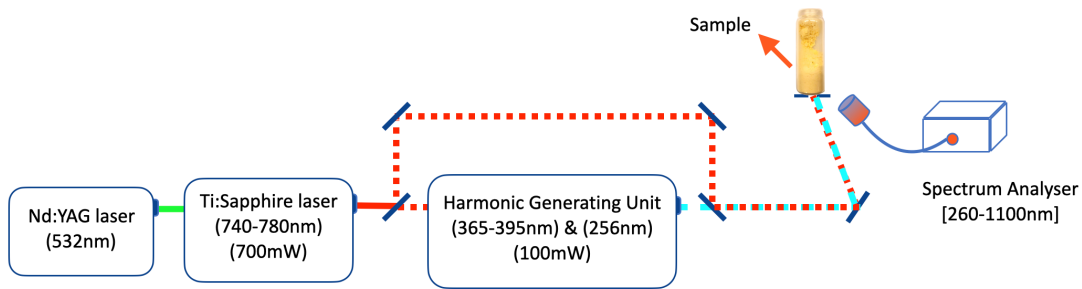


Figure 3.6: A schematic of the fluorescence setup allowing for both fluorescence and two-photon-fluorescence measurements. For two-photon fluorescence the beam is used directly without using the Harmonic Generating Unit (HGU), and an additional focusing lens is used below the sample opening. For measurements using the third harmonic an additional pinhole is used to clean up the laser beam, placed after the HGU.

As illustrated in the schematic, the excitation beam was generated using a tunable laser system, comprising a frequency-doubled, continuous wave Nd:YAG pump laser (Spectra-Physics Millennia Prime 10sJSPG), followed by a femtosecond-pulsed titanium-sapphire laser (Spectra-Physics Mode-locked Tsunami laser, Model 3941). The titanium-sapphire laser was capable of producing ultrashort pulses with durations below 100fs and offered a tunable wavelength range of 710–835nm, controlled via an internal prism-and-slit system within the lasing cavity. Depending on the selected wavelength, the output power of this laser varied between 1.2 and 1.7W.

The laser output from the Tsunami module was either used directly for upconversion fluorescence measurements or routed to a harmonic generation unit (Spectra-Physics Flexible Harmonic Generator GWO-23FS) to produce shorter excitation wavelengths. For two-photon fluorescence (TPF), the beam was used in its fundamental form, and an additional focusing lens was placed in front of the sample to ensure a high excitation power density—enhancing the efficiency of the nonlinear two-photon absorption process.

For one photon fluorescence, the frequency-doubling module was employed, featuring a nonlinear crystal to generate UV laser wavelengths. This configuration enabled the generation of excitation wavelengths in the range of 355–417nm, with an output power that can reach to 200–450mW, depending on the selected wavelength and the alignment of internal optical components. Additionally, third-harmonic generation (THG) was achieved by combining the frequency-doubled beam with a portion of the fundamental Tsunami laser output in a second nonlinear crystal,

producing excitation wavelengths in the 250–270nm range. The output power from the third-harmonic generation process was significantly lower, with a maximum of approximately 50mW. To improve spectral purity, a diaphragm was introduced after the THG stage to filter out residual light from the fundamental beam, thereby minimizing unwanted spectral contributions.

The laser beam, generated within one of the three available wavelength ranges, was directed towards the sample holder using a series of adjustable mirrors. To ensure precise alignment, a temporary alignment pinhole was placed at the sample position, allowing for fine-tuning of the beam’s trajectory onto a well-defined spot. The mirrors were adjusted accordingly to achieve optimal alignment.

To capture the fluorescence signal emitted from the illuminated area, a collimating lens (Avantes COL-UV/VIS, 6mm diameter, 8.7mm focal length) was positioned to collect the emitted light. This lens was connected via an optical fiber (Avantes FCB-UVIR600-2) to a spectrometer (AvaSpec 2048), capable of detecting fluorescence across the 260–1100nm range with a spectral resolution of 8nm.

The positioning of the collection optics was optimized using the principle of optical reciprocity: instead of connecting the fiber-end to the spectrometer, a lower power laser source was temporarily coupled into the fiber. The lens was then positioned in such a way that the light emerging from the fiber illuminated the alignment pinhole. Once the optimal collection geometry was established, the laser was decoupled from the fiber-end and the fiber was connected to the spectrum analyzer again.

Next, to minimize interference from direct diffuse reflection of the excitation beam, an optical filter was placed in front of the collimating lens. For one-photon-induced fluorescence measurements, a long-pass filter (EdgeBasicTM 405) was used to block the shorter excitation wavelengths, ensuring that only the fluorescence emission was detected. Conversely, for two-photon-induced fluorescence measurements, a short-wave pass filter (10SWF-650-B CB111, Newport) was employed to filter out longer wavelengths, preventing residual excitation light from influencing the detected fluorescence signal. Then the alignment pinhole was removed, and the system was ready for fluorescence measurements.

To monitor and maintain consistent excitation power, the Ophir Nova II Display power meter was used to measure the beam between measurements.

3.2.3 Fluorescence; Measurement Protocol

To conduct the fluorescence measurements, the laser source was first aligned and allowed to run for 20 minutes to reach stable operation. The spectrum analyzer was connected to a PC, and Avantes AvaSoft 8 software was used for data acquisition. Pixel smoothing was configured to reduce noise: for two-photon fluorescence, smoothing was set to 36 pixels; for one-photon fluorescence, it was set to only 6 pixels. All measurements were conducted in a dark environment to minimize ambient light interference.

Background measurements were periodically recorded to account for any external influences. The integration time for each measurement was manually set because fluorescence signals can rapidly diminish due to photobleaching, and automatic adjustment proved too slow. A fixed integration time (few milliseconds) was maintained for all measurements within the same experimental configuration to ensure consistency. However, when different excitation wavelengths or power levels were used, the integration time was adjusted accordingly to capture the full spectral response while avoiding detector saturation.

For each sample, 15 spectra were recorded. Between measurements, the sample was slightly repositioned to capture different regions within the vial, and after every five measurements, the sample was thoroughly mixed to maintain homogeneity. A dark spectrum was recorded after every 15 measurements. Extra care was taken to begin measurements immediately after placing the sample in the setup to minimize signal degradation caused by photobleaching. Fluorescence measurements were performed at excitation wavelengths of 395, 385, 375, 365, and 256nm, as well as via two-photon excitation at 790 and 750nm. The specific measurements done and their respective power levels are detailed in Sections 3.3.3.2 (two-photon fluorescence) and 3.3.3.1 (one-photon fluorescence).

Similar to the calibration performed for the reflectance measurements, a calibration was also implemented for the fluorescence measurements, enabling an absolute measurement of the fluorescence intensity. This was done using a transfer function that accounted for the wavelength-dependent transmittance of the optical fiber and the sensitivity of the spectrometer.

This transfer function was computed based on measurements done using a certified tungsten halogen light source (Avantes Avalight-CAL 0411016). More specifically, prior to our fluorescence measurements, the optical fiber and spectrometer were used to measure the emission of the calibration light source, with known spectral radiance, enabling to define a transfer function between the measured photons per wavelength and intensity per wavelength. The transfer function was saved as a calibration file, which was then loaded into AvaSoft to automatically calibrate the measurements by computing the absolute fluorescence intensity ($I_{sample}(\lambda)$), enabling direct comparisons between data collected using different setups.

The absolute fluorescence spectrum can be calculated through the following calibration equation:

$$I_{sample}(\lambda) = Cal_{data}(\lambda) \frac{Sample(\lambda) - Dark(\lambda)}{Ref_{cal}(\lambda) - Dark_{cal}(\lambda)} \frac{Int_{cal}}{Int_{sample}}, \quad (3.2)$$

where, $Cal_{data}(\lambda)$ is the certified intensity of the calibration light source in $\mu W/cm^2$. The term $Sample(\lambda)$ refers to the recorded fluorescence spectrum of the sample in analog-to-digital (A/D) counts, with $Dark(\lambda)$ representing the background noise recorded during the sample measurement. Similarly, $Ref_{cal}(\lambda)$ denotes the recorded spectrum of the calibration light source, and $Dark_{cal}(\lambda)$ accounts for the background noise during calibration. The integration times used for calibration and sample measurements are represented by Int_{cal} and Int_{sample} , respectively.

3.3 Spectroscopic Data Acquisition And Results

In this section we now describe the measurements conducted and present the collected spectral data. For clarity, this section is divided into two parts, focusing separately on reflectance and fluorescence measurements, each for both corn and wheat samples.

3.3.1 The Small Sphere Reflectance Measurements

The first dataset collected consisted of small-sphere reflectance measurements covering the full UV–VIS–NIR range, using the setup and measurement protocol described in Sections 3.2.1.1 and 3.2.1.2, respectively. In the following sections, we present these measurements for both corn and wheat samples. First however we briefly compare the spectra obtained for bottom compared to side measurements, to decide on an ideal measurement orientation.

3.3.1.1 Comparing The Bottom And Side Measurements

Before presenting the collected data in detail, we briefly examine the differences between bottom and side measurements to determine the optimal measurement position. This comparison is of interest as it proved challenging to ensure the samples were well mixed; the corn and wheat powders tended to coagulate, especially at the bottom of the vials, making homogenization difficult. Therefore, we aimed to investigate whether increased variability in the spectra would be observed when considering measurements at the bottom vs. side of the vial, which could indicate insufficient mixing.

To carry out this analysis, we focused on two specific samples: S01 and S04. Sample S01 contains only AFLA and is prepared from a single standard, ensuring homogeneity and eliminating the need for mixing. In contrast, S04 is a manually mixed sample containing both AFLA and DON, from two different standards (see Table 3.1). By comparing bottom and side measurements for both samples, we can assess whether poor mixing impacts the results.

For instance, if significant spectral differences between bottom and side measurements are observed for S04 but not for S01, this would suggest that bad mixing has occurred. Conversely, if similar differences appear in both samples, the variability is likely due to other factors, unrelated to mixing.

The results of this comparison are shown in Figure 3.7, which presents the mean and first standard deviation (std) of the 15 measurements per orientation per sample. Bottom measurements are shown as solid lines, and side measurements as dashed lines. A clear distinction is evident between bottom and side measurements for both samples.

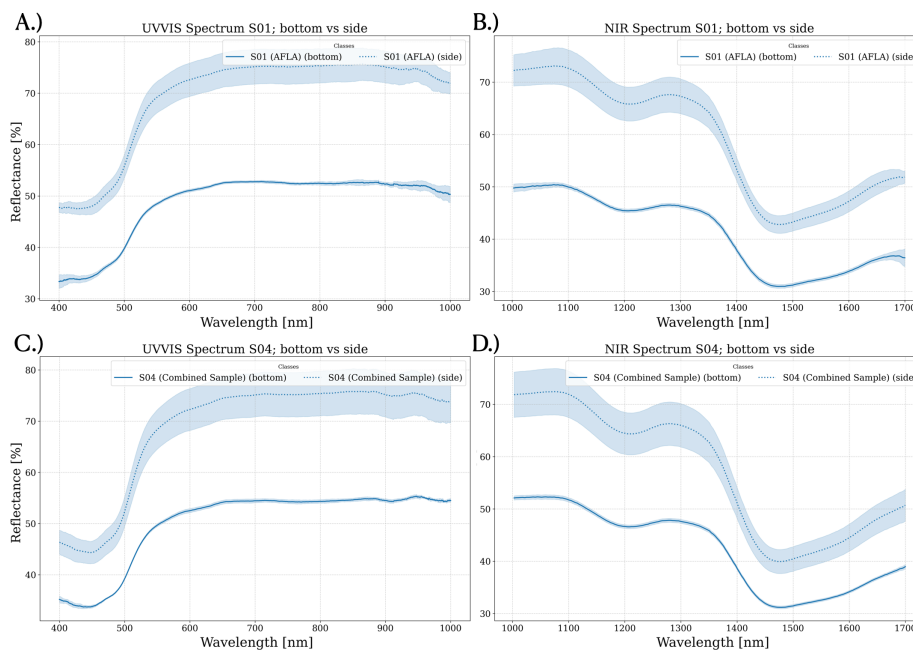


Figure 3.7: Mean reflectance spectra with standard deviation bands for samples S01 and S04, comparing bottom and side measurements. A.) UV-VIS and B.) NIR spectra for sample S01 (pure AFLA standard), and C.) UV-VIS and D.) NIR spectra for sample S04 (a mixed sample). Side measurements consistently exhibit greater spectral variability than bottom measurements, alluding to a problem with sample positioning.

Side measurements consistently show higher reflectance and, more notably, significantly greater variability (std-band) across both the UV–VIS and NIR regions—a pattern observed not only in these two samples but consistently across all samples. At first glance, this increased variability might be attributed to poor mixing or sample inhomogeneity. However, since the effect is also evident in S01—a pure, homogeneously distributed sample—inhomogeneity is unlikely the primary cause. Instead, the variability is most likely due to differences in sample positioning. When the vial is placed bottom-down, the aperture is fully covered, ensuring consistent illumination and reflection. In contrast, side placement does not fully cover the aperture, and the curved sides of the vial introduce significant variability in light reflection for even small differences in positioning. Maintaining a stable and repeatable position during side measurements was also more challenging, as the vial tended to roll or shift. We attempted to mitigate this using a custom holder, as shown in the setup in Figure 3.3, but side measurements still introduced considerable variability attributable to the vial shape and positioning, rather than the sample itself.

Beyond this inherent variability, we did not observe any clear trends indicating poor mixing—there was no significant additional variability in mixed samples compared to pure samples, suggesting that the extensive mixing conducted was effective and that the samples were likely well homogenized. Based on these findings, all subsequent analyses using both reflectance and fluorescence spectroscopy will be based solely on bottom measurements.

3.3.1.2 Small Sphere Measurements For Corn

Now we can look into the actual spectral data of the corn samples in more detail. Figure 3.8 shows the UV–VIS (Plot A) and NIR (Plot B) data for each sample (mean and std of the 15 measurements of each sample from the bottom).

These spectra exhibit considerable variability and overlap, making it difficult to visually associate individual spectra with specific samples. This also highlights the need for chemometric and ML-techniques to enable classification and regression based on this highly complex spectral dataset.

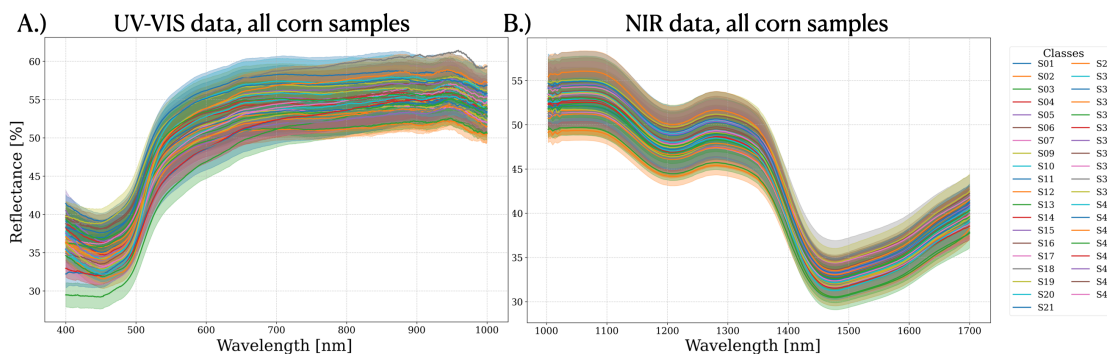


Figure 3.8: Mean reflectance spectra with standard deviation bands for samples S01 to S40. A.) shows the UV–VIS region, while B.) displays the NIR range. The complexity and overlap of the spectral data highlights the need for chemometric and ML approaches for effective data processing and interpretation.

The overall spectral shape—showing distinct minima (around 400–500nm and beyond 1400nm) and maxima (600–1100nm and 1300nm) extending from UV–VIS to NIR—is very similar to the spectral data for ground corn reference samples reported by Smeesters et al.[26] and Pettersson et

al. [115]. To explore the data in more detail, we can focus on the spectra of the pure standards: AFLA (S01), DON (S02), and ZEA (S03).

Figure 3.9 presents these results, showing the mean and standard deviation for the 15 spectra collected per sample. Although considerable overlap exists between the spectra—even in this simplified scenario of distinguishing between single-toxin samples—some meaningful differences can be observed.

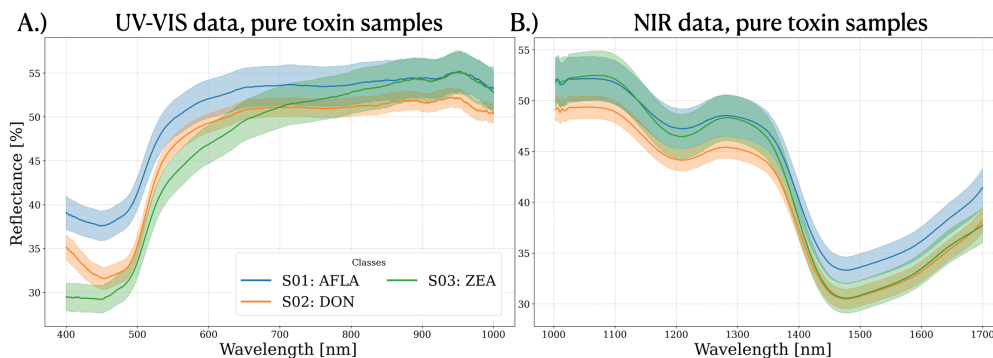


Figure 3.9: Mean reflectance spectra with standard deviation bands for samples S01 to S03. A.) shows the UV–VIS region, while B.) presents the NIR range. The spectra of the pure-toxin samples are visibly distinguishable, and the overall spectral shapes align with results reported by Smeesters et al. [26] and Pettersson et al. [115].

For example, in the UV–VIS range (Plot A), the clearest separation appears between 400–600nm, where the three samples can be distinguished from one another. Outside this range, substantial overlap remains. In the NIR region (Plot B), DON can be differentiated from AFLA and ZEA between 1000–1400nm, while AFLA stands out from ZEA and DON beyond 1400nm.

However, care should be taken, since the observed differences between these samples reflect not only the mycotoxin compositions but also potential matrix differences specific to each standard, unrelated to the toxins. Therefore, only by using the full series of composed samples—where matrix and mycotoxin compositions were decoupled, as discussed earlier in Section 3.1.2—can we ensure differentiation is based solely on mycotoxin content. In such cases, chemometric and machine learning techniques are essential to achieve reliable and accurate classification/regression.

Some complexities, that will be revisited later, are worth noting here. First, the pronounced differentiation between samples in the VIS region may be influenced by slight color differences between the samples, which are unrelated to mycotoxin content. Additionally, the spectral region around 1400nm can be problematic, as these wavelengths are strongly affected by humidity [137]. This is particularly relevant in industrial settings, where moisture content can vary depending on the processing stage (pre-harvest, harvest, or storage). As a result, classifiers trained on this region risk learning to detect moisture differences rather than mycotoxin-specific markers. For this reason, when training models we will have to investigate if these regions can be omitted while maintaining performance, to ensure we are not learning to differentiate based on these markers. This will be discussed in more detail later in Section 4.4.5.

3.3.1.3 Small Sphere Measurements For Wheat

Besides capturing the UV-VIS-NIR data for corn, we also captured the spectral data for the wheat samples. Figure 3.10 displays the obtained spectra (mean and std) for all the samples.

As evident from the figure, there is once again substantial spectral overlap between the samples, reinforcing the need for chemometrics and machine learning.

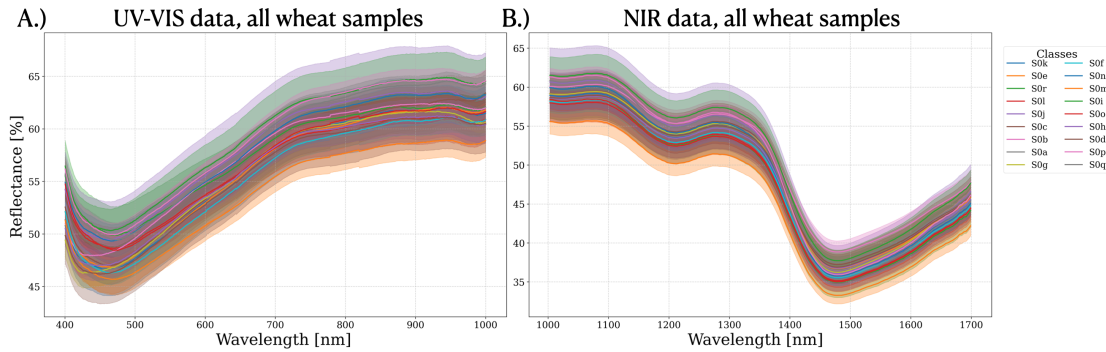


Figure 3.10: Mean reflectance spectra and first std band for the wheat samples (S0A–S0R). A.) shows the UV–VIS range, while B.) visualizes the NIR data.

In addition to analyzing all wheat samples together, we also compared the most contaminated (S0A) and the least contaminated (S0R) wheat samples. This comparison is shown in Figure 3.11. The most noticeable distinction appears in the 400–500nm region, where the blank sample (S0R) shows higher reflectivity. This trend also continues for the rest of the visible region and the start of the NIR region till 1000nm. This is consistent with findings by Alisaac et al. [138], who reported similar trends in ground wheat contaminated with *Fusarium* species. Interestingly, the std of the blank sample was also much higher than that of the contaminated sample.

In the range between 1000 and 1300nm, the spectra largely overlap, but in the region beyond 1400nm we again see a clearer distinction between the two samples. This aligns with the results of Beyer et al. [116], who found that the 1400–1900nm range offers the strongest differentiation for detecting deoxynivalenol content in wheat samples.

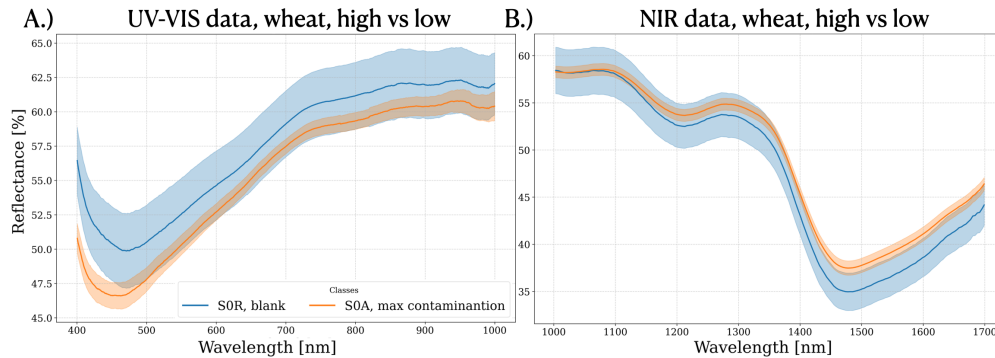


Figure 3.11: Mean reflectance spectra with standard deviation bands for the most contaminated (S0A) and least contaminated (S0R) wheat samples. A.) presents the UV–VIS range, while B.) shows the NIR range.

In addition to analyzing the wheat dataset independently, we also compared the reflectance profiles of the two matrices—wheat and corn. Figure 3.12 illustrates this comparison, showing the pure DON corn sample (S02) alongside the most heavily contaminated wheat sample (S0A). The wheat sample exhibits significantly higher reflectance across the entire spectral range. In

the VIS region especially, this is consistent with visual observations: the corn sample appeared more yellow and dull, whereas the wheat sample was noticeably whiter and brighter. Aside from overall intensity differences, the positions of the spectral minima and maxima were found to be similar between the two samples.

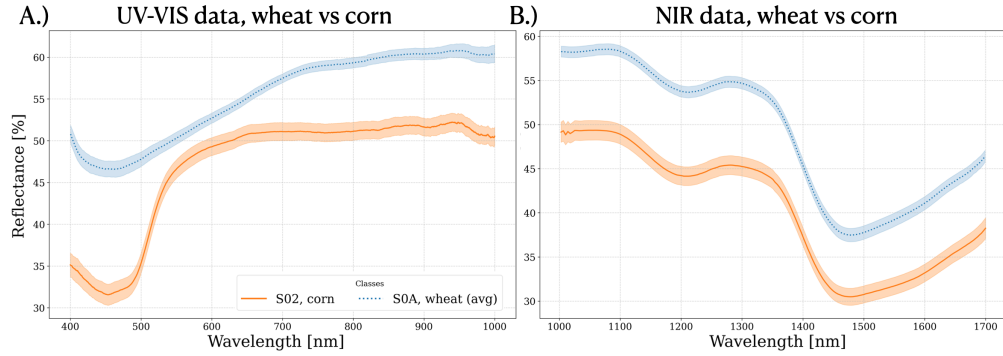


Figure 3.12: Mean reflectance spectra with standard deviation bands for corn sample S02 and wheat sample S0A. A.) shows the UV-VIS range, and B.) displays the NIR range. A clear distinction between the two spectra is evident, primarily due to a consistent baseline offset across the entire wavelength range. The positions of the maxima and minima are approximately aligned.

3.3.2 Large Sphere Reflectance Measurements

Along with the localized reflectance measurements, global reflectance measurements were also performed using the large integrating sphere setup and protocol described in Section 3.3.2. These measurements were conducted exclusively for the corn samples, aiming to investigate the impact of localized versus global measurements and assess whether illuminating only a small area is sufficient to capture reliable reflectance data for model training.

Figure 3.13 shows the resulting spectra (mean and std) for the pure toxin samples S01-S03 across both the UV-VIS (Plot A) and NIR (Plot B) spectral regions. Comparing these results with the small sphere measurements in Figure 3.9, we see that the overall spectral shapes are similar.

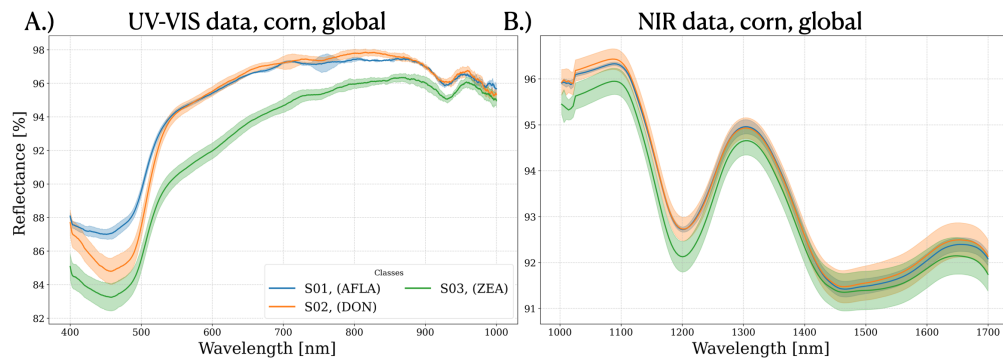


Figure 3.13: Large sphere reflectance spectra (mean with std) for the pure toxin samples: S01-S03. A.) the UV-VIS range and B.) the NIR range. Spectral differentiation is possible and most strongly pronounced in the UV-VIS range.

However, a much stronger difference is evident for the ZEA sample in the UV–VIS range under full-vial measurements. In the NIR region, ZEA also stands out more clearly, while DON and AFLA show more overlap. The most pronounced differentiation between samples appears in the UV region, where distinct variations in reflectance are strongest. Overall, the global measurements provide easier visual differentiation compared to the localized measurements.

3.3.3 Fluorescence Measurements

An extensive set of fluorescence measurements was conducted for both the corn and wheat samples. The setup and measurement procedures used are described in Section 3.2.2. As discussed earlier, a central challenge for in-situ mycotoxin detection, using fluorescence spectroscopy is selecting the optimal excitation wavelength—one that maximizes the fluorescence response of the toxins while minimizing background interference from the food matrix. Since we aim to detect multiple toxins simultaneously—a task not commonly addressed in the literature—it was necessary to investigate the most effective excitation wavelengths for each toxin. To this end, measurements were conducted at 790 and 750nm (two-photon excitation), as well as at 395, 385, 375, 365, and 256nm (one-photon excitation).

In the following sections, we present the results for both modalities, discussing one-photon fluorescence first, followed by two-photon fluorescence.

3.3.3.1 One-Photon Fluorescence

We began by exploring a range of excitation wavelengths for conventional single-photon fluorescence, aiming to optimize spectral contrast and sensitivity in mycotoxin detection. To this end, we examined fluorescence characteristics under varying excitation conditions.

Guided by promising results from previous studies [26, 30], we selected excitation wavelengths of 395, 385, 375, 365, and 256nm. While the narrow spacing between these wavelengths likely activates similar electronic transitions, subtle spectral shifts may still occur. These shifts could reveal additional sample-specific information and offer slight improvements in discriminative performance.

To ensure that comparisons reflected the influence of excitation wavelength rather than excitation power, we standardized the optical laser power at approximately 105mW for all excitation wavelengths, except for 256nm, where a reduced power of 50mW (or less) was used due to source limitations. The actual power levels monitored during the measurements are summarized in Table 3.2.

Wavelength [nm]	Corn Samples		Wheat Samples	
	Mean Power [mW]	Std [mW]	Mean Power [mW]	Std [mW]
395	105.6	3.7	108.3	4.3
385	104.9	9.5	103.8	1.2
375	104.6	5.4	107.0	0.2
365	107.2	11.0	104.3	1.4
256	50.3	2.3	18.9	1.8

Table 3.4: Measured power of the excitation beam during the fluorescence measurements, determined by measuring the beam power between every three samples measured. The power for the third harmonic is significantly lower due to lower conversion efficiency.

The resulting spectra (mean and std) per sample, are presented in Figures 3.14 and 3.16 for corn and wheat, respectively. In both figures we display the spectra for all excitation wavelengths explored (Plot *A*: 395nm, *B*: 385nm, *C*: 375nm, *D*: 365nm, *E*: 256nm).

We observe that as the excitation wavelength decreases, the fluorescence peak shifts progressively toward shorter wavelengths, consistent with the expected Stokes-shift. At the same time, substantial spectral overlap is evident across the datasets, highlighting the need for chemometric and machine learning techniques to achieve reliable differentiation in both classification and regression tasks.

Comparing the wheat and corn data, we see distinct differences: corn samples exhibit a broader emission peak with multiple small sub-peaks in the 450–550nm range, whereas wheat samples display a smoother, single peak within the same region. A clear chlorophyll peak is also visible around 700nm in both sample types, particularly at longer excitation wavelengths, and is more pronounced in wheat—likely reflecting a higher chlorophyll content compared to corn [26].

We can also observe that excitation at 256nm resulted in a noticeably weaker emission spectrum for both wheat and corn, due to the lower pump power used (approximately 50mW for corn and 20mW for wheat). For wheat in particular, this low excitation power led to a noisy spectrum. For both matrices, the spectra at 256nm show considerably more overlap than at the other excitation wavelengths, suggesting weaker discriminatory potential at this wavelength.

In addition to showing all spectral data in a single plot—which is difficult to interpret due to the large amount of overlapping spectra—we also present the spectra of just the three pure toxin samples: S01 (AFLA), S02 (DON), and S03 (ZEA). These are displayed in Figure 3.15 across all excitation wavelengths investigated. We observe that for excitations between 395 and 365nm (Plots *A–D*), the three samples are readily distinguishable, most obviously in the 400–550nm range, whereas at 256nm, differentiation becomes much less clear, inline with the trend observed prior.

Across all excitations, the DON-contaminated corn shows the weakest fluorescence signal, as expected since DON is non-fluorescent [49]. The remaining fluorescence signal can be attributed to the intrinsic fluorescence of the corn matrix. In contrast, the ZEA and AFLA samples display much stronger signals, indicating that the toxins themselves likely contribute directly to the fluorescence response. As a result, DON stands out most prominently compared to ZEA and AFLA. Considering an excitation of 395nm (Figure 3.15*A*) provides the clearest overall separation between all three toxins, while an excitation of 375nm (Figure 3.15*C*) most strongly differentiates the DON sample from AFLA and ZEA.

For AFLA on corn at 365nm excitation, literature reports a peak emission range of 400–500nm, with a maximum around 440nm [30, 38, 39], which aligns well with our observation (Figure 3.15*D*). For ZEA peak emission occurs at 450nm when excited at 314nm [50]. Although the excitation wavelength used in that study differs from those explored here, the emission peak observed under 365nm excitation closely matches these reported values.

We similarly examined the wheat samples in a simplified setting by focusing only on the spectra of the most, middle, and least contaminated samples to assess visual differentiation. Figure 3.17 presents this spectral data, using samples S0A (OTA: 8.3 μ g/kg, DON: 1000 μ g/kg, T-2: 49.0 μ g/kg, HT-2: 8.7 μ g/kg), S0I (OTA: 4.2 μ g/kg, DON: 500 μ g/kg, T-2: 24.5 μ g/kg, HT-2: 4.4 μ g/kg), and S0R (all: 0 μ g/kg), shown across all excitation wavelengths.

We observe that the clearest distinctions are achieved after excitation using 365 and 375nm, with moderate separability for excitation using 395nm. However, considering an excitation at 385nm and especially 256nm, spectral overlap hinders discrimination.

Interestingly, we also observe that the less contaminated samples exhibit stronger fluorescence signals. This is unexpected, given that the samples contain a mix of DON, OTA, T-2, and HT-2, with OTA in particular being a known fluorescent toxin that typically shows maximal emission around 460nm [44]. This suggests that the fluorescence signal may not be driven directly by the toxins themselves but rather by their effects on the wheat matrix—serving as an indirect proxy for mycotoxin presence—or by differences in matrix composition specific to the standards, unrelated to mycotoxin content.

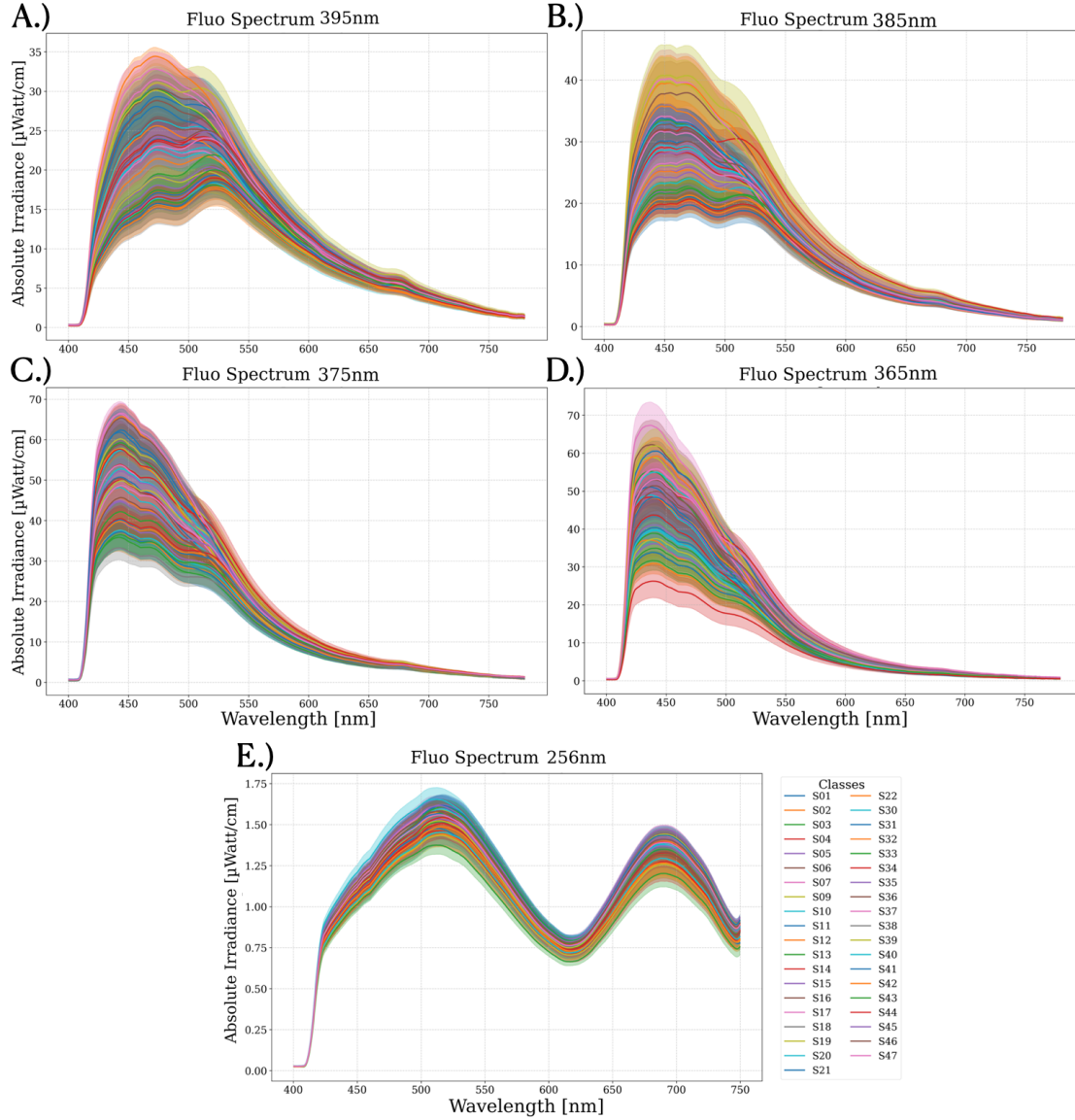


Figure 3.14: Fluorescence spectra (mean and std) for all **corn** samples obtained using excitation wavelengths of A.) 395, B.) 385, C.) 375, D.) 365, and E.) 256nm.

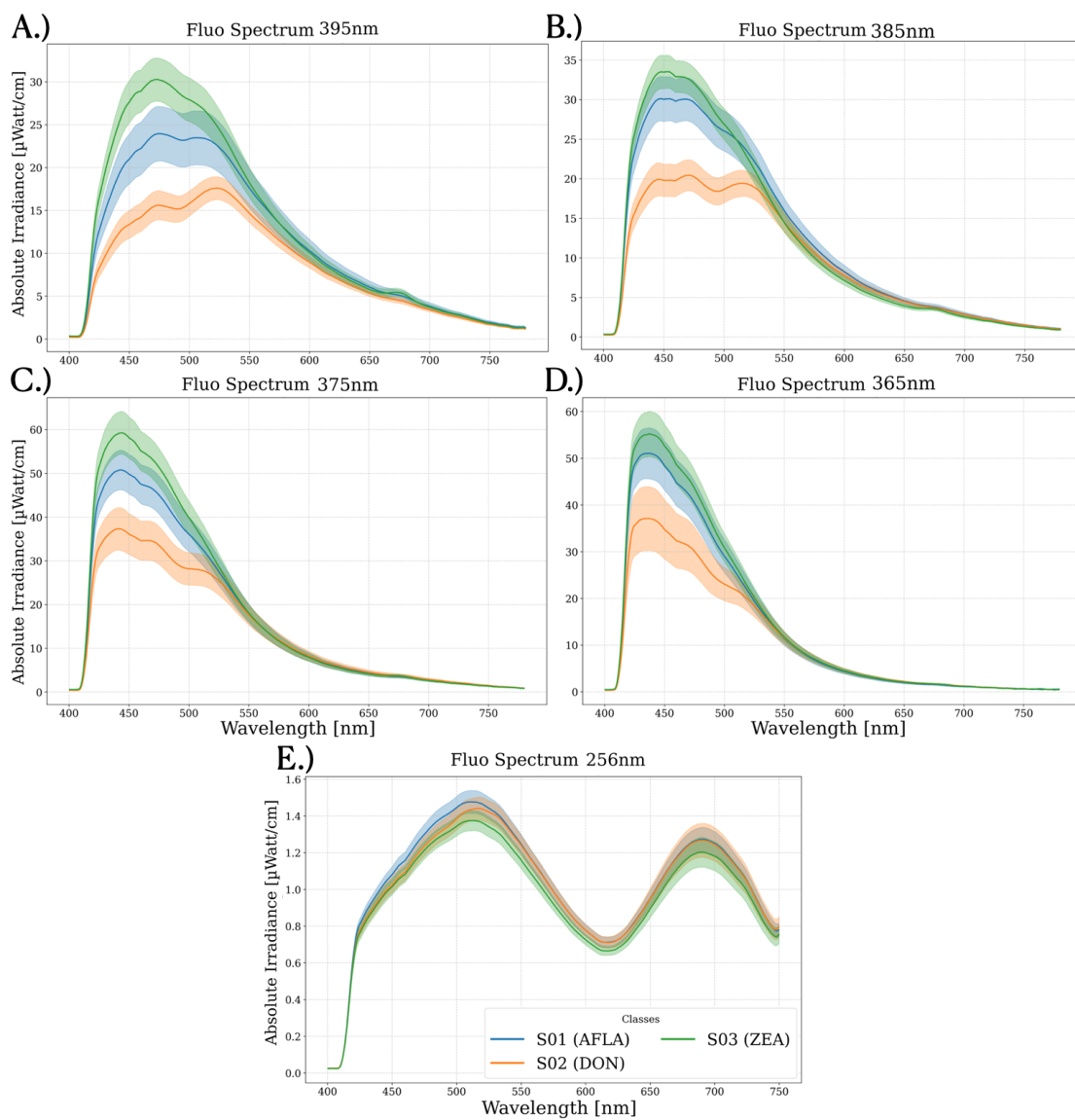


Figure 3.15: Fluorescence spectra (mean and std) for the pure-toxin **corn** samples (S01-S03), measured at excitation wavelengths of A.) 395, B.) 385, C.) 375, D.) 365, and E.) 256nm.

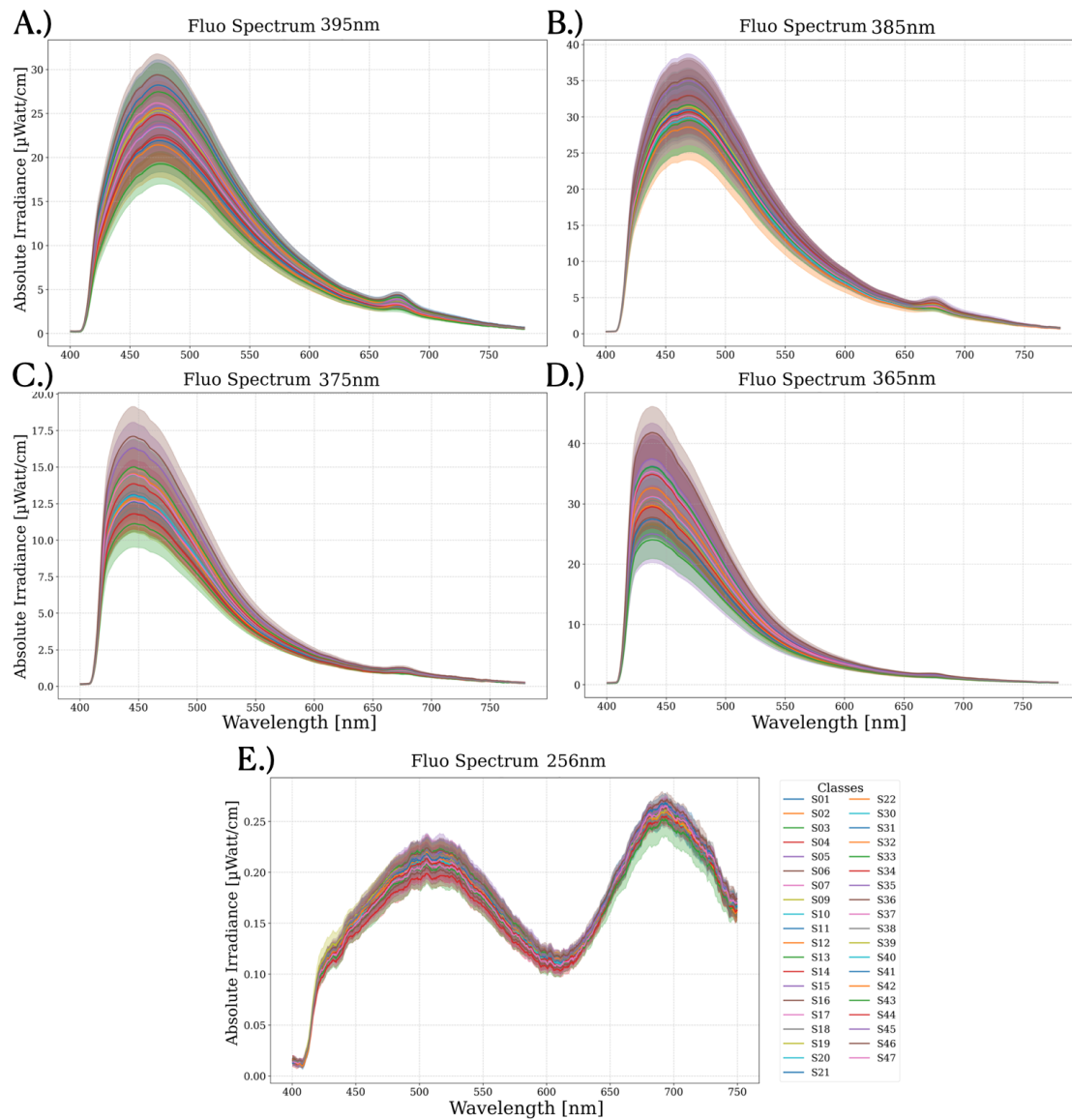


Figure 3.16: Fluorescence spectra (mean and std) for all **wheat** samples obtained using excitation wavelengths of A.) 395, B.) 385, C.) 375, D.) 365, and E.) 256nm.

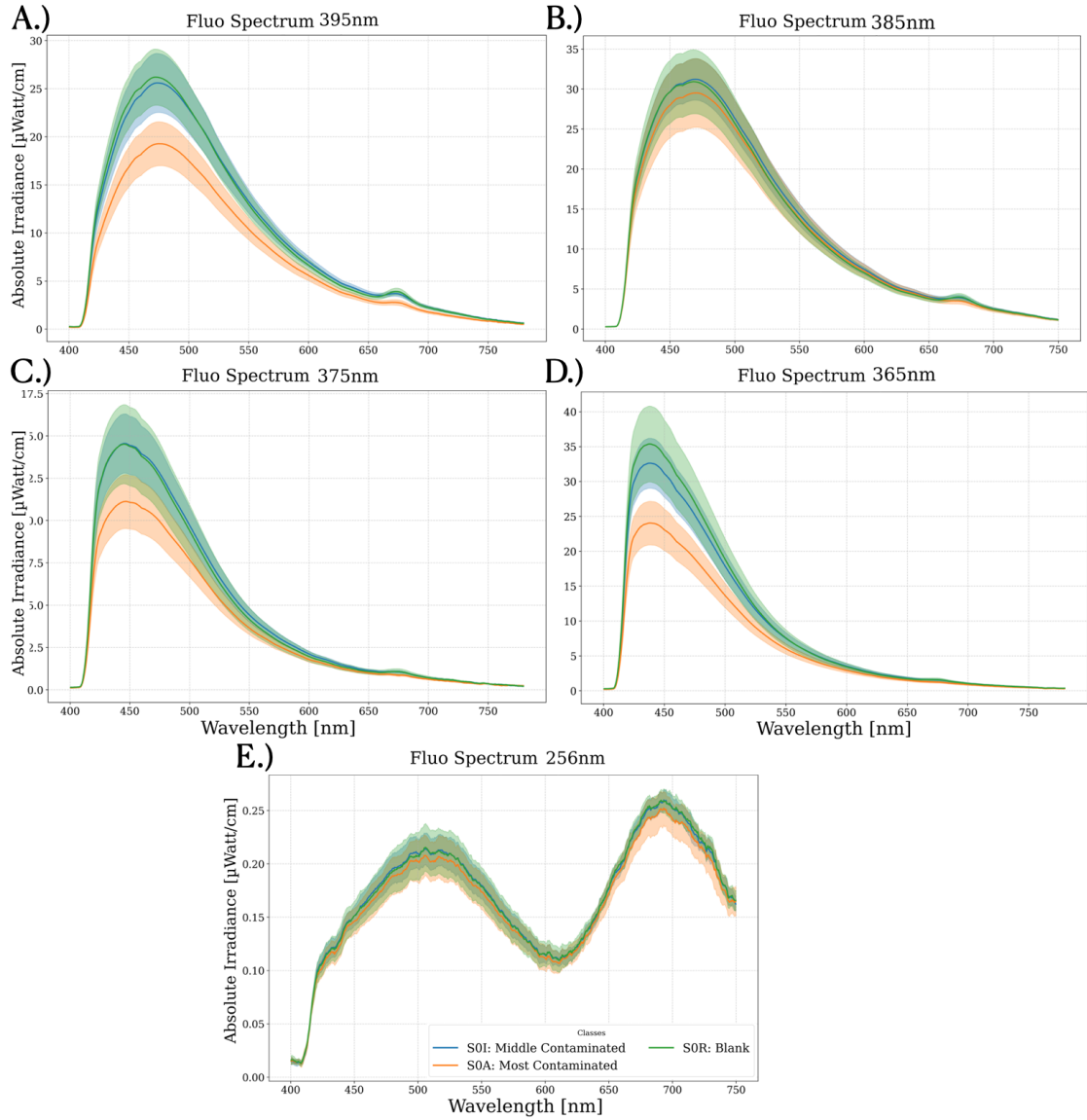


Figure 3.17: Fluorescence spectra (mean and std) for **wheat** samples S0A (highest contamination), S0I (middle contamination), and S0R (blank), at excitation wavelengths of A.) 395, B.) 385, C.) 375, D.) 365, and E.) 256nm.

Lastly, Figure 3.18 presents a direct comparison of corn and wheat fluorescence spectra under 395nm excitation. Wheat sample S0A (with the highest DON concentration at $1000\mu\text{g/kg}$) is compared to corn samples S21 and S25, which have similar DON levels of $1160\mu\text{g/kg}$ and $900\mu\text{g/kg}$, respectively. The wheat spectrum features a single sharp peak followed by a gradual decay, while the corn samples show a delayed onset and a distinct double-peak structure. Despite comparable DON concentrations, the spectral profiles differ markedly, reflecting differences in matrix composition.

These results further emphasize the strong matrix dependence of the fluorescence signals, even when toxin levels are aligned, highlighting the complexity associated with developing a single detection model capable of operating on different matrices.

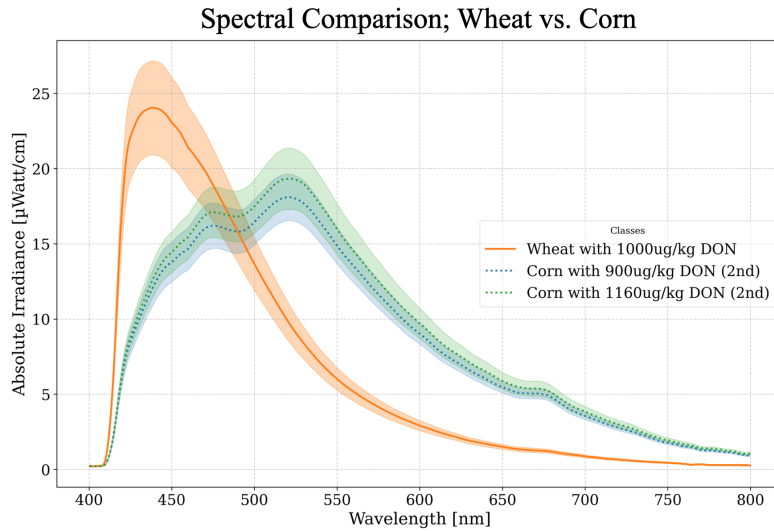


Figure 3.18: Comparison of 395nm fluorescence spectra for **wheat** sample S0A (highest DON concentration) and **corn** samples S21 ($1160\mu\text{g/kg}$) and S25 ($900\mu\text{g/kg}$) with similar DON levels. Despite comparable contamination levels, the spectra show substantial differences in overall shape, illustrating the strong influence of the sample matrix on the fluorescence response.

3.3.3.2 Two Photon Fluorescence

Besides one-photon fluorescence we also collected two-photon spectral data for a subset of the corn samples. Initial measurements were conducted at an excitation wavelength of 790nm with an excitation power of 1.4Watt. The resulting spectra (mean and std) for sample S01 (pure AFLA) are shown in Figure 3.19, Plots *A* and *B*. We initially focused on AFLA because two-photon fluorescence has previously been successfully applied for AFLA detection on corn kernels by Smeesters et al. [34].

In Plot *A*, which shows the full spectral range from 350 to 900nm, a dominant peak appears at 790nm. This peak corresponds to the excitation beam itself, indicating that some excitation light was reflected off the sample and leaked through the filter. No additional filter was available to further suppress or eliminate this leakage.

To gain clearer insight into the true fluorescence spectrum, we focused on the restricted range of 350–750nm—excluding the peak caused by leakage—as shown in Plot *B*. We ensured that no

detector saturation occurred during the measurements. Looking at this spectrum, we observed a noisy fluorescence signal with prominent maxima in the 550–650nm range, and some additional peaks around 350nm. Despite applying increased pixel smoothing for two-photon fluorescence (set to 36 pixels, compared to 6 for one-photon fluorescence), the signal remained noisy.

We can compare these spectra to those reported by Smeesters et al. [34]. Notably, we observe that the dominant peak region in our measurements is shifted to longer wavelengths compared to their reported emission maximum of $(503 \pm 19)\text{nm}$ for a 780nm excitation at 1510mW.

A second measurement was also conducted for the same sample using an excitation of 750nm. The corresponding spectra are shown in Plots *C* and *D* of Figure 3.19, again presenting both the full range and informative range (excluding the excitation beam leakage), respectively. Notably, the fluorescence signal was more pronounced and less noisy at this excitation wavelength, but looked very similar besides the reduced noise level.

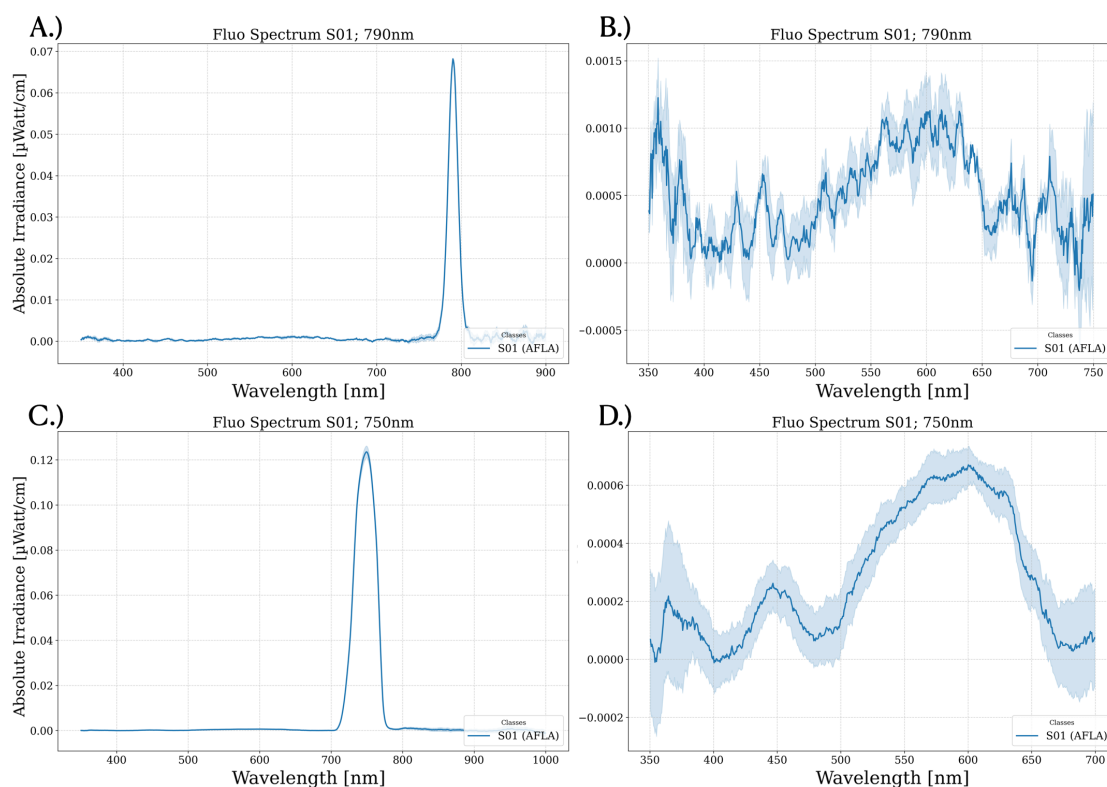


Figure 3.19: Two-photon fluorescence spectra for AFLA (sample S01) obtained with 790nm excitation (top row) and 750nm excitation (bottom row). *A.*) and *B.*) display the spectra with and without the excitation peak range for 790nm, while *C.*) and *D.*) present the corresponding spectra for 750nm excitation. A fluorescence signal can be seen, though it is weak and noisy.

As a final step, we also investigated whether two-photon fluorescence at 750nm—the excitation wavelength that previously performed best—could be used to differentiate between samples. To assess feasibility, we collected spectral data for the pure toxin standards S01 (AFLA), S02 (DON), and S03 (ZEA). The results (mean and standard deviation) are shown in Plot *A* of

Figure 3.20. We observed some differentiation, particularly in the UV range (350–400nm), with ZEA exhibiting the most distinct signal compared to the other two mycotoxins. However, overall, the spectra displayed a high degree of overlap.

Because two-photon fluorescence is highly dependent on excitation power, we also tested whether enhancing the signal using a focusing lens—thereby increasing illumination intensity—would improve differentiation. The spectra obtained with this focused setup are shown in Plot *B* of Figure 3.20. Here, a much clearer distinction between the samples is visible, with the most prominent differentiation occurring in the 550–650nm peak region. These results indicate that two-photon fluorescence, when using very high excitation powers, is capable of distinguishing between different mycotoxin types.

To conclude we can say that, while these results confirm that two-photon fluorescence can, in principle, be used for mycotoxin detection and discrimination, its practicality for real-world applications is limited. In our case, the concentration of AFLA tested was much lower than in previous studies, which contributed to the observed detection challenges. Only by using extremely high integration times (seconds) and irradiance (using the focusing lens) did we obtain distinguishable results. Such high irradiance requirements make this approach less suitable for in-line or in-field industrial applications—especially for handheld devices—thereby limiting its applicability for detecting AFLA levels close to the legal limit. For this practical reason, we do not consider two-photon fluorescence further in the remainder of this study.

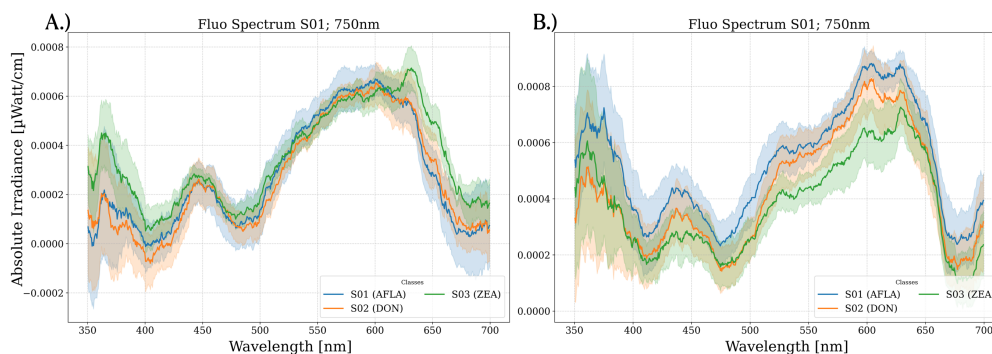


Figure 3.20: Two-photon fluorescence spectra for AFLA, DON, and ZEA (samples S01–S03) obtained using 750nm excitation. *A.*) shows the spectra without a focusing lens, while *B.*) presents the spectra with a focusing lens, demonstrating that increased illumination intensity improves the optical contrast.

3.4 Summary Of The Chapter

In this chapter, we detailed the experimental framework developed to investigate the feasibility of multi-mycotoxin detection in corn and wheat using optical spectroscopy. A diverse series of artificially synthesized, yet naturally contaminated samples was prepared using certified reference materials, allowing for precise control over toxin concentrations while maintaining matrix realism. For corn, 40 samples were composed from four standards (AFLA, DON, ZEA, and the mixed standard), while 18 wheat samples were generated using one contaminated and one blank standard.

Reflectance and fluorescence spectra were collected using purpose-built setups, covering the

UV–VIS–NIR range for both local and global reflectance, and fluorescence measurements at various excitation wavelengths, considering both one- and two-photon induced fluorescence.

Several key insights emerged from the collected spectral data. First, reflectance measurements across both crops revealed substantial spectral overlap between samples, with only modest visual differences. The NIR data (1400–1700nm) did provide great differentiation for the wheat samples, but for corn samples even the pure-toxin samples were hard to differentiate.

Second, the fluorescence measurements revealed distinct differences between samples, particularly within the 395–365nm excitation range. In corn samples, DON consistently produced the weakest fluorescence signal, consistent with its non-fluorescent nature, whereas ZEA and AFLA exhibited markedly stronger emissions. In wheat, fluorescence intensity did not correlate directly with contamination levels, indicating the detection of indirect, matrix-dependent effects that varied by standard. Excitation at 256nm yielded the weakest and noisiest signals overall, failing to produce clear differentiation between samples.

Finally, while two-photon fluorescence confirmed the theoretical feasibility of AFLA detection—especially after signal enhancement via a focusing lens—the practical utility of this modality was limited. Weak, noisy signals and the need for high excitation power and long integration times make two-photon fluorescence unsuitable for real-world, in-line, or handheld applications.

For completeness, Table 3.5 summarizes the spectral regions for both corn and wheat that visually exhibit the most pronounced differentiation between the various contamination types and levels. These regions might hold the most informative spectral information for model development, and feature selection, and this table can serve as a reference point. Cells in green indicate the data that is visually the most promising.

Dataset	Corn [nm]	Wheat [nm]
Fluo 395	420–550, 670	420–700
Fluo 385	420–530	670
Fluo 375	420–520	420–500
Fluo 365	420–500	420–500
UV-VIS	400–700	400–500
NIR	1000–1700	1400–1700

Table 3.5: Spectral regions visually showing the clearest differentiation between sample contamination types and levels, for both the corn and wheat datasets. For fluorescence, 670nm refers to the chlorophyll peak.

Chapter 4

Classification Model Development

This chapter presents the complete development and best performance of a Python-implemented optical spectroscopy-based classification model build to differentiate between various types of mycotoxin contamination on both corn and wheat. The chapter aims to: (i) clearly outline the methodological framework—including preprocessing, algorithm selection, and feature selection—and (ii) benchmark model performance against existing literature.

The chapter opens by revisiting the project’s starting point, describing the preliminary classification approach in place prior to this thesis, in Section 4.1. The performance of this early model is critically assessed, revealing key limitations that ultimately prompted a fundamental reevaluation of the model development and validation strategy.

Building on these insights, the chapter proceeds with a multidimensional optimization process to identify the most effective spectroscopic modality, preprocessing technique, and classification algorithm for each toxin–matrix combination. Once optimal models are established, feature selection is employed to reduce data dimensionality and simplify model complexity, with a focus on balancing predictive performance and system interpretability.

4.1 From Proof-of-Concept to Robust Validation

To set the stage for the model development work presented in this chapter, it is imperative to first summarize the project’s initial state. At the onset, Ir Magnus and Prof. Dr. Smeesters, in collaboration with Prof. Dr. Venturini from Zurich University of Applied Sciences, had already conducted preliminary experiments aimed at assessing the feasibility of detecting multiple mycotoxins using spectroscopic data. Their initial approach focused on a subset of the corn samples, which were grouped into seven distinct classes to reflect different contamination profiles. As detailed in Table 4.1, the classification scheme included single-toxin samples—pure AFLA (class 1), pure DON (class 2), and pure ZEA (class 3)—as well as mixtures: AFLA+DON (class 4), AFLA+ZEA (class 5), ZEA+DON (class 6), and a final group containing all three toxins (class 7). This preliminary framework was designed as a proof of concept to test whether a classification model could distinguish between simple and mixed contamination scenarios, allowing for multi-mycotoxin discrimination.

7-Class Distribution		
Class	Samples	Toxins
1	S01	AFLA
2	S02	DON
3	S03	ZEA
4	S04, S05	AFLA & DON
5	S06, S07	AFLA & ZEA
6	S08, S09	DON & ZEA
7	S10, S11, S12	AFLA, DON & ZEA

Table 4.1: Sample distribution for the initial 7-class classification scheme aimed at assessing the feasibility of developing a multi-mycotoxin detection framework.

Using this 7-class framework, a Linear Discriminant Analysis (LDA)¹ model was trained, using one-photon fluorescence data. The performance was evaluated via a Cross-Validation (CV) strategy. CV was selected to maximize use of the limited dataset while ensuring that validation was performed on unseen spectra, thus spectra that were not considered during the training. The used CV-scheme comprised four folds where 75% of the data (spectra) were allocated for training and 25% for validation, rotating the split four times so that every spectrum was used for validation exactly once.

Within this initial processing, as a key limitation was that the spectral data were randomly distributed across the folds, meaning that spectra from the same sample could appear in both the training and validation sets. This validation strategy was suboptimal resulting in data leakage, and mainly assessed the model’s ability to generalize across repeated measurements of the same physical samples, rather than testing its generalizability on entirely unseen samples—a critical flaw and a key aspect that will be tackled in this thesis enhancing the reliability of these results.

4.1.1 Reproducing The Initial Results

The initial step of this thesis project involved reproducing the existing classification pipeline developed by Magnus et al., with the primary objective of independently validating their earlier results. This provided a valuable opportunity to gain familiarity with the core aspects of the project—ranging from the underlying coding framework to the implementation of classification models and the nested-CV strategy as well as to benchmark the novel developed data-processing coding script in Python with the previously used data processing Matlab script of Indy Magnus.

A comparison between the initial Matlab spectral processing, and the newly developed Python processing script is presented for the fluorescence data, considering 395 and 385nm as excitation wavelengths, in Figure 4.1. Here we display two 7-class classification tasks. The data used comprised a set of fluorescence spectra, 5 per sample for each excitation, so 60 spectra total for an excitation of 395nm and similarly for 385nm. The spectra were preprocessed by trimming the range to 400–800nm, as indicated visually in the leftmost plots. The second and third plot show the 7-class Confusion Matrix (CM) achieved when using the new Python processing code and the previous Matlab script, respectively. Each CM visualizes the predicted versus actual class distribution: the highlighted diagonal elements indicate correct classifications, while off-diagonal elements represent misclassifications.

From these results, we observe that the 395nm fluorescence data yielded the highest classification performance, achieving an accuracy of $(65.3 \pm 1.7)\%$ —notably higher than the preliminary result of 52% reported by Magnus et al. The uncertainty represents the standard deviation across the four cross-validation folds, capturing variability due to data partitioning. In comparison,

¹LDA was selected based on their proven effectiveness in prior studies on spectroscopic mycotoxin detection in food matrices [30, 34, 125].

excitation at 385nm produced an accuracy of $(56.9 \pm 11.9)\%$, which is consistent within one standard deviation of the 41% obtained using the original MATLAB script. The performance differences and relatively large standard deviation are attributed to the limited training data and differences in random partitioning of spectra between the two implementations. Nonetheless, the overall classification trends were found to be consistently similar between both scripts confirming the successful reproduction of the original results.

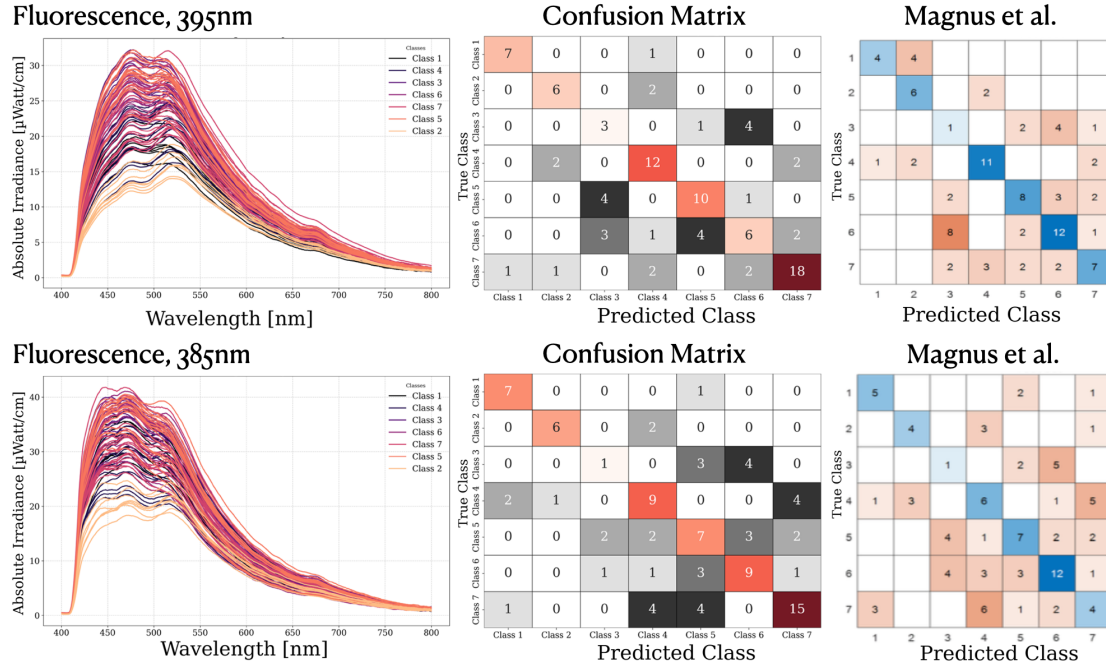


Figure 4.1: Classification performance of the initial LDA model for the 7-class problem defined in Table 4.1, using fluorescence data at 395 (top row) and 385nm (bottom row). For comparison, the original results obtained by Magnus et al. are also included, demonstrating closely matching performance.

4.1.2 Extending To Reflectance Data

After successfully reproducing the classification pipeline and validating the results for the fluorescence measurements, the next step was to extend the framework to include the UV-VIS and NIR reflectance measurements—both of which had not been investigated at the outset of this thesis. With only minor modifications (i.e. modifying the spectral input) to the existing classification pipeline, and using the same 7-classes, this could be implemented. The results obtained for the UV-VIS and NIR data are shown in Figure 4.2, in the top and bottom row respectively. Both spectral regions yielded comparably strong full-spectrum classification performance, achieving accuracies of up to $(87.7 \pm 9.4)\%$. This performance significantly exceeded that of the fluorescence-based models, indicating that reflectance data offered a more informative route for classification.

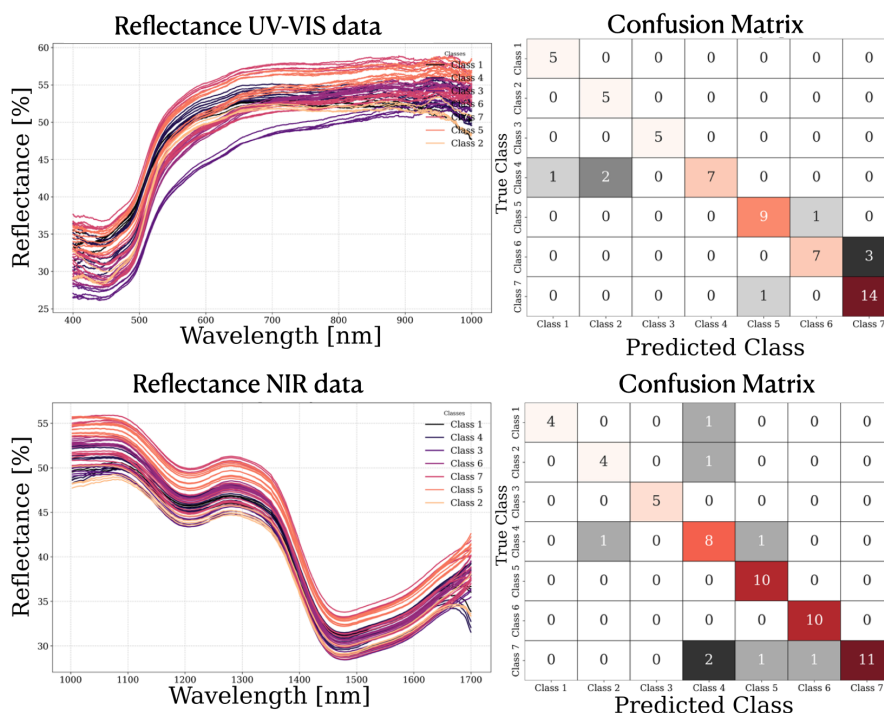


Figure 4.2: Classification performance of the initial model using LDA with CV on the reflectance data. The top row shows results for the UV–VIS spectral range, while the bottom row displays results for the NIR range. Both modalities significantly outperformed the fluorescence-based models (see Figure 4.1), highlighting the stronger discriminative power of reflectance spectroscopy for this classification task.

4.1.3 Reproducibility

The results presented in the previous sections demonstrated strong classification performance, particularly for the reflectance data. To further evaluate the model’s robustness and reproducibility, a new set of measurements was collected, now comprising 15 spectra per sample (as described in Chapter 3), for both fluorescence and reflectance modalities. Using this expanded dataset, a new LDA model was trained to assess reproducibility.

An example of the model performance for the UV–VIS reflectance data is shown in Figure 4.3. Here we can compare the original model performance (Plot A) to the new model trained on the new, expanded dataset (Plot B). Notably, the classification accuracy improved from (86.7 ± 9.4) to $(95.8 \pm 1.0)\%$.

This substantial performance improvement, mostly in terms of the reduced standard deviation, is likely attributable to the larger training dataset—15 spectra per sample versus the previous 5—which provided a stronger statistical foundation for model learning. Similar results were found using the NIR data and fluorescence data, indicating the benefit of using larger datasets for model development.

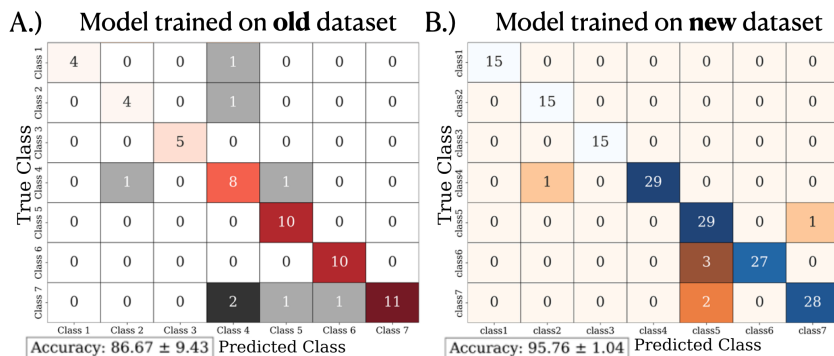


Figure 4.3: CM for UV–VIS-based 7-class classification using an LDA model. Plot A shows the performance of the original dataset, while Plot B presents the results obtained from the newly collected expanded dataset, indicating a significant improvement.

4.1.4 Generalization Challenges And Need For LOSO-CV

To assess the model’s ability to generalize to entirely unseen samples, a new data processing strategy was investigated. In this setup, a model was trained on the full dataset excluding one sample, and then tested on the spectral data of the left-out sample—ensuring true sample-level independence between training and testing. To maintain at least two samples per class in the original 7-class classification scheme, three additional pure samples were introduced: S23 (AFLA), S24 (DON), and S25 (ZEA), corresponding to classes 1, 2, and 3, respectively.

Although the models previously achieved promising CV accuracies of up to 96%, their performance under the leave-one-sample-out strategy painted a very different picture. In this setting, the model consistently failed to correctly classify unseen samples, highlighting severe overfitting: rather than learning to detect toxin-specific spectral patterns, the model had effectively memorized the spectral signatures of individual samples. This finding underscores a critical distinction—**measurements taken at different locations on the same physical sample cannot be considered equivalent to measurements on entirely new samples**. This key limitation had been overlooked in earlier approaches, where standard CV implicitly assumed independence between repeated measurements.

Figure 4.4, Plots A–C illustrate the classification of using the leave-one-out strategy on a particular samples (15 spectra per sample), representative instances of this lack of generalization. Figure 4.4 D shows the aggregate of this procedure repeated for every sample as test sample. We can see that the overall classification accuracy dropped sharply to just 8.5%², compared to the 95.8% achieved using standard, spectra-based CV.

The classifications are however, not entirely random, for example, class 1 (AFLA) is often misclassified as classes 4, 5, or 7—all of which also contain AFLA, suggesting the model correctly detects AFLA but fails to distinguish its presence in isolation. Similarly, class 2 (DON) is frequently predicted as class 4 (AFLA + DON), indicating that while the toxin is detected, the model confuses single-toxin cases with mixed-toxin profiles—making the misclassifications somewhat understandable.

²The main issue here is, of course, that there are only two samples per class, which was also a reason why standard CV was used in earlier analyses.

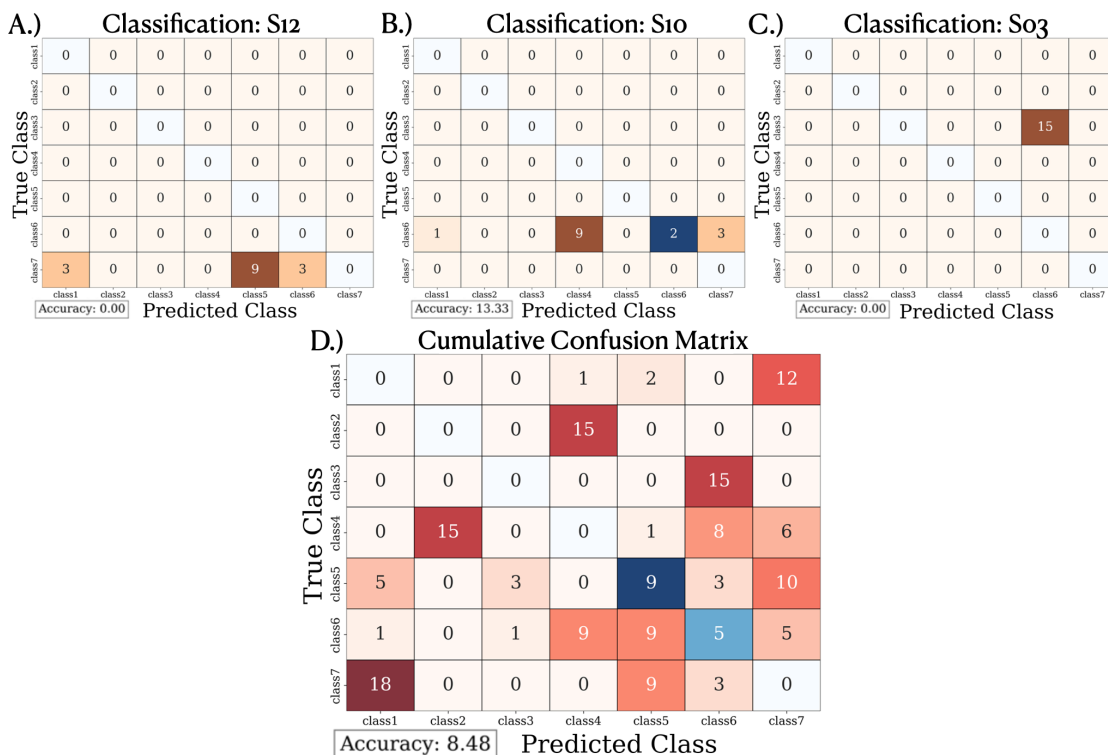


Figure 4.4: Plots A, B, and C illustrate the performance of LDA-models trained on all samples except one, which was then used as the validation sample. Plot D presents the aggregated performance across all such leave-one-sample-out validations, with each sample serving as the validation set exactly once.

To understand the root cause of this issue, it is crucial to revisit how the original cross-validation procedure was implemented. In that setup, the dataset—comprising all spectra—was randomly split into training and validation sets for each fold, without accounting for the fact that many spectra originated from the same physical sample. As a result, spectra from the same sample appeared in both the training and validation subsets. This introduced a critical flaw: the model was effectively being trained and tested on data derived from the same samples. Consequently, instead of learning generalizable spectral patterns linked to mycotoxin contamination, the model simply memorized sample-specific spectral characteristics. This “*leakage*” undermines the model’s ability to identify truly informative features related to contamination and explains its failure to accurately classify unseen samples during sample-independent testing.

To address this issue and ensure that the model learns to classify based on mycotoxin contamination rather than memorizing sample-specific spectral signatures, it is necessary to revise the validation strategy. A powerful alternative emerged from the very experiment that exposed the model’s limitations: the adoption of Leave-One-Sample-Out Cross-Validation (LOSO-CV), the theoretical foundations of which are discussed in Section A.4.1.

In LOSO-CV, the model is trained on the entire dataset except for one sample, which is held out for validation. This process is repeated iteratively such that each sample serves as the validation set exactly once. By validating on entirely unseen samples, LOSO-CV provides a much

more realistic estimate of the generalization performance. It explicitly prevents the model from overfitting to sample-specific characteristics and encourages the learning of generalizable spectral features linked to mycotoxin presence, rather than simply capturing variability within the corn matrix. Additionally, this method maximizes the use of available data for training—an important advantage given the limited number of samples.

For the remainder of the thesis, all model optimization and evaluation will be conducted using this Leave-One-Sample-Out Cross-Validation (LOSO-CV) strategy. More robust alternatives—such as nested cross-validation at the sample level or the use of an external hold-out set—would provide an even stronger assessment of model performance, however, these approaches are impractical here due to the limited amount of available data.

4.2 Classifier Development

With the LOSO-CV scheme established as the standard for model evaluation, the next phase of this work focuses on the systematic development and optimization of classification models. We opted to prioritize classification over regression initially, as this approach not only dominates the current scientific literature but also offers the most promising route toward practical, regulation-compliant systems.

Due to the structured synthesis of the corn samples—with independent mixtures of AFLA, DON, ZEA, and FUM³—initial model development focused exclusively on the corn dataset, the results of which are presented in Section 4.3. In contrast, the wheat dataset was composed using only two standards: one multi-mycotoxin contaminated and one blank. As a result, models developed for wheat samples inherently reflect the simultaneous presence of DON, T-2/HT-2, and OTA, without the ability to isolate the spectral contributions of each individual toxin. Classification of the wheat data is examined in Section 4.4, followed by the development of combined classifiers for DON detection across both wheat and corn, in Section 4.5.

The previously introduced 7-class classification scheme—which differentiated between samples containing AFLA, DON, ZEA, or their various binary and ternary combinations—was primarily used as a feasibility test. While it offered valuable insight into the spectral separability of multi-mycotoxin profiles, it is not directly applicable for regulatory use, where models must evaluate toxin concentrations in relation to established legal thresholds. From a food safety perspective, binary classification is of key interest, determining whether the contamination of all toxins falls below or above the respective legal limits.

An even better image can be ascertained using a three-class scheme. In this case, samples are classified as safe (clearly below the legal limit), risky (within a buffer zone around the limit), or unsafe (exceeding the limit). This “risky zone” serves as a tunable buffer that reduces false negatives (i.e., unsafe samples misclassified as safe) at the expense of increasing false positives (i.e., safe samples flagged as risky). The width of this buffer zone can be adjusted to balance safety concerns with food waste.

Implementing a three-class model for the direct detection of multiple mycotoxins poses a significant challenge, both due to the composition of the samples and the limitations of using a single classifier for simultaneous detection.

³FUM was not entirely independent in this dataset, as it was only available in a standard also containing AFLA, DON, and ZEA. As a result, FUM always co-occurred with these toxins. However, by mixing this standard with others containing varying concentrations of AFLA, DON, and ZEA, we were able to manipulate the relative ratios between the toxins. This approach was intended to prevent the model from simply learning the combined presence of multiple toxins, and instead encourage it to identify spectral features specific to FUM. Decoupling FUM from the specific matrix composition was, however, not possible, as only one standard with FUM was available.

Considering corn, no mycotoxin-free (blank) standard was available, making it difficult to populate the “safe” class, as reducing the concentration of one toxin through mixing inevitably increased the relative concentration of others. Second, training a single classifier to categorize samples as safe, risky, or unsafe provides only a broad contamination status. It fails to indicate which specific toxin exceeds the regulatory threshold and cannot distinguish cases where the classification is driven by the presence of a single toxin or the combined effect of multiple contaminants.

To address this, we adopted a modular classification strategy. Rather than building a single model to evaluate all toxins simultaneously, we trained four independent classifiers—both in binary and three-class configurations—each dedicated to one specific mycotoxin present in the corn dataset: AFLA, DON, ZEA, and FUM. Each model determines whether the concentration of its respective toxin falls below, near, or exceeds the defined threshold. This per-toxin approach provides more granular insight into contamination levels and, when the four models are applied in parallel, allows us to reconstruct a detailed contamination profile for each sample—clearly identifying which toxins are of concern and which remain within safe limits.

Furthermore one can easily expand such a combined model by integrating a new model trained independently for the detection of a new toxin. This modular strategy simplifies model development, improves transparency, and achieves the overarching goal of multi-mycotoxin detection, while maintaining alignment with food safety regulations and real-world application needs. The next subsection details how we structured the classification classes to support this strategy.

For wheat, a modular classification approach was not feasible due to the limited availability of standards—only a multi-toxin mixture and a blank were available. As a result, the entire wheat sample set consisted of varying concentrations of a fixed ratio of DON, OTA, and T-2/HT-2 toxins. Consequently, the data inherently restricts model development to the simultaneous detection of all four toxins, rather than distinguishing them individually.

4.2.1 Classification Classes

A key step in building robust classifiers is defining clear, meaningful class boundaries. Since our core objective is regulatory compliance, each toxin-specific model was designed at minimum as a binary classifier—distinguishing between samples that fall below or exceed the legal threshold (as detailed in Section 2.2). To gain deeper insight into model performance, particularly near decision boundaries, we also implemented a three-class system by adding a “risky” category for samples with toxin levels close to the threshold. This middle category allows the model to flag borderline cases that may require additional testing, balancing safety with operational efficiency.

It’s important to note that while the ideal training set would include samples across a wide concentration range, some limitations were unavoidable due to the nature of the available standards. For example, not all toxins were represented at concentrations exceeding legal thresholds. Nevertheless, if a model can accurately detect lower contamination levels, its performance is expected to scale to higher concentrations, as spectral features typically become more pronounced. In addition, we also considered the class balance into our models. To prevent inflated performance metrics—common when one class dominates the dataset—we adjusted class thresholds and sample distributions to achieve an as balanced dataset as possible within practical constraints.

Table 4.2 and Table 4.3 summarize the final thresholds for the binary and three-class models,

respectively. These cover all toxin-specific scenarios: AFLA, DON, ZEA, and FUM detection in corn as well as the combined multi-toxin detection in wheat, and cross-matrix DON detection across both corn and wheat. Where possible, legal limits were used directly—for example, $1250\mu\text{g}/\text{kg}$ for DON and $4\mu\text{g}/\text{kg}$ for AFLA (both on corn).

Two-Class Classification Thresholds

Corn	
Toxin	Threshold [$\mu\text{g}/\text{kg}$]
AFLA	2.88
DON	1250
ZEA	50
FUM	900
Wheat	
Reference	Threshold [$\mu\text{g}/\text{kg}$]
DON	500
Corn & Wheat (Combined)	
Reference	Threshold [$\mu\text{g}/\text{kg}$]
DON	500

Table 4.2: Two-class classification thresholds used for corn, wheat, and combined datasets. For corn, thresholds are set individually for AFLA, DON, ZEA, and FUM. For wheat, only DON concentrations are listed, but classification inherently reflects co-occurring OTA and T-2/HT-2 due to fixed mixing ratios.

Three-Class Classification Thresholds

Corn			
Toxin	Safe [$\mu\text{g}/\text{kg}$]	Risky [$\mu\text{g}/\text{kg}$]	Unsafe [$\mu\text{g}/\text{kg}$]
Total	< 2	$2 - 4$	> 4
AFLA			
DON	< 500	$500 - 1250$	> 1250
ZEA	< 25	$25 - 65$	> 65
FUM	$= 0$	$0 - 1000$	> 1000
Wheat			
Reference	Safe [$\mu\text{g}/\text{kg}$]	Risky [$\mu\text{g}/\text{kg}$]	Unsafe [$\mu\text{g}/\text{kg}$]
DON	< 330	$330 - 630$	> 630

Table 4.3: Three-class classification thresholds used for corn and wheat. For corn, thresholds are defined independently for AFLA, DON, ZEA, and FUM. The transition from “*risky*” to “*unsafe*” corresponds to the legal limits for AFLA and DON, while other thresholds were selected to ensure approximately balanced class distributions. For wheat, only DON concentrations are explicitly reported, but classification also reflects the proportional presence of OTA and T-2/HT-2, which co-occur in fixed ratios.

In the following sections, we detail the development of modular classifiers—beginning with the development of classification models for the detection of AFLA, DON, ZEA, and FUM in corn samples (Section 4.3), followed by the simultaneous detection of DON, OTA, and T-2/HT-2 in wheat (Section 4.4). Finally, in a third section, we explore cross-matrix classification of DON across both corn and wheat (Section 4.5).

Before diving into the model development results, however, we first present an overview of the overall strategy in Section 4.2.2. This is intended to provide clarity and structure, benefiting the understanding of the subsequent sections. Following that, Section 4.2.3 offers a brief discussion of the codebase used to implement the strategy.

4.2.2 The Model Development Strategy

To provide a clear overview of the modeling approach, Figure 4.5 presents a flowchart outlining the overall structure of the model development strategy.

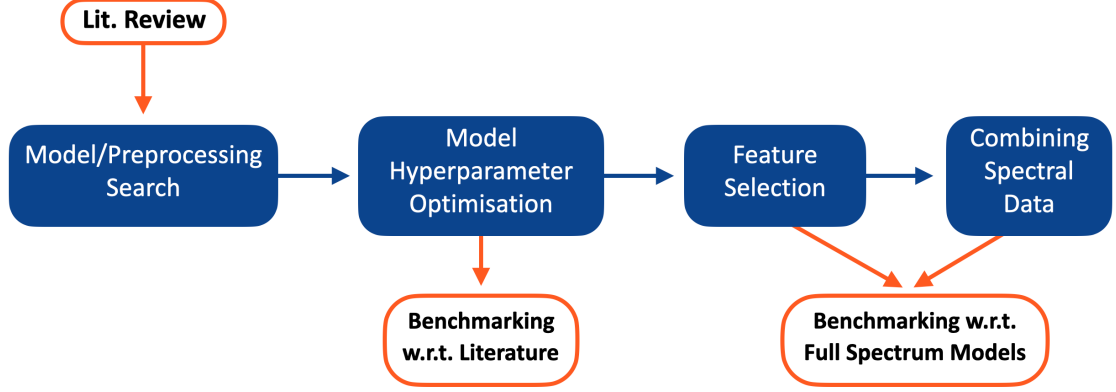


Figure 4.5: Flowchart summarizing the modeling framework. The process starts with a literature-informed model and preprocessing search, followed by hyperparameter optimization, feature selection, and finally, cross-platform data fusion. Each stage includes benchmarking to ensure robust evaluation.

The model development-process builds on the insights gathered from literature, as discussed in Section 2.5.3 and detailed further in Appendix A. To start, a comprehensive search is undertaken to explore a wide range of classification algorithms and preprocessing techniques identified within the literature review. These models are systematically benchmarked based on their effectiveness in detecting mycotoxins in corn (AFLA, DON, ZEA, FUM—each separately), in wheat (DON, OTA, and T-2/HT-2 simultaneously), and in a combined dataset for cross-matrix DON detection.

Once the top-performing classifiers are identified, a focused optimization phase is conducted to fine-tune key, aiming to enhance the predictive accuracy. The resulting best-performing models are then evaluated, not only for classification performance but also for computational efficiency, allowing us to assess the trade-offs between predictive power and practical resource requirements. A detailed error analysis and comparison against relevant literature are also provided to contextualize the results.

Building on this foundation, the chapter furthermore explores feature selection strategies aimed at reducing system complexity. Several leading techniques from the literature are systematically applied in combination with the best-performing full-spectrum classifiers to evaluate the trade-offs between dimensionality reduction and model performance. The most effective reduced-feature models—restricted to just 10 wavelengths—are benchmarked against both full-spectrum models and the state-of-the-art. In a final stage, we further investigate whether combining features across platforms (for instance, merging data from different fluorescence excitation wavelengths or integrating fluorescence and reflectance data) can improve classification accuracy. This cross-platform fusion is also benchmarked relative to the full-spectrum models.

With this structured roadmap in place, we now turn to a brief overview of the codebase underpinning the entire development process before delving into detailed results.

4.2.3 Notes On The Code Used For Model Development

Python was selected as the primary programming language, chosen for its open-source nature, extensive library ecosystem, wealth of publicly available code, and compatibility with generative AI tools for code development. The specific details of the code implementation, including the libraries used and the internal structure of individual functions are not covered in this section. Instead, the fully annotated codebase is provided separately, with documentation outlining the complete modeling pipeline and performance outputs for those interested [139]. Below a concise overview of the main structure of the code is provided.

The first step in developing a versatile codebase involved creating an efficient framework for loading, storing, and processing the spectral data from the raw ASCII files. This was accomplished using Pandas DataFrames, which provided a structured and flexible format for data manipulation. Code was then implemented to organize the extracted data into predefined classes and to subclassify individual spectra according to their sample of origin within each class. In addition, a suite of data preprocessing functions—such as trimming, smoothing, normalization, and ratio calculations (as described in Appendix A.1)—was integrated into the workflow.

These preprocessing steps were controlled through a custom central data construction function, which allowed users to specify the desired preprocessing and class distribution. To support exploratory analysis and interpretation, the framework also featured robust data visualization capabilities, enabling the generation of clear and informative spectral plots.

Beyond visualization, a flexible model training and validation function was developed. Particular emphasis was placed on ensuring the adaptability of this function to support future experimentation. The function accepted configurable inputs for both the classification model (e.g., LDA, KNN, etc.) and the feature selection method (e.g., SFS, MRMR), allowing for the seamless integration of various modeling techniques, ideal for model-search implementation. Performance validation was performed using the LOSO-CV scheme. Several additional functions were designed to visualize model performance, including confusion matrices and heatmaps for the visualization of model searches.

All these individual components were ultimately integrated into a central command function, which enabled the specification of any desired experiment with a single function call—from class distribution to preprocessing, classification algorithms and comparative experiments, as well as the desired performance metrics and visualizers.

4.3 Classification Of Mycotoxins On Corn

Following the model-development flow presented in Figure 4.5, with clearly defined class separation and a Leave-One-Sample-Out Cross-Validation (LOSO-CV) strategy, we now proceed to develop a classifier targeting the detection of AFLA, DON, ZEA, and FUM in corn samples. This model development entails a multidimensional optimization process, involving the selection of the optimal spectroscopic modality, data preprocessing strategy, and classification algorithm to identify the most effective detection platform.

A first comparative analyses will be to investigate what spectroscopic modalities, be it fluorescence (395, 385, 375, 365, 256nm) or reflectance (UV-VIS, NIR), or a combination thereof, are best suited for the detection of each toxin. Therefore, we will investigate the performance of the various models on each of the spectral datasets, and in a later stage a combination of these, evaluating the classification performance as function of the spectroscopic sensing method.

Next, we aim to optimize the data preprocessing steps. To this end, we investigate the most promising approaches identified in the literature, as discussed in Appendix A.1. Specifically, we consider the following preprocessing strategies: trimmed full-resolution spectral data (Section A.1.1), smoothed spectra (Section A.1.2), normalized data (Section A.1.3), and spectral ratio features (Section A.1.4), as well as combinations of these.

The fluorescence spectra are trimmed to the 400-800nm range, while for UV-VIS reflectance, the range is limited to 400-1000nm, and for NIR reflectance we use the 1000-1700nm wavelength range. Considering the smoothing, we consider smoothing values of 2, 4 and 6nm for fluorescence data (original resolution: 0.6nm⁴), 1, 3, and 5nm for UV-VIS data (original resolution: 0.3nm), and 8, 12, and 16nm for NIR data (original resolution: 4nm). Furthermore, we will explore different normalization techniques, including max-normalization and area-normalization, both applied to pre-smoothed data. Lastly, reflectance and fluorescence ratios are considered and computed from smoothed or area-normalized data, ratios here are interesting to divide away background influences.

Beyond optimizing the input data, we also investigate the most effective modeling approaches. Building on insights from the literature review in Section 2.5, a diverse set of classification algorithms was selected for comparative evaluation, based on their demonstrated performance in related spectroscopic applications. Models include: Linear Discriminant Analysis (LDA) (Appendix A.2.1.1), Quadratic Discriminant Analysis (QDA) (Section A.2.1.2), K-Nearest Neighbors (KNN) (Section A.2.1.4), Linear and Quadratic Support Vector Machines (LSVM & QSVM) (Section A.2.1.3), Random Forests (RF) (Section A.2.1.5), Artificial Neural Networks (ANN) (Section A.2.1.6), and Partial Least Squares Discriminant Analysis (PLS-DA) (Section A.2.1.7). Several of these models—particularly KNN, RF, and ANN—are known to be highly sensitive to hyperparameter settings. To ensure a fair and representative comparison, we adopted carefully chosen configurations based on established practices in literature.

For KNN, we evaluated five different neighborhood sizes (5, 8, 10, 12, and 15), consistent with commonly used values ranging between 7 and 12 [95]. For RF, the number of estimators is a critical factor influencing both performance and computational cost. We selected 100 estimators as a balanced choice that provides strong performance without excessive computational overhead. Considering ANN, we adapted a standard architecture using the ReLU activation function, the Adam optimizer, and a hidden layer of 100 nodes. Both single-layer and dual-layer ANN config-

⁴The initial resolution here is the resolution in the saved file, an interpolated resolution of the Avantes software.

urations⁵.

To systematically identify the optimal spectroscopic modality, preprocessing pipeline, and classification model, for each toxin the following section presents a detailed comparative analysis. This analysis is structured as a three-fold optimization process, wherein a range of classifiers are evaluated across multiple preprocessing techniques and spectroscopic platforms. For each configuration, classification accuracy is computed using LOSO-CV and visualized as heatmaps per modality to facilitate direct comparison. An important aspect of this approach is the deterministic nature of LOSO-CV. Since each model is trained on the spectra of all samples except one and validated on the excluded sample—repeating this process for each sample—there is no random data partitioning involved. Consequently, the use of a random seed, which was essential in earlier cross-validation schemes such as the one employed in the 7-class feasibility study, is no longer necessary.

The absence of the random seed is impacting the model evaluation, especially for models providing analytical solutions. These models, like LDA, will produce identical results across multiple runs, since their training process is deterministic under this validation framework. Consequently, for such models, it is not meaningful to report uncertainty in performance through repeated runs. In contrast, models like Random Forests (RF) and Artificial Neural Networks (ANN) involve stochastic training processes—due to factors such as randomized feature selection, initialization, and gradient-based optimization—which can lead to slight variations in performance across runs, even under LOSO-CV. Nevertheless, since LOSO-CV isolates evaluation to truly unseen samples, it remains the most stringent and realistic measure of generalized performance while keeping maximal data available for training.

Taking these technical considerations into account, we proceed with a detailed model search for classifying AFLA, DON, ZEA, and FUM in corn, using the three-class classification scheme outlined in Table 4.3. We first investigate fluorescence spectroscopy in Section 4.3.1, followed by an analysis of the reflectance data in Section 4.3.2.

4.3.1 Model-Search: Fluorescence

The first model-search was conducted for the three-class classification of AFLA in corn, using threshold values of 2 and 4 $\mu\text{g}/\text{kg}$ to define the safe, risky, and unsafe classes (see Table 4.3). The initial focus was on fluorescence data, evaluating all five excitation wavelengths—from 395nm down to 256nm. For each excitation modality, a model-versus-preprocessing comparison was performed, and the resulting classification accuracies are visualized as heatmaps in Figure 4.6. Overall, the classification performance for AFLA proved relatively modest. The highest accuracy achieved was 58%, using either the 395 or 385nm excitation. Accuracy dropped significantly at lower wavelengths, with the 256nm excitation yielding only 46%. These results fall short when compared to findings in existing literature, where classification accuracies for AFLA detection in corn commonly range from 80% to 90% for two-class models using thresholds of 20 $\mu\text{g}/\text{kg}$ [103, 105, 106, 113], and in a more recent study even up to 95.7% at a 10 $\mu\text{g}/\text{kg}$ threshold [93]. Several factors may account for this discrepancy. First, our use of a three-class classification scheme inherently complicates the task compared to the simpler binary classification commonly used in literature. Second, the thresholds applied in this study (2 and 4 $\mu\text{g}/\text{kg}$) are much stricter than those typically used, requiring the model to detect more subtle differences.

⁵In subsequent stages of the thesis, a more detailed hyperparameter analysis is also conducted for ANN and KNN to further explore the optimization potential of various architectures.

Additionally, care should be taken when interpreting the AFLA contamination level. The AFLA-contaminated samples were synthesized using two different AFLA standards, each with relatively large uncertainties in concentration. This likely introduces variability in the actual vs. labeled toxin levels of composed samples, shown in Table 3.2, hampering the model’s ability to obfuscate underlying spectral patterns and hinder the model’s ability to map spectral features to mycotoxin levels. A more detailed discussion of this complexity is given in Section 4.3.5.2.

Beyond finding 395 and 385nm excitations yielding the highest classification accuracies for AFLA, several consistent trends emerged regarding the most effective model/preprocessing combinations. Across all excitation settings, LDA consistently yielded strong results when applied to either low-resolution smoothed data (6nm) or normalized spectra, showing classification accuracies of up to 54% and 57%, respectively. Similarly, Linear Support Vector Machines (LSVM) showed strong performance when applied to high-resolution smoothed data (0.6nm) (57%) or when trained on spectral ratios derived from smoothed or normalized inputs (58%). Random Forests (RF) also demonstrated solid performance when used in combination with ratio-based preprocessing (up to 57%).

These three models—LDA, LSVM, and RF—consistently outperformed other algorithms in the classification of AFLA. Among them, LDA offered the added advantage of significantly faster training times, particularly when compared to LSVM and RF models, which became increasingly computationally intensive when paired with ratio-based features⁶.

Similar to the study for the detection of AFLA we also implemented three class model-searches for the detection of DON, ZEA, and FUM, employing thresholds as detailed in Table 4.3. The obtained heat maps for each of these comparative studies are shown in Figures 4.7, 4.8, 4.9, for DON, ZEA and FUM, respectively. To better illustrate the performance for all the toxins the best performance per excitation wavelength is shown in Table 4.4.

Toxin	395nm	385nm	375nm	365nm	256nm
AFLA	58%	58%	49%	50%	46%
DON	82%	80%	80%	81%	73%
ZEA	81%	71%	72%	72%	52%
FUM	73%	58%	69%	67%	62%

Table 4.4: **The best classification accuracy** for each toxin for each fluorescence excitation from 395 to 256nm. Results are derived from the model-searches shown in Figures 4.6, 4.7, 4.8, and 4.9.

As shown in Table 4.4, DON classification achieved the highest overall accuracy among the four toxins, reaching up to 82%. However, it is important to emphasize that no direct comparison with existing literature is possible in this case. DON is a non-fluorescent toxin, and to our knowledge, no previous studies have attempted its detection using fluorescence-based methods. The observed classification performance likely stems from indirect spectral features—such as matrix alterations induced by mycotoxin biosynthesis and metabolism—rather than direct fluorescence emissions.

Among the tested excitation wavelengths, those between 395 and 365nm yield the highest accuracies for DON, with results consistently ranging between 80% and 82%. In contrast, excitation

⁶A more detailed analysis of the computational demands associated with each model type will be presented later in Section 4.3.3.

at 256nm performs significantly worse, confirming the trend observed with AFLA. The best-performing models for DON classification were once again LDA, LSVM, and RF. LDA achieved top performance (82%) when applied to low-resolution smoothed data (6nm) or normalized spectra. LSVM performed best using high-resolution trimmed data or smoothed data (up to 82%), while RF achieved optimal results exclusively when used with ratio-based features (81%).

For ZEA and FUM, the 395nm excitation again proved to be the most effective, with classification accuracies of 81% and 73%, respectively. As with the other toxins, LDA, LSVM, and RF remained the top-performing models across these tasks, showing consistent preference for these classifiers when paired with the aforementioned preprocessing techniques.

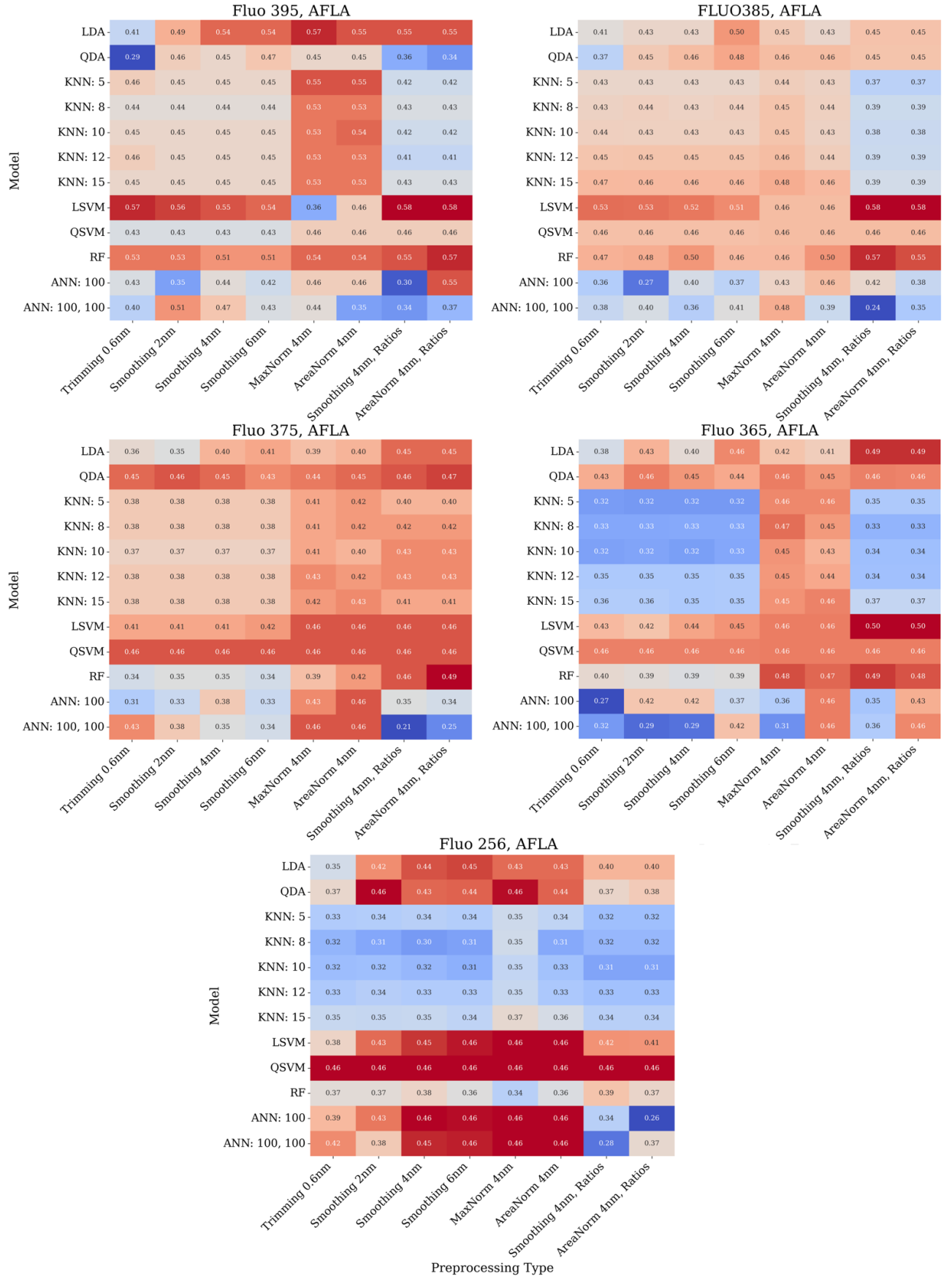


Figure 4.6: Classification accuracy heatmaps for each fluorescence modality, illustrating the results of the model/preprocessing search for AFLA detection on corn using a three-class classification scheme. Thresholds of 2 and $4\mu\text{g}/\text{kg}$ were used to define safe, risky, and unsafe classes. Overall classification performance is lower than typically reported in the literature. The best-performing models—LDA, LSVM, and RF—achieved their highest accuracy using trimmed, smoothed or ratio-based data.

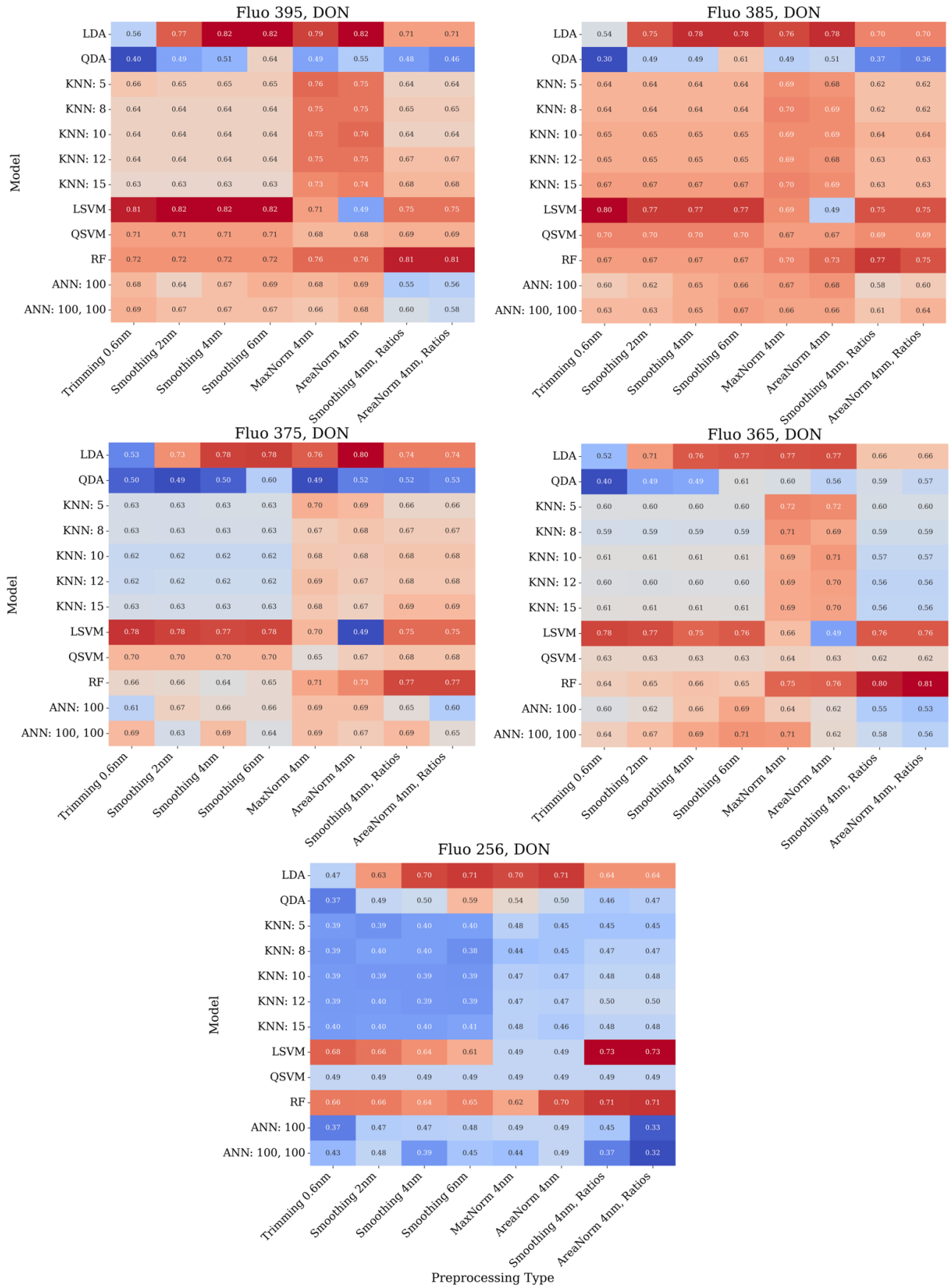


Figure 4.7: Classification accuracy heatmaps for each fluorescence modality, illustrating the results of the model/preprocessing search for DON detection on corn using a three-class classification scheme. Thresholds of 500 and 1250 $\mu\text{g}/\text{kg}$ were used to define safe, risky, and unsafe classes. Overall classification performance is great, reaching up to 82%. The best-performing models—LDA, LSVM, and RF—achieved their highest accuracy using trimmed, smoothed or ratio-based data.

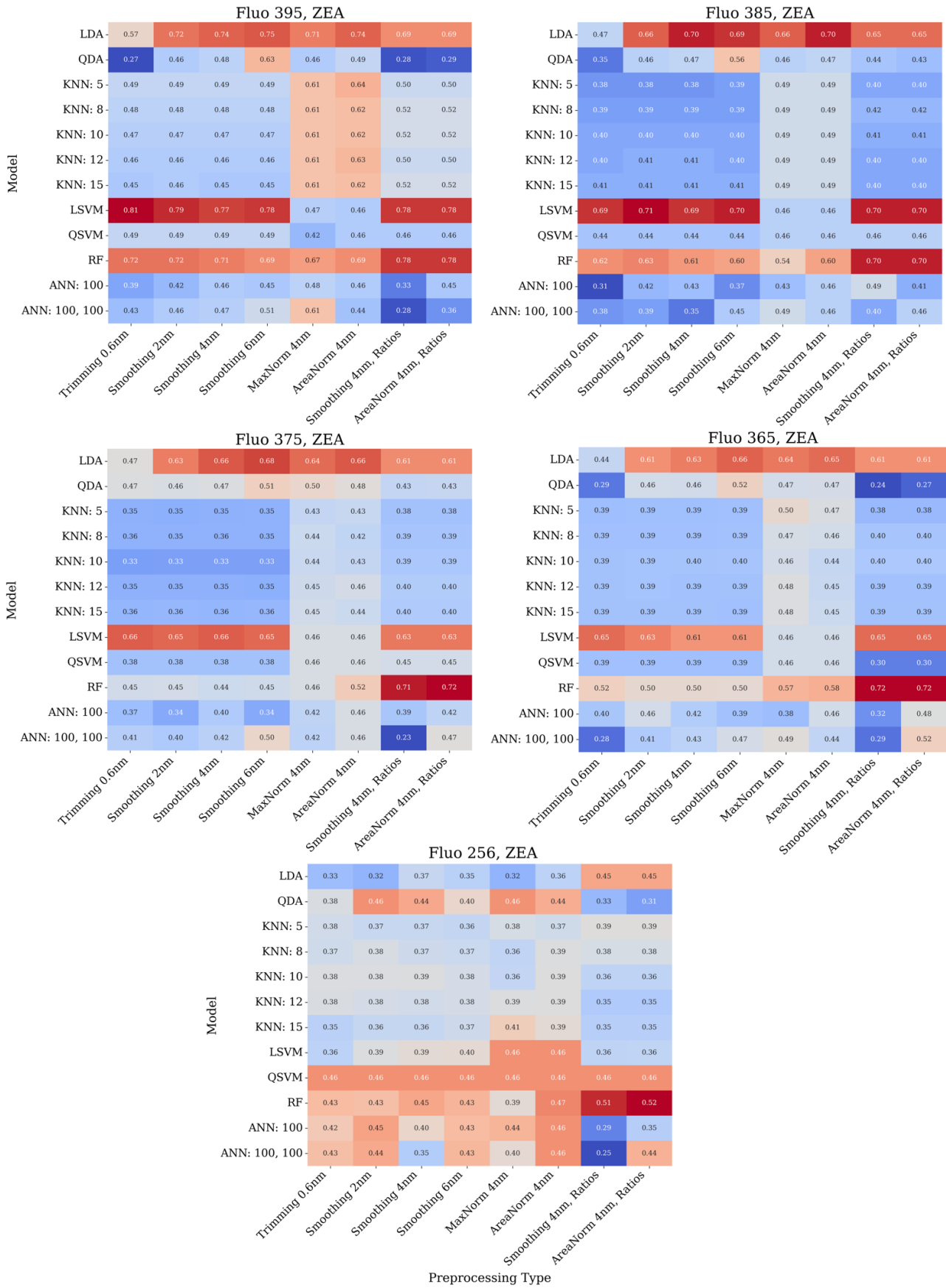


Figure 4.8: Classification accuracy heatmaps for each fluorescence modality, illustrating the results of the model/preprocessing search for ZEA detection on corn using a three-class classification scheme. Thresholds of 25 and 65 $\mu\text{g}/\text{kg}$ were used to define safe, risky, and unsafe classes. Overall classification performance was good reaching up to 81%. The best-performing models—LDA, LSVM, and RF—achieved their highest accuracy using trimmed, smoothed or ratio-based data.

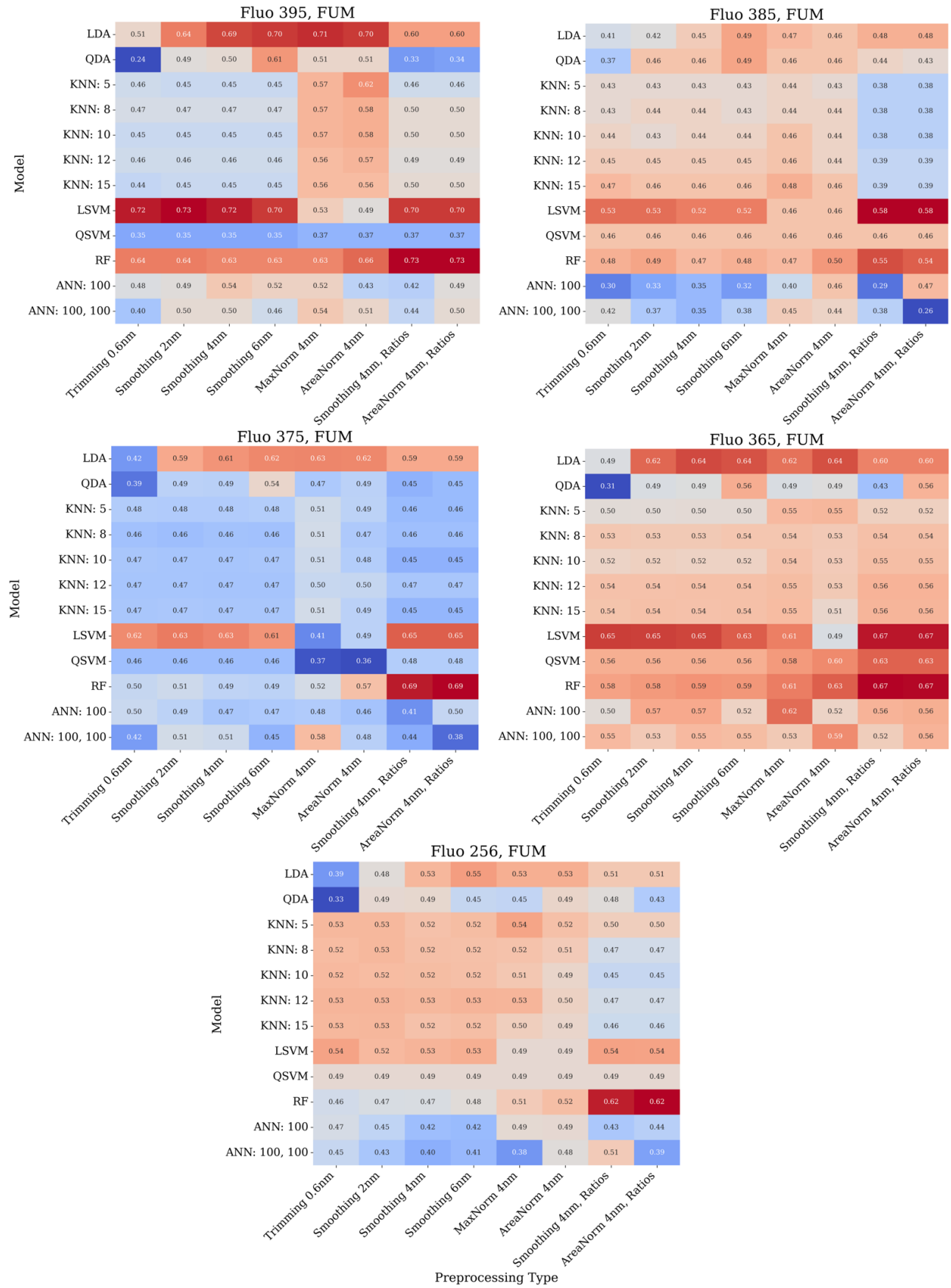


Figure 4.9: Classification accuracy heatmaps for each fluorescence modality, illustrating the results of the model/preprocessing search for FUM detection on corn using a three-class classification scheme. Thresholds of 0 and 1000 μ g/kg were used to define safe, risky, and unsafe classes. Overall classification performance was good reaching up to 71%. The best-performing models—LDA, LSVM, and RF—achieved their highest accuracy using trimmed, smoothed or ratio-based data.

4.3.1.1 Hyperparameter Sensitive Models

As discussed previously in Section 4.3, certain classifiers—most notably K-Nearest Neighbors (KNN) and Artificial Neural Networks (ANN)—are particularly sensitive to the choice of hyperparameters. While both models showed occasional promising results during the initial model-search, their performance generally fell short with respect to the ones of LDA, LSVM, and RF. This suggested that a more thorough exploration of their respective hyperparameter spaces might lead to improved results.

To investigate this, we conducted a more extensive hyperparameter optimization for these models. Given that DON detection using fluorescence at 395nm yielded the highest classification accuracy overall, this setup was selected as the case study for the expanded search. In the original model-search, the best classification accuracies achieved were 69% for ANN and 76% for KNN. For ANN we explored different activation functions (ReLU, tanh, logistic) using a network with one layer of 100 nodes, and different network structures (5 layers of 100 nodes, 5 layers of 1000, 2 layers of 1000 and 3 of 100) using the ReLU activation function. For KNN we explored more neighbors ranging from just 1 to 30. For both cases, different preprocessing methodologies were explored, including different smoothing levels and normalization.

The outcomes of this refined hyperparameter analysis are presented in Figure 4.10. For the ANN, the highest performance was achieved using the original configuration—trimmed data combined with the ReLU activation function—reaching an accuracy of 74%. This reflects a 5% improvement compared to the earlier result shown in Figure 4.7. However, this increase is attributed to the stochastic nature of neural network training rather than a change in hyperparameter configuration. No significant performance gain was realized as a result of modifying the network architecture.

In the case of KNN, the hyperparameter search did not yield any further improvement beyond the previously observed 76%. However, we did find that KNN performed noticeably better when trained on area-normalized data compared to smoothed data, highlighting the preference of KNN for this preprocessing method. Nevertheless, performance plateaued at 76%, confirming that KNN is not ideal for this classification task, compared to LDA, LSVM, and RF.

4.3.1.2 Evaluation Of PLS-DA For DON Classification

Up to this point, the classification models evaluated included LDA, QDA, KNN, LSVM, QSVM, RF, and ANN. However, one model that was not yet been included is Partial Least Squares Discriminant Analysis (PLS-DA) (see Section A.2.1.7). Although PLS-DA did not initially emerge as a focus during the literature review, its relevance became more apparent as key regression model (Section 5.1.3). Motivated by these findings, we explored the classification performance of PLS-DA as well. To this end, a PLS-DA model was implemented in Python using PLSRegression as a foundation. The model was applied to the classification of DON using fluorescence data at 395nm, given the other models works best when considering this excitation wavelength, and a model search was conducted to compare configurations using varying numbers of PLS-components.

The resulting accuracy heatmap is shown in Figure 4.11. The model achieved a peak accuracy of 85% using 7–10 components and area-normalized spectra smoothed to a 3nm resolution. This accuracy exceeds that of any classification model tested thus far. However, as will be discussed in Section 4.3.4, similar performance can eventually be reached using LDA after appropriate optimization. As main drawback, PLS-DA shows to be less suited for use with feature selection methods, which are applied later in this study. Therefore, despite its strong standalone

performance, PLS-DA was excluded from further analysis in favor of models that integrate more seamlessly with the complete analytical pipeline.

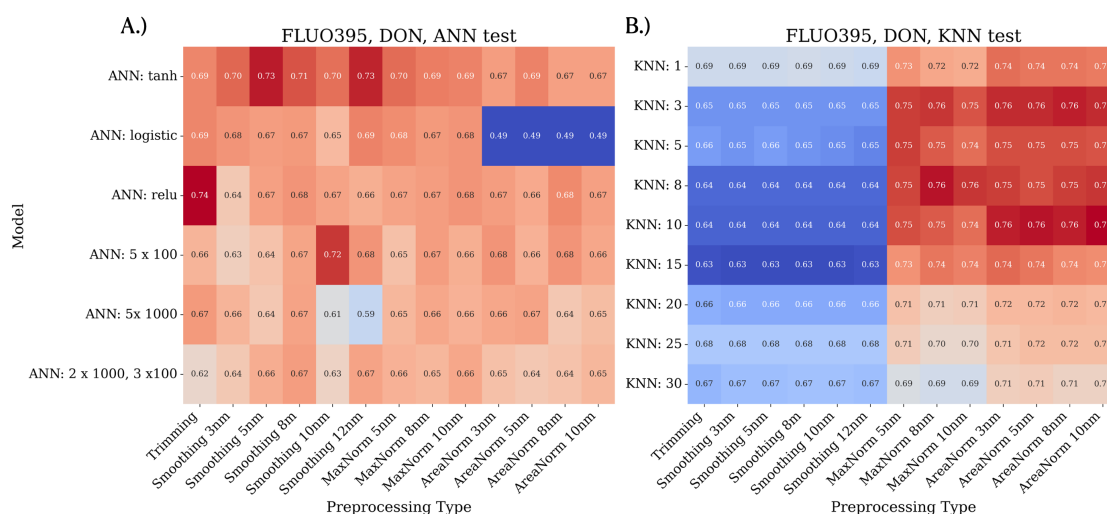


Figure 4.10: Hyperparameter search results for A.) ANN and B.) KNN models in the classification of DON at 395nm excitation. KNN consistently performs better when trained on normalized data, though further tuning does not yield performance improvements beyond the initial baseline. ANN performance did not improve as a result of architecture optimization.

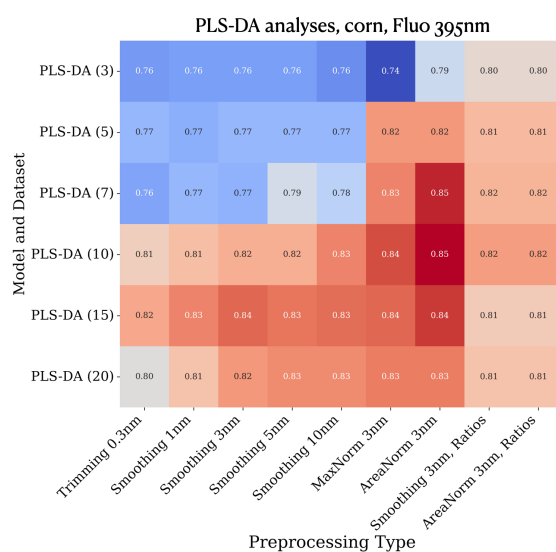


Figure 4.11: Classification accuracy of PLS-DA models for DON detection using fluorescence data at 395nm with area normalization and 3nm spectral smoothing. The heatmap shows accuracy as a function of the number of PLS-components. The model achieved peak performance at 85% accuracy using 7–10 components, exceeding the results of all previously tested classification models under comparable conditions.

4.3.2 Model-Search: Reflectance

In line with the approach taken for the fluorescence datasets, a model-search was conducted for the reflectance data. This data included both UV–VIS and NIR spectral ranges, acquired using small and large integrating sphere setups (Section 3.2.1). The specifics of the models and preprocessing steps used in this search were previously detailed in Section 4.3.

The classification results for each spectroscopic modality are presented in Figures 4.12, 4.13, 4.14, and 4.15, corresponding to AFLA, DON, ZEA, and FUM, respectively. To better visualize the highest-performing configurations, the best classification accuracies for each toxin and modality are summarized in Table 4.5. Looking at this information it becomes immediately apparent that classification based on the reflection spectra perform substantially worse compared to those achieved using 395nm fluorescence data (Table 4.4).

However, a slight performance improvement can be observed when using data from the large integrating sphere, suggesting that global reflectance measurements offer enhanced discriminative power relative to localized reflectance.

Despite this improvement, classification accuracy for reflectance data still lags behind that of the fluorescence-based models. The best overall result was obtained for DON using the UV–VIS large sphere setup, where an LDA model trained on trimmed data achieved an accuracy of 79%. This outperformed all other reflectance-based configurations across all toxins. LSVM models yielded the second-best performance for DON, reaching 77% accuracy with the same trimmed data input.

Toxin	UV–VIS	NIR	UV–VIS (Large Sphere)	NIR (Large Sphere)
AFLA	50%	48%	55%	51%
DON	67%	66%	79%	54%
ZEA	61%	55%	60%	50%
FUM	50%	52%	63%	51%

Table 4.5: Best classification accuracy for each toxin using reflectance data, comparing small and large integrating sphere setups for both UV–VIS and NIR modalities separately. While performance improves slightly when using the larger sphere, overall results remain lower than those obtained using fluorescence data.

Given the significantly lower performance observed for reflectance data—often 10–20% below that of fluorescence-based classification—we decided to proceed exclusively with the fluorescence data for mycotoxin detection in corn.

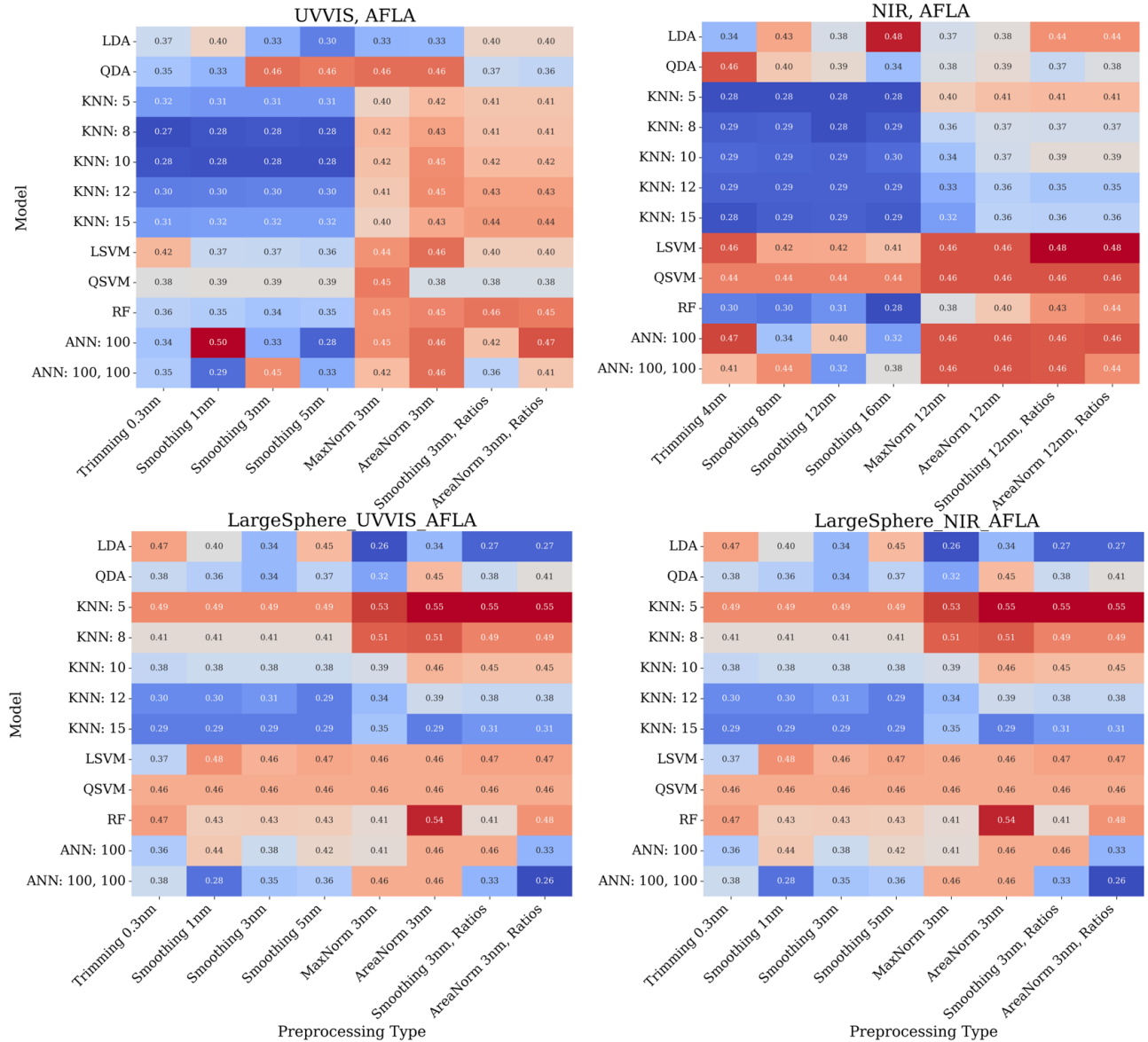


Figure 4.12: Classification accuracy heatmaps for each reflectance modalities, illustrating the results of the model/preprocessing search for AFLA detection in corn using a three-class classification scheme. Thresholds of 2 and $4\mu\text{g}/\text{kg}$ were used to define safe, risky, and unsafe classes. Overall classification performance was poor compared to the results obtained using fluorescence (Figure 4.6).

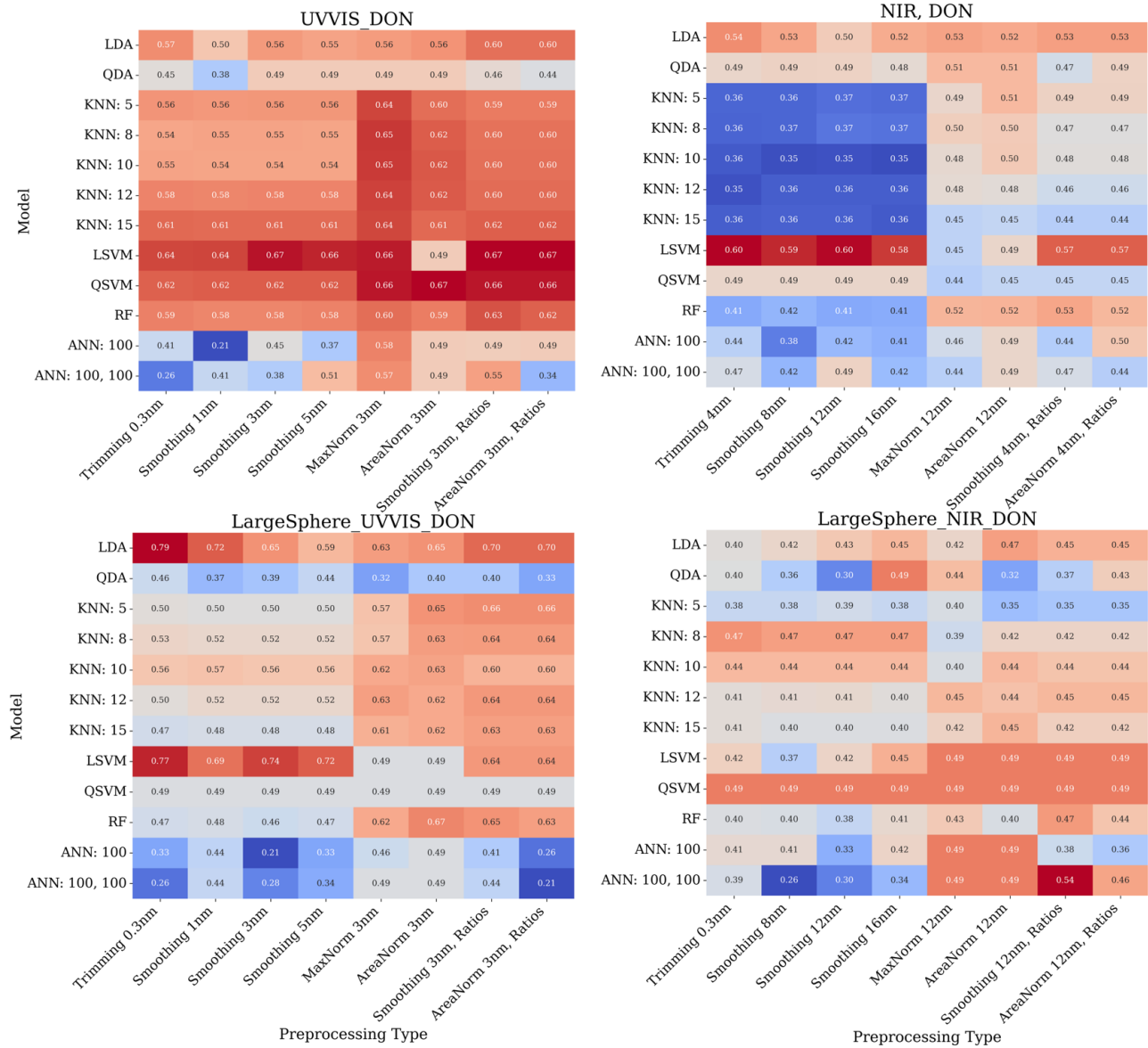


Figure 4.13: Classification accuracy heatmaps for each reflectance modalities, illustrating the results of the model/preprocessing search for DON detection in corn using a three-class classification scheme. Thresholds of 500 and 1250 $\mu\text{g}/\text{kg}$ were used to define safe, risky, and unsafe classes. Overall classification performance was poor compared to the results obtained using fluorescence (Figure 4.7), only large sphere UV-VIS measurements yielded reasonable results.

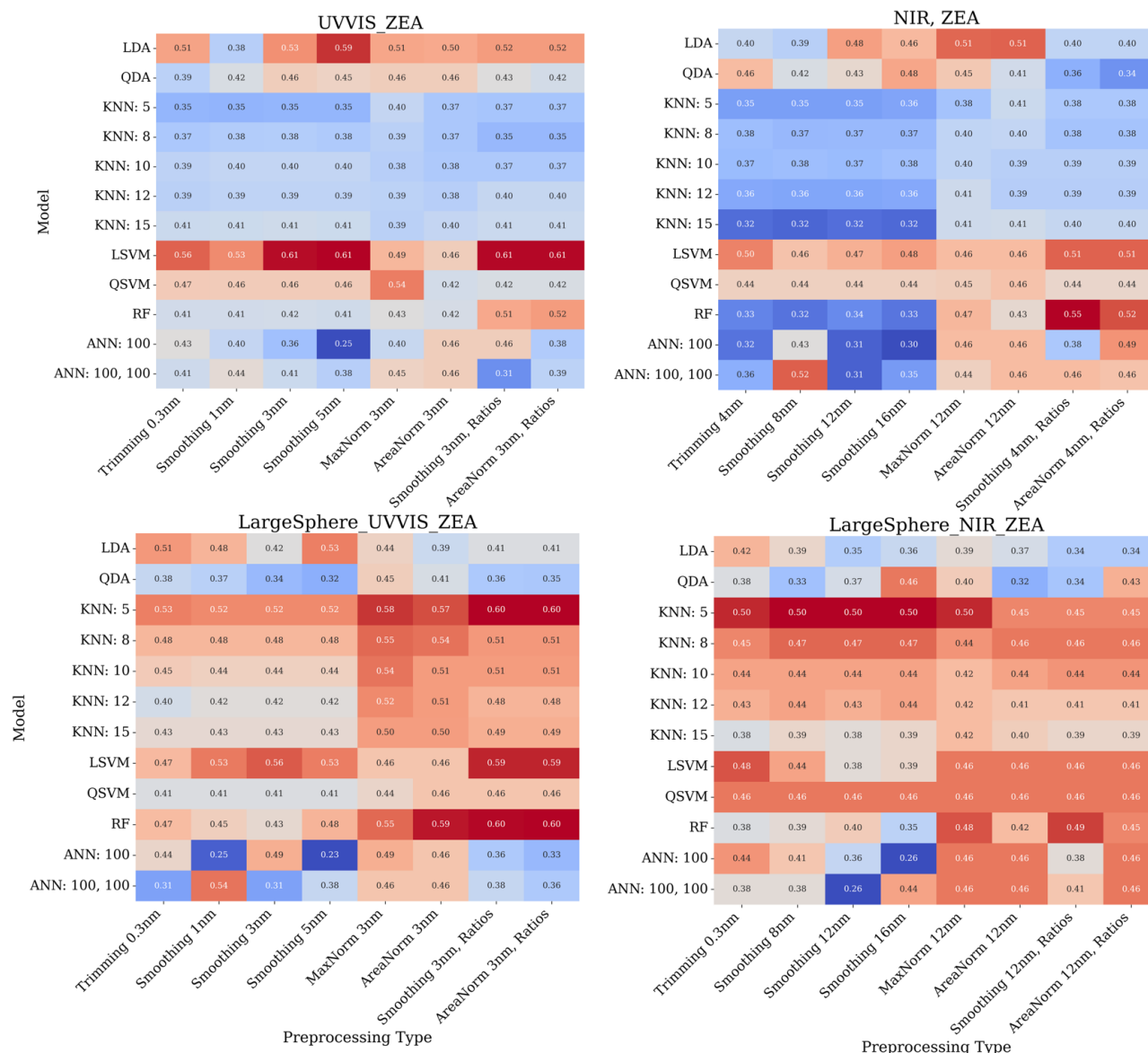


Figure 4.14: Classification accuracy heatmaps for each reflectance modalities, illustrating the results of the model/preprocessing search for ZEA detection in corn using a three-class classification scheme. Thresholds of 25 and 65 $\mu\text{g/kg}$ were used to define safe, risky, and unsafe classes. Overall classification performance was poor compared to the results obtained using fluorescence (Figure 4.8).

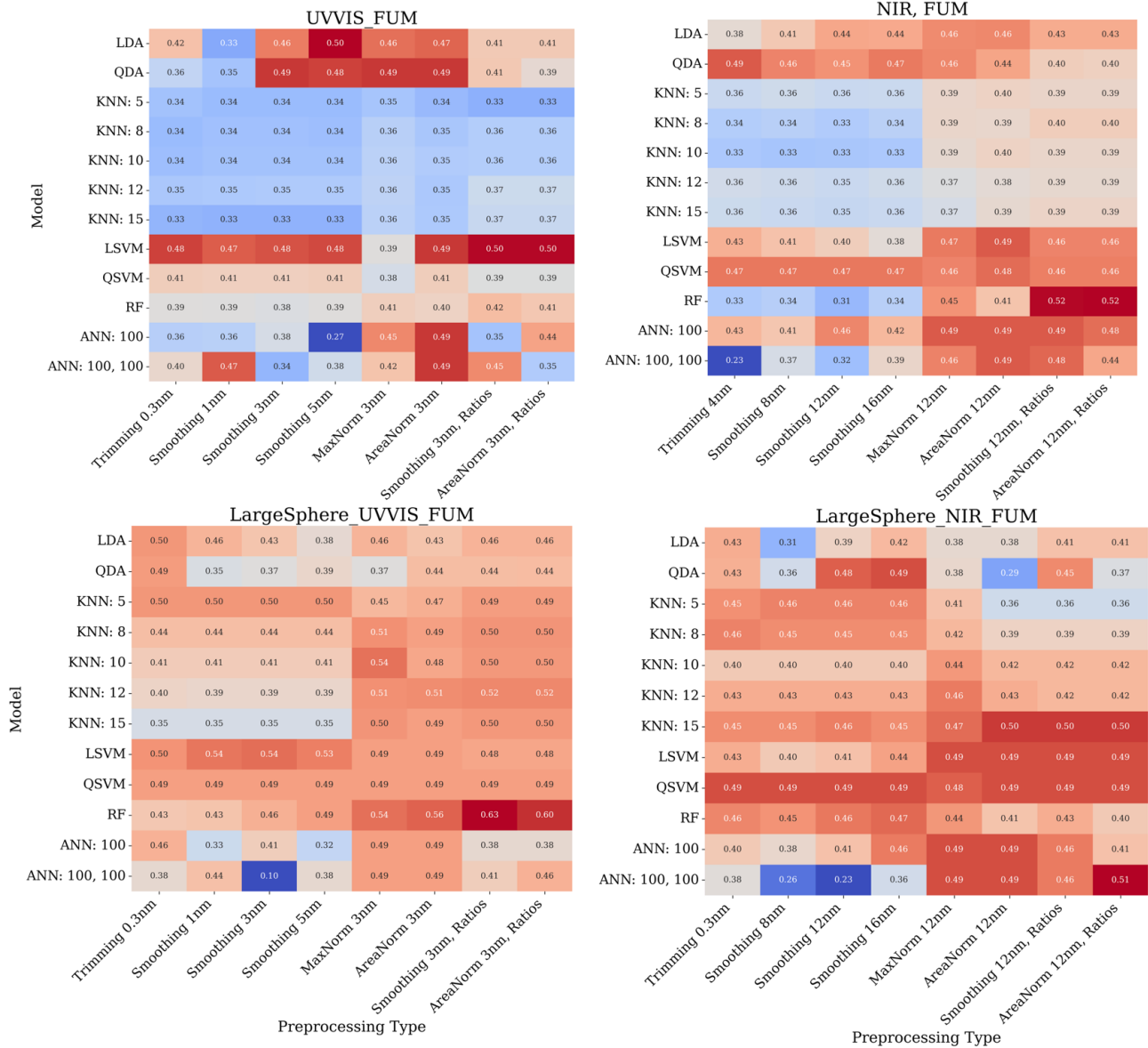


Figure 4.15: Classification accuracy heatmaps for each reflectance modalities, illustrating the results of the model/preprocessing search for FUM detection in corn using a three-class classification scheme. Thresholds of 0 and 1000 $\mu\text{g}/\text{kg}$ were used to define safe, risky, and unsafe classes. Overall classification performance was poor compared to the results obtained using fluorescence (Figure 4.9).

4.3.3 Computational Cost Comparison

Given the inherently high dimensionality of spectroscopic data, computational efficiency plays a critical role in the practical deployment of machine learning models—especially when incorporating feature selection techniques such as Sequential Feature Selection, which are computationally intensive. Even modest differences in model runtime can scale non-linearly and significantly impact total training time.

To assess the computational demand of the classification models evaluated thus far, a systematic runtime analysis was performed using fluorescence data at 395nm smoothed to 4nm resolution. Each model was trained and evaluated three times under identical conditions. To account for variability due to processor load and thermal throttling, models were executed in randomized order, and both the mean and standard deviation of execution times were recorded.

The results, depicted in Figure 4.16, reveal that LDA has the lowest computational demand, followed closely by KNN and LSVM. By contrast, RF and ANN exhibited significantly higher training times. For these measurements a MacBook Air 2019 was used with 1.6GHz Dual-Core Intel Core i5. While actual runtimes will naturally depend on the specific hardware used, the relative differences between models offer meaningful insight into their computational efficiency.

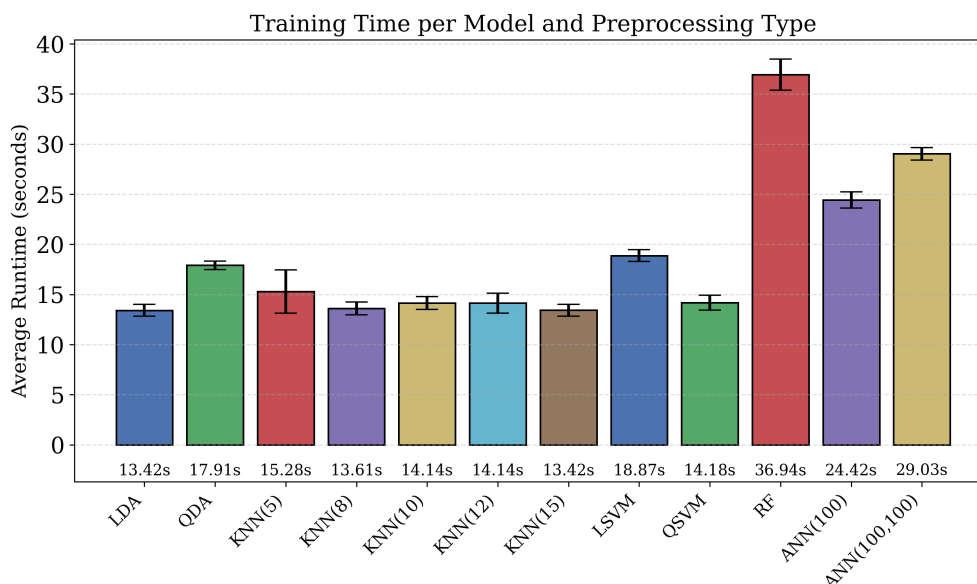


Figure 4.16: Average training time (in seconds) for each classification model evaluated on the corn dataset using fluorescence data at 395nm (smoothed to 4nm resolution). Each model was run three times to account for variability due to system load, with error bars indicating the standard deviation. LDA exhibits the lowest computational demand, while RF and ANN require significantly longer training times.

It should be noted that this initial comparison used uniform preprocessing (smoothed spectra at 4nm), whereas model-specific preprocessing preferences—identified in the model-search heatmaps, shown in Figures 4.6 through 4.9—indicate a more nuanced picture. LDA generally performs best on low-resolution smoothed data, LSVM benefits from high-resolution or ratio-based spectra, and RF yields optimal results with spectral ratios. These preprocessing techniques

substantially affect both the dimensionality of the data and the computational load, particularly in the context of iterative feature selection.

To further examine this, we evaluated the training time of the four best-performing classifiers—LDA (smoothed), LSVM (trimmed and ratio-based), and RF (ratio-based)—using their respective optimal preprocessing pipelines. The results, shown in Figure 4.17, clearly illustrate that LDA maintains a significant computational advantage, requiring less than half the time of LSVM (trimmed) and nearly an order of magnitude less than LSVM or RF when used with spectral ratios.

This finding has significant implications for feature selection. Since methods like SFS scale non-linearly with both the number of features and the time required to train a model, the use of RF or LSVM with ratio-based preprocessing becomes prohibitively expensive. Consequently, LDA emerges not only as a performant but also computationally optimal choice, particularly well-suited for workflows involving feature reduction and iterative model refinement.

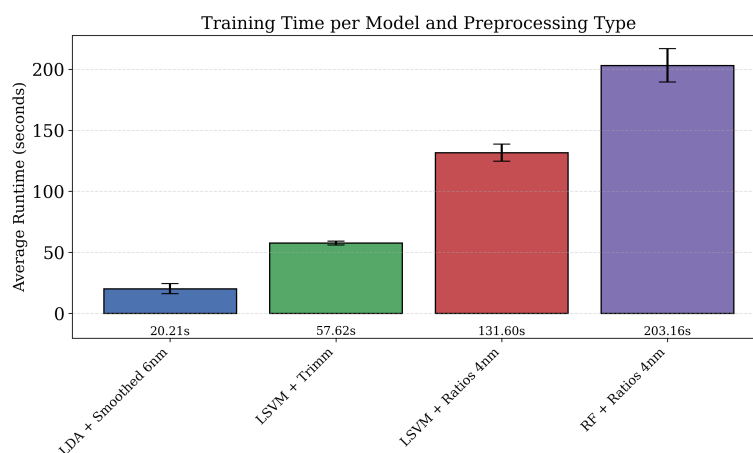


Figure 4.17: Training time comparison of the four best-performing classification models using their respective optimal preprocessing strategies on the corn dataset using fluorescence data at 395nm. LDA (with low-resolution smoothed data) shows the highest computational efficiency, while LSVM and RF models using ratio-based preprocessing exhibit significantly higher training times—up to tenfold longer. Error bars represent the standard deviation across three runs per model, accounting for variability due to hardware performance.

4.3.4 Model Optimization

From the model-search detailed in the previous section, fluorescence data—particularly with 395nm excitation—emerged as the most effective for classification across all toxins. The top-performing models were consistently Linear Discriminant Analysis (LDA) using smoothed data at reduced resolutions, Linear Support Vector Machines (LSVM) using trimmed and smoothed data, and Random Forests (RF) trained on ratio features derived from smoothed or normalized spectra. The best classification accuracies achieved were 58% for AFLA (LSVM, Ratios), 82% for DON (LDA, Smoothing 6nm), 81% for ZEA (LSVM, trimmed), and 73% for FUM (RF, Ratios).

In this section, we aim to further optimize the best-performing models—LDA, LSVM, and RF—by fine-tuning their respective hyperparameters and expanding the preprocessing search implemented in the previous phase. We first focused on optimizing LDA by investigating the impact of different solvers: SVD, which is suitable for datasets with fewer features than samples (as is the case here), and LSQR, which is preferred when regularization⁷ is desired. We tested LDA with both unregularized ("*None*") and automatically regularized ("*auto*") settings (Section A.2.1.1).

For LSVM, we explored the influence of the regularization parameter, C , which controls the trade-off between maximizing the margin and minimizing classification error (Section A.2.1.3). The values tested were $C = 0.001, 0.1, 1$, and 10 , with $C = 1$ the default value used. In the case of Random Forests, we examined the effect of the number of estimators (i.e., the number of decision trees in the ensemble), testing 50, 100, and 200 estimators (Section A.2.1.5). Additionally, we compared two commonly used split criteria: (1.) Gini impurity, defined by Equation A.23, which measures the likelihood of incorrect classification of a randomly chosen element, and (2.) entropy defined by Equation A.24, which is based on information gain and is often more sensitive to class purity in nodes.

Similar to the previous evaluations, different preprocessing methods were evaluated for each model. For LDA and LSVM, we tested trimmed and smoothed spectral data at various smoothing levels. We also applied area-normalization to LDA inputs across multiple smoothing settings and investigated the use of spectral ratios for LSVM. For RF, we tested various ratio combinations derived from smoothed and area-normalized spectra at different resolutions.

The results of the model optimization for AFLA, DON, ZEA, and FUM classification in corn using 395nm fluorescence data are presented in Figures 4.18, 4.19, 4.20, and 4.21, respectively. Each figure shows classification accuracy heatmaps for the three classifiers under investigation—LDA, LSVM, and RF—across various hyperparameter and preprocessing configurations. Similar studies were also done for all other excitation wavelengths, but yielded similar results and performance below that of 395nm.

Considering LDA, we observed that regularization generally led to better performance. In addition, when applying regularization, the spectral preprocessing only showed to minimally impact the performance. In contrast, when applying SVD, the model favored smoothed data at higher smoothing levels (i.e., lower spectral resolution). Compared to the original model-search results, the LDA performance improved slightly: DON accuracy increased from 82% to 85%, ZEA from 75% to 77%, and FUM from 70% to 72%, yielding a consistent 2–3% improvement, AFLA did not improve.

For LSVM, no strong overall trend was observed, though a regularization parameter of $C = 1$ generally produced the best results for DON, ZEA, and FUM. Higher values such as $C = 10$ could also yield competitive accuracy, but at a significantly higher computational cost. Smoothed data consistently performed well, though the optimal smoothing level varied: for DON, performance was stable across smoothing levels, while for ZEA and FUM, higher smoothing degraded accuracy slightly—but provided faster training times. Notably, LSVM performance remained largely unchanged after optimization, maintaining the same peak accuracies as before (less than 1% change).

In the case of RF, performance also showed minimal sensitivity to hyperparameter changes or preprocessing methods. Accuracy varied only within a few percentage points across all config-

⁷Regularization in LDA refers to the introduction of a shrinkage parameter to stabilize the estimation of the covariance matrix, which can improve model generalization, particularly when the number of features is high relative to the number of samples.

urations. However, RF models were substantially more computationally demanding, especially since they were trained on high-dimensional ratio-based inputs. For this reason, higher smoothing levels—which reduce dimensionality—are preferred to reduce computation time. Overall, RF performance improved slightly by about 2% after optimization.

The best classification accuracies obtained after optimization are summarized in Table 4.6. When compared to the initial model-search results in Table 4.4, it is clear that optimization provided only modest performance gains. For example, DON classification improved from 82% to 85%, while other toxins saw either minor improvements or no significant change.

Toxin	LDA	LSVM	RF
AFLA	57%	61%	59%
DON	85%	82%	82%
ZEA	77%	81%	80%
FUM	72%	74%	75%

Table 4.6: **Best classification accuracies** for each toxin using the optimized LDA, LSVM, and RF models based on 395nm fluorescence data. Results are derived from the model optimization study in Figures 4.18 (AFLA), 4.19 (DON), 4.20 (ZEA), and 4.21 (FUM).

Based on these results, we conclude that extensive model optimization had limited impact on the performance. However, it did provide valuable insight into model behavior. In particular, we found that LDA with regularization or SVD solver, is remarkably robust to smoothing and can maintain high performance even at lower spectral resolutions, making it more computationally efficient.

Given the high computational demands of RF—especially with ratio-based features—we exclude it from further analysis, as in the next sections SFS will be explored, which iteratively trains a multitude of models, increasing the computational load drastically.

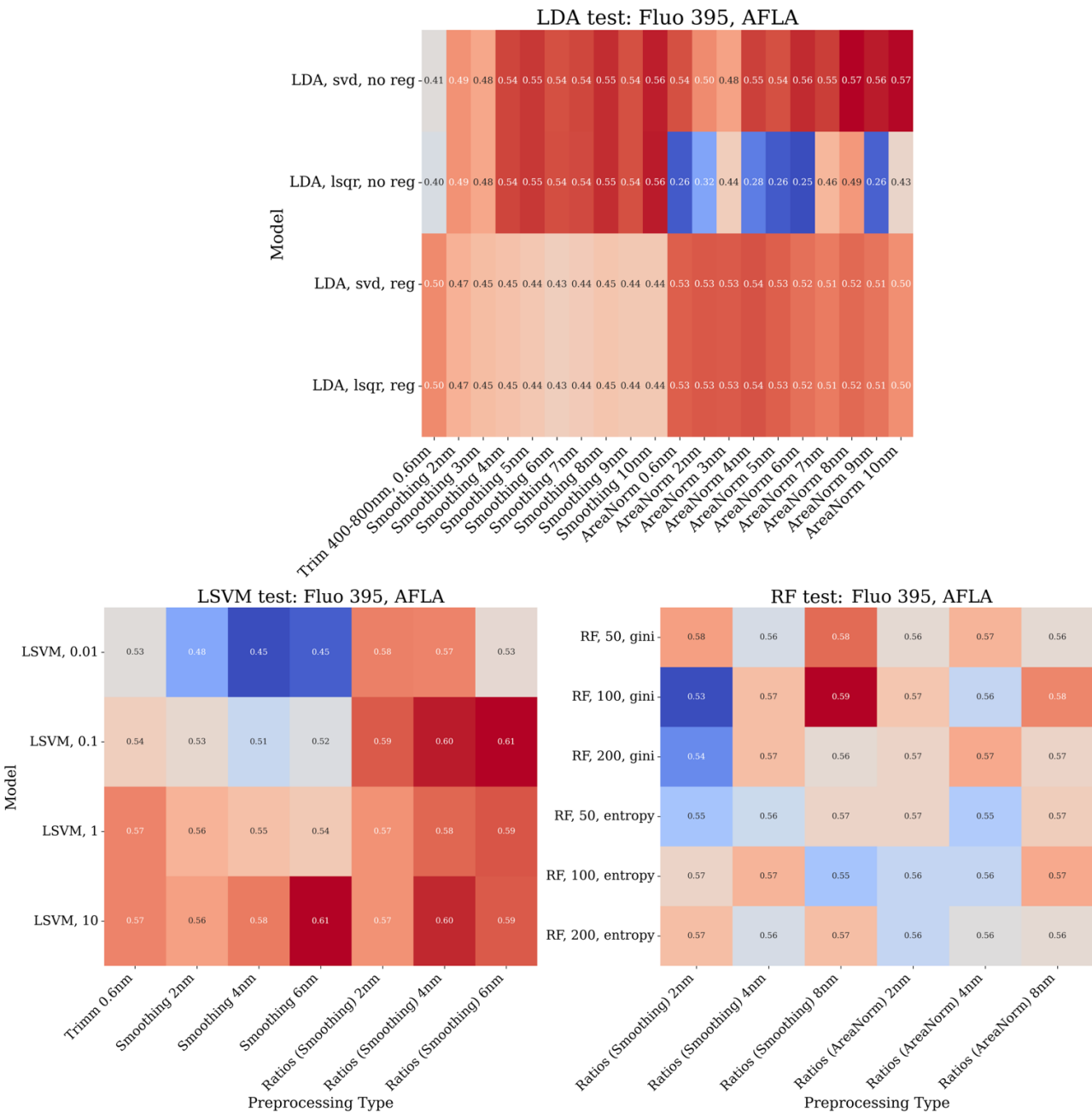


Figure 4.18: AFLA model optimization results for LDA, LSVM, and RF using 395nm fluorescence data. Various hyperparameters were explored in combination with refined preprocessing techniques. The optimization led to slight improvements in classification performance compared to the initial results presented in Figure 4.6.

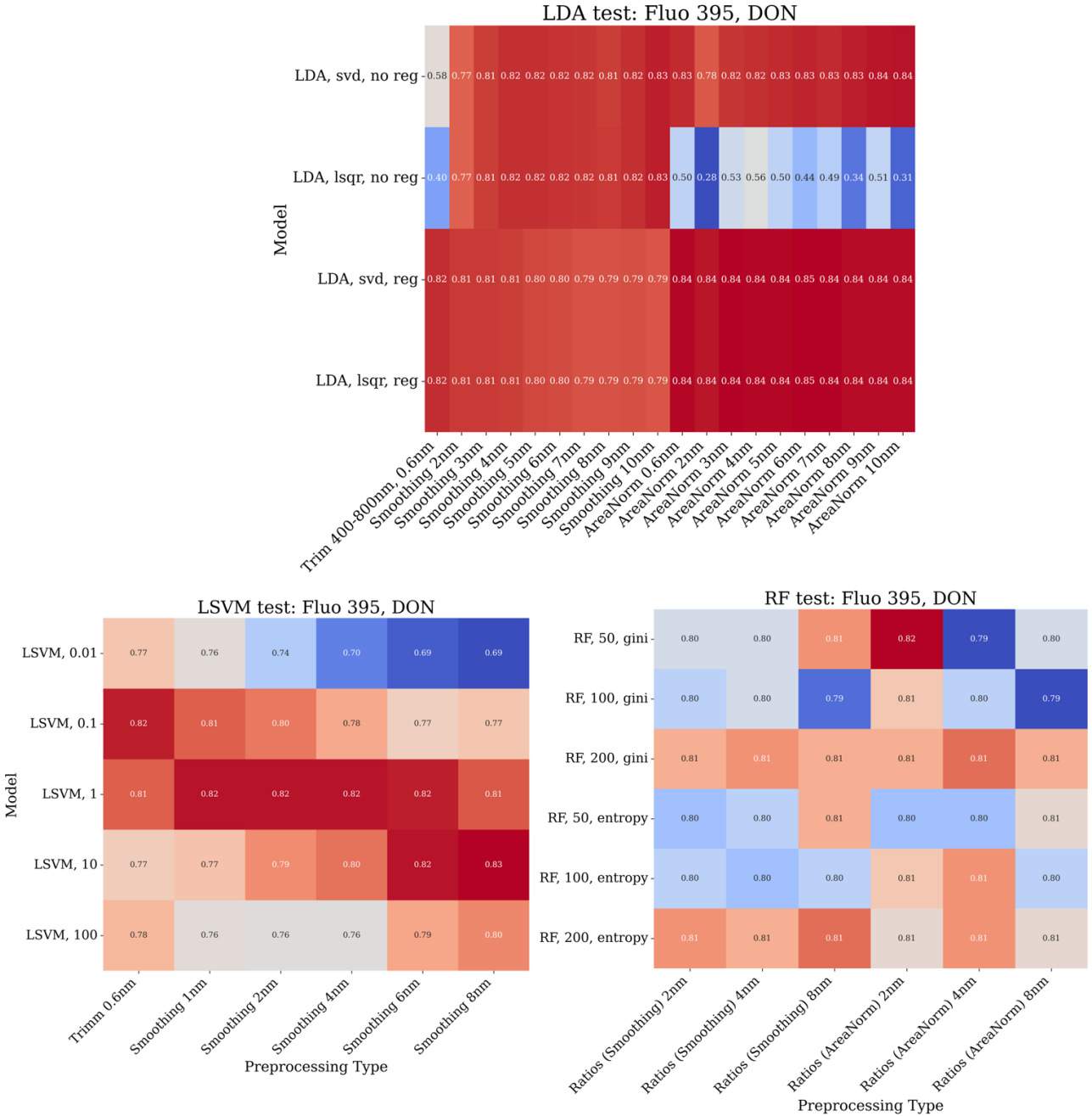


Figure 4.19: DON model optimization results for LDA, LSVM, and RF using 395nm fluorescence data. Various hyperparameters were explored in combination with refined preprocessing techniques. The optimization led to slight improvements in classification performance compared to the initial results presented in Figure 4.7.

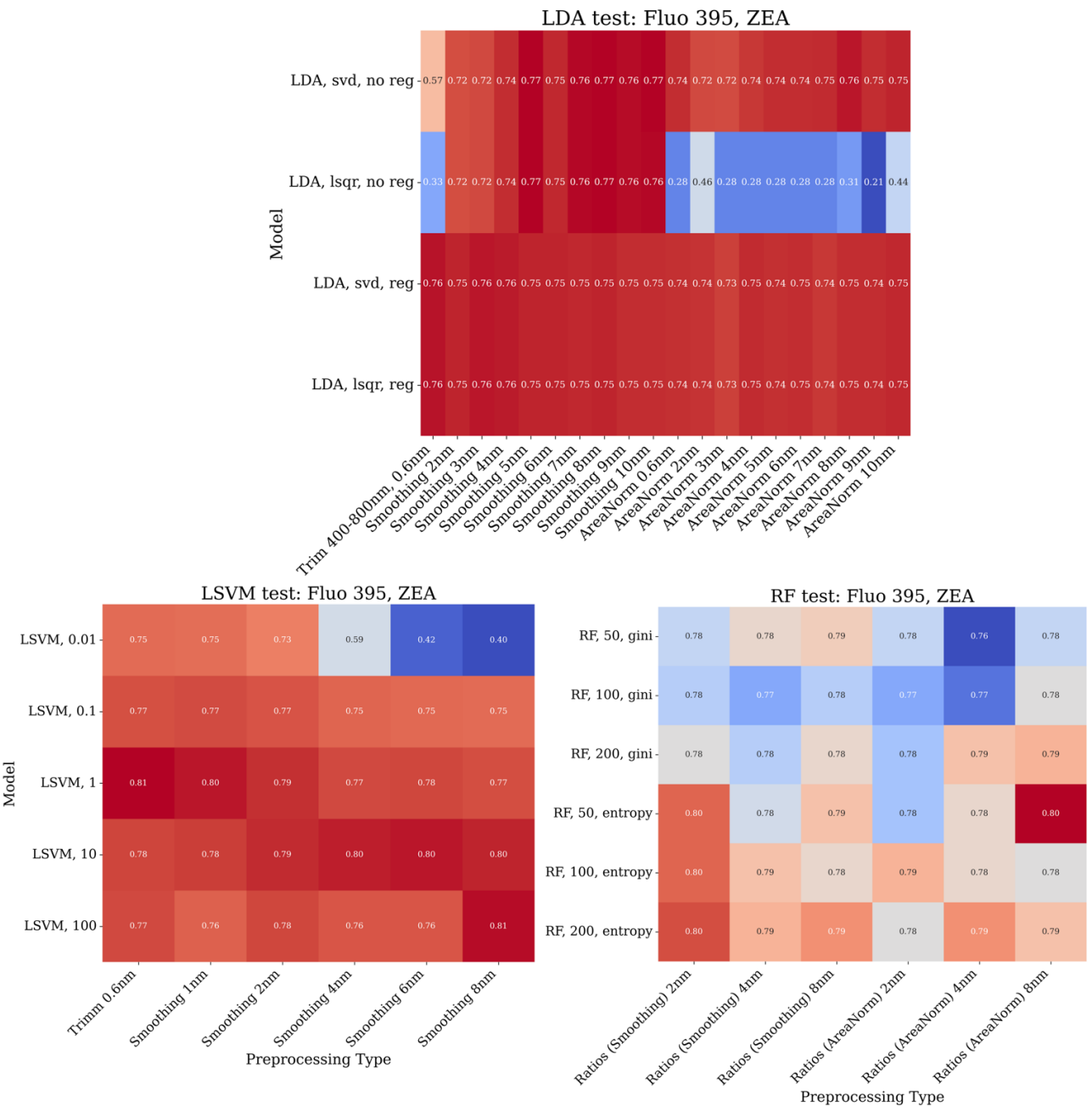


Figure 4.20: ZEA model optimization results for LDA, LSVM, and RF using 395nm fluorescence data. Various hyperparameters were explored in combination with refined preprocessing techniques. The optimization led to slight improvements in classification performance compared to the initial results presented in Figure 4.8.

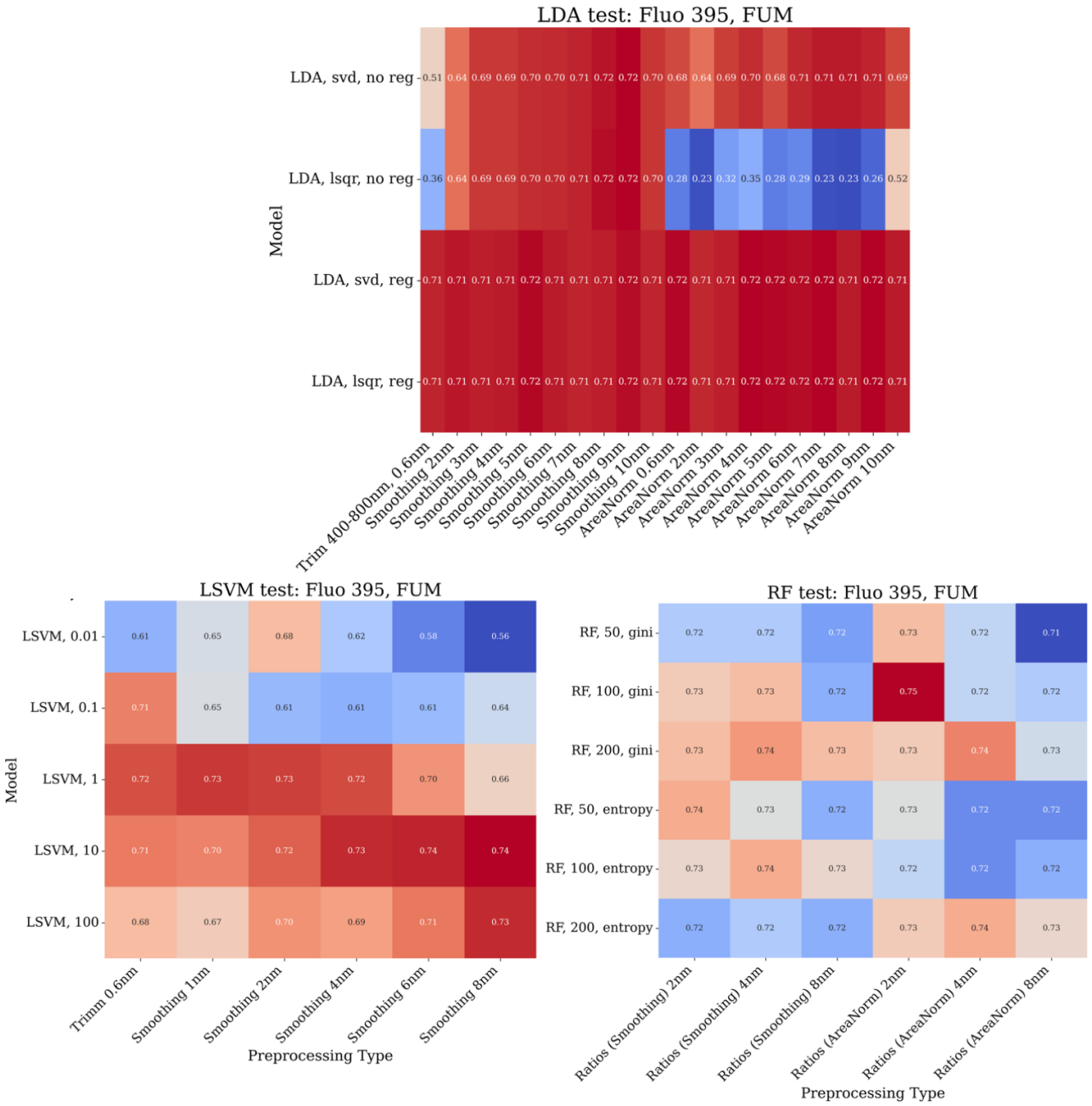


Figure 4.21: FUM model optimization results for LDA, LSVM, and RF using 395nm fluorescence data. Various hyperparameters were explored in combination with refined preprocessing techniques. The optimization led to slight improvements in classification performance compared to the initial results presented in Figure 4.9.

4.3.5 Feature Selection

In the previous sections, we thoroughly explored the process of identifying optimal classifiers for the detection of AFLA, DON, ZEA, and FUM in corn. Up to this point, all models were trained using the full preprocessed spectral data as input. However, one of the key objectives of this work is to reduce the number of spectral features (wavelengths) used by the models to a small, informative subset—ideally between 8 and 10. As previously discussed, such a reduction would significantly enhance the practicality of these models for integration into real-time food sorting systems and portable, in-field detectors.

To achieve this dimensionality reduction, we evaluated feature selection algorithms like Sequential Feature Selection (SFS, Section A.3.2), and Minimum Redundancy Maximum Relevance (MRMR, Section A.3.3). While PCA is effective at reducing dimensionality, it does so by generating new features as linear combinations of the original wavelengths. Since our goal is to retain interpretability and identify specific, physically meaningful wavelengths, we focus here on SFS and MRMR, which directly select actual wavelengths rather than combined features.

In this section we analyze both dimensionality reduction methods and investigate the trade-off between classification accuracy and feature reduction. To ensure a fair comparison, feature selection was applied to the two best-performing and most computationally efficient classifiers identified earlier: LDA (without regularization, using the SVD solver) and LSVM (with $C = 1$). For both models, smoothed data at 8nm resolution and area-normalized data at 8nm were used. Feature selection was performed on the top fluorescence datasets considering 395nm to 365nm excitation, and the models were evaluated for varying numbers of selected features, ranging from 1 to 30. The classification performance across these conditions is visualized as accuracy heatmaps in Figure 4.22, showing model accuracy as a function of the number of selected features for both MRMR and SFS.

Running a single heatmap for feature selection proved to be highly time-consuming: approximately 2 hours were required to generate a full MRMR heatmap, while a complete SFS map took around 22 hours due to its iterative nature. As a result, SFS was executed fully only for the DON dataset. For AFLA, ZEA, and FUM, a reduced subset of SFS configurations were evaluated to confirm whether similar performance trends could be observed. This reduced coverage leads to the empty spaces seen in Figure 4.22. Besides, we observed that integrating SFS with LDA resulted in significantly faster and more effective feature selection than with LSVM, since an LDA model trains much faster, and consistently achieves higher accuracy with fewer selected features.

Beyond the stark difference in computational demands, a clear distinction in performance emerged. SFS consistently outperformed MRMR, particularly when working with a small number of selected features. While MRMR required more than 20 features to approach the performance of full-spectrum models, SFS achieved near-maximum accuracy with only 10 features.

For example, in the case of DON classification, MRMR reached 75% accuracy using 10 features, whereas SFS achieved 83% with the same number. At 30 features, MRMR reached 83%, and SFS slightly surpassed this with 85%, which closely aligns with the performance of the full-spectrum model. These findings demonstrate that SFS is far superior to MRMR when the objective is to reduce input dimensionality while maintaining high accuracy—especially in scenarios where only a limited number of features can be used. SFS's superiority, however, comes at a high computational cost. This is inherent to its sequential nature: in each step, SFS evaluates all possible feature additions to identify the one that improves performance the most, repeating this process until the desired number of features is selected. This results in a combinatorial increase in training iterations, especially for models with longer training times.

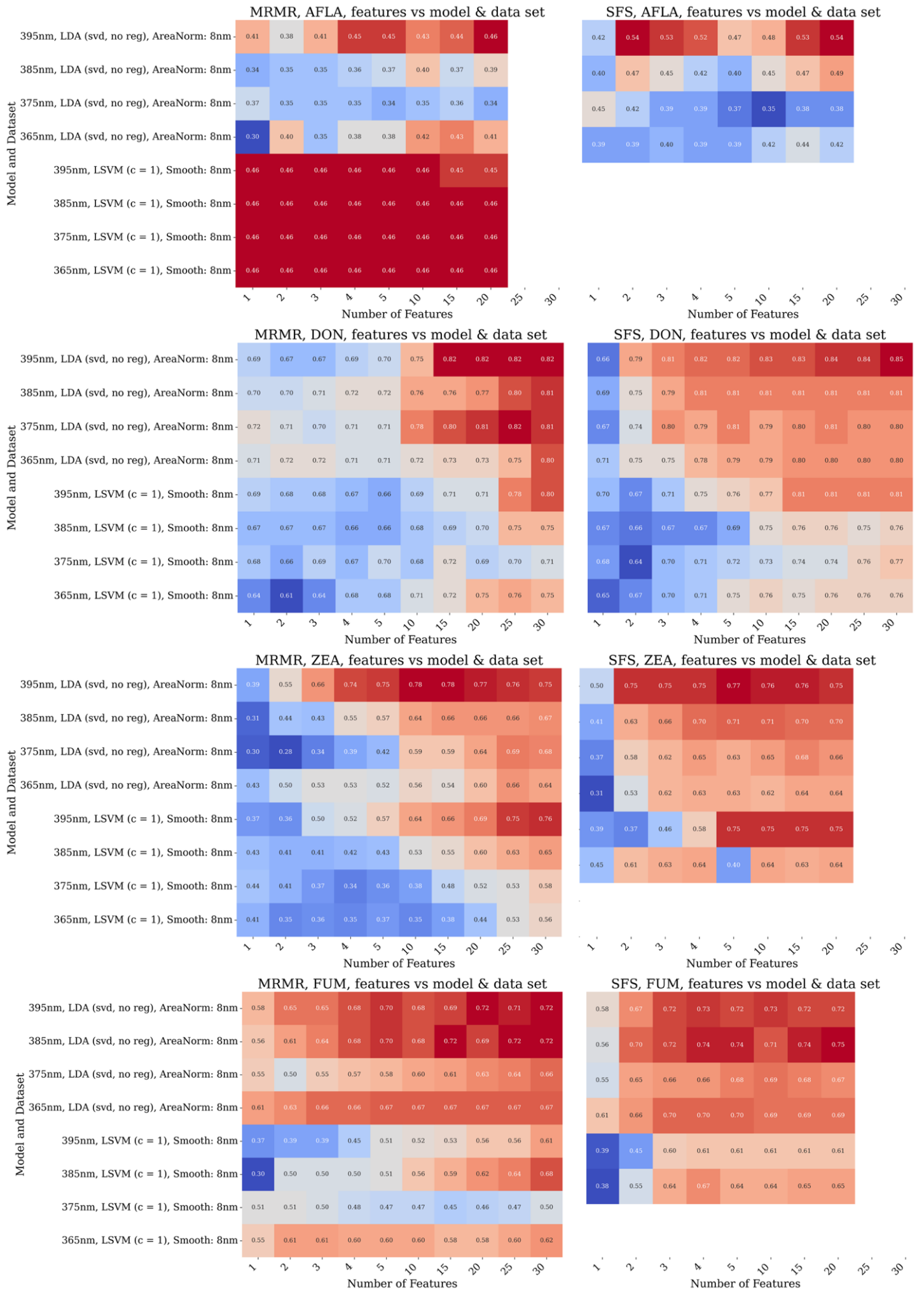


Figure 4.22: Accuracy heatmaps comparing feature selection performance using MRMR (left) and SFS (right). Each heatmap shows classification accuracy as a function of the number of selected features, across different model-preprocessing-spectroscopic modality combinations. SFS clearly outperforms MRMR when selecting a small number of features, achieving near full-spectrum accuracy with as few as 10 wavelengths.

4.3.5.1 10 Feature Models

We now investigate how well the LDA classifier (with smoothing 8nm) performed for AFLA, DON, ZEA, and FUM detection, when using Sequential Feature Selection (SFS) to reduce the spectral input to only the 10 most informative wavelengths.

An inner two fold CV was used to realize the feature selection. Due to the inherent randomness in the SFS process—where different subsets of features may be selected in each run—each model was trained and evaluated 5 times. This allows us to capture the variability introduced by the feature selection step and report both the average accuracy and standard deviation. The final results are presented in Table 4.7.

Toxin	395nm [%]	385nm [%]	375nm [%]	365nm [%]
DON	48.00 ± 1.37	45.05 ± 1.58	34.35 ± 1.78	42.24 ± 0.89
DON	83.08 ± 0.31	80.07 ± 0.66	80.05 ± 1.15	79.66 ± 0.88
ZEA	74.97 ± 0.21	70.70 ± 0.73	67.27 ± 0.73	63.70 ± 0.96
FUM	71.35 ± 0.96	72.19 ± 0.39	66.10 ± 1.60	70.19 ± 0.70

Table 4.7: Classification accuracy with standard deviation for AFLA, DON, ZEA, and FUM using LDA models trained on 10 features, same as shown in Figure 4.22, across different fluorescence excitation wavelengths.

From the table, we observe that the 395nm excitation again provides the best classification performance across all toxins, consistent with results obtained using full-spectrum data. Moreover, the classification accuracy using only 10 selected features is only slightly lower—typically by around 2%—compared to the full-spectrum models. For example, DON reaches an accuracy of $(83.08 \pm 0.31)\%$, just below the 85% obtained after full model optimization. Similarly, ZEA and FUM achieve $(74.97 \pm 0.21)\%$ and $(71.35 \pm 0.96)\%$, respectively, compared to 77% and 72% with the full spectral data. Only for AFLA do we observe a significant performance decrease from 57% to just $(48.00 \pm 1.37)\%$.

Overall, the results demonstrate that SFS enables a substantial reduction in model complexity while largely preserving classification performance for models that preformed well on full spectral data.

4.3.5.2 Complexity For AFLA Detection

The classification performance for AFLA has consistently been poor—starting from the initial model-search and persisting through to the current stage. One contributing factor previously discussed is the use of extremely challenging classification thresholds of 2 and $4\mu\text{g}/\text{kg}$, which are significantly below those typically reported in literature (e.g., $10\mu\text{g}/\text{kg}$ or higher). This increased difficulty in distinguishing between such low concentrations likely plays a role in the underwhelming model performance.

Another important issue, not yet explored in detail, concerns the composition of the AFLA-contaminated samples. Specifically, two different standards were used to synthesize the samples: a pure AFLA standard, with a nominal concentration of $(4.8 \pm 1.9)\mu\text{g}/\text{kg}$, and a MIX standard, with $(6.0 \pm 2.5)\mu\text{g}/\text{kg}$ of total AFLA. The pure standard details only AFB₁, with AFB₂, AFG₁, and AFG₂ present below $1\mu\text{g}/\text{kg}$, whereas the MIX standard reports a combined concentration of AFB₁, AFB₂, AFG₁, and AFG₂.

These values highlight a major concern: the relative uncertainty in these concentrations is extremely high. Given the small absolute concentrations involved, even minor deviations result in large relative errors. For example, if two samples are composed to contain $2\mu\text{g/kg}$ AFLA using the two different standards, and one of them deviates by just one standard deviation, the actual concentration could differ by over 50%. In practice, this means that two nominally identical samples—one composed from the pure standard and the other from the MIX—could have substantially different true concentrations. This label inaccuracy severely limits the model’s ability to learn reliable spectral patterns, as the relationship between spectra and ground-truth labels becomes distorted.

To investigate whether this composition-based variability was indeed responsible for the poor performance, we repeated the classification experiment using only the samples composed from the pure AFLA standard, thereby eliminating the variability introduced by mixing sources⁸. The results are shown in Figure 4.23: the confusion matrix on the left corresponds to the model trained on all AFLA-contaminated samples, using full spectral information, while the right shows the performance when restricted to samples composed solely from the pure standard, also using the full spectrum. The difference is striking. Classification accuracy improved from 56.7% to 72.6%—a substantial gain that strongly suggests the inconsistencies in sample composition were a key factor behind the poor performance observed earlier.

However, this improvement comes at a cost. By restricting the dataset to samples composed using only one standard, the number of available training samples is significantly reduced. This limits the statistical robustness of the performance evaluation and introduces greater uncertainty into the model’s generalizability. For this reason we won’t consider AFLA going forward for the three-class classification. Nevertheless, these findings underscore the importance of consistent and well-characterized sample composition, especially when working at low contamination levels and attempting high-resolution classification.

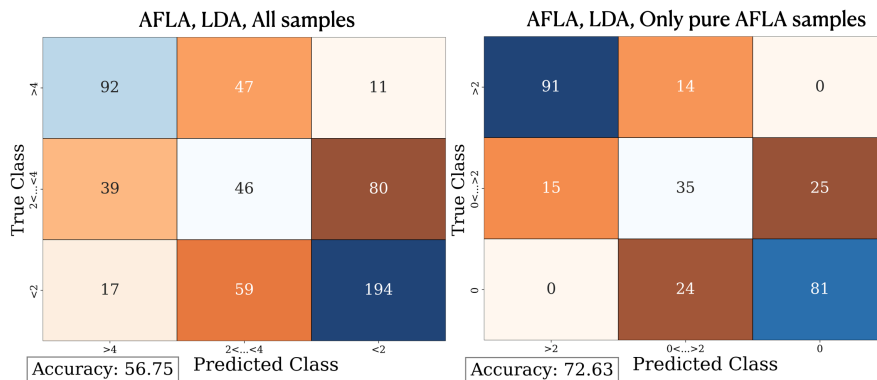


Figure 4.23: Classification performance for AFLA using a three-class classification scheme with thresholds set at 2 and $4\mu\text{g/kg}$. The left confusion matrix shows the results when using all available AFLA-contaminated samples, while the right matrix reflects performance when restricted to samples composed solely using the pure AFLA standard, full spectral information was used in both cases.

⁸A similar experiment was conducted using only the samples containing AFLA from the MIX standard. However, the sample distribution in this subset was highly imbalanced, with one class being overrepresented. This led to a biased training process and poor model performance, as the classifier was unable to effectively learn the distinctions between the AFLA-specific classes.

4.3.6 Combining Complementary Data

So far, we have demonstrated successful classification of DON, ZEA and FUM in corn using fluorescence spectroscopic data. Through model-search and optimization, LDA, LSVM, and RF emerged as the top-performing classifiers. However, considering the trade-off between classification accuracy, computational cost, and compatibility with feature selection, LDA was selected as the most suitable model. Using Sequential Feature Selection (SFS), we were able to reduce the spectral input to just 10 features while still achieving approximately 95% of the accuracy obtained with full-spectrum models. This significant dimensionality reduction greatly simplifies both model complexity and potential hardware implementation. In addition, our comparison across optical modalities showed that fluorescence consistently outperformed reflectance-based approaches, with the 395nm excitation providing the highest classification accuracy for all four toxins.

Up until this point, each spectroscopic modality was considered independently. However, the complementary nature of the various modalities (fluorescence at different excitations and different reflectance ranges) presents an opportunity to improve performance by integrating their combined information. One straightforward approach—concatenating the full spectra from multiple modalities—was tested but proved computationally impractical, particularly when combined with SFS, due to the large increase in data dimensionality.

To address this, we adopted a more efficient fusion strategy: first applying SFS independently to each of the four best-performing fluorescence modalities (395–365nm) to select the top 10 features per modality. These 40 features were then combined into a new, unified dataset for model training. The classification results from this fused feature set are shown in Figure 4.24, where LDA models were trained using smoothed data (8nm resolution).

We achieved classification accuracies of 84.5%, 79.1%, and 78.0% for DON, ZEA, and FUM, respectively. For DON, this performance matches the accuracy previously obtained using the full spectrum at 395 nm, while for ZEA and FUM, it exceeds the full-spectrum results achieved with LDA. These findings suggest that fusing a small number of carefully selected features from complementary fluorescence excitations can enhance performance beyond what is possible with any individual modality alone.

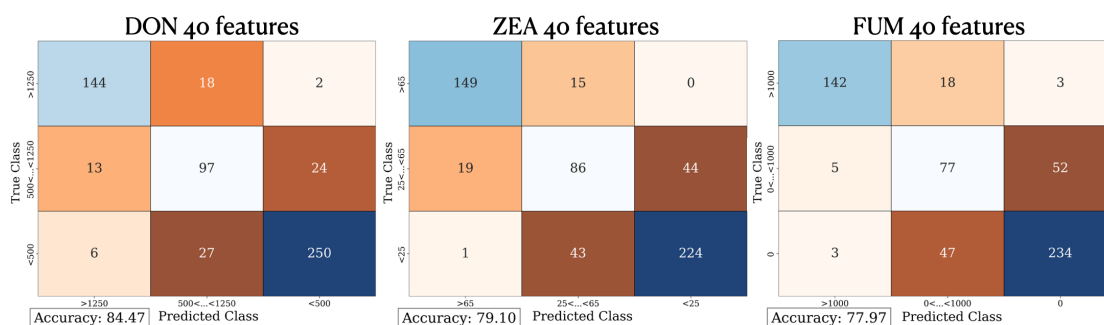


Figure 4.24: Confusion matrices showing the classification performance for DON, ZEA, and FUM using LDA models trained on a fused feature set. Each model was trained using data smoothed at 8nm resolution and the top 10 features selected via SFS from each of the four fluorescence excitation wavelengths (395–365nm), resulting in a combined input of 40 features.

While the fusion of 40 features from complementary fluorescence modalities significantly improved classification performance, this feature set remains relatively large for practical imple-

mentation. To further reduce model complexity, we employed an additional round of Sequential Feature Selection (SFS) to select the top 10 features from across the previously combined set of 40 features. The overall structure of this optimized modeling pipeline is illustrated schematically in Figure 4.25 A.

In this framework, the top 10 features are first selected independently from each fluorescence modality (395–365nm) using LDA and SFS. These 40 selected features are then combined, and a second round of LDA with SFS is applied to identify the most informative 10 features across the fused dataset. The final LDA classifier is then trained using these 10 globally selected features. Figure 4.25 B shows the classification performance as a function of the number of selected features in the final feature reduction step for DON, ZEA, and FUM.

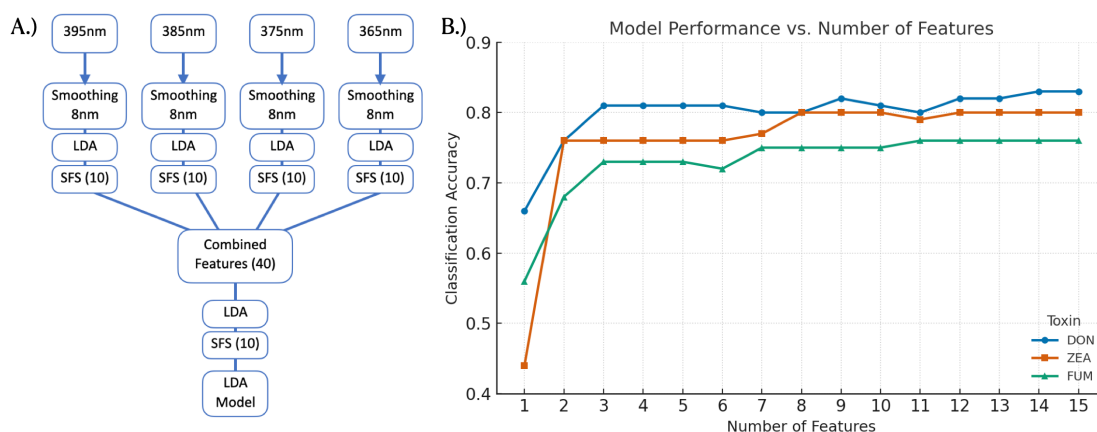


Figure 4.25: A.) Schematic overview of the two-stage feature selection and classification pipeline. First, the top 10 features are selected from each fluorescence modality (395–365nm) using SFS and LDA. These 40 features are then merged, and a second SFS step selects the best 10 features across the combined dataset, used to train the final LDA model. B.) Classification accuracy as a function of the number of selected features during the second SFS stage for DON, ZEA, and FUM.

We focus on the accuracy achieved when selecting only 10 features from the previously combined set of 40, and compare it with three other configurations: (i) the full-spectrum model at 395nm, (ii) the best-performing 10-feature model using only the 395nm data, and (iii) the model trained using all 40 fused features across modalities. This comparative analysis is summarized in Table 4.8.

For DON, the highest classification accuracy (85%) is obtained using the full-spectrum data from the 395nm excitation. Among the models limited to 10 features, the best performance (83%) is achieved by selecting features solely from the 395nm modality. The cross-platform fusion model—selecting 10 features from the combined 40—yields a slightly lower performance of 81% for DON⁹. In contrast, for ZEA and FUM, the best classification accuracies among all 10-feature models are achieved using the cross-platform approach. Selecting 10 features from the fused set improves performance to 80% for ZEA and 75% for FUM, exceeding the accuracy of even the full-spectrum models.

⁹Interestingly, the final SFS procedure does not recover the 10 most informative features from the 395nm set when selecting from the full pool of 40 features, despite this subset achieving the best individual performance. This is likely due to the greedy, iterative nature of SFS, which adds features one at a time based on local optimization at each step. As a result, globally optimal combinations can be missed.

Model Configuration	DON	ZEA	FUM
All features (395nm)	85%	77%	72%
Top 10 features (395nm)	83%	76%	73%
40 fused features (395–365nm)	84%	79%	78%
Top 10 of 40 fused features	81%	80%	75%

Table 4.8: Comparison of classification accuracy for DON, ZEA, and FUM across various model configurations. Results include full-spectrum models, SFS-selected features from a single modality (395nm), fused features from four fluorescence excitations, and the final reduced 10-feature model from the fused set.

Based on these findings, we conclude that the optimal feature-reduced model for DON consists of 10 features selected solely from the 395nm fluorescence data. For ZEA and FUM, however, the best performance is achieved by fusing features from multiple modalities and selecting the top 10 from the combined pool. In the next section, we explore these best-performing models in greater detail, presenting their performance for both three-class and two-class classification tasks, and providing an interpretation of the selected wavelengths.

4.3.6.1 The Best 3-Class Classifiers

In the following subsections, we provide a detailed evaluation of the best-performing 10-feature classification models for DON, ZEA and FUM. These models were selected based on their optimal trade-off between classification performance and model simplicity, as outlined in Table 4.8. Each subsection presents the confusion matrix, analyzes classification trends and errors, and discusses the spectral features selected through SFS.

4.3.6.1.1 3-Class DON-Classifiers

The best-performing 10-feature, 3-class, DON-classifier was an LDA/SFS model trained using features exclusively selected from the 395nm fluorescence dataset, with 8nm smoothing applied. The confusion matrix corresponding to this classification is presented in Figure 4.26. Here we find an overall classification accuracy of 83%.

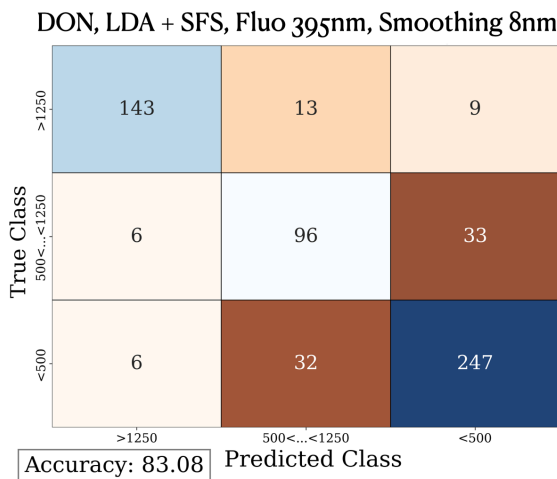


Figure 4.26: Confusion matrix for the three-class classification of DON using the LDA model with features selected via SFS from the 395nm fluorescence data, preprocessed with 8nm smoothing. The results show high overall accuracy with very few second-order errors.

A closer inspection of the results reveals that only a small number of spectra were misclassified in a critical way—i.e., labeling unsafe samples as safe or vice versa. Most misclassifications were first-order errors (e.g., classifying a(n) (un)safe sample as risky or a risky one as (un)safe), rather than second-order errors that span two categories. In fact only 2.6% of the dataset was involved in second-order misclassifications, while approximately 14.4% resulted in first-order errors. This is a promising outcome, indicating that the model is highly effective at distinguishing safe from unsafe samples, which is critical for regulatory and practical applications.

The specific wavelengths selected by SFS¹⁰ are listed in Table 4.9. These features span the entire spectral range, with a notable concentration between 400–550nm—corresponding to the region where the DON signal visually shows the most pronounced differentiation from ZEA and FUM, as illustrated in Figure 3.15.

Feature	λ [nm]
1	772
2	496
3	488
4	520
5	568
6	528
7	560
8	504
9	400
10	640

Table 4.9: Top 10 wavelengths selected by SFS for the DON classifier using 395nm fluorescence data and LDA. These wavelengths represent the most informative features contributing to the three-class classification model. Note that these wavelengths are the center point of the smoothed 8nm spectral information actually selected.

Interestingly, the first feature selected was at 772nm, near the upper edge of the considered spectrum. Although this region does not visibly distinguish DON from other toxins, its selection likely reflects a limitation of the SFS algorithm: during the initial iteration, each candidate feature is evaluated in isolation, and all models perform poorly because they attempt to predict a multi-class outcome using a single wavelength-feature. As a result, the selection of the first feature tends to be inconsistent across folds, without a strong, repeatable preference for any particular wavelength. Only in later iterations—once the algorithm begins forming feature combinations—do clearer patterns emerge, and the most informative wavelengths are consistently identified.

To further analyze model performance, we also examined accuracy on a per-fold basis to identify individual samples that were frequently misclassified. While most samples were classified correctly across the most of the 15 spectra per sample, several outliers emerged, as highlighted in Figure 4.27.

Notably, sample S10 posed the greatest classification difficulty. Despite a DON concentration of 1320 μ g/kg—above the legal threshold—S10 accounted for all second-order errors in which unsafe samples were misclassified as safe¹¹. Additional problematic samples included S14, S09, and S24.

¹⁰Wavelengths were chosen based on the 10 most frequently selected features across all folds of the LOSO-CV procedure.

¹¹As will become evident in later sections, sample S10 frequently appears as an outlier across multiple experiments. Its consistent misclassifications raises concerns about the sample's composition. Although the author did not personally prepare this sample, a possible explanation could be a mislabeling or error during preparation,

Sample S14 has a DON concentration of $440\mu\text{g/kg}$ —just below the $500\mu\text{g/kg}$ threshold separating the safe and risky classes. It is therefore not surprising that this sample was frequently misclassified as risky. Under the LOSO-CV approach, the model is trained without access to this specific sample and does not explicitly “*know*” the threshold. Instead, it learns from the distribution of contamination levels present in the training set. When S14 is used as the validation sample, it becomes the closest point to the decision boundary, yet the model has only seen samples with higher or lower concentrations during training. As such, classifying S14 inconsistently—as either safe or risky—can be understood as a natural result of interpolation. The model attempts to generalize based on surrounding data, and this behavior does not necessarily indicate a true classification error, but rather reflects the inherent ambiguity near the threshold. Samples S09 and S24, however, presented more concerning misclassifications. Both had DON concentrations of approximately $880\mu\text{g/kg}$ (well within the risky range), yet were misclassified as safe for all spectra. These were, however, the only risky-class samples composed using the pure DON standard, whereas most other samples containing this standard were in either the safe or unsafe class. It is therefore possible—though not confirmed—that the model grouped these samples based on their matrix composition rather than true DON content. This suggests that some residual sample matrix effects may still influence classification, especially when the training data lacks sufficient variation within specific concentration zones.

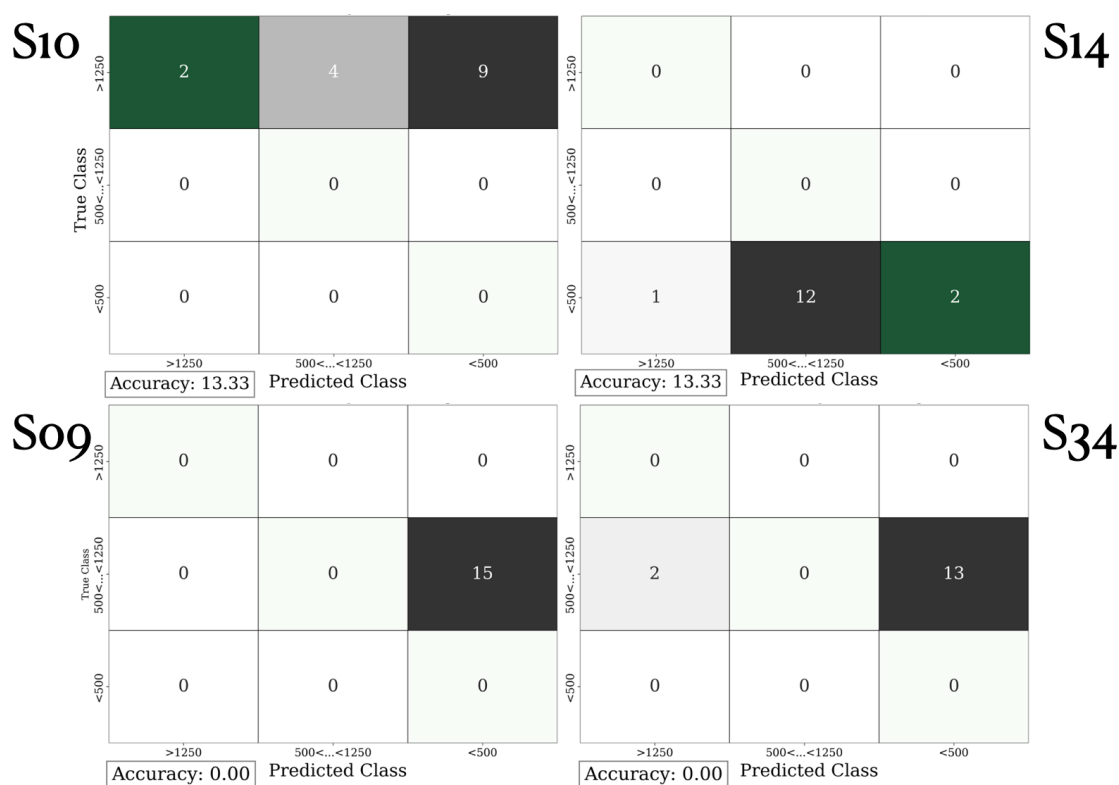


Figure 4.27: Most significant misclassified samples in the three-class DON classification task on corn. Samples S09, S10, S14 and S24, show repeated misclassifications, including second-order errors.

especially considering that other samples with similar DON concentrations, such as S05, were correctly classified.

4.3.6.1.2 3-Class ZEA & FUM-Classifiers

The top-performing 10-feature models for ZEA and FUM were the models using fused datasets spanning 395-365nm fluorescence data. Figure 4.28 presents the confusion matrices for both models. In both cases, second-order misclassifications are minimal. In the case of ZEA, the most prominent misclassifications occurred for samples S19, S38, and S10. Sample S19 lies close to the unsafe threshold with a concentration of $70.3\mu\text{g/kg}$, just above the cutoff of $65.0\mu\text{g/kg}$, making its classification inherently ambiguous—this kind of error is expected at class boundaries. Samples S38 and S10 fall clearly within their respective classes, and the cause of their misclassification remains unclear¹². For FUM, the most significant classification errors involve samples S17 and S34, with contamination levels of 500 and $832\mu\text{g/kg}$, respectively. These were misclassified as “safe” despite clearly falling in the “risky” range (thresholds set at 0 and $1000\mu\text{g/kg}$). These errors are more difficult to explain and may reflect limitations in the model’s ability to detect moderate contamination levels for FUM.

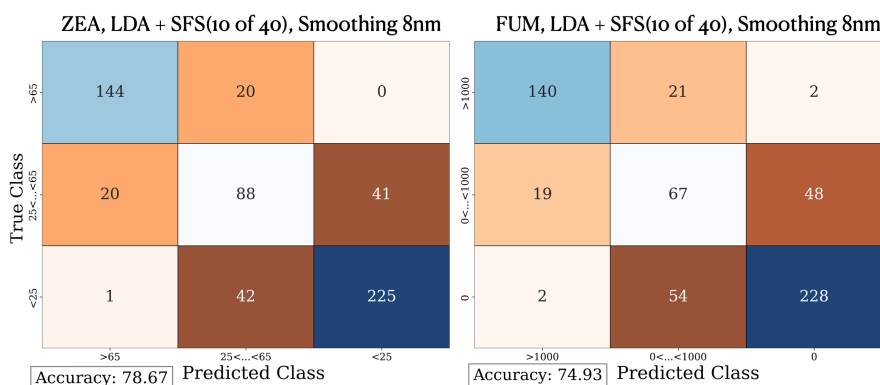


Figure 4.28: Confusion matrix for the three-class classification of ZEA (left) and FUM (right) using the LDA model with 10 features selected via SFS from the the top 40 features from across 395-365nm fluorescence data, preprocessed with 8nm smoothing. The results show high overall accuracy with very few second-order errors, indicating the model’s strong ability to distinguish between critical contamination levels.

Examining the selected wavelengths listed in Table 4.10, we observe that the majority of features originate from the 395nm fluorescence dataset, followed by the one of 385nm excitation. For ZEA, there is also one contribution from the 375nm set, while for FUM, several informative features originate from the 365nm data.

Notably, the selected wavelengths for ZEA and FUM tend to cluster above 600nm, contrasting with those selected for DON, which were primarily located in the 400–550nm range. This might indicate that we are differentiating based on the chlorophyll fluorescence signal as this is the only noticeable spectra feature in this range at approximately 670nm. Though visually we can not get a clear idea of why this spectral region can classify the samples so well. The observed shift from the 400-550nm region to the +600nm region, does suggest that for different toxins different spectral regions carry the most discriminative information.

This difference may be partly explained by the underlying fluorescence properties of the toxins: ZEA is a fluorescent mycotoxin and exhibits intrinsic emission in the 400–500nm range, which is also reflected in the selected wavelengths. In contrast, FUM is non-fluorescent and its classifi-

¹²It is worth noting that sample S10 also caused misclassifications in the DON model, suggesting that the issue may be sample-specific rather than model-related, though this cannot be confirmed with certainty.

cation likely depends on secondary effects, such as alterations in the sample matrix, for example on the chlorophyll-characteristic regions.

Rank	ZEA		FUM	
	Excitation Wavelength [nm]	λ [nm]	Excitation Wavelength [nm]	λ [nm]
1	395	712	385	648
2	385	720	385	496
3	385	640	365	632
4	395	704	385	792
5	395	672	365	672
6	385	448	395	672
7	375	404	395	496
8	395	648	365	616
9	385	624	385	627
10	385	672	365	664

Table 4.10: Top 10 spectral features selected by SFS for ZEA and FUM classification models from the top 40 cross-platform pooled features. Each feature is characterized by its corresponding dataset (fluorescence excitation wavelength in nm) and its central wavelength. The rank indicates the order in which the features were selected one by one.

4.3.6.2 The Best 2-Class Classifiers

Up to this point, we have extensively analyzed the performance of the best-performing three-class models for the classification of DON, ZEA, and FUM in corn. However, a key limitation remains: three-class results cannot be directly compared to the vast majority of studies in the literature, which almost exclusively report results from binary (two-class) classification tasks. To enable such a comparison and benchmark our models accordingly, we now turn our attention to the development of two-class classification models.

The threshold values used for this binary classification are presented in Table 4.2. Using these thresholds, we evaluated the performance of the best 10-feature models developed using LDA in combination with SFS, trained on the 395nm fluorescence dataset with 8nm smoothing. The results for all four toxins—AFLA, DON, ZEA, and FUM—are presented in Figure 4.29. Across all toxins, two-class models clearly outperform their three-class counterparts, as also highlighted in Table 4.11.

This gain in performance, however, comes at the cost of reduced interpretability: the intermediate “*risky*” class used in the three-class scheme provided valuable insight during model development and allowed for a more nuanced risk assessment. Nevertheless, from a practical or industrial perspective, a binary classification system—indicating whether a sample is safe or unsafe—is often sufficient and more easily implemented.

Next we can detail the results for the four toxins, AFLA, DON, ZEA and FUM, and thereafter make the comparison to literature.

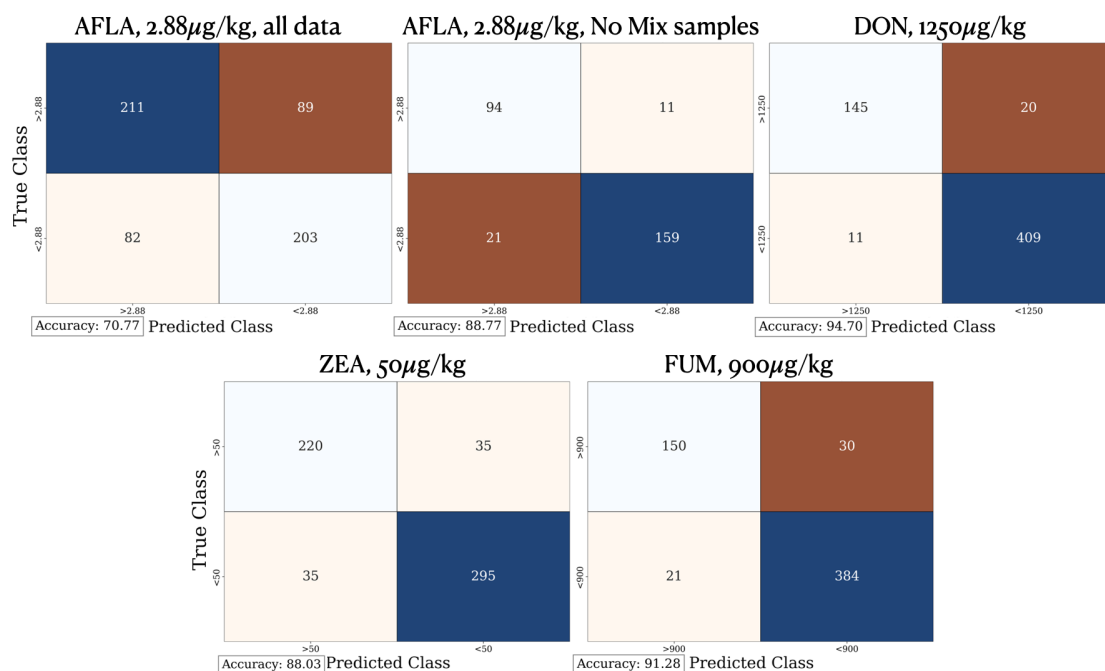


Figure 4.29: Confusion matrices for the two-class classification of corn samples using an LDA model, considering fluorescence data with excitation at 395nm, smoothed at 8nm resolution, and reduced to 10 features via SFS. Each matrix corresponds to one toxin—AFLA, DON, ZEA, and FUM—with threshold values indicated above each plot. For AFLA, two cases are distinguished: one using all available samples, and another using only samples composed with the pure AFLA standard. The latter yields notably better performance, suggesting that inconsistencies between standards may have affected classification accuracy.

Type	AFLA	DON	ZEA	FUM
3 Classes	72.6%	83.0%	78.7%	75.0%
2 Classes	88.8%	94.7%	88.7%	91.3%

Table 4.11: Best classification accuracy for each toxin using 10-features from the 395nm fluorescence data, shown for both two-class and three-class classification tasks. Performance improves significantly when switching to binary. For AFLA we show the case only considering sample composed by mixing the pure standards.

AFLA: Unlike the previous sections focused on 3-class classification, AFLA is now included in the analysis, as the reduced complexity of a binary classification task enables effective model training with fewer samples. The first plot in Figure 4.29 shows the performance for AFLA classification using all samples and a threshold at 2.88 $\mu\text{g/kg}$, chosen to ensure balanced class sizes. A classification accuracy of 70.8% is obtained. However, when restricting the dataset to only those samples composed using the pure AFLA standard (excluding the mixed standard), classification accuracy increased significantly to 88.8%, as shown in the second plot of Figure 4.29. This further supports the earlier hypothesis that the large relative uncertainty between the AFLA concentrations in both standards severely hampered model performance, but for a two class

differentiation this effect is less severe.

DON: For DON, we used the legal threshold of $1250\mu\text{g/kg}$. The resulting model achieved an excellent classification accuracy of 94.7%. As before, some of the major sources of misclassification include sample S10, as well as other edge-case samples close to the decision boundary.

ZEA and FUM: For ZEA and FUM, thresholds of $50\mu\text{g/kg}$ and $900\mu\text{g/kg}$ were used, respectively. This yielded classification accuracies of 88.0% for ZEA and 91.3% for FUM. The most prominent classification errors again involved edge samples near the threshold and certain problematic samples previously identified in the three-class classification task, including S10.

We now compare the performance with results reported in state-of-the-art literature. For AFLA (excluding samples composed with the MIX standard) we achieved a classification accuracy of 88.7%. In comparison, literature reports: 84.0% using ANN [113], 87.7% using SVM [105], 87% using ANN [114], and 78.2% using LDA [103], all with a threshold of $20\mu\text{g/kg}$. A more recent study by Kim et al. [93] reached 95.7% using a $10\mu\text{g/kg}$ threshold.

While our accuracy falls slightly below this recent benchmark, it is noteworthy that our model used an even stricter threshold of $2.88\mu\text{g/kg}$. However, the limited number of samples in our AFLA set, and AFLA derived solely from one standard, reduces our confidence in this result, as there remains a possibility that the classifier is recognizing spectral characteristics of the matrix rather than AFLA concentration itself.

For DON, we achieved a classification accuracy of 94.7% using the legal threshold of $1250\mu\text{g/kg}$. This is a particularly strong result, especially given the absence of prior fluorescence-based DON classification studies in the literature—likely due to the toxin’s non-fluorescent nature. Nonetheless, our model was able to detect DON indirectly, likely leveraging spectral signatures associated with matrix alterations induced by mycotoxin metabolism. For comparison, Kos et al. [99] reported an accuracy of 79% using UV–VIS spectroscopy and decision trees, with a higher threshold of $1750\mu\text{g/kg}$. These findings suggest that fluorescence-based methods may offer a promising and potentially superior alternative for DON detection, despite its lack of inherent fluorescence.

For ZEA, our model reached an accuracy of 88% using a threshold of $50\mu\text{g/kg}$ —well below the regulatory limit of $350\mu\text{g/kg}$. This is comparable to the 93.3% accuracy reported by Wang et al. [102], who used hyperspectral fluorescence imaging (HSI) for maize kernels and a higher threshold. Our approach, however, relies on a far simpler setup using spectral information at only discrete wavelengths, and uses a lower threshold for detection.

Finally, for FUM, we obtained an accuracy of 91.3% using a $900\mu\text{g/kg}$ threshold, also below the legal limit of $4000\mu\text{g/kg}$. FUM is also non-fluorescent and literature only reported reflectance results. A UV-VIS study by Stasiewicz et al. [98] achieved 77% accuracy using a $1000\mu\text{g/kg}$ threshold. Another recent work by Kim et al. [28] reported 95.7% accuracy using shortwave infrared (SWIR) spectroscopy, though no specific threshold was provided. Our performance is thus comparable, and illustrates that fluorescence can also be used for its detection.

In summary, despite the use of lower concentration thresholds and only a handful of wavelengths, our fluorescence-based models demonstrated performance that is **comparable to—or in some cases even surpasses—the ones that are reported in the literature for all four mycotoxins**. Importantly, the models presented thus far are each trained to detect a single toxin, yet they do so effectively even in the presence of multiple co-occurring contaminants. This effectively allows for a **modular approach enabling straightforward integration of multiple**

models to enable multi-mycotoxin detection, with accuracies per toxin as specified in Table 4.11. This result goes beyond the state-of-the-art and shows that effective multi-mycotoxin monitoring on corn using fluorescence spectroscopy is feasible.

4.3.7 Conclusion

In this section, we systematically developed and optimized classification models for the detection of four major mycotoxins—AFLA, DON, ZEA, and FUM—in corn. Leveraging fluorescence spectroscopy across five excitation wavelengths and reflectance spectroscopy in both UV–VIS and NIR ranges, we explored a comprehensive modeling pipeline: from model selection and hyperparameter tuning to feature selection and multi-modal data fusion. Throughout this process, a three-class classification scheme was employed to distinguish between safe, risky, and unsafe contamination levels, providing a more nuanced and informative framework compared to traditional binary classification approaches. Finally, two-class classification models were developed following the same strategy, and their performance was compared with findings reported in the literature.

Some interesting findings were that reflectance spectroscopy underperformed across all toxins and was therefore excluded from further optimization steps. Even in its best configurations—using a large integrating sphere—reflectance remained inferior to fluorescence. Among all fluorescence modalities, excitation at 395nm consistently delivered the highest classification performance.

In the model selection process, LDA, LSVM, and RF emerged as the top-performing classifiers, with LDA offering the best overall balance between classification accuracy, computational efficiency, and compatibility with feature selection techniques such as Sequential Feature Selection. Using LDA/SFS to select the 10 most informative wavelengths we realized models that closely matched full-spectrum performance. For example, DON was classified with 83% accuracy using only 10 fluorescence features at an excitation of 395nm, compared to 85% for the full-spectrum model. ZEA and FUM achieved 80% and 75%, respectively, when fusing and reducing features from multiple fluorescence modalities.

Despite these successes, the study also revealed key limitations—most notably in AFLA detection, where performance was significantly hindered by high variability in sample composition and uncertainty in the concentration of commercially provided AFLA standards. Classification accuracy improved dramatically (from 57% to 73% in 3-class, and from 71% to 89% in 2-class) when limiting the analysis to only samples prepared with the pure standard and not the mixed standard. This emphasizes the critical importance of well-characterized training samples, especially when working with low contamination thresholds and complex multi-class frameworks.

When benchmarked against literature, the final two-class models demonstrate comparable or superior performance despite the use of lower toxin thresholds and using only a handful of wavelengths. This highlights the strong practical potential of fluorescence spectroscopy coupled with machine learning models for real-time, portable, or inline multi-mycotoxin detection systems.

4.4 Classification Of Mycotoxins On Wheat

In the previous section, we extensively detailed the development of classification models for the independent detection of AFLA, DON, ZEA, and FUM in corn. This development included identifying the optimal model architecture per toxin, applying feature selection to reduce the spectral dimensionality to just 10 key wavelengths, and integrating features across multiple spectroscopic modalities to further enhance performance. The final models reached classification accuracies

close to or exceeding those reported in literature, despite the use of lower contamination thresholds and relying on only a select number of spectral features. In this section, we now present the analogous model development pipeline for the multi-mycotoxin on wheat.

The wheat samples were prepared by mixing two standards—a contaminated one (containing DON, T-2/HT-2, and OTA) and a blank—to create a nearly equidistant series of contamination levels. As a result, all samples contain proportional mixtures of these toxins, allowing only for the detection of their simultaneous presence. Details of sample composition are provided in Table 3.3. The dataset was divided into three classes of equal size using DON concentration thresholds set at 330 and 630 $\mu\text{g}/\text{kg}$. Spectral data from both fluorescence (excitation wavelengths: 395, 385, 375, 365, and 256nm) and reflectance measurements (UV-VIS and NIR) were utilized, as described in Section 3.3.

For model development, the same structured methodology was applied: an initial model screening phase evaluated various classifiers and preprocessing techniques across all spectroscopic platforms. This was followed by model optimization, involving the fine-tuning of hyperparameters and preprocessing pipelines for the top-performing classifiers. Finally, dimensionality reduction was explored through feature selection techniques.

4.4.1 Model-Search

The model-search strategy for mycotoxin classification in wheat mirrored the approach detailed earlier for corn, as described in Section 4.3.1. We evaluated the same set of models across the various spectroscopic modalities, using the same preprocessing strategies. The classification accuracy heatmaps for the fluorescence and reflectance datasets are shown in Figures 4.30 and 4.31, respectively.

As observed before, LDA, LSVM, and RF consistently emerged as the top-performing classifiers. LDA achieved its best results using smoothed data at lower spectral resolutions. LSVM performed best using trimmed data or high-resolution smoothed spectra (2-3nm), while RF yielded top performance when trained on spectral ratios derived from smoothed and normalized data. The trade-off between classification accuracy and computational complexity was again evident, positioning LDA and LSVM as the most practical and computationally efficient models for continued development.

Beyond the classifier trends, we also analyzed performance across different spectroscopic platforms. Table 4.12 presents the best classification accuracy for each model and modality. In line with the corn results, fluorescence after 385 and 395nm excitation proved to be the most informative excitation wavelengths. Excitations at 375 and 365nm performed moderately well, whereas 256nm showed poor classification results, consistent with prior findings across all mycotoxins studied on corn.

Somewhat¹³ unexpectedly, we observed that reflectance-based modalities—UV-VIS and NIR, and particularly the full-range UV-VIS-NIR, (constructed by concatenating the UV-VIS and NIR spectra as illustrated in Figure 3.10)—performed exceptionally well. In fact, classification accuracy using reflectance data surpassed that of any fluorescence modality, a stark contrast to the results found for corn, where reflectance yielded relatively poor model performance. This might be caused by the inherent differences in the matrix composition of corn and wheat and

¹³A deliberate spelling error left in, to recapture the reader’s attention in the midst of this already extensive report.

the relative effect of the toxins on its composition. Another explanation could relate to the sample composition: since all wheat samples were synthesized by mixing only two standards, it is possible that the reflectance-based models primarily learn matrix-specific differences rather than toxin-specific spectral features as both could not be decoupled during sample preparation, as discussed in Section 3.1.2.

Given the limitations in the available sample data, no definitive conclusion can be drawn on this matter. Consequently, future additional validation measurements will be necessary to define if the observed classification performance can be attributed to the mycotoxins alone, or if the matrix differences are impacting the measurements. For now, it is important to keep in mind that the classification models might be influenced by both the mycotoxin presence as the concurrent matrix differences.

Modality	LDA [%]	LSVM [%]	RF [%]
Fluo 395	73	77	76
Fluo 385	77	81	80
Fluo 375	61	64	70
Fluo 365	66	68	74
UV-VIS	68	79	65
NIR	93	87	85
UV-VIS-NIR	96	84	85

Table 4.12: Classification accuracy for the best-performing models in the three-class classification of wheat, across various spectroscopic platforms.

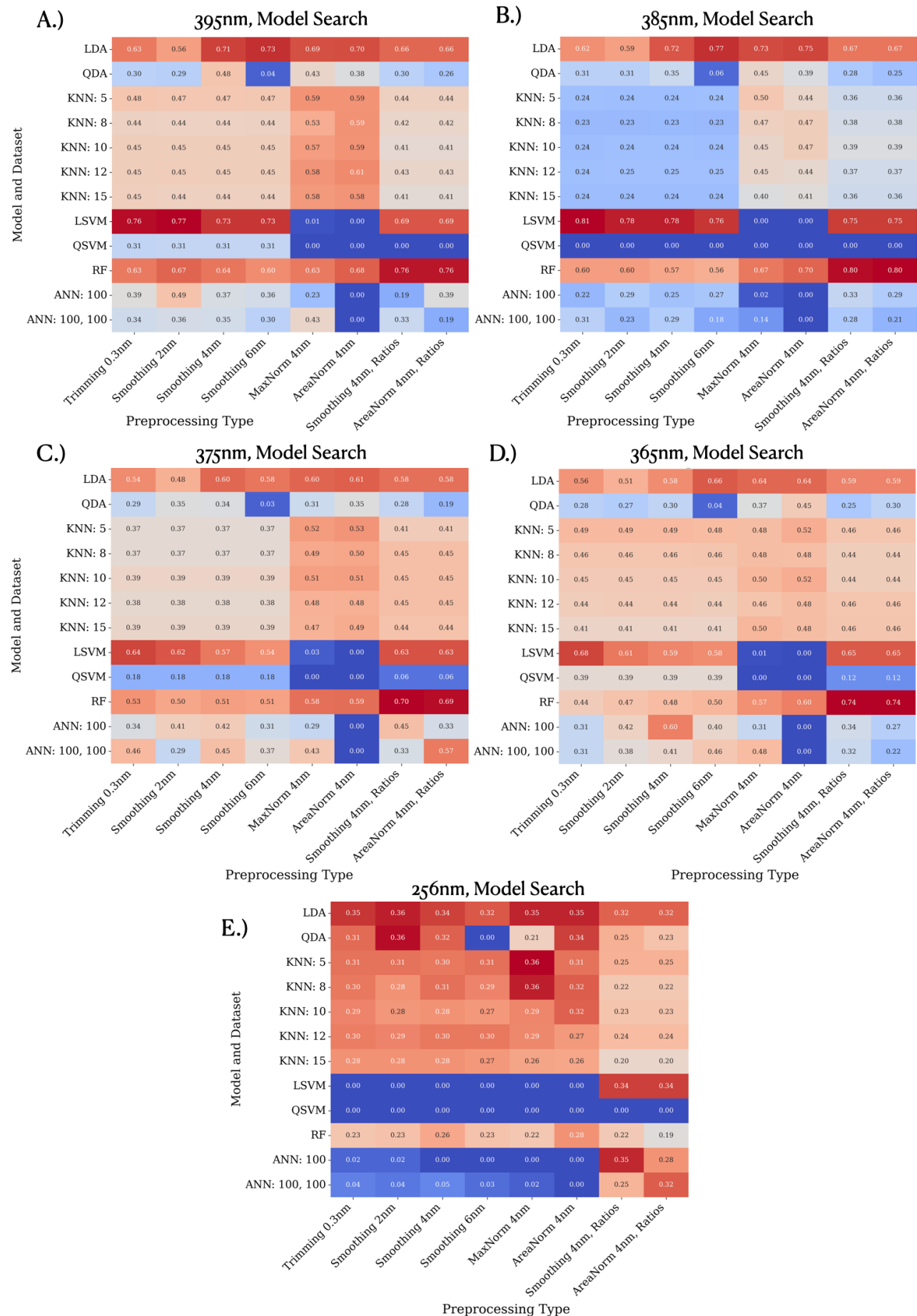


Figure 4.30: Classification accuracy heatmaps for each fluorescence modality, illustrating the results of the model/preprocessing search for the simultaneous detection of DON, T-2/HT-2 and OTA on wheat, using a three-class classification scheme. Thresholds of 330 and 630 $\mu\text{g/kg}$ for DON were used to define safe, risky, and unsafe classes. Similar trends are observed to the classification of corn and the best-performing models are LDA, LSVM, and RF.

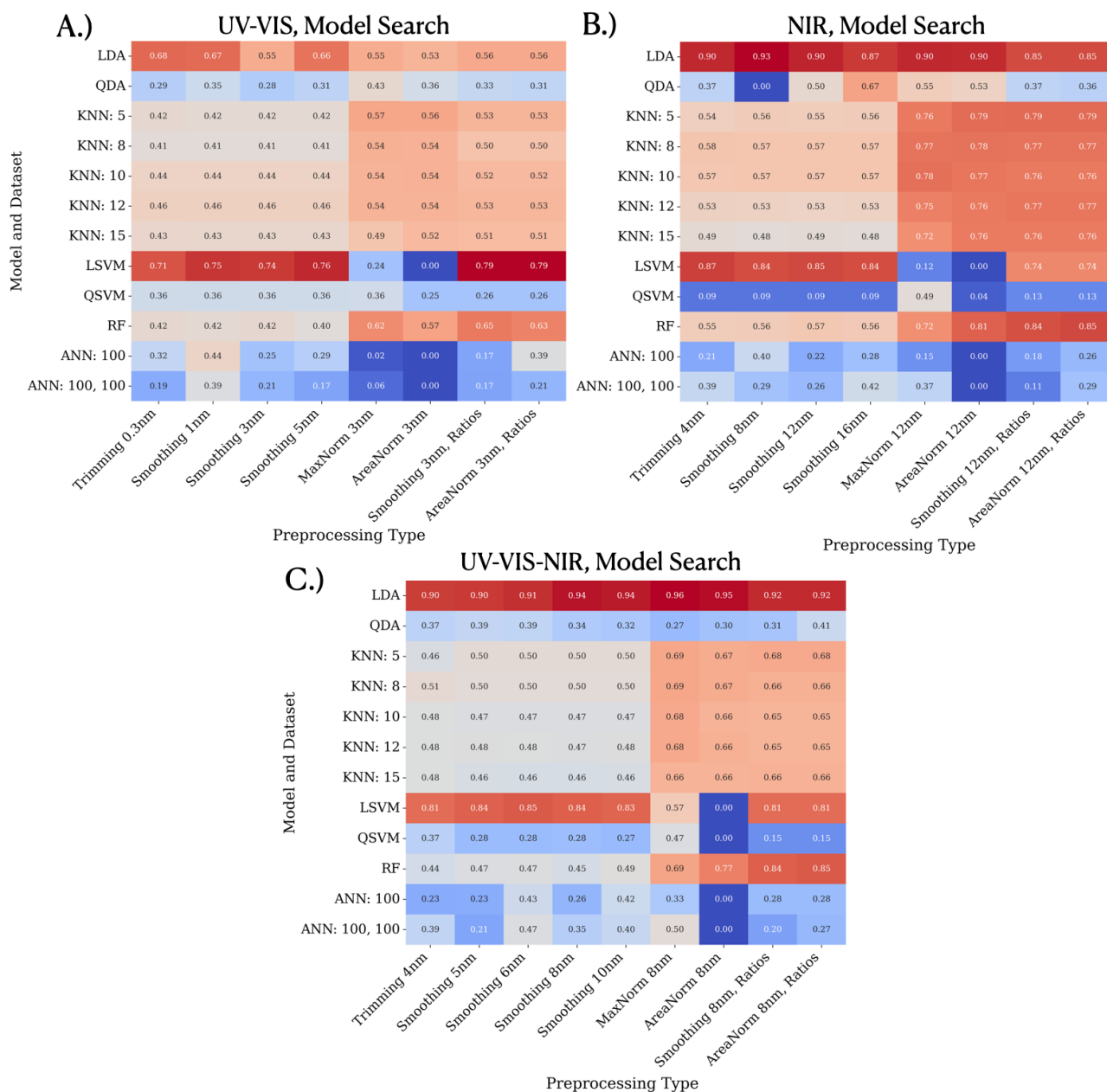


Figure 4.31: Classification accuracy heatmaps for each reflectance modality, illustrating the results of the model/preprocessing search for the simultaneous detection of DON, T-2/HT-2 and OTA on wheat, using a three-class classification scheme. Thresholds of 330 and 630 $\mu\text{g/kg}$ for DON were used to define safe, risky, and unsafe classes. The best-performing models are LDA, LSVM, and RF. Classification accuracy using reflectance data is higher than using fluorescence data (Figure 4.30), a stark contrast to the results found for corn.

4.4.2 Model Optimization

The initial model-search already demonstrated excellent performance in classifying wheat samples—particularly when using reflectance data, as shown in Table 4.12. Building on these results, we further refine the LDA, LSVM, and RF classifiers through hyperparameter tuning and the optimization of preprocessing strategies. For a detailed overview of the hyperparameters and methodological procedures employed, we refer the reader to the corresponding section on corn classification (Section 4.3.4). For illustrative purposes, the optimized classification results for the 395nm fluorescence data are presented in Figure 4.32.

A key observation here is that the LDA model exhibited a substantial improvement following the optimization (using high smoothing (6-10nm) and SVD), with classification accuracy rising from 73% to 80%—a notable gain of 7%. In contrast, LSVM and RF showed only marginal sensitivity to hyperparameter tuning, with improvements typically limited to the 1–2% range. This trend—where LDA benefits most from optimization while LSVM and RF are minimally affected by hyperparameter changes—was consistently observed across all other fluorescence datasets as well as the reflectance data. While the full set of optimization plots is omitted here for brevity, the best-performing results for each modality are summarized in Table 4.13, which presents classification performance before and after optimization across all measurement platforms.

Modality	LDA [%]	LDA* [%]	LSVM [%]	LSVM* [%]	RF [%]	RF* [%]
Fluo 395	73	80	77	77	76	78
Fluo 385	77	82	81	83	80	81
Fluo 375	61	71	64	67	70	71
Fluo 365	66	75	68	73	74	77
UV-VIS	68	76	79	78	65	66
NIR	93	94	87	91	85	87
UV-VIS–NIR	96	99	84	86	85	88

Table 4.13: Comparison of classification accuracy before and after (*) model optimization for three-class classification of wheat samples across various spectroscopic platforms. Results are shown for LDA, LSVM, and RF models. LDA is found to show significant improvements after optimization, LSVM and RF are not much affected.

From this overview, we observe that optimization yields clear performance gains, with LDA showing the most significant improvement, reaching up to 99% accuracy when using the combined UV–VIS–NIR reflectance data. Considering LDA’s consistently strong performance, as well as its significantly lower computational cost compared to LSVM and RF (especially when paired with computationally expensive feature selection methods like SFS), we again select LDA as the preferred model for subsequent feature reduction and final classifier development on wheat.

For LDA we found that using either smoothed or area-normalized data at low resolutions 6–14nm yielded the best performance, though there is no clear trend to the optimal resolution as it differed between datasets (see Figures 4.30, and 4.31).

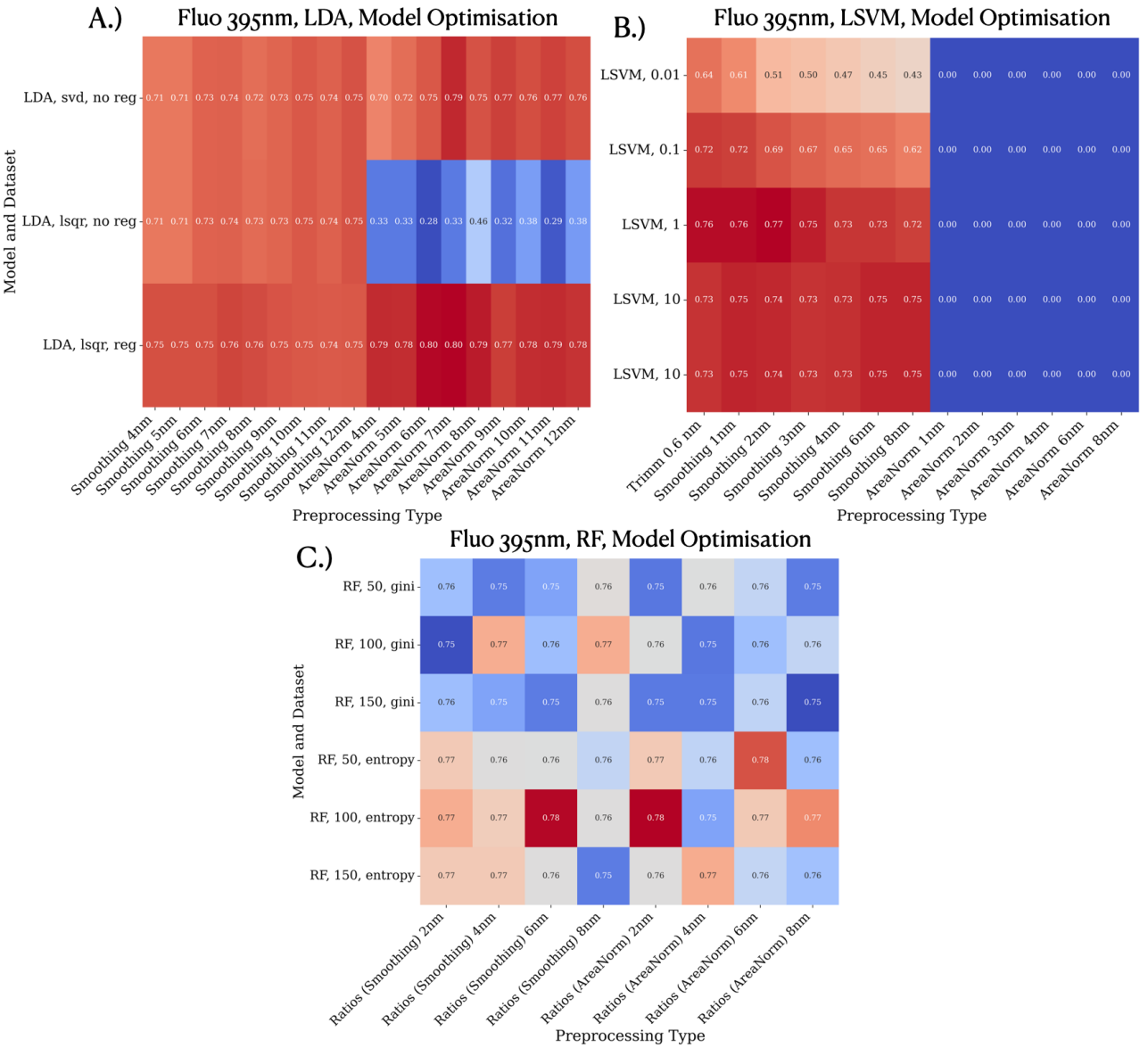


Figure 4.32: Model optimization results for LDA, LSVM, and RF applied to the 395nm fluorescence data for wheat classification. A range of hyperparameters and preprocessing strategies were evaluated to improve model performance. Notably, LDA showed a substantial improvement—up to 10%—compared to its initial performance reported in Table 4.12. LSVM and RF exhibited minimal gains, with limited sensitivity to hyperparameter or preprocessing adjustments. Similar results were found for the other spectroscopic platforms.

4.4.3 Feature Selection

Now that the optimal models have been identified, we proceed to investigate the trade-off between model complexity—achieved through feature reduction—and classification performance. Figure 4.33 presents the accuracy heatmaps for both MRMR and SFS feature selection methods applied to LDA models across the various datasets, using the best-performing preprocessing configurations determined during the model optimization step.

As observed previously in the corn dataset (see Section 4.3.5), SFS consistently outperforms MRMR, particularly when selecting fewer than 20 features. This further confirms SFS as the superior method for selecting the most informative features, albeit at the cost of increased computational demand.

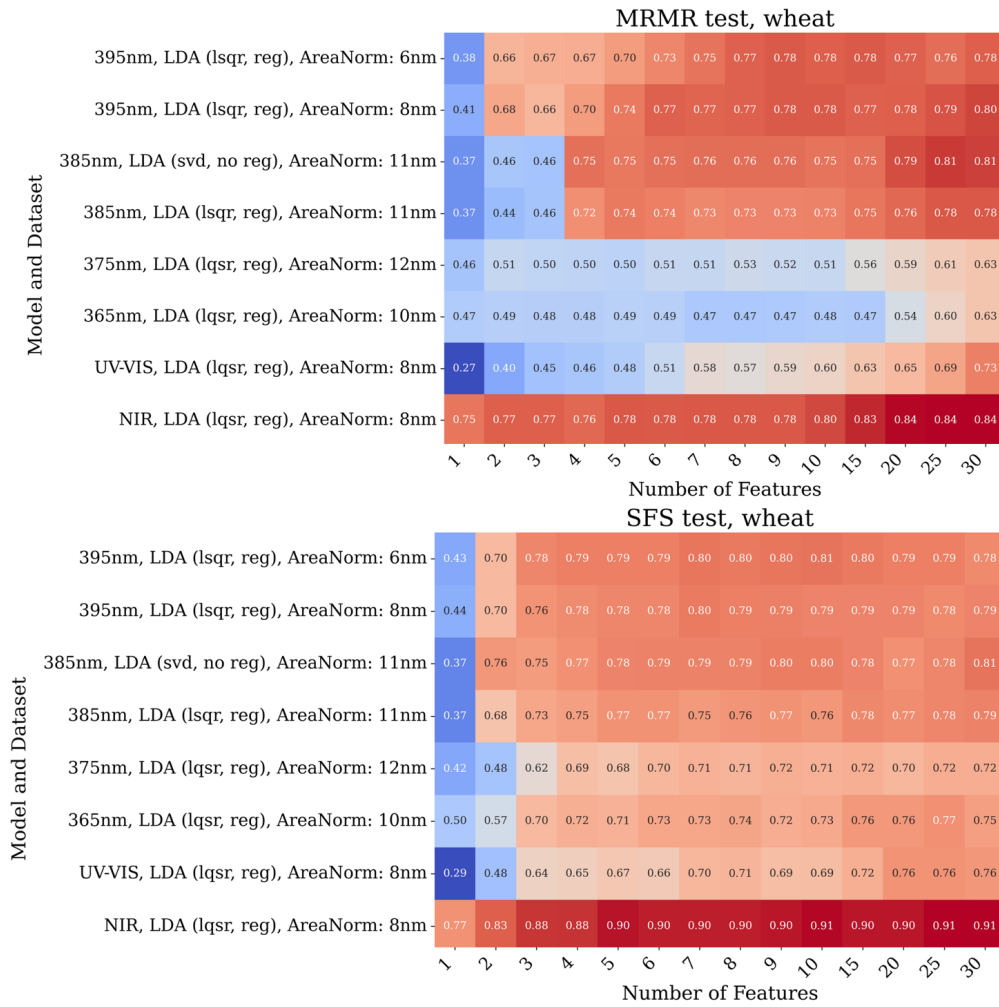


Figure 4.33: Accuracy heatmaps comparing MRMR (top) and SFS (bottom) feature selection methods applied to LDA models across various spectroscopic platforms for wheat classification. Each heatmap shows classification accuracy as function of the number of selected features. SFS consistently outperforms MRMR, particularly when selecting fewer than 20 features. The NIR reflectance data yields the best overall performance, followed by fluorescence data at 385 and 395nm. These results confirm that SFS is more effective for low-dimensional feature selection despite its higher computational cost.

Across all datasets, we found that LDA models trained on 8nm area-normalized data achieved the best results, making it the preprocessing method of choice moving forward. Based on this, we compared the classification performance of full-spectrum LDA models to those trained on only the top 10 features selected using SFS. These results are summarized in Table 4.14.

Modality	All features [%]	10 features [%]
Fluo 395	79.3	78.6
Fluo 385	78.4	78.0
Fluo 375	66.3	68.5
Fluo 365	73.7	75.6
UV-VIS	75.6	69.3
NIR	92.2	90.7
UV-VIS-NIR	90.7	88.2

Table 4.14: Comparison of classification accuracy for wheat samples using models trained on the full spectrum (same info as shown in Table 4.13) versus the 10 most informative features selected via SFS. The NIR dataset yielded the highest overall performance. A slight decrease in accuracy (approximately 2%) was observed across all modalities when reducing the input from the full spectrum to 10 selected features.

From Table 4.14 we see that models using only 10 selected features achieve performance levels within 2–3% of their full-spectrum counterparts, demonstrating a significant reduction in model complexity with minimal performance loss. The highest accuracy was achieved using the NIR reflectance data, reaching 92.2%. The confusion matrix corresponding to this best performing NIR-model is shown in Figure 4.34.

Notably, there are no second-order misclassifications, and only a very limited number of samples were incorrectly classified at all. Among those misclassified were samples S0L and S0M—both boundary samples around the safe-risky threshold (see Table 3.2 for the concentration values of the samples). Given the nature of LOSO-CV, the model cannot learn the exact decision boundary but must infer it from the remaining data. Thus, these “errors” are expected edge cases rather than genuine model failures. Overall, the model shows exceptionally strong performance, especially when excluding borderline cases.

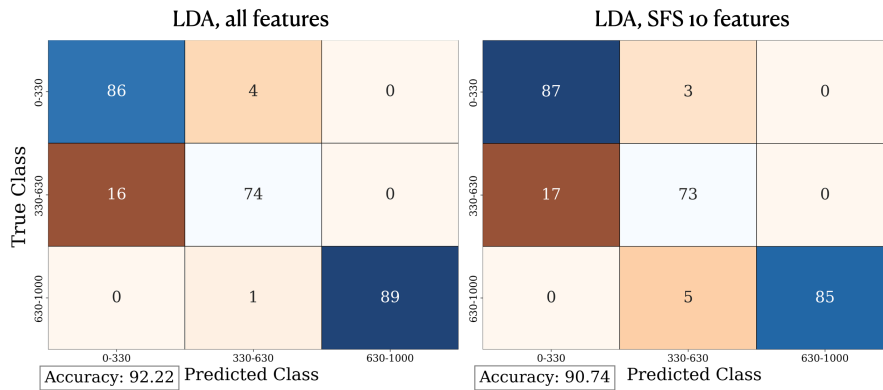


Figure 4.34: Confusion matrix of the best-performing LDA model for wheat classification using all (left) and only the top 10 features (right) from the NIR reflectance data. The model was trained on area-normalized spectra, smoothed at 8nm. Performance of the mode using 10 features was 2% below that of the model using full spectral information.

4.4.4 Combining Complementary Data

As previously implemented for the corn dataset in Section 4.3.6, we also investigated whether combining spectral information from multiple modalities could enhance classification performance for wheat. Following a similar approach to that illustrated in Figure 4.25 A, LDA models were trained using a merged set of features selected via SFS across all available spectral datasets.

This process consisted of two stages: first, selecting the top 10 most informative features from each individual dataset; second, either training an LDA model on the combined feature set or applying SFS once more to reduce the merged set to a final selection of 10 features.

For wheat, we explored three data fusion strategies: (1) combining all fluorescence datasets, (2) combining all reflectance datasets, and (3) combining both fluorescence and reflectance datasets. The results of these experiments are summarized in Table 4.15, which compares the classification accuracy for each individual dataset (using both the full spectral range and the top 10 selected features) alongside the fused models.

The features selected from the combined datasets for each of the three scenarios are presented in Table 4.16. Interestingly, integrating data from multiple modalities did not result in substantial performance gains; improvements were generally marginal, typically on the order of just 1–2%. For the fluorescence data, features were drawn across the 385–365nm datasets, with the largest proportion coming from the 365nm data, followed by 385 and 375nm. Notably, almost all of the selected wavelengths were above 600nm, mostly clustered between 650 and 700nm—corresponding to the chlorophyll peak region.

However, this chlorophyll fluorescence emission is most pronounced in the 395nm excitation data, with decreasing prominence in the 385 and 375nm datasets, as illustrated in Figure 3.17, which shows spectra of the most, middle, and least contaminated samples. The fact that many features were selected from the 365nm data—where the chlorophyll peak is only faintly discernible—is unexpected, especially if this peak is preferred for classification.

The classifier does, whoever, appear to rely heavily on this spectral region, suggesting that it may be capitalizing on chlorophyll content rather than direct mycotoxin-specific markers. Otherwise, mycotoxin contamination might indirectly affect chlorophyll levels, thereby serving as an indirect marker. However, given that the current wheat samples are derived from just two standards, no conclusive interpretation can be made. As such, additional validation measurements are necessary to enable generalization, and to confirm the observed spectral differences genuinely reflect mycotoxin contamination.

In the other two classification cases, the majority of selected features originated from the NIR data, aligning with the previous observations that indicated NIR spectroscopy as the most performant technology. Most features were selected in the 1000–1200nm range, with additional selections appearing in the 1600–1700nm region.

Interestingly, a number of the selected spectral features originated from the 1400–1600nm wavelength range, which is also typically associated with moisture absorption. This could indicate that the moisture content influences the classification, especially since all wheat samples are prepared from just two standards. However, this considers a limited subpart of the selected wavelengths, while additionally a similar moisture content for the different samples is expected. The performance of the cross-platform merged-data models did not surpass that of the models using only NIR data. This indicates that integrating multiple modalities does not offer a meaningful advantage in this case. Given the added complexity and computational cost of merged models, focusing solely on NIR data appears to be the most efficient and effective approach.

Performance	All Features [%]	10 Features [%]
Fluorescence Data		
Fluo 395	79.3	78.6
Fluo 385	78.4	78.0
Fluo 375	66.3	68.5
Fluo 365	73.7	75.6
Combined	81.4	79.6
Reflectance Data		
UV-VIS	75.6	69.3
NIR	92.2	90.7
Combined	91.1	89.3
All Data		
All Combined	89.3	88.9

Table 4.15: Classification performance for the wheat dataset using LDA with all spectral features vs. only the top 10 features selected using SFS. Results are shown for each fluorescence and reflectance modality, as well as for various combinations of spectral data. Only a minor or no improvement is realized when combining spectral datasets.

Rank	Fluorescence only		Reflectance only		All data combined	
	Dataset	λ [nm]	Dataset	λ [nm]	Dataset	λ [nm]
1	Fluo 385	722	NIR	1632	NIR	1632
2	Fluo 365	666	NIR	1688	NIR	1688
3	Fluo 385	794	NIR	1160	NIR	1160
4	Fluo 385	690	NIR	1056	NIR	1056
5	Fluo 385	714	UV-VIS	984	UV-VIS	984
6	Fluo 365	682	NIR	1080	Fluo 395	504
7	Fluo 375	698	NIR	1000	NIR	1080
8	Fluo 375	722	NIR	1384	NIR	1384
9	Fluo 365	722	NIR	1072	NIR	1072
10	Fluo 365	594	NIR	1664	NIR	1664

Table 4.16: Top 10 spectral features selected by SFS for the classification of wheat, using only fluorescence data, only reflectance data (UV-VIS and NIR), and all data combined. Each feature is identified by the dataset (modality or excitation wavelength) and its corresponding central wavelength. The rank indicates the order in which the features were selected one by one.

4.4.5 Experiment: Restricting The Spectral Range

As a final experiment on the wheat dataset, we investigated the impact of restricting the spectral ranges used for classification—motivated by practical considerations relevant to real-world deployment of these techniques.

In fluorescence data, for example, wavelengths beyond 600nm are strongly affected by chloro-

phyl fluorescence, which can introduce significant interference—particularly in crops with high chlorophyll content, such as pistachios [30]. This interference stems from the fact that chlorophyll levels can vary widely depending on crop growth, and storage conditions. Such variability introduces additional spectral noise and complicates the detection of mycotoxin-specific patterns. To improve robustness, a reliable mycotoxin detection system should ideally avoid this spectral region.

A similar practical consideration applies to reflectance data. In industrial settings, such as in-line food sorting systems, lights are typically present, emitting in the visible part of the spectrum. As a result, relying on reflectance in the visible range could make the detection process more sensitive to ambient light interference. To mitigate this, we restricted the UV-VIS data to the 750–1000nm range (effectively the low NIR region). Additionally, for the NIR reflectance dataset, we excluded wavelengths above 1400nm, as this region is heavily influenced by water absorption, another factor that varies greatly between crops depending on moisture level. Therefore, only the 1000–1400nm range was retained for NIR.

With these wavelength range in place, we retrained the models and compared their performance to the full-spectrum counterparts. The results are presented in Table 4.17. For fluorescence, we observed a substantial performance drop—approximately 20%—when wavelengths beyond 600nm were excluded. For reflectance data (using the combined UV-VIS-NIR dataset), excluding either the visible or the +1400nm range individually led to a minor drop of around 5%. However, excluding both ranges simultaneously caused a much larger decrease in performance, of approximately 14%.

These findings suggest that, for the wheat samples, certain spectral regions—particularly those associated with chlorophyll and water absorption—are important for the classification task. This may be an indication that part of the model’s discriminative power stems from detecting matrix-specific effects, or the influence of the mycotoxins on these matrix elements, rather than solely the mycotoxin content. Nevertheless, restricting the spectral range also inherently reduces the number of available features, potentially discarding informative wavelengths linked to the mycotoxins.

Modality	Full Range [%]	Limited Range [%]
Fluo 395	79.33	51.32
Fluo 385	78.38	39.58
Fluo 375	66.30	48.47
Fluo 365	73.69	52.54
UV-VIS-NIR (full)	90.74	–
(no VIS)	–	87.04
(no 1400+)	–	85.56
(no VIS, 1400+)	–	77.04

Table 4.17: Comparison of classification accuracy for wheat using LDA models across full vs. restricted wavelength ranges. For fluorescence, performance drops substantially when excluding wavelengths above 600nm. For reflectance, modest declines are observed when omitting the visible or high NIR (> 1400nm) regions individually, with a much larger drop when both are excluded.

4.4.6 Conclusion

In this section, we demonstrated the feasibility of classifying wheat samples contaminated with a multi-mycotoxin mix (DON, T-2/HT-2, OTA), using both fluorescence and reflectance spectroscopy. For this we employed the previously developed model development pipeline—comprising model-search, optimization, feature selection, and data fusion—now applied to wheat data as oppose to corn data.

Among the evaluated classifiers, LDA again emerged as the most practical choice, achieving top accuracy, while maintaining low computational overhead. The highest classification performance was obtained using NIR reflectance data, with models trained on just 10 selected wavelengths reaching up to 92.2% accuracy. Performance could be further improved to over 95% when excluding edge samples near threshold boundaries, suggesting that the models are both robust and highly informative.

It is also important to note that a direct comparison to existing literature is not straightforward, as this study focuses on the simultaneous detection of a specific mixture of mycotoxins. More specifically, the spectral patterns observed here reflect the combined presence of DON, T-2/HT-2, and OTA, rather than isolated signatures of individual toxins. This overlap makes it difficult to disentangle the specific contributions of each mycotoxin, and therefore, the results presented cannot be directly compared to studies that target the detection of single, isolated compounds. Despite this, the overall classification performance achieved is remarkably high, compared to the state-of-the-art listed in Table 2.09, especially when considering the use of the reduced feature sets.

Although we explored data fusion across multiple spectral modalities, no significant improvement was observed with respect to the single use of the NIR reflectance data. We also investigated the impact of restricting the spectral range to support practical deployment. This analysis highlighted the importance of specific wavelength regions—particularly those influenced by chlorophyll (in fluorescence) and water absorption (in reflectance).

Throughout all of these analyses, the considered wheat samples were prepared by mixing two standards—one containing a mix of all mycotoxins and one blank. Consequently, the mycotoxins co-occurred in all samples, imposing that the individual spectral contributions of a specific toxin cannot be disentangled, while being supplemented with matrix-specific characteristics. Therefore, it remains uncertain whether the classification models are detecting signals directly associated with the toxins, or if they are instead relying on indirect correlations with matrix composition that coincides with contamination by the nature of the sample composition. This ambiguity underscores the need for future studies employing more independent sample-series, composed without mixing reference materials, to better isolate spectral features related to solely the mycotoxins content from confounding matrix markers.

4.5 Classification Of Mycotoxins On Corn And Wheat

As previously mentioned, this study extends beyond the development of models for mycotoxin detection on individual matrices such as corn and wheat. We also investigated a cross-matrix classification approach aimed at detecting DON (deoxynivalenol) across both corn and wheat samples using a single unified classifier.

The motivation behind this approach is to explore the potential of cross-platform models which, if successful, could generalize beyond individual matrix scenarios and eliminate the need for retraining or redesigning models for each specific crop type. Such versatility would greatly enhance the scalability and practicality of mycotoxin detection systems in real-world applications. Depending on the model, this approach could also increase confidence that the classifier is detecting the presence and effects of DON itself, rather than relying on matrix-specific markers for classification.

Given the increased complexity of combining two distinct crop matrices—each with different spectral characteristics and DON concentration ranges (0–2200 $\mu\text{g}/\text{kg}$ for corn and 0–1000 $\mu\text{g}/\text{kg}$ for wheat)—a three-class classification scheme was deemed impractical due to imbalanced class sizes and limited sample numbers. Instead, we implemented a binary classification model, defining the threshold for DON contamination at 500 $\mu\text{g}/\text{kg}$, as outlined earlier in Table 4.2, to ensure a roughly equal distribution of samples across the safe and unsafe classes for both matrix-type samples.

In the sections that follow, we once again apply the full model development pipeline: beginning with model-search (Section 4.5.1) and optimization (Section 4.5.2), followed by feature selection (Section 4.5.3) and final model evaluation (Section 4.5.4)—this time applied to the combined corn and wheat dataset for DON detection.

4.5.1 Model-Search

As was the case in previous detection model development, the first step involved an extensive search across model and preprocessing combinations. The results of this model-search are presented in Figures 4.35 and 4.36, corresponding to fluorescence and reflectance data, respectively. The 256nm excitation was excluded from the fluorescence data, given its consistently poor performance on both the corn and wheat datasets.

Among the models evaluated, LDA, LSVM and RF once again emerged as the top performers, consistent with previous findings. Interestingly, K-Nearest Neighbors (KNN) also showed notably strong performance in this combined two-class classification setting, a trend not observed in earlier experiments. This improved performance may be attributed to two main factors: first, the simplified nature of the binary classification task, and second—and more importantly—KNN’s likely ability to effectively distinguish between the two sample types (corn and wheat) and then to classify contamination levels within each type separately. This is in contrast to other methods, that might tend to generalize the DON detection across matrices.

In terms of preprocessing preferences, previously observed trends were reaffirmed. LDA performed best with smoothed data at lower spectral resolution (8–10nm), LSVM favored high-resolution (2–4nm) smoothed or normalized inputs, and RF continued to benefit from the use of spectral ratios. Newly observed in this setting, KNN achieved its best performance with area-normalized or max-normalized data.

To better compare model performance across modalities and classifiers, the best classification results from each configuration are summarized in Table 4.18. From this overview, it becomes clear that all models and modalities achieve similar performance, typically around 80% accuracy. As expected, the 395 and 385nm fluorescence excitations deliver the highest classification accuracies, consistent with results found for corn and wheat individually. The combined reflectance data (UV-VIS + NIR) also performs well, matching the fluorescence-based models in accuracy.

Dataset	LDA [%]	KNN [%]	LSVM [%]	RF [%]
Fluo 395	81	82	83	83
Fluo 385	81	80	85	83
Fluo 375	78	80	82	83
Fluo 365	80	79	81	84
UV-VIS	67	83	70	82
NIR	81	73	81	79
UV-VIS-NIR	83	83	83	84

Table 4.18: Classification accuracy for the combined corn and wheat DON detection task using two-class classification. Results are shown for four best performing classifiers (LDA, KNN, LSVM, and RF) across the various spectroscopic datasets.

Given the high computational cost associated with applying SFS to LSVM and RF—as established in earlier sections—we exclude these models from further analysis. Instead, we proceed with LDA (due to its consistent accuracy and computational efficiency) and now also include KNN, which has shown promising results in the combined classification task. In the next section, we further optimize these models by performing hyperparameter tuning and exploring refined preprocessing strategies to maximize classification performance.

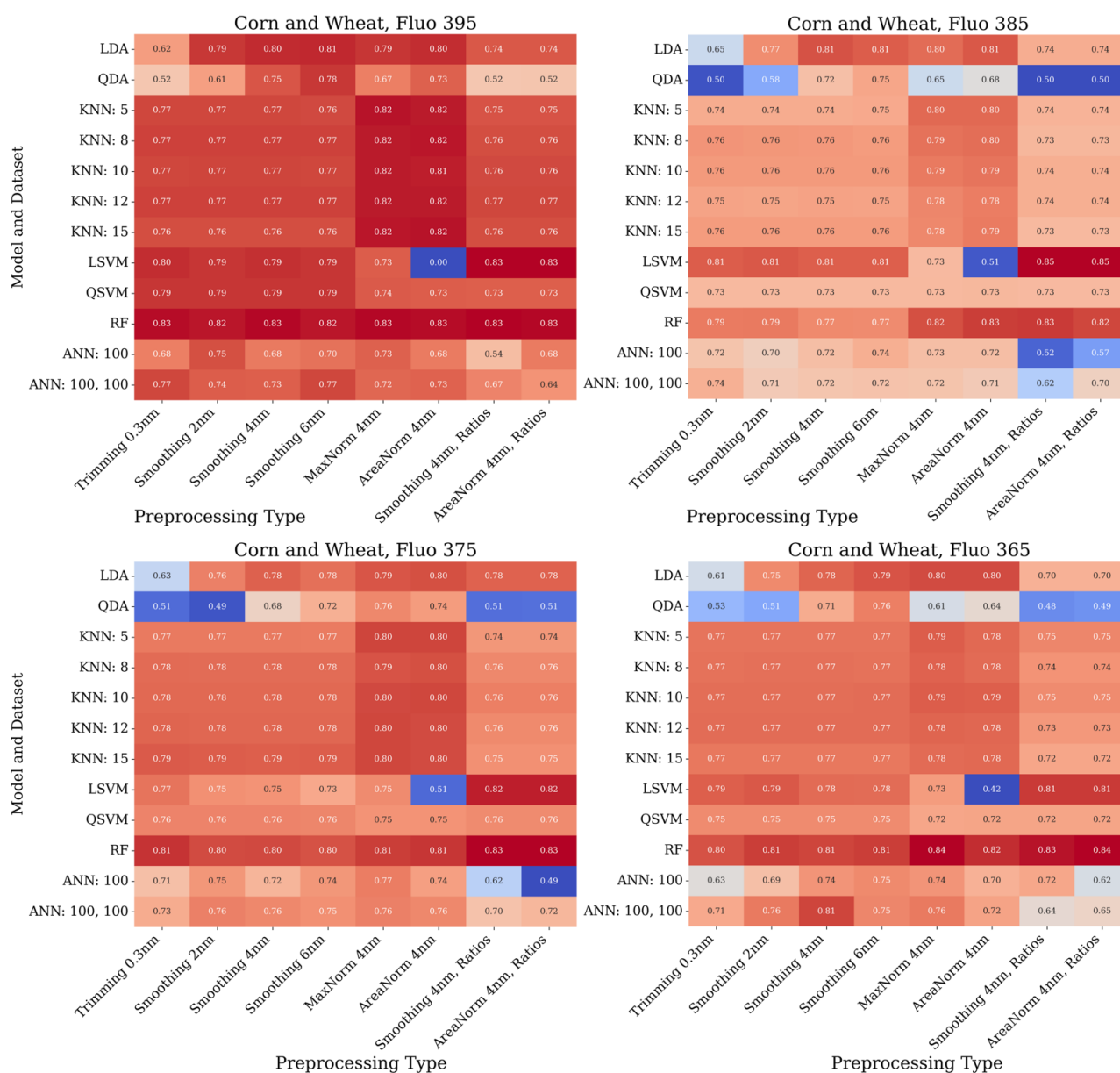


Figure 4.35: Classification accuracy heatmaps for each fluorescence modality, illustrating the results of the model/preprocessing search for the detection of DON on wheat and corn combined, using a two-class classification scheme with a threshold of $500\mu\text{g/kg}$. Similar trends are observed to the classification of corn and wheat individually, now also finding KNN as a good performing model.

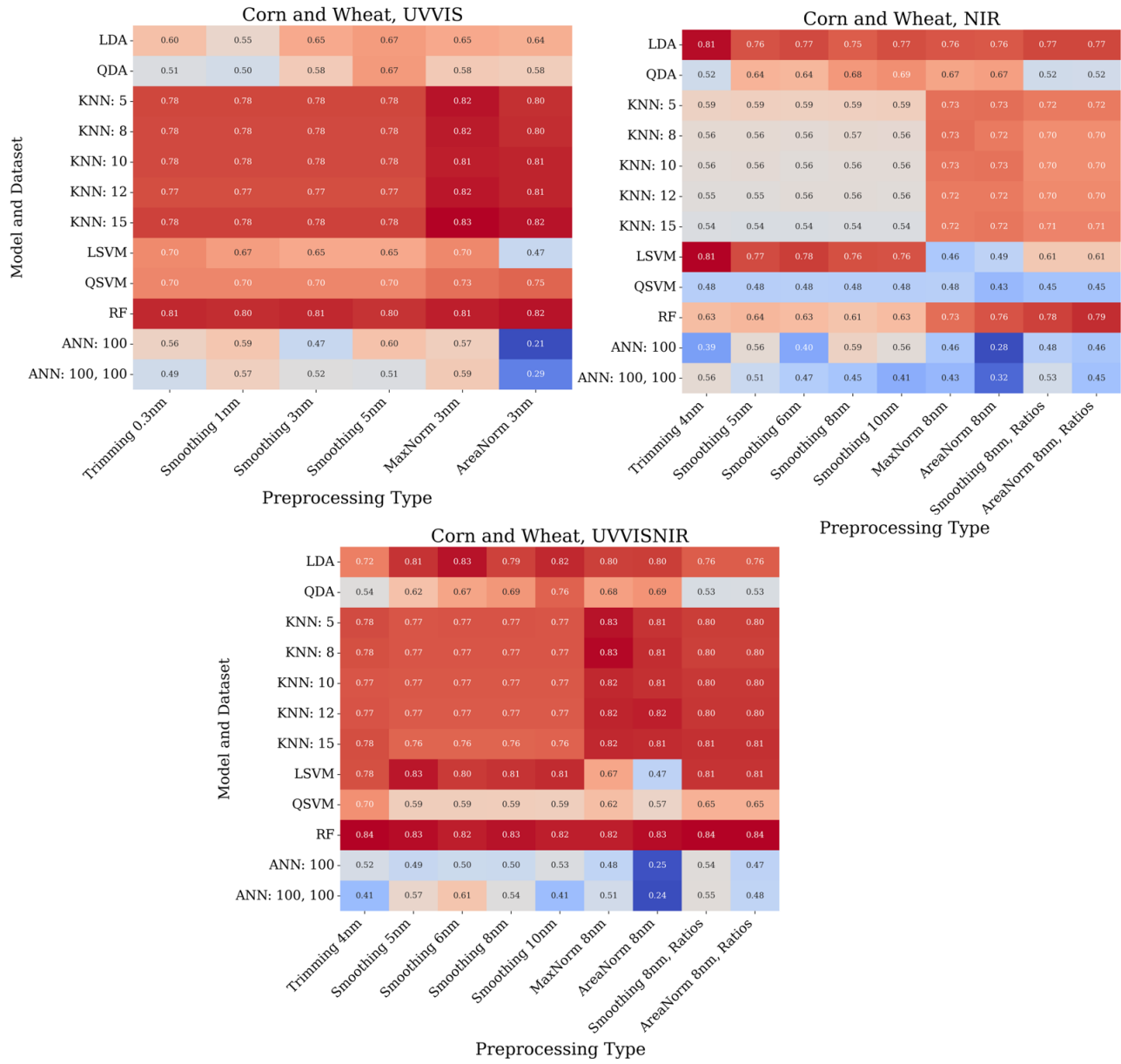


Figure 4.36: Classification accuracy heatmaps for each reflectance modality, illustrating the results of the model/preprocessing search for the detection of DON on wheat and corn combined, using a two-class classification scheme with a threshold of $500\mu\text{g}/\text{kg}$. Similar trends are observed to the classification on corn and wheat individually, now also finding KNN as an optimal model.

4.5.2 Model Optimization

Next, we further optimized the LDA and KNN models for the detection of DON across both matrices combined. For LDA, we again assessed the impact of regularization and compared the performance of the SVD and LSQR solvers. Additionally, we refined the preprocessing pipeline, with a particular focus on smoothing and area-normalization—both of which had previously demonstrated strong results. For KNN, a comprehensive hyperparameter search was conducted by varying the number of neighbors from 3 to 40 to identify the optimal configuration. This optimization process was performed individually for each dataset included in the study.

An example of the resulting accuracy heatmaps is shown in Figure 4.37, highlighting the model optimization for both the LDA and KNN models, on the 395nm fluorescence dataset. For KNN, we generally observed that using a small number of neighbors—typically around 3—yielded the best performance, although this was not consistent across all datasets. For preprocessing, KNN consistently performed best when using area-normalized input.

In the case of LDA, the LSQR solver tended to deliver more stable and consistent results, albeit at a higher computational cost. In contrast, the SVD solver achieved comparable performance while requiring significantly less computation, but proved more sensitive to the choice of preprocessing, particularly the smoothing level. LDA showed no strong preference between smoothing and area-normalization, except when using LSQR and no regularization, then smoothing is favored.

Overall, the performance improvements achieved through this optimization stage were relatively modest, with only minor gains compared to the original model-search results.

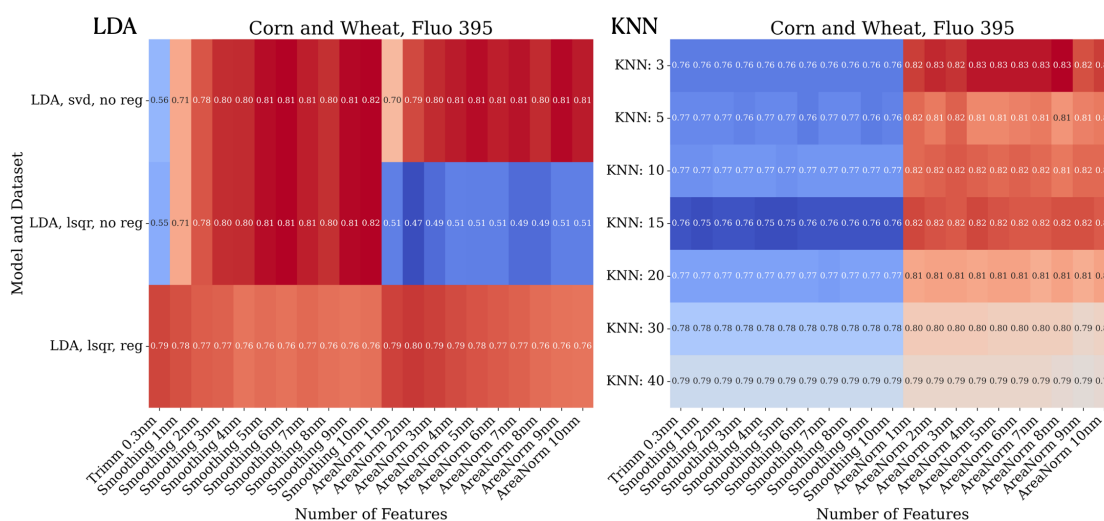


Figure 4.37: Model optimization results for LDA and KNN applied to the 395nm fluorescence data for the combined corn and wheat DON classification task. A range of hyperparameters and preprocessing strategies were explored to enhance model performance.

The updated classification results are summarized in Table 4.19. When compared to the initial performance values reported in Table 4.18, we find that model accuracy increased by only 1–2%. Performance between LDA and KNN was nearly equivalent across most datasets, with the exception of the UV–VIS dataset, where KNN outperformed LDA by a substantial margin of 15%,

and a large contrast for NIR, where LDA outperforms KNN with 7%.

Dataset	LDA [%]	LDA * [%]	KNN [%]	KNN * [%]
Fluo 395	81	82	82	83
Fluo 385	82	81	80	80
Fluo 375	78	79	80	81
Fluo 365	80	80	79	79
UV-VIS	67	68	83	83
NIR	81	81	73	74

Table 4.19: Classification accuracy for combined corn and wheat DON detection before and after (*) model optimization. Performance improvements were modest, with gains of only 1–2% compared to the initial results.

4.5.3 Feature Selection

As a next step, we now proceed to implement the feature selection. The results using both MRMR and SFS are presented in Figure 4.38. As expected, MRMR proves to be computationally efficient but underperforms at low feature counts, while SFS consistently yields superior classification accuracy—especially when only a limited number of features are used.

Interestingly, we find that NIR data combined with LDA struggles to maintain high accuracy when using fewer features, requiring over 30 features to reach optimal performance. In contrast, the KNN model (with $k = 3$) based on the 395nm excitation fluorescence dataset performs exceptionally well with minimal input, achieving 82% accuracy using just 3 wavelengths. The KNN model based on the 385nm fluorescence dataset follows closely behind, while UV-VIS and NIR modalities again lag behind in the low-feature regime.

These findings suggest that fluorescence data, particularly at 395nm, is far more suitable for compact, efficient models in this combined classification task, whereas reflectance-based modalities require significantly more spectral information to perform well.

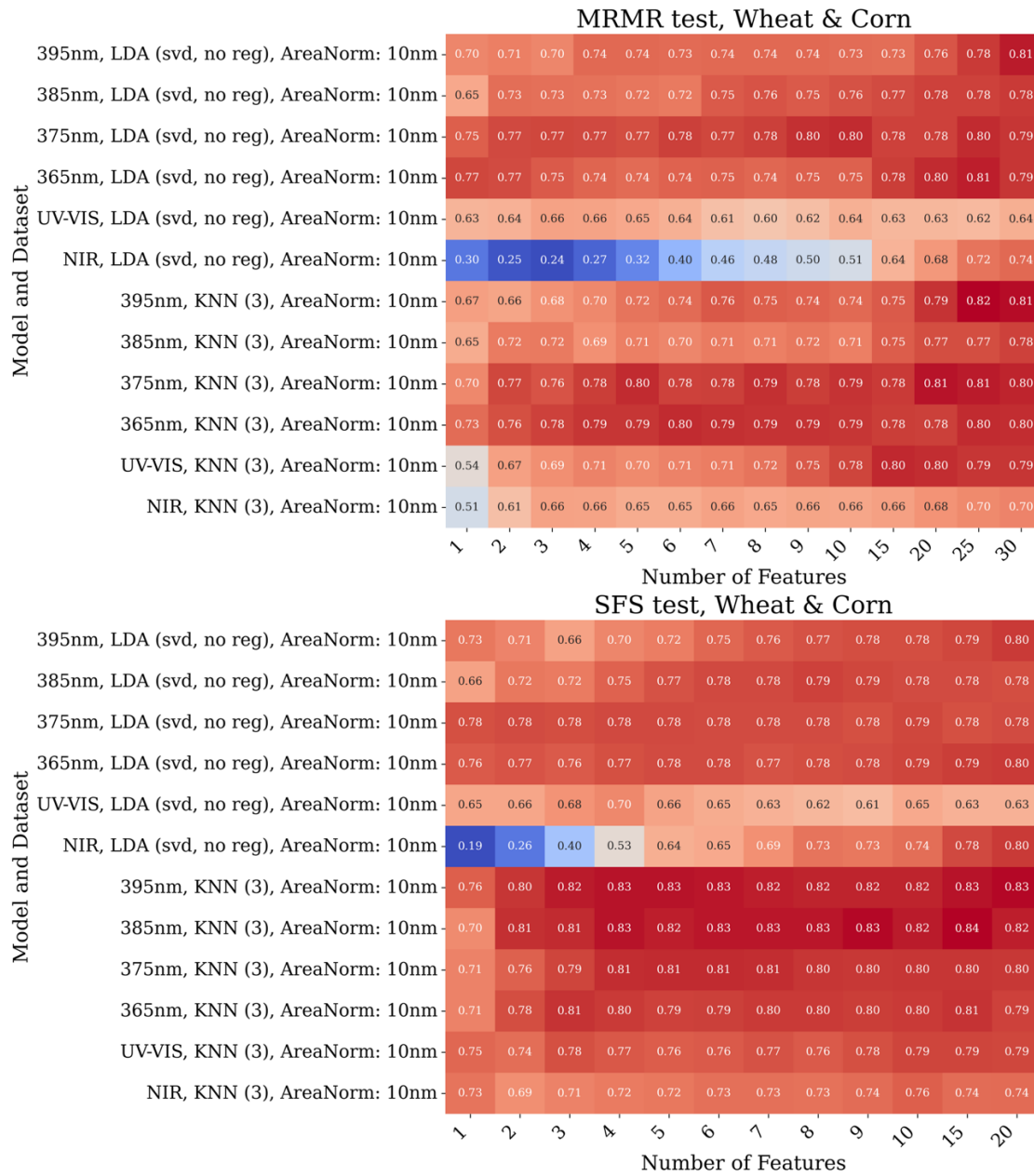


Figure 4.38: Accuracy heatmaps comparing MRMR and SFS feature selection methods applied to LDA and KNN models for the combined detection of DON on wheat and corn across different spectroscopic modalities. Each heatmap illustrates classification accuracy as a function of the number of selected features. As expected, SFS consistently outperforms MRMR, particularly in the low-feature regime (fewer than 10 features).

4.5.4 Best Model & Conclusion

To conclude the section on cross-matrix DON classification, we present the best-performing 10-feature model trained on the combined wheat and corn dataset. This model uses only the top 10 features selected via SFS from the 395nm excitation fluorescence data, preprocessed with area-normalization and smoothed to a 10nm resolution.

The model used was a 3 neighbor KNN model, and the corresponding confusion matrix is shown in Figure 4.39. The final classification accuracy reached was 82%—lower than the 95% achieved on corn alone, but notably, this model works across both matrices using a strict threshold of $500\mu\text{g/kg}$.

To better understand the model’s limitations, we examined the misclassified samples. For corn, the following samples had at least 5 out of 15 spectra misclassified: S04, S09, S10, S14, S16, S27, S32, S33, S39, and S40. For wheat, only two samples—S0H and S0I—met this criterion.

While there are more corn samples (#40) than wheat samples (#18), the imbalance in misclassification remains disproportionate, suggesting that the model struggles more with corn. This is perhaps expected, as the corn samples were composed using four different standards, making them inherently more complex than the wheat samples, which were derived from only two.

Investigating the specific errors we can state the following: sample S10 once again emerges as an outlier and the other misclassifications can largely be explained by their proximity to the classification threshold of $500\mu\text{g/kg}$. Specifically, for wheat, S0H and S0I were both next to the threshold, with concentrations of 576 and $500\mu\text{g/kg}$, respectively—making them borderline cases. For corn S04, S14, S16, S32, S33, and S39 also had concentrations near the threshold, explaining the misclassifications.

However, some corn samples such as S09, S27, S31, and S40 were misclassified by the majority of their spectra, despite having concentrations well above or below the threshold (see Tables 3.1 and 3.2 for the details of the samples). The reasons for these misclassifications remain unclear, as no evident pattern or justification could be found, other than a failure of the models predictive power.

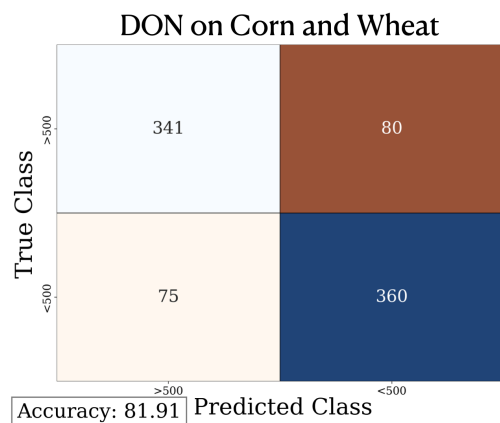


Figure 4.39: Confusion matrix for the two-class classification of DON across combined wheat and corn samples. The model shown is a KNN classifier, trained using the top 10 features selected by SFS from the 395nm fluorescence data. Spectra were preprocessed using area-normalization and smoothed to a 10nm resolution.

In summary, this section has demonstrated the feasibility of developing a unified, cross-matrix classification model for DON detection on both corn and wheat using fluorescence spectroscopy. Despite the added challenge of combining two distinct matrices—each with unique spectral profiles and mycotoxin concentration distributions—the resulting model achieved strong perfor-

mance, reaching an accuracy of 82% using only 10 selected features at a threshold of $500\mu\text{g/kg}$. This performance, while worse than that of the corn and wheat-only models, underscores the potential of cross-platform classifiers for practical applications where model generalizability is crucial. Importantly, the use of fluorescence data at 395nm, coupled with area-normalized and smoothed preprocessing, proved to be the most robust configuration across all tested modalities. SFS once again emerged as a powerful method for dimensionality reduction, enabling highly compact models with minimal loss in accuracy.

Overall, these results provide strong evidence that spectroscopic detection of DON can be made more versatile and scalable through unified classification approaches—paving the way toward more efficient, real-world implementation in food safety monitoring across diverse agricultural products.

4.6 Chapter Conclusion

In this chapter, we systematically developed and validated classification models for the in-situ detection of mycotoxins on food, using both fluorescence and reflectance spectroscopy. The study spanned two matrices, corn and wheat, employing a structured pipeline that included model-search, hyperparameter optimization, feature selection, and cross-platform data fusion. We further extended the approach by investigating cross-matrix classification, training unified models capable of detecting DON across both corn and wheat. A variety of machine learning models were explored, with a strong emphasis on dimensionality reduction using Sequential Feature Selection (SFS) to achieve compact yet accurate classifiers.

For corn, fluorescence spectroscopy consistently outperformed reflectance-based methods, with excitation at 395nm emerging as the most informative modality. Among classifiers, LDA provided the best trade-off between accuracy and computational efficiency, particularly when coupled with SFS to reduce input to just 10 key wavelengths. The best models achieved classification accuracies up to 95% (binary) and 83% (three-class) for DON, with similar trends observed for ZEA and FUM, but not for AFLA.

For wheat, reflectance spectroscopy—particularly NIR—proved highly effective, outperforming fluorescence. LDA remained the preferred classifier, with optimized models achieving up to 92% accuracy using only 10 selected features for DON. Attempts to combine data from multiple modalities yielded limited improvements, suggesting that NIR data alone is sufficient for reliable classification. However, wheat samples were fabricated from only two standards, making it difficult to disentangle spectral markers of mycotoxins from matrix-specific effects, yielding only limited confidence for these results.

Finally, in the cross-matrix case-study, a unified DON classifier was developed using combined corn and wheat data. The best-performing model—a KNN classifier trained on 10 SFS-selected features from 395nm fluorescence—achieved an accuracy of 82%. While this falls below the performance of matrix-specific models, it demonstrates the feasibility of creating generalizable classifiers. Together, these findings highlight the strong potential of fluorescence and reflectance spectroscopy, coupled with machine learning, to deliver efficient and versatile tools for mycotoxin detection—while also underscoring the critical importance of well-controlled sample fabrication for model robustness and generalizability.

Chapter 5

Regression Model Development

Thus far, our analysis has focused exclusively on the development and evaluation of classification models. These models have demonstrated the feasibility of both qualitative and semi-quantitative mycotoxin detection by categorizing samples according to predefined classes. However, classification inherently lacks the ability to provide precise quantitative estimates. At best, it allows us to determine whether a sample fits within a specific range—for example, whether contamination levels fall above or below certain legal thresholds. This leaves significant uncertainty: a sample classified as “safe” could either be entirely toxin-free or contain contamination just below the legal limit—two very different scenarios in terms of regulatory risk and food safety.

To address this limitation and gain a more detailed insight into the contamination level of individual samples, we turn to the development of regression models. Regression allows for continuous prediction of mycotoxin concentrations, offering a more granular and informative perspective that supports better decision-making, especially in borderline or risk-prone cases.

In this chapter we present the results of the development of regression models for the detection of DON on wheat, corn and both matrices combined. Throughout this chapter we follow an completely analogous modeling pipeline as was used during classification model development, encompassing model-search, optimization, and feature selection. As key difference, we now focus on using regression models where the aim is to map the spectral information to a continuous concentration rather than a class-based grouping of the data.

As discussed in Section 3.10, numerous regression models have been explored for predicting mycotoxin concentrations in food products based on spectroscopic measurements. However, these models are often trained on datasets comprising contamination levels that span a much broader range than the legally relevant thresholds, and the relative uncertainties in their predictions frequently exceed these limits, preventing regulatory compliance.

In this section, we experimentally implement and evaluate a range of regression models under more stringent and realistic concentration conditions, aiming to assess the feasibility of accurate mycotoxin quantification within legally relevant thresholds. **Notably, our contamination levels are chosen to reflect realistic field conditions rather than artificially high concentrations, making the task more challenging but also more relevant for practical applications.** In addition, we specifically target the development of a DON regression model that can generalize across different sample matrices, namely both wheat and corn, thereby increasing its applicability. Furthermore, the model is designed to remain effective even in the presence of co-occurring mycotoxins. In contrast to

most existing literature that focuses on optimizing detection accuracy or lowering the limit of detection under controlled conditions, our goal extends beyond this: we aim to develop a more robust and generalizable model capable of handling matrix variability and co-contaminations.

As a first step, we investigated the development of a regression model considering the wheat samples (Table 3.3), motivated by two key factors. Previously, the classification results for wheat demonstrated exceptionally high accuracy—reaching up to 99% using UV-VIS-NIR data with LDA models—indicating strong underlying spectral markers. Furthermore, the wheat samples were synthesized using nearly equidistant toxin concentration levels, making these dataset well-suited for regression analysis. This evenly distributed design minimizes the risk of bias from local overfitting and allows the model to capture broader, global concentration trends more effectively.

The development of the regression model followed an analogous methodology as the classification models. Consistent with the classification workflow, Leave-One-Sample-Out Cross-Validation (LOSO-CV) is used throughout, ensuring unbiased performance assessment while maximizing the use of available data (Section 4.1.4). The development process starts with a comprehensive model-search to identify the most informative combination of spectroscopic datasets, regression algorithm, and preprocessing strategy, followed by the optimization of the best-performing configurations. The results for wheat are presented in Section 5.1, for corn in Section 5.2, and for corn and wheat combined in Section 5.3. Subsequently, the exploration of feature selection techniques to simplify the spectral input and evaluate the trade-off between model complexity and predictive accuracy is explored in Section 5.4.

5.1 Regression Model For Mycotoxin Detection On Wheat

A wide range of regression algorithms are available. Within this section, we assume the reader has a basic background on the available algorithms. An in-depth overview of the key considered regression models is given in Appendix A.2.2.5. Partial Least Squares Regression (PLSR) is among the most frequently used and successful models for spectroscopic mycotoxin prediction. Its ability to handle high-dimensional, collinear data makes it particularly well-suited for applications involving complex spectral measurements. In our study, we evaluate PLSR models with 2, 3, 5, 7, 10, and 15 PLS-components¹. In addition, we test a range of alternative models, including Support Vector Regression (SVR) (Appx A.2.2.1), Random Forest Regressors (RFR) (with 100 and 200 estimators, Appx A.2.2.2), Artificial Neural Networks Multi-Layer Perceptron Regressors (ANN-MLPR) (single and multi-layer configurations, Appx A.2.2.3), and sparsity-inducing linear alternatives like Lasso and Elastic Net regressors (Appx A.2.2.4).

Each model will be tested across multiple spectroscopic datasets, including fluorescence using 395, 385, 375, and 365nm excitations, as well as UV-VIS and NIR reflectance data. A comprehensive set of preprocessing techniques will be employed, including spectral trimming, smoothing at different resolutions, normalization, and ratios. Spectral ranges and resolution-specific preprocessing parameters mirror those used in the classification workflow: fluorescence spectra are trimmed to 400–800nm, UV-VIS reflectance to 400–1000nm, and NIR reflectance to 1000–1700nm. Smoothing resolutions of 2, 4, or 6nm (fluorescence), 1, 3, or 5nm (UV-VIS), and 8, 12, or 16nm (NIR) are explored.

¹PLS-components are latent variables (e.g., linear combinations of wavelengths) that capture the directions of maximum covariance between the predictor variables (e.g., spectral data) and the response variable (e.g., toxin concentration), enabling effective dimensionality reduction and prediction.

To illustrate the complexity of the spectral data of the wheat samples we can look once more at Figures 3.10 and 3.16 showing the reflectance and fluorescence data, respectively, on a per sample (concentration) basis.

From these plots, it becomes evident that there is substantial spectral overlap between the spectra at varying concentrations. This underscores the necessity of machine learning-based regression models to uncover subtle spectral patterns associated with mycotoxin content.

5.1.1 Model-Search

To identify the optimal combination of spectroscopic dataset, regression algorithm, and preprocessing strategy, a comprehensive model-search was performed, aiming the simultaneous prediction of DON, T-2/HT-2, and OTA concentrations in the wheat spectral data. The key performance metrics used to evaluate the models were the coefficient of determination (R^2) and the Root Mean Squared Error (RMSE) (refer to Appendix A.2.2 for more details). R^2 -values² were visualized as heatmaps, following the same approach used earlier for classification accuracy.

The results of the model-search are presented in Figure 5.1, illustrating the performance of various model-preprocessing combinations across all datasets. Partial Least Squares Regression (PLSR) consistently emerged as the top-performing method. Configurations utilizing between 5 and 15 PLS-components typically achieved the highest accuracy, particularly when paired with area-normalization or ratio-based preprocessing. In addition to its strong predictive performance, PLSR also demonstrated significantly faster computation times. A detailed comparison of the computational efficiency of all regression models evaluated is provided in Section 5.1.2.

Lasso and Elastic Net regressors also performed well, particularly when used with ratio features, though they incurred higher computational costs. Random Forest Regressors (RFR), using both 100 and 200 estimators, showed solid performance as well, again benefiting from ratio preprocessing but requiring substantially more computational resources.

Among the spectroscopic datasets, NIR reflectance emerged as the most informative for regression—mirroring trends previously observed in classification (Section 4.3). This was followed closely by the 385 and 395nm fluorescence datasets. A summary of the best R^2 -values achieved for PLSR, Elastic Net, Lasso, and RFR across each dataset is presented in Table 5.1. The highest performance was achieved with NIR data ($R^2 = 0.96$), followed by Fluo 385nm ($R^2 = 0.88$) and Fluo 395nm ($R^2 = 0.87$).

Dataset	PLSR	Elastic Net	Lasso	RF
Fluo 395	0.87	0.86	0.87	0.85
Fluo 385	0.88	0.88	0.86	0.87
Fluo 375	0.84	0.81	0.85	0.80
Fluo 365	0.87	0.82	0.87	0.82
UV-VIS	0.85	0.82	0.87	0.84
NIR	0.96	0.94	0.94	0.92

Table 5.1: R^2 -values for DON, T-2/HT-2, and OTA regression on wheat using different regression models and spectroscopic datasets. PLSR generally performs best, with NIR spectroscopy yielding the highest accuracy across all models.

²An R^2 value can be negative when the model’s predictions are worse than simply predicting the mean of the target variable. This occurs when the residual sum of squares exceeds the total sum of squares, indicating that the model introduces more error than it explains.

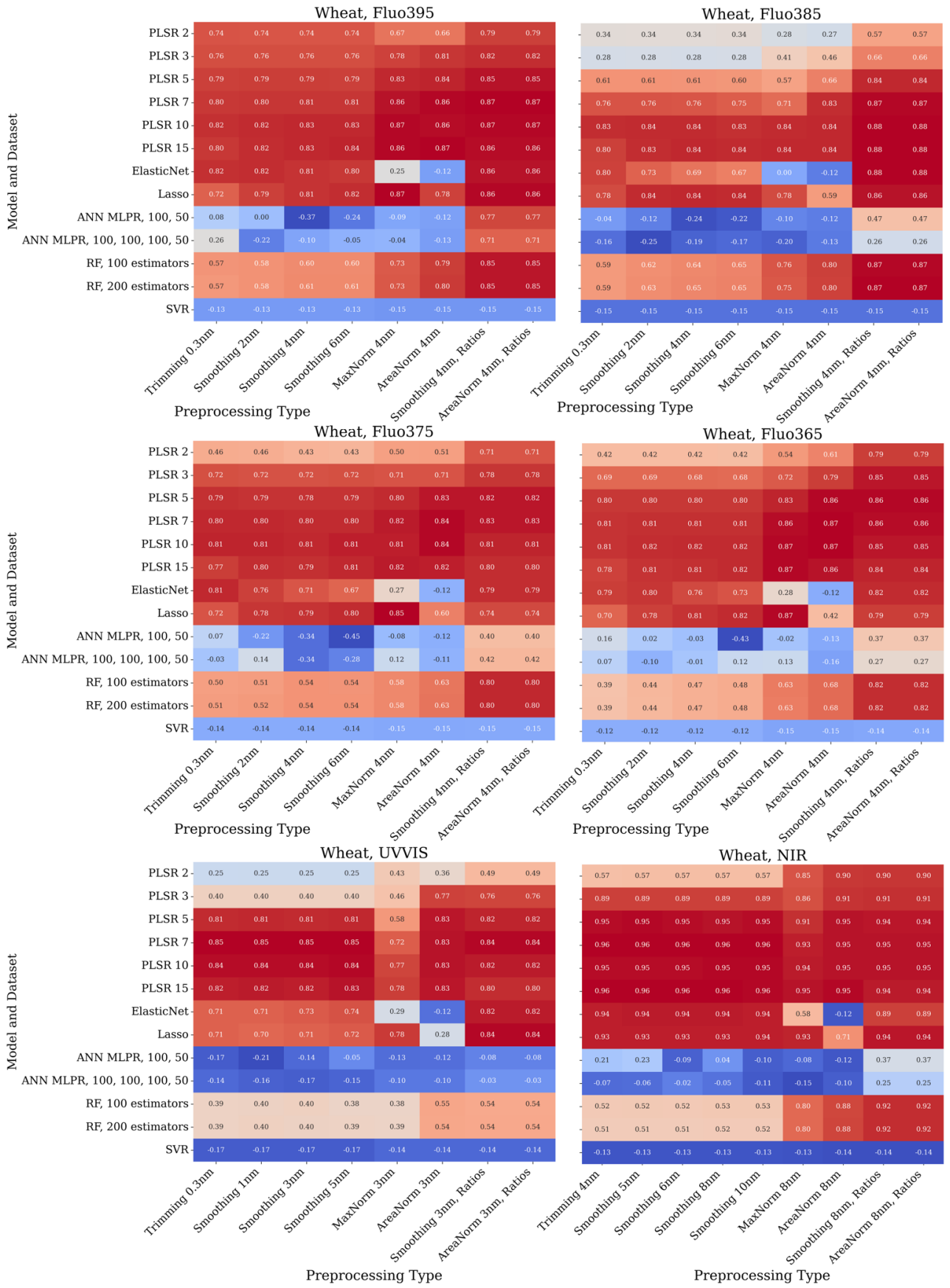


Figure 5.1: Regression R^2 -heatmaps for each fluorescence excitation wavelength (395, 385, 375, and 365nm) and reflectance modality (UV-VIS and NIR), showing the outcome of the model/preprocessing search for simultaneous prediction of DON, T-2/HT-2, and OTA concentrations on wheat.

5.1.2 Computational Cost Comparison Of Regressors

Just as a computational cost comparison was performed for the classification models (see Section 4.3.3), a similar analysis was conducted for the regression models. For this, we employed the wheat dataset and used fluorescence data at 395nm, smoothed to a 4nm resolution—matching the preprocessing pipeline used during classifier evaluation, but now applied in a regression context. Each regression model was executed three times in randomized order, and both the mean and standard deviation of runtime were recorded. This approach was designed to account for hardware-related effects, such as thermal throttling and background system load, thereby providing a more robust estimate of computational demand.

The results, shown in Figure 5.2, reveal that PLSR is the most computationally efficient model, followed closely by Lasso and Elastic Net, which perform similarly. In contrast, models such as RFR and ANN-MLPR exhibit significantly higher computational requirements.

It is important to note that, for consistency, all models were evaluated using the same preprocessing approach. However, as observed in the R^2 -heatmaps in Figure 5.1, not all models perform optimally with this particular preprocessing. Models such as Lasso and Elastic Net require ratio-based preprocessing for optimal performance, which substantially increases computational demand. Thus, PLSR emerges not only as the top-performing regression model in terms of predictive accuracy, but also as the most computationally efficient, making it especially well-suited for large-scale or feature-reduction-intensive applications.

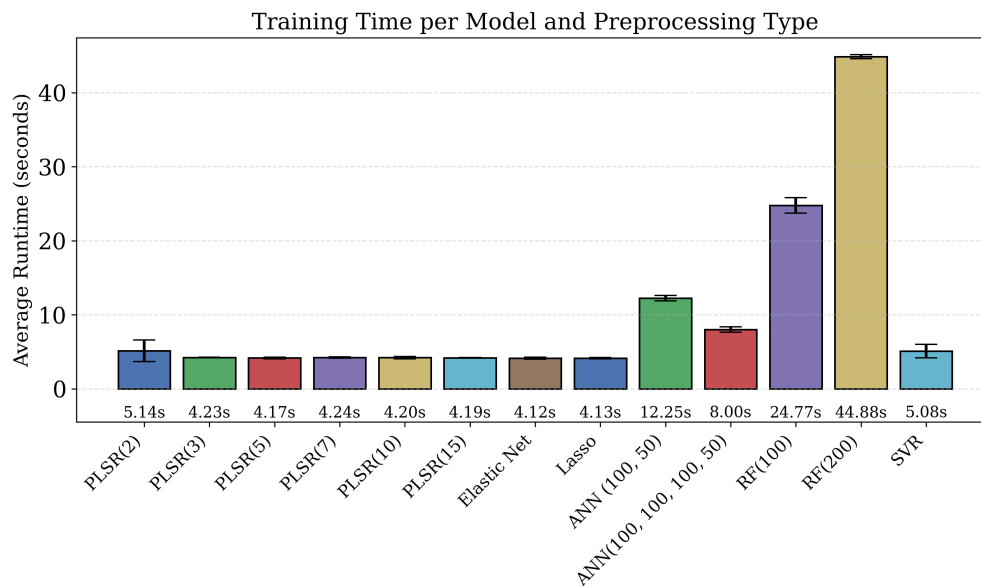


Figure 5.2: Average training time (in seconds) for various regression models evaluated on the wheat dataset using fluorescence data at 395nm (smoothed to 4nm resolution). Each model was executed three times in randomized order to account for hardware-induced variability, with error bars representing the standard deviation. PLSR demonstrates the lowest computational demand, followed closely by Lasso and Elastic Net. In contrast, RFR and ANN-MLPR require significantly longer training times.

5.1.3 Model Optimization

The model-search confirmed findings from the literature, identifying PLSR as the optimal regression approach. Specifically, the best results were achieved using 5–15 PLS-components in combination with area-normalization as the preprocessing strategy.

To further refine these results, a detailed model optimization was conducted, fine-tuning the number of PLS-components, while in parallel, refining the preprocessing strategy. The results of this optimization are presented in Figure 5.3.

We observe that the optimization step yielded only marginal improvements—only around 1–2%. A comparison of the original and optimized model performances is summarized in Table 5.2. Overall no strong trend emerged in terms of optimal preprocessing for the fluorescence datasets, while reflectance data showed a slight preference for smoothed over area-normalized data. A balanced and consistently effective configuration was achieved using PLSR with 9 components applied to 10nm area-normalized data. The performance of this setup is also reported in Table 5.2. As such, this configuration will be used in the following sections for further analysis.

Dataset	Original	Optimized	PLSR 9
Fluo 395	0.87	0.88	0.87
Fluo 385	0.88	0.91	0.84
Fluo 375	0.84	0.86	0.85
Fluo 365	0.87	0.88	0.88
UV-VIS	0.85	0.85	0.85
NIR	0.96	0.97	0.95

Table 5.2: Comparison of model performance (R^2) before and after optimization (columns 2 and 3). The final column presents the performance of the PLSR model with 9 components applied to area-normalized data at 10nm resolution—a configuration that consistently delivered strong results.

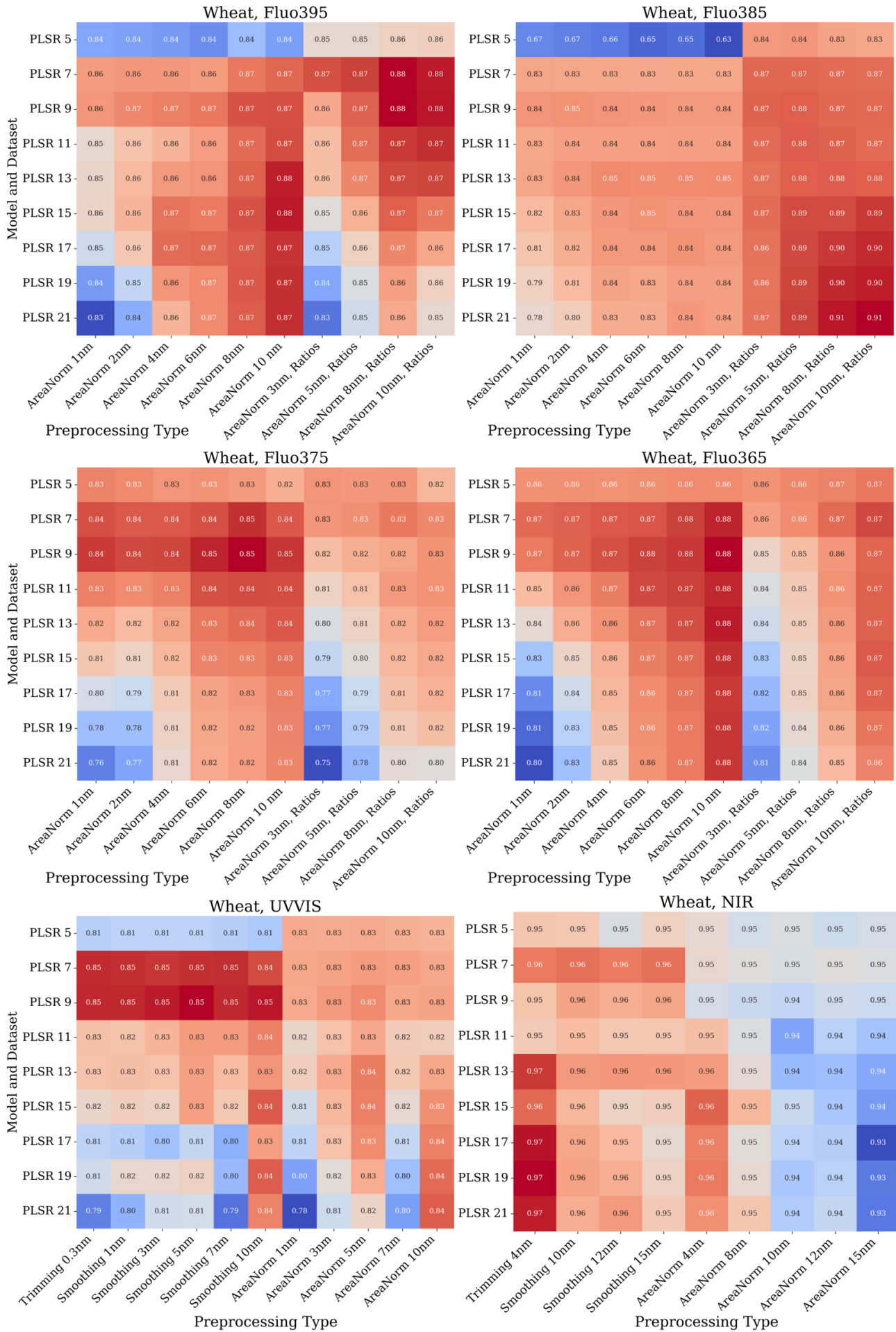


Figure 5.3: Regression R^2 -heatmaps for each fluorescence (395, 385, 375, and 365nm) and reflectance (UV-VIS and NIR) modality, showing the outcome of the model and preprocessing optimization for simultaneous prediction of DON, T-2/HT-2, and OTA concentrations in wheat.

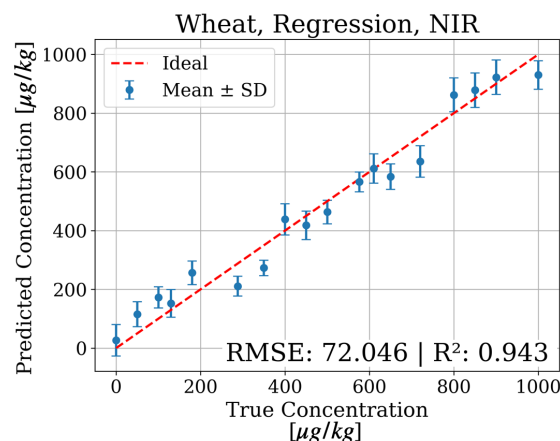
5.1.4 The Best Model: Mycotoxins On Wheat

Having evaluated a broad range of regression models, preprocessing strategies, and spectroscopic modalities, we now present the best-performing model for the simultaneous quantification of DON, T-2/HT-2, and OTA in wheat. This model, a PLSR with 9 components, uses the NIR reflectance data preprocessed using area-normalization at 10nm. The performance ascertained is summarized in Figure 5.4, which shows the predicted concentrations versus the actual concentrations per sample. Each data point represents the average predicted value across 15 spectra for a given wheat sample, with error bars indicating the standard deviation.

The model achieves an R^2 -value of 0.94 and a RMSE of $72.1\mu\text{g/kg}$, indicating a strong predictive fit. We also calculated the LOD and LOQ using equations A.56 and A.57, respectively. The LOD was found as $231.9\mu\text{g/kg}$, and LOQ as $702.9\mu\text{g/kg}$. Both values that fall below the legal threshold of $1250\mu\text{g/kg}$ indicating we can detect below the legal threshold.

Compared to literature this is a very high R^2 -value considering we were working with a significantly lower concentration range. Note however that we detect the combined effect of DON, T-2/HT-2, and OTA and not only DON. While the model performance is highly encouraging, care should be taken since also inherent matrix differences specific to the original standards might impact the measurements. To ensure a robust model, further validation measurements are required.

Figure 5.4: Regression performance of the best model: PLSR with 9 components, applied to NIR reflectance data—preprocessed with 10nm area-normalization. Each point corresponds to one sample, based on the average prediction across the 15 spectra per sample. Error bars indicate the standard deviation of predictions. The model indicates a strong performance even within a concentration range below the legal limit.



5.2 Regression Model For Mycotoxin Detection On Corn

Building on the promising results obtained for wheat regression, the next logical step is to investigate whether comparable predictive performance can be achieved for mycotoxin detection in corn. Here the effects of individual toxins and matrix composition were more effectively decoupled during sample synthesis. Thusly a successful implementation in this context would significantly strengthen confidence in the model’s ability to make mycotoxin-specific predictions. As discussed in Section 4.3.7, DON yielded the highest classification performance; ergo, we now turn our attention to the regression analysis of DON concentrations in corn.

The same model development pipeline was employed to implement this: starting with a comprehensive model-search comparing multiple regression algorithms, followed by targeted model optimization to enhance performance. In the model-search, as was the case for wheat regression, we found PLSR as the superior model, offering both high accuracy and computational efficiency. Subsequent model optimization led to minor improvements in performance. The R^2 -heatmaps from this optimization process are shown in Figure 5.5, and the best-performing model–preprocessing combinations are summarized in Table 5.3.

From the heatmaps, it is evident that regression performs poorly on the reflectance data (UV-VIS and NIR), mirroring the weak classification performance observed earlier (see Section 4.3.7). In contrast, fluorescence data yields considerably better results, with the best R^2 of 0.83 achieved using the 395nm dataset and a PLSR (21 components) model, using either area-normalized or ratio-based³ preprocessing (both smoothed to 10nm resolution). The remaining fluorescence datasets all performed worse and showed notable differences in preprocessing preference. Given that 395nm fluorescence with area-normalization achieved the highest performance, we proceed with this configuration for subsequent analysis.

Dataset	Area-Normalized [R^2]	Ratios [R^2]
Fluo 395	0.83	0.83
Fluo 385	0.73	0.74
Fluo 375	0.78	0.82
Fluo 365	0.78	0.83
UV-VIS	0.58	–
NIR	0.48	–

Table 5.3: Comparison of regression R^2 -values for DON detection on corn using PLSR. Results are shown for both area-normalized and ratio-based preprocessing at 10nm resolution. Fluorescence at 395nm yields the highest and most consistent performance, while reflectance-based datasets (UV–VIS and NIR) perform poorly, consistent with classification findings.

³Ratio-based preprocessing was not applied to the UV-VIS and NIR datasets due to their high dimensionality, which rendered the computations infeasible under the given resource constraints.

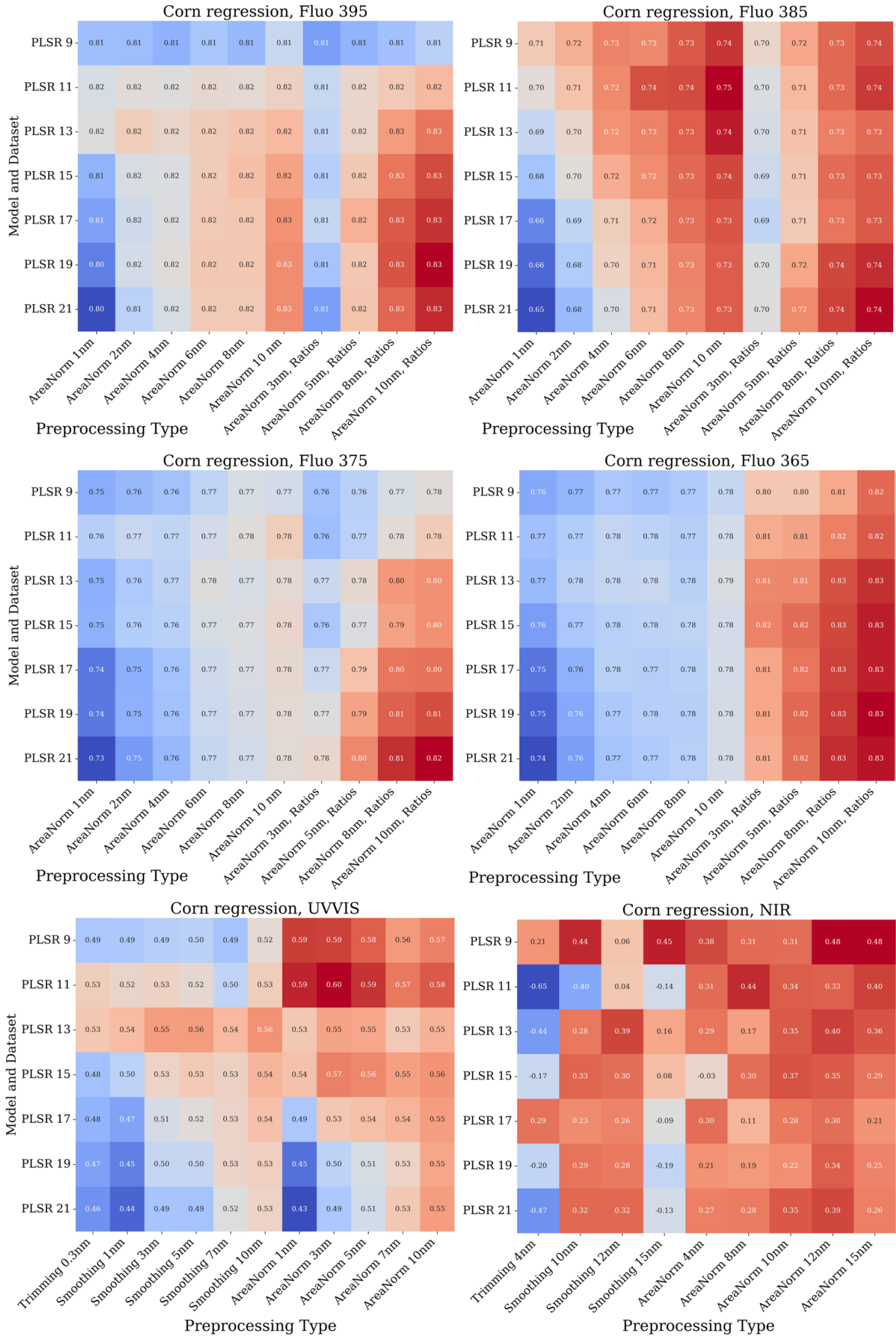


Figure 5.5: Regression R^2 -heatmaps for each fluorescence (395, 385, 375, and 365nm) and reflectance (UV-VIS and NIR) modality, showing the outcome of the model and preprocessing optimization for the prediction of DON on corn.

5.2.1 The Best Model: DON On Corn

Based on the results of the model search and subsequent optimization, the best-performing regression model for DON detection in corn was achieved using smoothed and area-normalized 395nm fluorescence data at a resolution of 10nm, employing a PLSR with 21 components. This configuration provided the highest predictive accuracy and computational efficiency among all approaches evaluated.

The predicted versus actual DON concentrations across all corn samples, is visualized in Figure 5.6. Each data point represents the average prediction of the 15 spectra per sample, with error bars indicating the standard deviation. The model achieved an R^2 -value of 0.83 and an RMSE of $295.6\mu\text{g}/\text{kg}$. We thus find more moderate results for the case of DON detection on corn compared to regression on the wheat samples.

Compared to literature we achieve a lower R^2 -value, as can be seen by comparing these results with those reported in Table 2.10. Our RMSE is however significantly better than many of the reported studies, which is likely due to the fact that we trained a model for sample spanning a lower concentration range. The fact that we worked with a lower concentration range than generally presented in literature also makes the regression analysis more challenging.

Using the average RMSE of the DON-free samples as an estimate of the standard deviation (σ), along with the slope of the regression line, we calculated the limit of detection (LOD) and the limit of quantification (LOQ). The LOD was found to be $1017.4\mu\text{g}/\text{kg}$, and the LOQ $3083.2\mu\text{g}/\text{kg}$. The LOD thus still falls under the legal threshold, but LOQ is already more than twice the legal limit, indicating that operating within a legal framework is already more challenging.

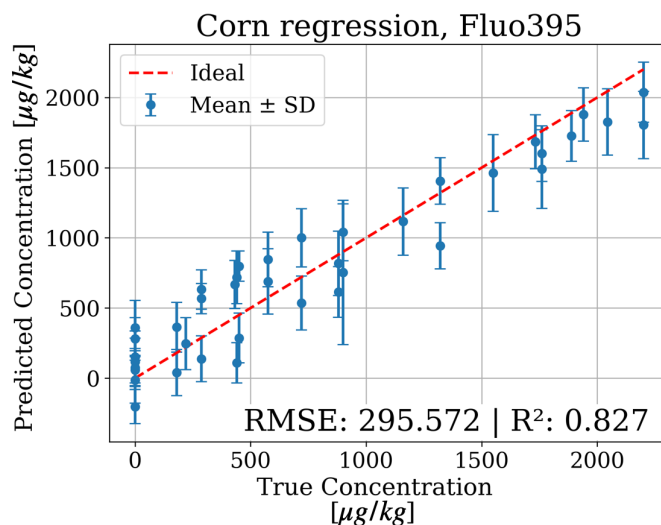


Figure 5.6: Regression performance of the best model for corn, a PLSR with 21 components applied to 395nm fluorescence data, preprocessed with 10nm area-normalization. The plot shows predicted versus actual concentrations of DON. Each point corresponds to one sample, based on the average prediction across the 15 spectra, error bars indicate the first standard deviation.

5.3 Regression Model For Mycotoxin Detection On Corn And Wheat

Lastly, paralleling our classification efforts, we also explored whether it is feasible to develop a unified regression model capable of detecting DON across both corn and wheat samples simultaneously. A successful cross-matrix model would offer significant advantages in terms of generalizability, and ensure we are detecting DON specific characteristics rather the matrix specific markers.

Once again we started by conducting a comprehensive model-search, evaluating multiple regression algorithms and preprocessing strategies across the spectroscopic datasets. The results are shown in Figure 5.7. Interestingly, a shift in model preference is observed compared to previous experiments. In this case, Random Forest Regression (RFR) models—particularly those using area-normalization or ratio-based preprocessing—consistently outperform other approaches. While PLSR previously dominated in both the wheat and corn datasets individually, it ranks second on the combined dataset and performs notably worse than RFR.

This shift can be attributed to the two-matrix nature of the classification challenge. Each matrix (corn and wheat) possesses its own unique chemical composition and resulting spectral profile. It is also likely that the spectral markers linked to mycotoxin presence differ between matrices. RFR is well-suited to this scenario, as its ensemble of decision trees can implicitly separate the two types of spectra and independently evaluate mycotoxin content within each group. In contrast, PLSR is a linear method that lacks the same flexibility to model matrix-specific relationships separately, making it less effective in this more heterogeneous setting.

To better highlight the comparative performance between RFR and PLSR, Table 5.4 presents the best R^2 -scores achieved by each model across all datasets. Notably, fluorescence data at 365nm yields the highest overall R^2 -value at 0.83, using the RFR and area-normalized data, at a 4nm resolution. In contrast, PLSR consistently falls short, by approximately 10%, across all fluorescence modalities. The models based on the reflectance data show generally a much poorer performance.

Dataset	RFR [R^2]	PLSR [R^2]
Fluo 395	0.77	0.70
Fluo 385	0.77	0.65
Fluo 375	0.76	0.76
Fluo 365	0.82	0.76
UV-VIS	0.53	0.51
NIR	0.40	0.41

Table 5.4: Comparison of regression R^2 -values for DON detection on combined corn and wheat datasets. RFR models using area-normalized data outperform PLSR across all spectroscopic platforms.

Following the model-search, additional optimization procedures were carried out for both RFR and PLSR models, including hyperparameter tuning and refined preprocessing variations. However, no significant improvements in performance were observed beyond the initial model-search results. Overall, these findings are promising, as they demonstrate that regression models can be trained to generalize across different matrices. If such models—trained to detect mycotoxins across a variety of foodstuffs—could also generalize to previously unseen materials, it would significantly simplify real-world deployment. A single regression model capable of handling multiple agricultural products could streamline implementation and reduce the need for matrix-specific calibration. While this aspect was not investigated in the present study, it represents a potentially fruitful direction for future research.

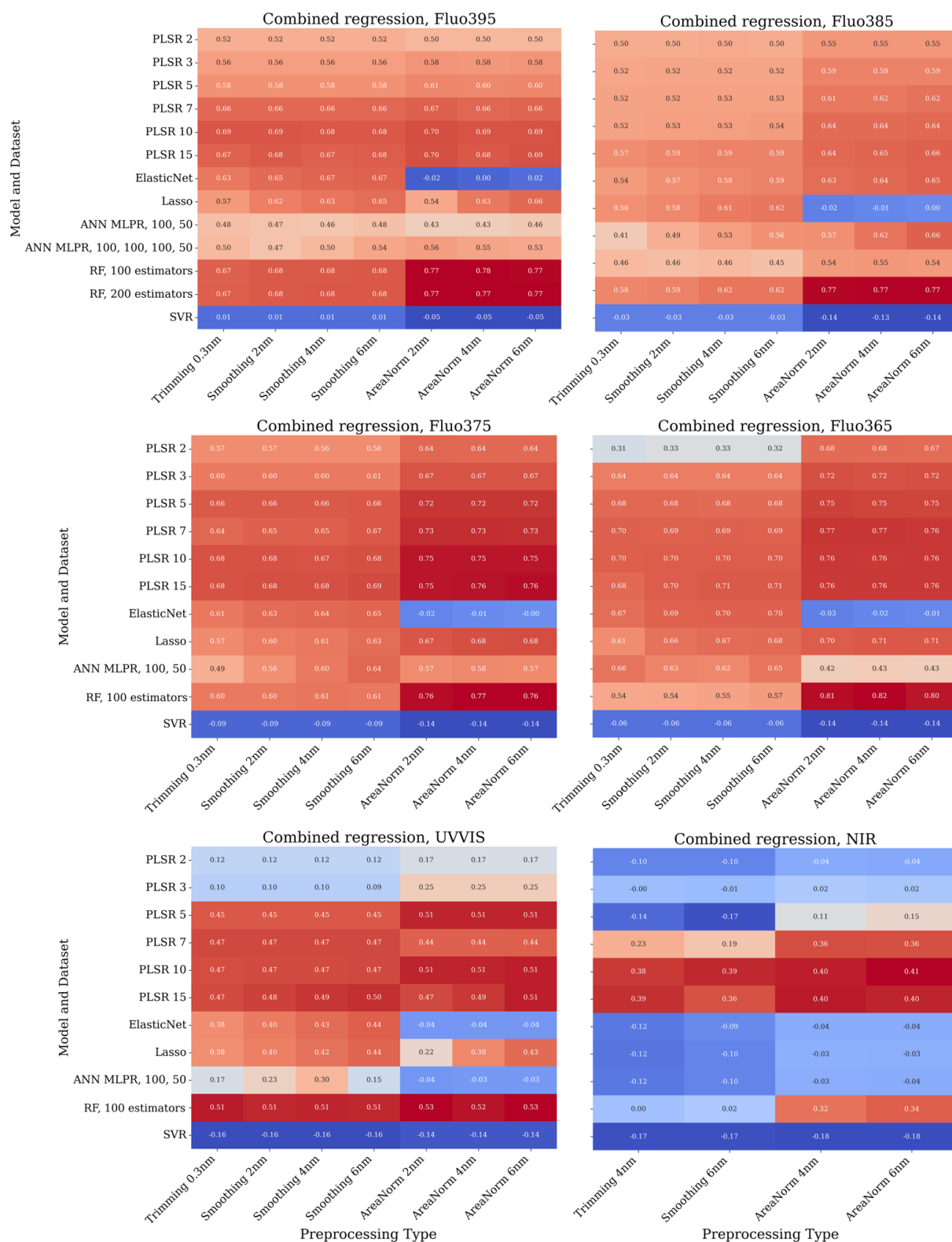
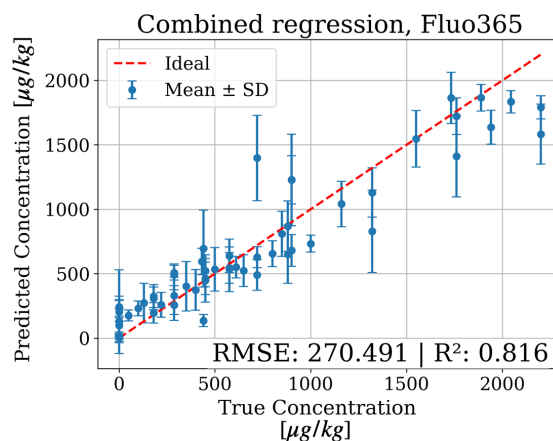


Figure 5.7: Regression model-search results for DON detection on corn and wheat matrices. RFR using ratios and area-normalized data consistently achieves the highest R²-values.

5.3.1 The Best Model: DON On Wheat And Corn

Having conducted a comprehensive evaluation of various regression algorithms and preprocessing strategies, we now present the best-performing model for the cross-matrix regression task. This model—a RFR with 100 estimators—was trained on area-normalized fluorescence data at 365nm, smoothed to a 4nm resolution. Its predictive performance is illustrated in Figure 5.8. The model achieved an R^2 -value of 0.82 and an RMSE of $270.5\mu\text{g/kg}$, closely matching the performance obtained for corn, but worse than the performance for wheat. This suggests that the challenge of detecting DON across both matrices is comparable to that for corn-only detection. However, the figure also reveals a notable imbalance in the distribution of sample concentrations, with the majority of samples clustered at lower DON levels. This imbalance results in an under-representation of higher concentrations during training, thereby limiting the model’s ability to accurately predict at the upper end of the concentration range. The model shows to be become biased towards localized trends in the lower concentration region, reducing its generalizability across the full spectrum of contamination levels.

Figure 5.8: Regression performance of the best cross-matrix model: RFR, 100 estimators, applied to fluorescence data at 365nm, preprocessed using area-normalization with 4nm smoothing. Each point represents the average prediction over the 15 spectra per sample, with error bars indicating the standard deviation.



To further assess the model’s sensitivity, we computed the Limit of Detection (LOD) and Limit of Quantification (LOQ) using the RMSE from the lowest concentration level as the estimate of σ . This yielded a LOD of $870.4\mu\text{g/kg}$ and a LOQ of $2637.7\mu\text{g/kg}$. The LOD thus falls below the legal threshold of $1250\mu\text{g/kg}$, while the LOQ does not meet this criterion.

Overall, the combined model’s performance did not decline relative to the corn-only model, indicating that cross-matrix monitoring is indeed feasible using a single regressor. This also reinforces confidence that the model is genuinely learning to predict DON concentration independent of the sample matrix.

5.4 Feature Selection

As a final component to this study, we investigated the feasibility of applying feature selection techniques—specifically Sequential Feature Selection (SFS) and Minimum Redundancy Maximum Relevance (MRMR)—to reduce the dimensionality of the spectroscopic input in the regression tasks. The aim here was to evaluate whether a carefully selected subset of spectral features can achieve comparable predictive performance to full spectrum regression, thereby simplifying model complexity and improving industrial applicability.

To keep the analysis computationally manageable while still relevant, we applied these feature

selection techniques to the most successful and computationally efficient case: the regression of the wheat samples using a PLSR model with 9 components and area-normalized 395nm fluorescence data, smoothed to a 10nm resolution.

However, applying feature selection to PLSR showed to present a technical challenge. Algorithms like SFS operate iteratively, evaluating model performance as features are added one by one. In the initial stages—when only a small number of features (e.g., one, two, or three) have been selected—training a PLSR model with a fixed number of components (e.g., nine) becomes problematic. PLSR requires at least as many input features as components to construct its latent variables, which are linear combinations of the original features. In practice, even more features may be needed, as the selected features are often not sufficiently linearly independent. This means that during the early iterations of feature selection, the dimensionality is too low to support the desired number of components, leading to model failures. Consequently, this incompatibility makes PLSR unsuitable for use with iterative feature selection techniques like SFS, in line with the problem encountered for PLS-DA as seen in Section 4.3.1.2.

To address this, we adapted the modeling procedure by dynamically adjusting the number of PLS-components during feature selection. Specifically, we began with PLSR(1)—PLSR with only one PLS-component—when selecting the first feature, switching to PLSR(2) once two latent variables can be formed, and continue this procedure incrementally until reaching the predetermined number of components or maximal number of selected wavelength features.

While this workaround introduces some inconsistency—since features selected under PLSR(1) may not be optimal in the context of a PLSR(9) model—it provides a practical and computationally viable strategy to rank features step-by-step. This approximation enabled us to proceed with feature selection without violating the underlying assumptions of the regression algorithm. The resulting performance, in the form of R^2 -heatmaps for both MRMR and SFS, is shown in Figure 5.9. Here we observe that MRMR requires over 30 features to reach full-spectrum performance, while SFS achieves comparable accuracy using only 7 features—reaffirming its greater effectiveness for feature selection. NIR data also once again emerges as the most informative modality, consistent with previous findings for full-spectrum regression.

In a next phase we also repeated the same analysis for the DON-regression on corn, with results shown in Figure 5.10. Here we observed the same trend: SFS consistently outperformed MRMR, achieving full-spectrum performance using approximately 10 features.

Finally, we applied feature selection to the combined corn and wheat regression task. In this scenario, Random Forest Regression emerged as the best-performing model—despite being significantly more computationally demanding. Running MRMR on this model required approximately a full day to complete. Given the significantly higher computational demands of SFS, implementing this method was deemed impractical. The estimated runtime exceeded one week, and the available hardware resources were insufficient to support such extended processing time. Nevertheless, based on trends observed across the wheat and corn datasets, we reasonably expect that SFS would outperform MRMR in this context as well. The obtained results for MRMR are shown in Figure 5.11.

Here we observe that using only 10 features, the model achieves performance approximately 10% below that of the full-spectrum baseline. While including more features—around 30 or more—or using SFS which is expected to likely improve results and close this gap, but at the expense of excessive computational resources.

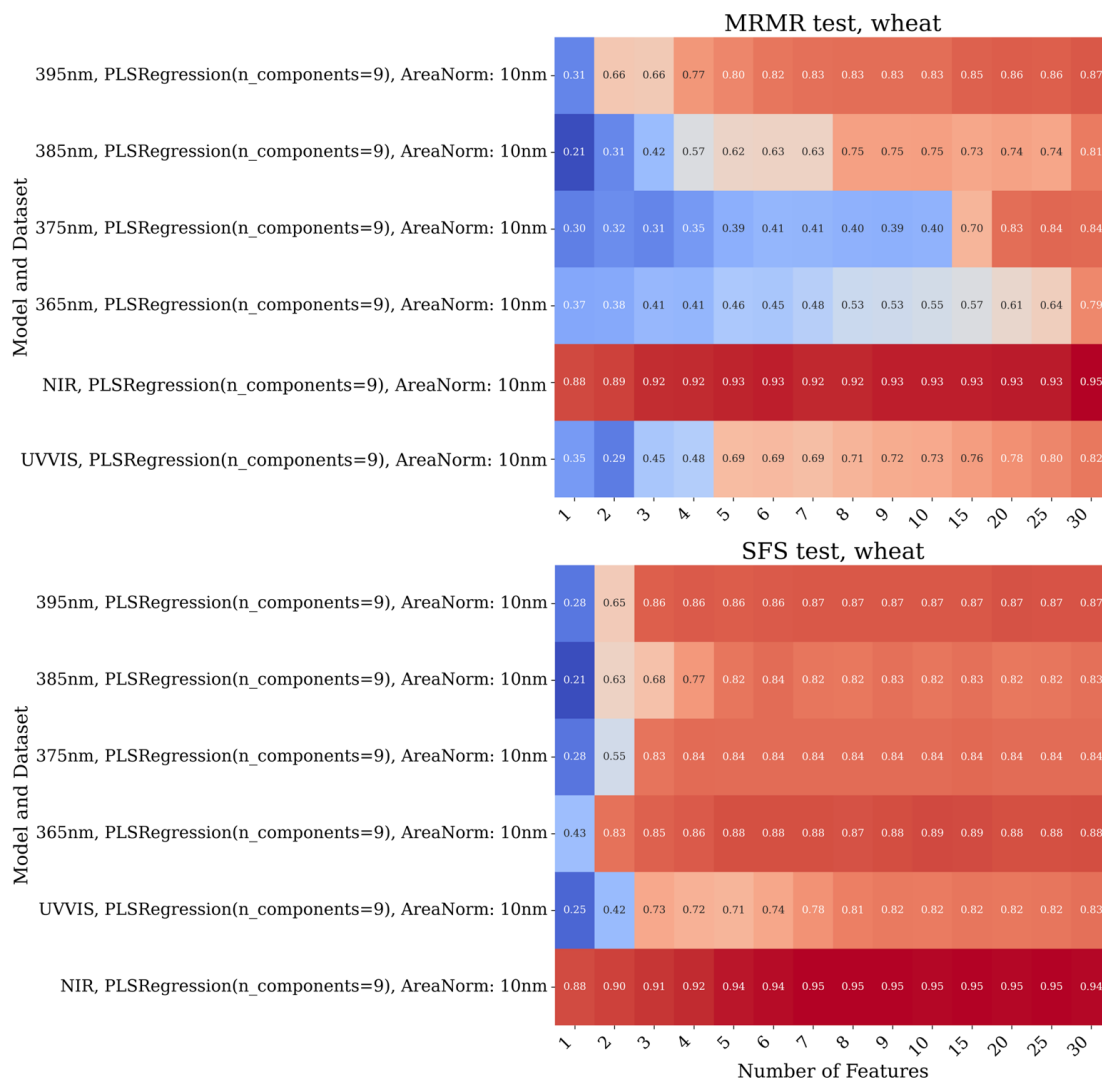


Figure 5.9: R^2 -heatmaps comparing MRMR (top) and SFS (bottom) feature selection methods applied to PLSR (9), for the wheat samples, run on area-normalized data at 10nm resolution, across all spectroscopic modalities. As expected, SFS consistently outperforms MRMR—particularly in the low-feature regime (fewer than 10 features)—demonstrating its superior ability to identify informative subsets of spectral features.

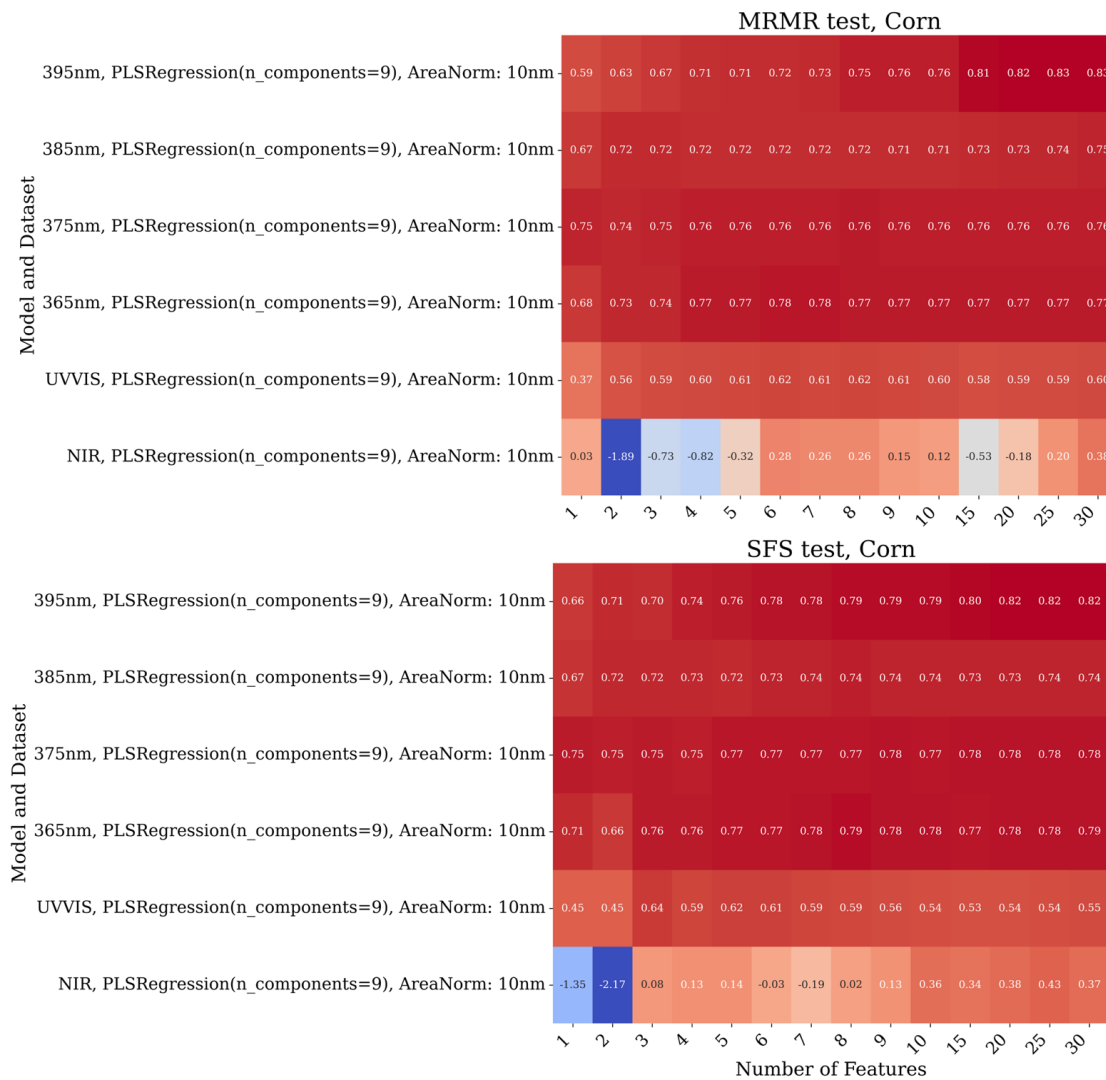


Figure 5.10: R^2 -heatmaps comparing MRMR (top) and SFS (bottom) feature selection methods applied to PLSR (9) on area-normalized data at 10nm resolution, across all spectroscopic modalities, targeting the regression of DON on corn. As expected, SFS consistently outperforms MRMR—particularly in the low-feature regime (fewer than 10 features)—demonstrating its superior ability to identify informative subsets of spectral features.

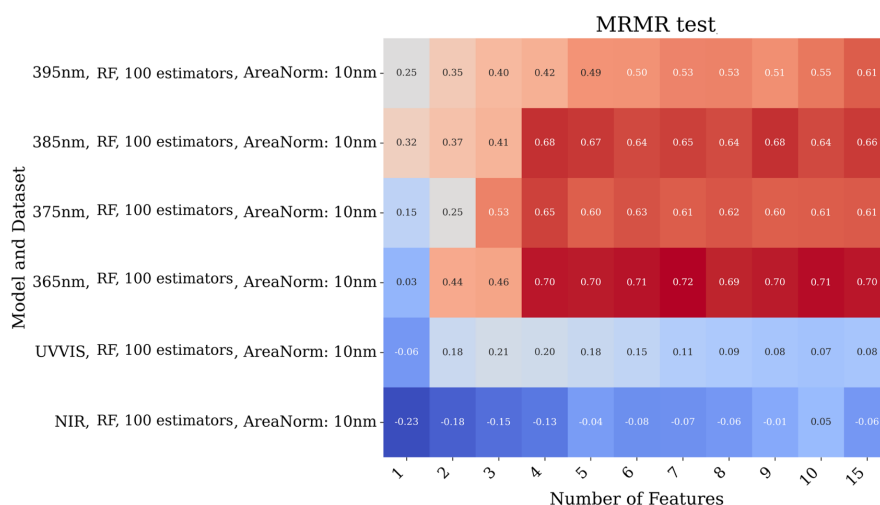


Figure 5.11: R^2 -heatmap of MRMR feature selection, applied to RFR (100 estimators) using area-normalized data at 10nm resolution, across all spectroscopic modalities, aiming for the regression of DON on corn and wheat combined. Using 15 features we stay 10% below full spectrum performance. Creating this heatmap required over a day of compute.

5.5 Conclusion: DON Detection Using Regression Models

In this chapter, we extended the spectroscopic analysis of mycotoxins from classification to quantitative regression, aiming to predict precise contamination levels of DON in wheat⁴, corn, and across both matrices. We applied a structured pipeline—mirroring the classification workflow—encompassing model search, optimization, feature selection and data merging. Various regression models were evaluated, including PLSR, SVR, RFR, and others, across fluorescence and reflectance spectroscopy data, with preprocessing strategies carefully tailored to each modality. Our objective was to assess not only predictive performance but also the feasibility of dimensionality reduction for practical deployment.

For wheat, regression performance was highly promising: the best model—a PLSR trained on NIR reflectance data—achieved an R^2 of 0.94 and an RMSE of $72.1\mu\text{g/kg}$, with an LOD well below legal thresholds, demonstrating strong predictive power even within a narrow and realistic concentration range. Feature selection via SFS further confirmed that near-full performance could be retained with as few as 7–10 features, underscoring the potential for low-complexity, high-accuracy models. Considering the corn samples, the application of regression models appeared to be more challenging. The top model—a PLSR using 395nm fluorescence—achieved an R^2 of 0.83 and a higher RMSE of $295.6\mu\text{g/kg}$. While respectable, the LOQ exceeded legal limits, indicating that while detection is feasible, precise quantification, with high confidence, remains difficult under the current model. Future work might improve on these results by investigating the integration of spectroscopic data of complementary platforms for model development. Additionally having larger more diverse datasets with more uniform sample distribution over the concentration span could improve on these results.

When attempting a unified regression model across both corn and wheat, Random Forest Regres-

⁴This concerned the simultaneous detection of DON, T-2/HT-2, and OTA.

sors outperformed PLSR, achieving an R^2 of 0.82. This finding reflects the added complexity of cross-matrix regression and the benefit of non-linear models in such heterogeneous contexts. The strong performance of the combined model also reinforces confidence that it is genuinely learning the DON content itself, rather than matrix-specific spectral patterns. In addition, and for the first time to our knowledge, it showcases the quantification of DON using fluorescence spectroscopy in a multi-matrix regression setting—exceeding the state-of-the-art findings reported in the literature.

Chapter 6

Validating Model Performance

In the preceding chapters, we systematically developed both classification and regression models, covering the entire pipeline—from model selection and optimization to spectral feature reduction. The final models demonstrated competitive performance, achieving detection accuracies competing with literature benchmarks, but now in a multi-toxin environment. Thus far, however, model development and evaluation relied exclusively on the Leave-One-Sample-Out Cross-Validation (LOSO-CV) strategy, a strategy, which was the preferred choice given the limited availability of training samples.

Furthermore all samples were composed by mixing a handful of reference standards, meaning that the samples remained related not only by their mycotoxin contamination, but also by their matrix composition. As a consequence the broader question of generalization to entirely unseen material remains unresolved.

To address this, the following sections describe a final evaluation conducted within this thesis, in which independently sourced reference materials—distinct from those used in any previous part of the study—are employed to create new DON-contaminated wheat and corn samples. The aim being to evaluate how well a model trained on the original dataset generalizes to these new, entirely independent validation samples.

6.1 Validation Samples

To assemble a set of fully independent validation samples, four certified reference materials were purchased from Trilogy Analytical Laboratory [140]. These standards were prepared following the same protocol as those provided by Romer Labs [130], as described in Section 3.1.1.

Two wheat and two corn samples, each contaminated with varying concentrations of DON were bought. For corn, standards DC-20030 and DC-20024 contained DON at concentrations of $(0.5 \pm 0.1)\mu\text{g}/\text{kg}$ and $(2.0 \pm 0.2)\mu\text{g}/\text{kg}$, respectively. For wheat, standards DW-20013 and DW-20010 were used, with DON levels of $(0.4 \pm 0.1)\mu\text{g}/\text{kg}$ and $(1.8 \pm 0.1)\mu\text{g}/\text{kg}$, respectively.

Using these standards, a series of new validation samples was composed as detailed in Table 6.1. These samples are fully independent from the samples considered in the previous chapters, enabling external model validation. Importantly, and in contrast to the previously considered samples, these new validation samples contained only DON and no other mycotoxins, allowing for a more specific assessment of the model’s ability to quantify DON in isolation. Visually, the samples were indistinguishable from those prepared earlier, as illustrated in Figure 6.1.

Independent Validation Samples: Composition and DON Content					
Sample	DON [$\mu\text{g}/\text{kg}$]	DC-20024 [g]	DC-20030 [g]	DW-20010 [g]	DW-20013 [g]
S51	2000 ± 200	10	0	0	0
S52	1250 ± 112	5	5	0	0
S53	500 ± 100	0	10	0	0
S54	1800 ± 100	0	0	10	0
S55	400 ± 100	0	0	0	10
S56	1100 ± 71	0	0	5	5

Table 6.1: Overview of the independently sourced validation samples, composed from four certified standards with varying DON concentrations. Each entry indicates the sample's total DON level and the precise mass of each reference material used. The yellow rows indicate corn samples, the beige rows wheat samples.

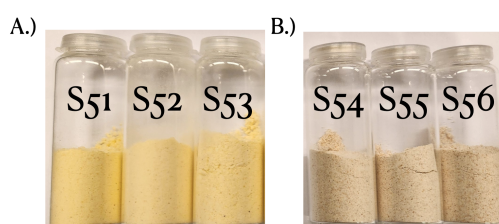


Figure 6.1: A.) The corn samples S51-S53, B.) The wheat samples S54-S56.

6.2 Spectral Data Of The Validation Samples

After sample preparation, we conducted 15 measurements per sample, capturing both the full UV-VIS-NIR reflectance spectra and fluorescence at an excitation wavelength of 395nm. The same measurement procedures and setups as detailed before in Section 3.2.1 and 3.2.2, were used. Figure 6.2 shows the obtained spectra for the new validation (dashed line) wheat (Plots *A* to *C*) and corn (Plots *D* to *F*) samples, compared to the spectral data of the original samples used for model training (full line), shown on a per class basis, using a threshold of $500\mu\text{g}/\text{kg}$ for DON on wheat and $1250\mu\text{g}/\text{kg}$ for DON on corn.

For wheat, a pronounced difference is observed in the fluorescence spectra at 395nm excitation (Plot *A*). The validation samples exhibit a higher overall fluorescence intensity and a more prominent chlorophyll emission peak compared to the training data. Furthermore, the spectral distinction between contamination classes has diminished, with increased overlap observed. These changes suggest a potential shift in matrix composition. Importantly, the new samples contain only DON, while the original training set included co-contaminants such as OTA, T-2, and HT-2. In addition to the natural variation between different wheat samples, also this difference in mycotoxin composition may partly explain the observed spectral deviations.

Turning to the wheat reflectance data (Plots *B* and *C*), we observe in the visible region (Plot *B*) a strong spectral divergence, due to a color differences between the samples. In the NIR range (Plot *C*), there is stronger spectral overlap between the new and original samples, however, we still observe a slight deviation starting around 1400nm, which may be attributed to differences in moisture content between the two sample sets.

Considering the corn measurements, presented in Plots *D* to *F*, we observe a much stronger spectral match between the new measurements and the original data. For the fluorescence measurements the new spectra fall largely within the first standard deviation region. Nevertheless, the new spectra display more prominent double peaks in the 470–530nm region and show no chlorophyll peak¹. Suggesting that there is significant matrix variation.

The UV-VIS-NIR data for corn shows better overall consistency with the original measurements compared to wheat. Still, the visible region again exhibits the largest spectral deviation, followed by the +1400nm region—also here implying that these regions may capture more information related to background matrix composition rather than mycotoxin-specific signatures.

In summary, the spectral data of the new samples exhibit notable deviations—likely originating from differences in matrix composition and the lack of co-contaminants. These deviations introduce additional complexity, making external validation using these samples particularly challenging. A thorough assessment of model generalization to this independent validation set will be carried out in Section 6.3.

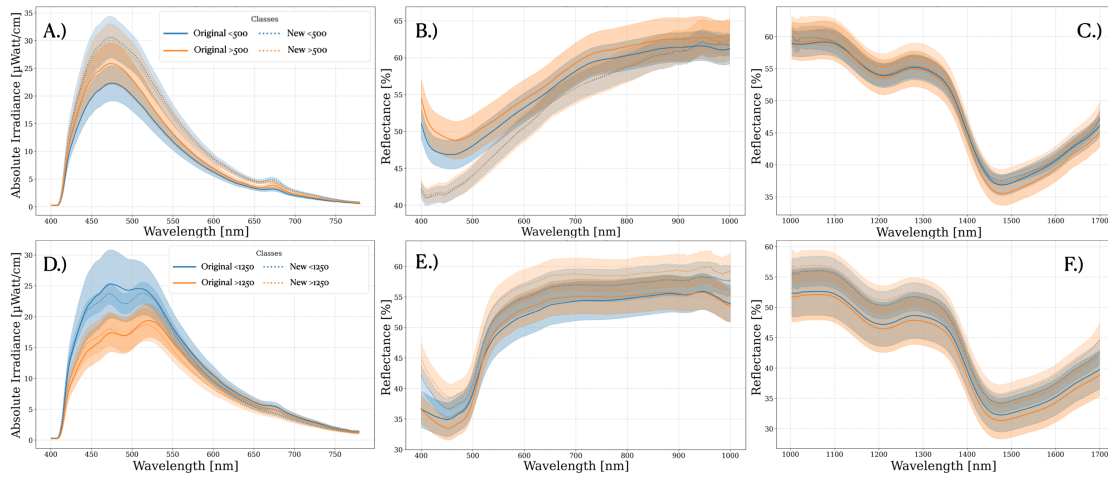


Figure 6.2: A.) Fluorescence spectra (excitation at 395nm) of the new wheat validation samples (S54–S56), compared to the original training set for both contamination classes ($< 500\mu\text{g}/\text{kg}$ and $> 500\mu\text{g}/\text{kg}$ DON). Dotted lines represent the new samples, while solid lines show the original data. The first standard deviation band is included. A distinct spectral shift is observed between the new and original measurements. B.) and C.) Full UV–VIS–NIR reflectance spectra of the same wheat samples, highlighting noticeable discrepancies in the VIS region. D.) Fluorescence spectra (395nm excitation) for the new corn validation samples (S51–S53), shown alongside the original training data, DON thresholds of $< 1250\mu\text{g}/\text{kg}$ and $> 1250\mu\text{g}/\text{kg}$. The new samples largely fall within the first standard deviation range of their respective contamination classes. E.) and F.) Full UV–VIS–NIR reflectance spectra for the corn samples, comparing new and original measurements.

¹The lack of a chlorophyll peak here, also suggesting that chlorophyll fluorescence may be a confounding variable, not necessarily indicative of mycotoxin presence. Though previously useful in distinguishing classes, it may not generalize well to unseen data where chlorophyll content differs substantially.

6.3 Validating Full Spectrum Classifiers

With the new spectral data acquired, the next step involved evaluating the generalizability of the best-performing models—trained on the complete training set (all measurements of the original samples)—by validating their performance on the newly obtained, independent validation samples. We investigated binary LDA classifiers for DON detection, considering thresholds of $500\mu\text{g/kg}$ for DON in wheat and $1250\mu\text{g/kg}$ for DON in corn.

Model performance for DON on wheat—using fluorescence spectra at 395nm as well as UV–VIS, NIR reflectance data—is shown in Figure 6.3. All spectral data was preprocessed by area normalization and smoothing to 10nm resolution. The top row, Plots A–C, illustrate classification accuracy on the training data, while the bottom row, Plots D–F, present performance on the new validation samples. Plots A and D show the result for the Fluorescence data, Plots B and E for UV-VIS and Plots C and F for NIR data. The model consistently performs really well for the training data (data seen during training), but fails to perform well on the validation samples (data not seen during training).

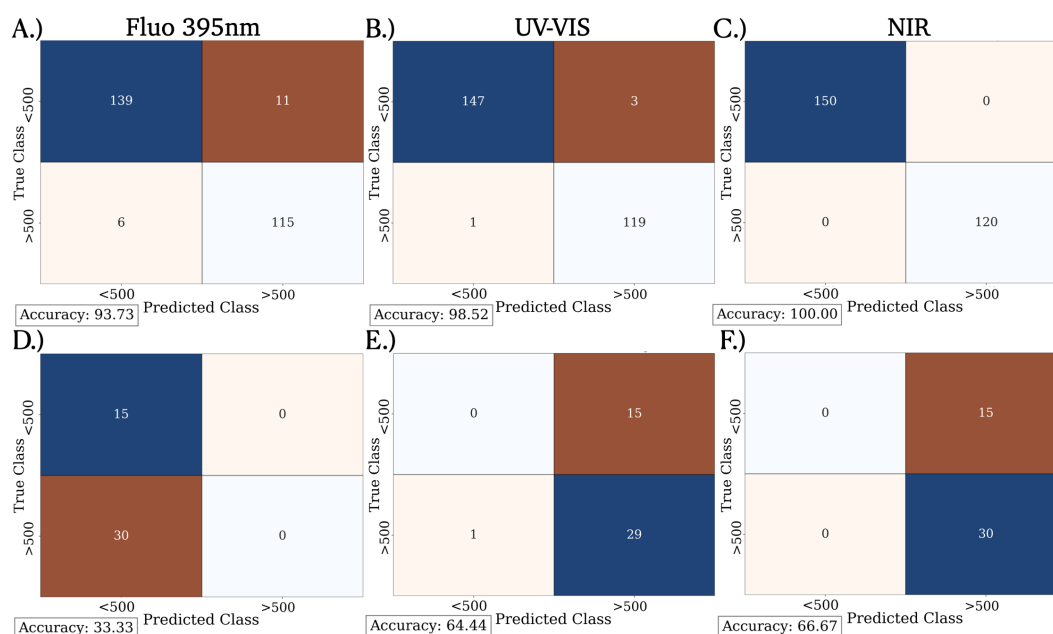


Figure 6.3: Model performance for DON classification in wheat using full-spectrum data. Top row (A–C): Predictions on training data using A.) fluorescence spectra (395nm excitation), B.) UV–VIS reflectance, and C.) NIR reflectance. Bottom row (D–F): Predictions on external validation samples for the same modalities.

Reflectance-based classification results indicated that all validation samples were identified as exceeding the $500\mu\text{g/kg}$ threshold, despite one sample containing only $400\mu\text{g/kg}$. In contrast, classification based on the fluorescence data tends to classify all samples as having a DON contamination below $500\mu\text{g/kg}$. This poor performance may be due to the fact that the model was trained to detect a combination of DON, OTA, and T-2/HT-2, whereas the validation samples contain only DON. As a consequence, some of the spectral markers the model relied on during training may be absent, limiting its ability to detect DON in isolation.

The observation that all fluorescence-based predictions fall below the threshold may, for example, suggest that the model heavily depends on OTA-associated spectral features—features that are not present in the new, pure DON validation samples, and hence all samples are classified as in the low concentration class.

Another reason for the consistently poor performance might be the limited natural variability in the training dataset, which was composed by mixing only a small number of reference materials. This narrow range of matrix compositions restricts the model's exposure to the diversity encountered in real-world samples. As a result, the model may struggle to accurately recognize DON in sample matrices that deviate significantly from those it was trained on. This would suggest that, at present, the model lacks the necessary robustness and has not been trained on a sufficiently diverse dataset to support reliable detection under naturally variable conditions. Consequently, for the current model to achieve broader applicability, re-training would likely be required for different wheat varieties from distinct geographical regions.

A final possible explanation is that the model, when using UV–VIS and NIR data, may in fact be classifying the samples correctly. Notably, both high-contamination samples were accurately identified. The apparent misclassification of the low-contamination sample could be due to its actual DON concentration exceeding the $500\mu\text{g}/\text{kg}$ threshold. This is possible given the sample concentration of $(500 \pm 100)\mu\text{g}/\text{kg}$, which is only one std from the threshold. And of course there is also the uncertainty of the training samples themselves, adding additional uncertainty on the respective true concentrations of the training and validation samples.

Similarly, we implemented classification models for DON detection on corn. The results are presented in Figure 6.4, where the top row, Plots A–C, once again show the performance on the training set, and the bottom Plots D–F, shows the performance on the external validation set.

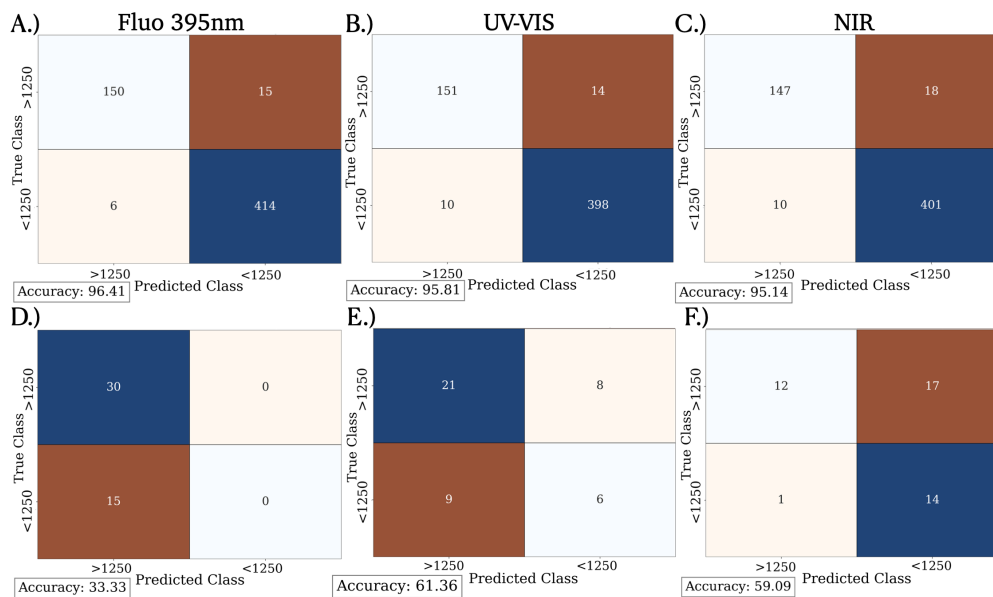


Figure 6.4: Model performance for DON classification in corn using full-spectral data. Top row (A–C): Predictions on training data using A.) fluorescence spectra (395nm excitation), B.) UV–VIS reflectance, and C.) NIR reflectance. Bottom row (D–F): Predictions on external validation samples for the same modalities. As with wheat, performance is strong on the training data but poor on unseen samples.

As was the case for wheat, the model performs well on the training data, but fails to generalize effectively to the unseen, unrelated validation samples.

When using fluorescence data, all validation samples are incorrectly classified as exceeding the $1250\mu\text{g}/\text{kg}$ threshold, even though the lowest concentration sample contains only $500\mu\text{g}/\text{kg}$ DON. In this case, we cannot attribute the misclassification to sample concentration uncertainty. Once again, it is possible that the model—trained on a sample series composed of only four standard materials—does not sufficiently capture the natural variability of the matrix, resulting in poor generalization to entirely unseen samples.

Considering the corn samples, the model was trained to distinguish DON from other co-occurring toxins such as AFLA, ZEA, and FUM. In the training dataset, all samples were mixtures containing all four toxins, and no blank matrix was available. As a result, decreasing the concentration of DON inherently increased the relative proportions of the other toxins. Consequently, the model may have learned to infer DON presence indirectly—by detecting the relative absence of AFLA, ZEA, and FUM—rather than identifying DON-specific spectral features. This would limit its effectiveness in identifying DON when it appears in isolation as is the case for these validation samples. This hypothesis aligns with the observed misclassifications, where all validation samples are incorrectly classified as exceeding the DON threshold. This aligns with the fact that in the original training data, high-DON samples consistently lacked significant levels of the other toxins—just like the new pure-DON validation samples—further supporting the idea that the model may be recognizing toxin composition patterns rather than DON directly.

For the UV–VIS and NIR reflectance data, model performance was slightly better; however, several misclassifications still occurred. Overall, these results suggest that models trained solely on the original dataset struggle to generalize effectively to the new, independent samples. This may indicate that the current models do not yet sufficiently capture the natural variability present in real-world samples.

One potential strategy to enhance model robustness and reduce sensitivity to matrix variability is to restrict the model to only the most informative spectral features. In this context, we limit the spectral range used for classification to a small subset of wavelengths, thereby minimizing the influence of uninformative spectral regions on model performance.

Accordingly, the following section briefly evaluates classifier performance when constrained to the ten most important wavelengths, as previously identified through Sequential Feature Selection (SFS) in Chapter 4.

6.4 Feature Based Classifiers

Next, models were trained using the training dataset, now employing only the 10 most informative spectral features, selected by SFS. By restricting the spectral input to a limited number of highly informative wavelengths, the aim was to reduce model sensitivity to background or matrix-related variations present in non-informative spectral regions. The performance of these reduced-input models was then assessed on the validation data, as shown in Figure 6.5.

For wheat Figure 6.5, Plots A–C, model performance remains unchanged compared to the full-spectrum classifiers shown in Figure 6.3. This could indicate that background variability may not be the primary confounding factor, as reducing the spectral range does not improve DON classification.

Instead, the performance limitations may arise from the presence of co-occurring toxins, as well as residual effects of natural variation—particularly differences in chlorophyll content. Additionally, potential discrepancies between the actual DON concentrations and the certified reference values in the validation and training samples may also contribute.

For corn, the results obtained using the reduced feature models are shown in Figure 6.5, Plots *D–F*. Compared to the full-spectrum models in Figure 6.4, no improvement was observed when using fluorescence or UV–VIS data. However, for the NIR data, performance did improve: both the high and low contamination samples were correctly classified.

However, it should be noted that here the threshold-level sample at $1250\mu\text{g}/\text{kg}$ was excluded from this evaluation, and thus only two validation samples were available for this analysis, which limits the reliability and statistical significance of these results.

Additionally, we investigated whether merging the top 10 features from all fluorescence excitations and reflectance modalities could further improve DON sensing in corn. The resulting model performed similarly to the NIR-only model, correctly classifying the same two samples. This suggests that the merged-spectrum model is also promising. Nevertheless, due to the limited number of available validation samples, definitive conclusions cannot yet be drawn. More extensive validation with a larger and more diverse sample set will be necessary to robustly assess the model’s ability to detect DON in corn under realistic conditions.

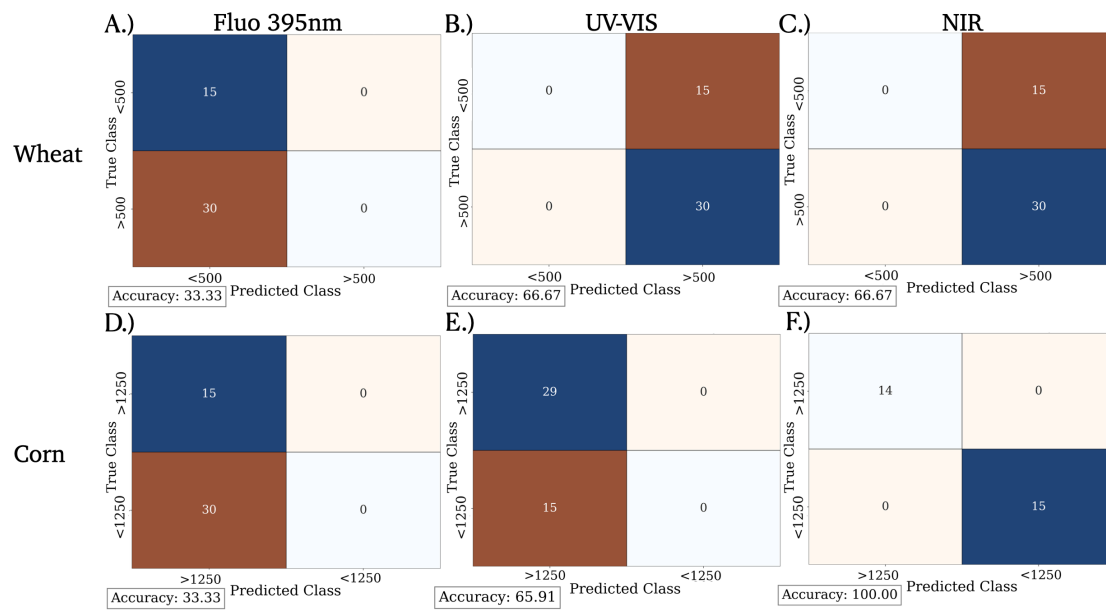


Figure 6.5: Model performance for DON classification on wheat and corn using only 10 features. The top row (*A–C*), shows the results for wheat, the bottom row (*D–F*) shows the results for corn. We observe no improvement for wheat, and for corn performance only improves using NIR data.

6.5 Conclusion On Model Generalizability

This chapter assessed the generalization capacity of the developed classification models by validating them on a newly created set of independent wheat and corn samples. We found that full-spectrum models yielded poor performance on these validation samples, an outcome that may be explained by several factors.

First, the limited diversity of the training data, which was composed from only two wheat and four corn standards, likely constrained the model's exposure to natural matrix variability. As a result, the models may have learned representations that do not extend well to samples with different compositional profiles. Second, the wheat training data included only samples contaminated with multiple toxins—namely DON, OTA, and T-2/HT-2—whereas the validation samples contained only DON. This discrepancy could imply that the models were relying, at least partially, on spectral features associated with co-occurring toxins. Third, uncertainty in the certified DON concentrations of the validation and training samples introduces ambiguity in evaluating the correctness of the model's predictions, as true concentrations can deviate from the assigned labels during both training and validation.

Similar issues were observed for corn. The full-spectrum models again failed to classify the new validation samples accurately, even though these samples were clearly differentiated by DON levels.

Here, a notable factor is the way the training data was constructed. Since no blank matrix was available, lowering the DON concentration in training samples required increasing the concentration of other toxins (AFLA, ZEA, FUM). This may have led the model to associate the presence of DON with the absence of these other toxins, rather than learning direct spectral signatures of DON itself. As a result, the model failed to detect DON when it appeared alone in the new samples, at least when considering the full spectral ranges.

To address the natural variation, and minimize its influence on the spectral classification algorithm, the evaluation was repeated after application of a feature selection step, limiting the considered spectral range to the top 10 wavelengths. Reducing the number of input features did not improve the classification performance for the wheat samples. However, for corn, a marked performance improvement was observed when using the 10 key features selected via SFS from the NIR data. This finding suggests that constraining the spectral input to the most relevant wavelengths may enhance model robustness by minimizing background interference and mitigating the effects of natural matrix variability. While these last results are promising, the limited number of validation samples—especially for corn—makes it difficult to draw definitive conclusions.

In summary, these findings demonstrate that the current models, though effective under cross-validation, do not yet generalize reliably to fully independent sample sets. This limitation is likely due to insufficient training diversity, confounding effects from co-contaminants, and uncertainties in the certified concentrations of the reference materials. While feature selection and data fusion approaches offer some improvement—particularly for corn—further progress will require the inclusion of more diverse training data encompassing a wider range of sample matrices, blank references, and isolated toxin conditions. This extended dataset appears to be required to enable building models capable of robust, real-world mycotoxin detection across naturally variable sample compositions.

Chapter 7

Concluding Remarks & Future Perspectives

This thesis set out to address a critical challenge in agrifood safety: the non-destructive detection of multiple mycotoxins in staple crops using optical spectroscopy and machine learning. Conventional chemical methods, while sensitive and well-established, are destructive, costly, and unsuitable for real-time or high-throughput screening. In contrast, non-invasive spectroscopic platforms—combined with data-driven predictive models—offer a promising pathway toward scalable, in-line detection.

Accordingly, this study pursued three core objectives. First, the development of classification models for the in-line sorting of samples according to regulatory thresholds. Second, the construction of regression models to enable more granular, quantitative assessment—particularly valuable for resolving borderline cases or tracking contamination trends over time. Third, the application of dimensionality reduction techniques to transition from full-spectrum models to those relying on a discrete subset of highly informative spectral features, thereby reducing system complexity and enhancing industrial applicability.

These objectives were addressed through a systematic pipeline encompassing sample synthesis, data acquisition, spectroscopic platform evaluation, preprocessing, model training, feature selection, and rigorous performance benchmarking using validation schemes designed to promote generalizability.

7.1 Summary Of Results

This thesis delivered several key contributions that jointly advance the state-of-the-art in non-destructive mycotoxin detection. First, we developed classification models for the detection of the most common and harmful mycotoxins on corn and wheat in multi-toxin settings. For this purpose an extensive model comparison was done—including Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), and Artificial Neural Network (ANN)—across both fluorescence (at various excitation wavelengths) and UV-VIS-NIR reflectance modalities. A critical innovation was the adoption of Leave-One-Sample-Out Cross-Validation (LOSO-CV), ensuring that model performance reflected generalizability to unseen samples—an essential condition for real-world deployment.

From the model benchmarking, LDA consistently emerged as the most effective classifier when considering the trade-off between accuracy and computational efficiency. In binary classification tasks, the LDA model achieved up to 95% accuracy for deoxynivalenol (DON) detection in corn under multi-toxin conditions using fluorescence data excited at 395nm. Similarly, high classification accuracies were obtained for other toxins: 89% for aflatoxin (AFLA) and zearalenone (ZEA), and 91% for fumonisin (FUM)—all based on thresholds set well below the respective regulatory limits. In wheat, we achieved classification accuracies of up to 99% using full-range UV–VIS–NIR reflectance data for the detection of a mixture of DON, ochratoxin A (OTA), and T-2/HT-2. However, due to the simplified sample synthesis—based on only two reference standards—it was not possible to disentangle the spectral contributions of individual toxins. This introduces ambiguity as to whether the models detected true toxin-specific markers or merely spectral characteristics of the standards themselves.

Beyond matrix-specific detection on only corn or wheat, we also investigated the feasibility of cross-matrix models for DON detection. An LDA model trained on both corn and wheat samples achieved 82% accuracy in binary classification using a $500\mu\text{g}/\text{kg}$ threshold, highlighting the potential for unified detection frameworks capable of operating across heterogeneous matrices. Finally, we explored whether merging data from different spectroscopic modalities—such as combining reflectance and fluorescence data or fluorescence data from multiple excitation wavelengths—could enhance classification performance. Within the considered test case, the results showed limited improvement when considering multi-modal data fusion.

A second major contribution of this work was the development of regression models for mycotoxin quantification in multi-toxin environments, specifically targeting contamination ranges at or below regulatory limits. The objective was to assess whether the required levels of sensitivity and specificity could be met for regulatory-compliant detection. A range of predictive models—including Partial Least Squares Regression (PLSR), Random Forest Regression (RFR), Least Absolute Shrinkage and Selection Operator (Lasso), Elastic Net Regression (Elastic Net), and Support Vector Regression (SVR)—were evaluated.

PLSR models emerged as the best performing models, particularly when using 9 latent components. For wheat the best model performance was found employing the NIR reflectance data, achieving an R^2 of 0.94 and an Root Mean Squared Error (RMSE) of just $72.1\mu\text{g}/\text{kg}$, in predicting the combined presence of DON, OTA, and T-2/HT-2. The corresponding Limit Of Detection (LOD) and Limit Of Quantification (LOQ), w.r.t. DON, were calculated as $231.9\mu\text{g}/\text{kg}$ and $702.9\mu\text{g}/\text{kg}$, respectively—both well below the legal threshold. Note however that this model technically detects the combined effects of the four toxins, and not just DON.

For corn, regression modeling focused specifically on DON detection within a multi-toxin environment. Using fluorescence data excited at 395nm, the best-performing PLSR model achieved an R^2 of 0.83 and an RMSE of $295.6\mu\text{g}/\text{kg}$, with a LOD of $1017.4\mu\text{g}/\text{kg}$, again demonstrating sub-threshold sensitivity and confirming the feasibility of quantitative detection under realistic, spectrally complex conditions. Regression was also realized using a RFR model on a cross-platform basis, for both corn and wheat realizing performance on par with the results for corn alone.

A final key contribution of this work was the integration of algorithmic feature selection techniques in the modeling pipeline for complexity reduction. Both Sequential Feature Selection (SFS) and Minimum Redundancy Maximum Relevance (MRMR) were evaluated as dimensionality reduction strategies. While MRMR proved to be computationally efficient, it was significantly outperformed by SFS in scenarios where only a limited number of features were to be selected.

Notably, using just 7–10 spectral bands (post-smoothing), selected via SFS, we were able to preserve predictive performance comparable to that of full-spectrum models, while significantly simplifying system complexity. This finding highlights the potential for constructing efficient, low-complexity systems suitable for real-time mycotoxin detection in industrial (food sorting) applications.

Table 7.1 summarizes the performance of the best-performing classification and regression models, using either the full spectral input or only the top 10 most informative features selected via SFS, for both the corn and wheat matrices separately as well as in combination.

Classification Accuracy			
Matrix and Data	Toxin	Full Spectrum [%]	Top 10 SFS [%]
Corn, Fluo 395 nm	AFLA	73	–
	DON	85	83
	ZEA	77	75
	FUM	72	71
Wheat, NIR	DON, OTA, T-2/HT-2	96	91
Corn + Wheat, Fluo 395 nm	DON	83	–
Regression R^2			
Matrix and Data	Toxin	Full Spectrum [%]	Top 10 SFS [%]
Corn, Fluo 395 nm	DON	83	79
Wheat, NIR	DON, OTA, T-2/HT-2	96	95
Corn + Wheat, Fluo 365 nm	DON	82	71

Table 7.1: Performance of the best-performing classification (3 classes, thresholds in Table 4.3, LDA) and regression (PLSR) models for corn and wheat, using both full-spectrum input and the top 10 features selected via SFS.

Summarized, this thesis advances the non-destructive optical detection of harmful mycotoxins by demonstrating unprecedented performance in multi-mycotoxin sensing. Through the synergistic integration of optical spectroscopy and machine learning, the proposed platform achieves classification and quantification accuracies compliant with European regulatory limits, paving the way toward scalable, real-time food safety monitoring.

Key contributions of this work include:

- **Pioneering multi-mycotoxin detection on staple crops:** To the best of our knowledge, this thesis presents the first successful implementation of a multi-mycotoxin classification framework that achieves high accuracies on both corn and wheat matrices, including co-contaminated samples.
- **Development of advanced spectral data processing pipelines:** Novel strategies were introduced to combine complementary spectroscopic modalities—fluorescence and reflectance—enabling performance beyond the state-of-the-art, which traditionally focuses on single-modality, single-matrix detection.

- **Demonstration of Sequential Feature Selection (SFS) for robust data reduction:** This thesis validates SFS as an effective technique to reduce spectral dimensionality to as few as 10 wavelengths without significant performance loss. By reducing complexity without sacrificing accuracy, the proposed methods align with the practical constraints of in-line and in-field implementation.
- **A modular and scalable detection architecture:** The framework supports both classification and regression across individual and combined matrices and for a variety of toxins in co-occurrence, enabling flexible integration for food safety screening in diverse industrial contexts.

7.2 Limitations And Future Perspectives

While this thesis has demonstrated the feasibility of using optical spectroscopy and machine learning to detect mycotoxins at trace concentrations—using both classification and regression models in multi-toxin environments—several limitations remain that constrain its generalizability and industrial applicability.

First and foremost, the study relied on a limited number of reference standards, featuring limited natural variability. In the case of wheat, only two standards were used, which did not allow for decoupling the spectral contribution of individual mycotoxins. As a result, all wheat-based models effectively learned to detect a fixed mixture of DON, OTA, and T-2/HT-2, providing no insight into the model’s ability to identify any toxin in isolation. For corn, a more diverse dataset was constructed using four independent standards, which included AFLA, DON, and ZEA from two different source materials. This allowed for a more robust experimental design, wherein the concentrations of individual mycotoxins could be varied independently of other toxins, and decoupled from the specific matrix composition of an individual standard. These results therefore offer higher confidence that the model detected toxin-specific features, rather than relying on confounding effects from co-occurring toxins or matrix-specific spectral patterns.

However, even in this improved setup, we remained limited by the small number of available standards. For instance, while the model could distinguish between DON and other toxins, we could not test its ability to differentiate DON contamination from an uncontaminated matrix without the presence of other co-contaminants—a significant limitation when considering practical deployment.

Future studies should prioritize the construction of sample series based on a larger and more diverse set of independent reference materials. Rather than generating samples through mixing, it is recommended to source naturally contaminated samples from different geographic regions to increase natural variability and reduce the risk of model overfitting to matrix-specific markers like chlorophyll content or moisture level. Expanding the diversity of the training data in this way would not only improve generalizability but also make models more resilient to unseen data, which is essential for real-world applications.

However, obtaining reliable naturally contaminated samples is difficult. Only a limited number of homogeneous, well-characterized samples are available, as contamination in nature is typically invisible and highly localized. Moreover, current chemical analysis methods such as Liquid Chromatography-Tandem Mass Spectroscopy (LC-MS/MS) require destructive sampling, which renders those samples unsuitable for subsequent optical measurements—further complicating the collection of valid ground-truth data for spectroscopy-based model development.

Another critical step for industry-aligned development is to move beyond ground grain samples toward individual kernel-level detection. In practical applications—such as in-line sorting or handheld scanners—individual kernels are the unit of analysis, not homogenized powders. However, this presents a major challenge: chemical methods like LC-MS/MS can only determine batch-level contamination and are unable to provide ground truth for individual kernels, due to the heterogeneous distribution of mycotoxins. To overcome this, future work should aim to collect large batches of naturally contaminated grains with known batch-wide concentrations, then scan individual kernels to develop probabilistic models that can predict contamination levels at the single-kernel level based on population statistics.

In addition to expanding datasets and moving to kernel based samples, the validation strategy also requires refinement. While this work employed a rigorous LOSO-CV approach to mitigate overfitting and simulate unseen data conditions in a data constrained environment, a more realistic benchmark for industrial deployment would involve testing models on entirely independent validation sets composed of unrelated samples—not derived from mixing known standards. This would more accurately reflect operational variability and improve confidence in real-world model robustness.

A major technical contribution of this thesis was the implementation of feature selection strategies aimed at reducing model complexity. SFS-based models retained predictive performance comparable to full-spectrum models. This reduction in dimensionality is a promising step toward hardware-efficient, industry-ready solutions—particularly in real-time sorting systems with strict latency and memory constraints.

Future improvements in this area could involve hardware-informed feature selection pipelines, where only wavelengths that are easily implementable in low-cost optical systems are considered. For instance, avoiding regions like the chlorophyll fluorescence peak (670nm) may be advisable if, in future work with independent samples, this region is found to correlate more with matrix composition than mycotoxin concentration. Similarly, selecting spectral bands with sufficient spacing could enable simpler optical filtering and multi-channel detection architectures. Likewise, excitation sources for fluorescence should be restricted to wavelengths compatible with cost-effective, compact light sources such as LEDs or low-power diode lasers. Our findings already indicate that fluorescence excitation at 395–385nm consistently yields the strongest results, making it a compelling candidate for future industrial implementation.

In summary, while the current study has laid a strong foundation by demonstrating that compact models based on selected spectral features can achieve regulatory-grade detection accuracy, significant work remains to transition from laboratory proof-of-concept to a deployable, industry-ready solution. This includes expanding sample diversity, shifting toward whole-kernel analysis, improving validation strategies with truly independent datasets, and optimizing feature selection with hardware constraints in mind. With these steps, the development of robust, real-time, and cost-efficient systems for mycotoxin detection becomes a realistic objective.

Appendix A

Theoretical Background On ML For Mycotoxin Detection

This appendix provides a theoretical overview of the machine learning and chemometric methods commonly used for spectroscopic mycotoxin detection throughout this work. While the main body of this thesis focuses on the practical implementation and evaluation of these models, the following sections offer foundational background on key concepts—including data preprocessing, mathematics behind the models, training procedures, validation strategies, feature selection techniques, and the key performance metrics. This material is intended to support readers less familiar with the underlying algorithms and to clarify the methodological choices made throughout the experimental chapters.

A.1 Spectral Data And Preprocessing Steps

The initial step in building both regression and classification models is to collect spectral data from samples with known target compositions. For our study, we obtained spectral data from corn and wheat samples with verified mycotoxin levels for AFLA, DON, ZEA, FUM, T-2, HT-2, and OTA. A detailed description of the sample preparation and spectral measurements conducted is provided in Section 3.1.

Computationally, each spectrum, x_i , is represented as $\{x_{i_1}, x_{i_2}, \dots, x_{i_n}\}$, where n denotes the number of features (or wavelengths) in the spectrum. By concatenation this representation can be expanded to encompass data from multiple spectroscopic platforms. Each spectrum is paired with a label y_i that specifies the known mycotoxin concentration or class for regression and classification models respectively. For binary classification, y_i is set to either 0 or 1, indicating whether the sample falls below or exceeds a predetermined threshold—classifying it as safe or unsafe. More complex classifications are also possible; for instance, a three-class model could label samples as safe, risky, or unsafe.

All spectral data combined is represented by a matrix X , with dimensions $l \times n$, where l corresponds to the number of spectra in the dataset. The labels y_i are stored in a vector Y with l entries.

Before proceeding with model training, it is crucial to apply preprocessing techniques to enhance the quality of raw spectral data. Spectroscopic measurements are susceptible to unwanted

variations, including background noise, scattering effects (in fluorescence measurements), unintended fluorescence (in reflectance measurements), calibration errors, and environmental influences. Effective preprocessing mitigates these variations while preserving the key spectral features necessary for mycotoxin detection, thereby significantly boosting model robustness and accuracy [30].

Several preprocessing techniques are commonly employed to enhance spectral data quality. Basic methods include trimming the spectra to focus on the most informative wavelength range, smoothing to reduce noise, and normalizing to correct for variability in light sources and sensors [31]. More advanced algorithms, such as Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV), address issues like light scattering, noise, and baseline shifts [108]. Techniques like Savitzky-Golay filtering can compute first- and second-order derivatives to accentuate subtle variations in absorbance or fluorescence while minimizing baseline drift [125].

An additional preprocessing strategy involves computing the ratios of spectral intensities across different wavelength combinations. This approach identifies the most informative ratios and effectively serves as an alternative normalization method, compensating for variations in illumination power, sample-to-detector distance, and other factors when only a discrete subset of wavelengths is available [33].

Dimensionality reduction methods—including Principal Component Analysis (PCA), Minimal Redundancy Maximal Relevance (MRMR), and Sequential Feature Search (SFS)—further refine the data by selecting a subset of key spectral features, thereby decreasing dimensionality, accelerating training and validation, and reducing sensitivity to noise. These feature selection approaches are discussed in detail in Section A.3.

Since the optimal preprocessing technique is not known a priori, it is standard practice to experiment with multiple methods and their combinations, ultimately selecting the preprocessing steps that yield the best model performance for the specific implementation. For our application, it is important to note that we cannot employ all of the advanced preprocessing techniques mentioned previously, as our goal is to develop a model that operates using only a discrete subset of spectral features, selected via a feature selection scheme. Consequently, methods that require access to the full spectral range for correction—such as Savitzky-Golay filtering, MSC, or SNV—are not applicable. In this work, we therefore restrict the data preprocessing techniques to range trimming, smoothing, normalization, and ratios, each of which is discussed in more detail the following subsections.

A.1.1 Trimming

Spectral trimming is a commonly employed preprocessing technique that simplifies analysis by restricting the spectral range to exclude noise-dominated regions. This process involves mathematically removing uninformative or unstable segments from each spectrum x_i , thereby reducing dimensionality and enhancing overall signal quality.

Typically, the edges of the spectral range are the most unstable due to limitations in the light source and reduced spectrometer sensitivity at extreme wavelengths. These edge regions often exhibit elevated noise levels and decreased signal reliability, which can adversely affect model performance. By eliminating these unstable sections, spectral trimming ensures data consistency and focuses the analysis on the most relevant and informative spectral features.

A.1.2 Smoothing

A second commonly applied preprocessing technique is data smoothing, which aims to reduce noise and enhance spectral clarity by lowering the spectral resolution. Mathematically, this involves modifying each measured spectrum by either removing certain spectral features or aggregating adjacent features through stepwise averaging, thereby reducing the total number of features considered, and removing high frequency noise.

Formally, given a spectrum, x_i , such smoothing is performed by grouping consecutive features into bins and computing their mean:

$$\tilde{x}_{i_j} = \frac{1}{k} \sum_{m=(j-1)k+1}^{jk} x_{i_m}, \quad j = 1, 2, \dots, \frac{n}{k}, \quad (\text{A.1})$$

where x_{i_m} is the m -th feature of the spectrum, k represents the bin size (i.e., the number of original features averaged together), and $\tilde{x}_{i_j}$ is the j th feature in the resulting smoothed spectrum with a reduced number of features (n/k).

By applying data smoothing, we remove high-frequency noise and improve the signal-to-noise ratio, which can lead to more robust model performance. Additionally when combining various spectral techniques we can restrict the resolution of each platform to the same value, to work with equal resolution, allowing for direct comparisons between spectroscopic platforms.

A.1.3 Normalization

A third widely used preprocessing method is normalization, which aims to standardize spectral intensity across all spectra. Several normalization techniques exist, and this study will compare their effectiveness.

A commonly used approach is max-normalization, where each spectrum is divided by its maximum intensity value, ensuring that the highest peak is scaled to 1. Mathematically, for a given spectrum x_i , max-normalization is defined as:

$$\tilde{x}_i = \frac{x_i}{\max(x_i)} \quad (\text{A.2})$$

where $\tilde{x}_i$ represents the normalized spectrum, and $\max(x_i)$ is the maximum intensity value within x_i . The advantage of max-normalization is that it scales all spectra to have the same peak intensity, preventing models—particularly neural networks—from learning biases related to absolute spectral magnitude. Without normalization, models may overemphasize spectra with higher absolute intensities relative to those with lower intensities, resulting in biased training and predictions.

However, a notable drawback of this method is its sensitivity to noise. If a spectrum contains a noisy peak, max-normalization will scale the entire spectrum based on that outlier, potentially distorting the relative intensities of relevant spectral features. This can make affected spectra unrecognizable to the model, introducing model errors. To mitigate this, max-normalization can be used in combination with smoothing to reduce the effect of noise. This will however reduce the spectral resolution, and might lose valuable information, so alternatively a range of other normalization schemes can be used.

One option is to use normalization by division of a constant value, V . This constant can be

a randomly chosen value, or for example one can use the mean of the maximal value of all considered spectra. mathematically we get

$$\tilde{x}_i = \frac{x_i}{V}. \quad (\text{A.3})$$

Another widely used normalization technique is area-normalization, where each spectrum is normalized by division by the total integrated area under the curve. This method ensures that all spectra have the same overall intensity, making it particularly useful for compensating for variations in illumination conditions. Mathematically, given a spectrum x_i , area-normalization is computed as:

$$\tilde{x}_i = \frac{x_i}{\sum_{j=1}^n x_{i_j}}, \quad (\text{A.4})$$

where $\sum_{j=1}^n x_{i_j}$ represents the total sum of all spectral intensity values within the spectrum. The resulting normalized spectrum $\tilde{x}_i$ has an area of 1, meaning that the overall spectral energy is conserved across all spectra.

As evident from the formula, this normalization approach is not limited to continuous spectra; it can also be effectively applied to discrete spectral datasets, where only a finite number of spectral features are measured. This makes area-normalization a versatile method suitable for a wide range of spectroscopic applications, including those with sparse or low-resolution data.

A.1.4 Ratios

An alternative approach, particularly well suited for applications using discrete wavelengths, is ratio-based normalization. Instead of normalizing intensities based on absolute values, this method computes intensity ratios between all possible wavelength pairs, effectively capturing relative spectral variations rather than absolute intensity differences. This approach serves as a discrete alternative to traditional normalization methods while also expanding the feature space by incorporating additional spectral relationships into the model.

Mathematically, given a spectrum $x_i = \{x_{i_1}, x_{i_2}, \dots, x_{i_n}\}$, the set of all possible intensity ratios is computed as:

$$r_{i,j,k} = \frac{x_{i_j}}{x_{i_k}}, \quad \forall j, k \in \{1, 2, \dots, n\}, j \neq k, \quad (\text{A.5})$$

where $r_{i,j,k}$ represents the ratio between the intensity at wavelength j and wavelength k for spectrum x_i . This transformation generates a new feature space consisting of $n(n-1)/2$ unique ratio features, significantly increasing the dimensionality of the dataset.

A.2 Machine Learning Models

In the realm of spectral analyses for mycotoxin detection, two powerful methodologies stand out: chemometrics and machine learning. Chemometrics employ mathematical and statistical techniques to extract meaningful information from complex spectral data. This approach excels at managing high-dimensional, noisy, and correlated datasets, facilitating the interpretation, classification, and quantification of chemical properties. In contrast, Machine Learning (ML) develops data-driven algorithms that identify patterns within spectral data to make accurate predictions, particularly excelling at capturing complex, nonlinear relationships for both classification and regression tasks. Both chemometric and ML methods can be classified as supervised or unsupervised, depending on whether or not labeled training data is used.

In unsupervised learning, the model is provided with only the input data X and must identify underlying patterns or structures without reference to predefined labels (Y). Common unsupervised techniques include clustering algorithms—such as k-means, and hierarchical clustering—which group similar samples together, and dimensionality reduction methods like Principal Component Analysis (PCA), which transform the data into a lower-dimensional space while preserving its most informative features [35]. Conversely, supervised learning trains a model using both the input data X and the corresponding output labels Y . By learning the mapping between X and Y , the model becomes capable of making accurate predictions on new, unseen data [108].

Supervised learning itself is further subdivided into classifiers and regression models. The distinction between the two lies in the nature of their predictive outcome. Classification models predict discrete categorical labels, assigning an observation to a predefined class, while regression models predict continuous numerical values, like concentrations. Mathematically, given a dataset consisting of spectral measurements, x_i , a classification model learns a mapping function:

$$f : x_i \rightarrow \{C_1, C_2, \dots, C_k\}, \quad (\text{A.6})$$

where C_k represents a finite number of class labels, such as contamination levels or categories as "safe", "risky", or "unsafe". In contrast, a regression model estimates a continuous function:

$$f : x_i \rightarrow y_i, \quad (\text{A.7})$$

where y_i represents the concentration.

From a food safety perspective, classification is often more practical than regression because regulatory guidelines typically specify strict contamination thresholds, making it sufficient to determine whether a sample falls below or above these limits [125]. Classification models effectively categorize samples as safe or unsafe based on these predefined thresholds, facilitating rapid and automated screening. These models, furthermore, require less training data because a simplified mapping to discrete categories is learned, whereas regression models must approximate a continuous function and are more vulnerable to background noise which can compromise their robustness in real-world applications [125]. However, regression models offer the advantage of providing quantitative information rather than merely qualitative or semi-quantitative results, yielding detailed estimates of contamination levels that offer greater insight into the specific degree of contamination and provide increased certainty regarding sample safety—a critical benefit for risk assessment and informed decision-making.

Given the potential and benefits of both model types, the following sections will highlight the best-performing methodologies from the literature, covering both classification and regression platforms.

A.2.1 Classification Models

Starting with classification models is often advantageous because they require less training data due to their simplified mapping to discrete categories. Since regulatory compliance generally hinges on determining whether a sample falls above or below a specific threshold, even a simple two-class model can suffice. Consequently, classification models are not only easier to implement but also provide a solid foundation for meeting regulatory standards, which explains the current research focus in this area.

As summarized in Table 2.09, a wide variety of classification algorithms have been explored with varying degrees of success. While many studies evaluate a single model, others compare multiple classifiers to identify the optimal solution, an approach that mitigates the risk of selecting a suboptimal model. In this study, we adopt this comparative strategy to evaluate and identify the most accurate, robust, and practical classification techniques for mycotoxin detection.

Among the models employed in literature are techniques such as Linear and Quadratic Discriminant Analysis (LDA & QDA) [30, 103], Linear and Quadratic Support Vector Machines (LSVM & QSVM) [105], K-Nearest Neighbors (KNN) [104], and Partial Least Squares Discriminant Analyses (PLS-DA) [100]. In addition, more advanced machine learning techniques—such as Decision Trees (DT) [107], Random Forests (RF) [95], Artificial Neural Networks (ANN) [99], Deep Neural Networks (DNN) [99]—have been explored, though they are less frequently successfully applied in the literature.

In the following subsections a concise discussion on each of these classifiers is given, presenting their working principles, advantages and limitations. This theoretical discussion is based on the book *"The Elements of Statistical Learning: Data Mining, Inference, and Prediction"* by Hastie et al. [141], course notes: *"AI for chemical Engineering"* by Jaewook Nam thought at Seoul National University, and the prior thesis work of Indy Magnus: *"Combining Spectroscopy And Machine Learning Processing For Precise Sample Classification"* [125].

A.2.1.1 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a Bayesian classification algorithm based on Bayes' theorem [125]. It assumes that the probability distribution of each class follows a multivariate Gaussian distribution with a class-specific mean and a shared covariance matrix. The goal of LDA is to find a linear decision boundary that maximizes class separability while minimizing variance within each class. Mathematically, for a given input feature vector x_i (spectrum), the posterior probability of class C_k is given by Bayes' theorem:

$$P(C_k|x_i) = \frac{P(x_i|C_k)P(C_k)}{P(x_i)}, \quad (\text{A.8})$$

where: $P(C_k|x_i)$ is the posterior probability of class C_k given the observation x_i , $P(x_i|C_k)$ is the likelihood, representing the probability of observing x_i given class C_k , $P(C_k)$ is the prior probability of class C_k , and $P(x_i)$ is the marginal probability of x_i across all classes. Assuming that each class follows a Gaussian distribution, the likelihood function can be expressed as:

$$P(x_i|C_k) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x_i - \mu_k)^T \Sigma^{-1}(x_i - \mu_k)\right), \quad (\text{A.9})$$

where μ_k is the mean vector of class C_k , Σ is the shared covariance matrix across all classes, and d is the dimensionality of the input space.

Since the covariance matrix is assumed to be the same for all classes, the decision boundary between any two classes is linear, leading to a linear discriminant function:

$$\delta_k(x_i) = x_i^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log P(C_k) \quad (\text{A.10})$$

A new sample x_i is assigned to the class C_k that maximizes the discriminant function $\delta_k(x_i)$, resulting in a linear decision boundary that separates different class regions.

A key advantage of LDA is its closed-form analytical solution, which makes it computationally efficient [125]. Besides standard LDA, there is also Regularized LDA (RLDA). RLDA extends the standard approach by introducing a regularization parameter to stabilize the covariance estimates, particularly when the number of features is high relative to the number of samples. This regularization blends the within-class scatter matrix with a scaled identity matrix, preventing overfitting and ensuring that the covariance matrix remains well-conditioned. As a result, RLDA can offer improved generalization and robustness, especially in situations with highly correlated features or limited data.

A.2.1.2 Quadratic Discriminant Analysis (QDA)

Quadratic Discriminant Analysis (QDA) extends Linear Discriminant Analysis by relaxing the assumption that all classes share the same covariance matrix. While LDA assumes a common covariance structure across all classes, QDA allows each class C_k to have its own covariance matrix Σ_k , making it better suited for non-linearly separable data.

Given the same Bayesian framework as LDA, QDA differs in its class-conditional likelihood function, which now incorporates class-specific covariance matrices:

$$P(x_i|C_k) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1}(x_i - \mu_k)\right). \quad (\text{A.11})$$

Since Σ_k varies for each class, the discriminant function for QDA becomes:

$$\delta_k(x_i) = -\frac{1}{2} \log |\Sigma_k| - \frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1}(x_i - \mu_k) + \log P(C_k). \quad (\text{A.12})$$

Unlike LDA, where the decision boundaries are linear, QDA results in quadratic decision boundaries, allowing for more flexible class separation. A key disadvantage of QDA is an increased risk of overfitting, particularly when the number of training samples is small.

A.2.1.3 Support Vector Machines (SVM)

Support Vector Machines (SVM) are powerful and versatile machine learning algorithms that can be applied to both classification and regression tasks [125]. They are particularly well-suited for high-dimensional data, even when the number of training samples is relatively small compared to the number of features. Unlike probabilistic models such as LDA and QDA, SVM is a geometric method, relying on the concept of maximizing the margin between classes to improve classification robustness.

SVMs are naturally suited for binary classification, where they construct an optimal separating hyperplane in feature space, ensuring maximum separation between classes. Multi-class classification can be achieved through one-vs-one or one-vs-all extensions.

For linearly separable data, SVMs construct a linear decision boundary in the form of a hyperplane. This hyperplane divides the feature space into two regions corresponding to different class labels. Mathematically the hyperplane is described by

$$w^T x_i + b = 0, \quad (\text{A.13})$$

here w is the normal vector to the hyperplane and b the intercept or bias term. The key principle of Linear SVM is maximizing the margin, defined as the distance between the hyperplane and the closest training samples, known as support vectors. This ensures robustness against small

variations in the data. This optimization problem for a perfectly separable dataset (hard-margin SVM) is formulated as:

$$\min_{w,b} \frac{1}{2} \|w\|^2, \quad (\text{A.14})$$

subject to:

$$y_i(w^T x_i + b) \geq 1, \quad \forall i = 1, \dots, N \quad (\text{A.15})$$

where y_i denotes class labels, and N is the number of training samples.

Since real-world data is rarely perfectly separable, soft-margin SVM introduces slack variables ξ_i to allow for misclassified points while penalizing large deviations:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i, \quad (\text{A.16})$$

subject to:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i, \quad (\text{A.17})$$

with C a regularization parameter controlling the trade-off between maximizing the margin and minimizing classification errors.

Overall LSVM tends to only perform well when data is linearly separable and struggles with complex nonlinear relationships. To handle nonlinear data we can extend the model to quadratic support vector machines (QSVM), where we introduce a nonlinear decision boundary by mapping the original feature space into a higher-dimensional space where a linear separation is possible. Instead of explicitly transforming data this kernel trick is used to compute inner products in this transformed space without explicitly computing feature mappings.

A common choice for QSVM is the quadratic kernel function, defined as:

$$K(x_i, x_j) = (x_i^T x_j + c)^2, \quad (\text{A.18})$$

where c is a hyperparameter controlling the polynomial offset. The decision function for QSVM is given by:

$$f(x) = \sum_{i=1}^N \alpha_i y_i K(x_i, x) + b, \quad (\text{A.19})$$

where α_i are Lagrange multipliers, optimized using sequential minimal optimization, and $K(x_i, x)$ is the kernel function used to compute the similarity between new observations and support vectors. Using QSVM allows also for nonlinear class separation, distributions with curved decision boundaries, making it more suitable for complex spectroscopic data. But it is computationally more expensive and can, in similar fashion to QDA, lead to overfitting.

A.2.1.4 K-Nearest Neighbors (KNN)

The K-Nearest Neighbors (KNN) algorithm is a non-parametric, instance-based learning method used for both classification and regression tasks. Unlike other classifiers such as LDA, QDA, and SVM, which build a global decision boundary based on training data, KNN is a lazy algorithm, meaning it does not explicitly construct a model during training, it also doesn't make any prior assumptions on the data. Instead, classification is performed based on the similarity between a new sample and its nearest neighbors in the training dataset based on some kind of distance metric.

Mathematically KNN works as follows. Given a dataset $(x_i, y_i)_{i=1, \dots, L}$, where x_i are feature vectors (spectra) and y_i are class labels, KNN classifies a new sample, x' , by computing the distance between x' and all training samples. Then the algorithm proceeds to select the k nearest training samples (neighbors) based on the computed distances, and assigns the class most frequently occurring among these to the new sample. The most common distance metric used in KNN is the Euclidean distance, given by:

$$d(x', x_i) = \sqrt{\sum_{j=1}^n (x'_j - x_{i_j})^2}, \quad (\text{A.20})$$

for this equation $d(x', x_i)$ is the distance between sample x' and training sample x_i , n is the number of features per spectrum (wavelengths), and x'_j and x_{i_j} represent the j -th feature of the test and training samples, respectively. The predicted class $\hat{y}$ for a new observation x' is determined by majority voting among its k -nearest neighbors:

$$\hat{y} = \arg \max_C \sum_{i \in \mathcal{N}_k} \mathbb{I}(y_i = C), \quad (\text{A.21})$$

with $\mathcal{N}_k$ the set of k nearest neighbors, $\mathbb{I}(\dots)$ an indicator function that returns 1 if the condition is true and 0 otherwise, and C represents a class label.

For weighted KNN, the influence of each neighbor is weighted by its distance from x' , giving more importance to closer neighbors:

$$\hat{y} = \arg \max_C \sum_{i \in \mathcal{N}_k} w_i \mathbb{I}(y_i = C), \quad (\text{A.22})$$

where $w_i = \frac{1}{d(x, x_i) + \epsilon}$ ensures closer neighbors contribute more to the decision.

The performance of KNN is highly dependent on the choice of k . Small k (e.g. k equal to 1-2) makes the model highly sensitive to noise and outliers, while larger values smooth the decision boundary, making the classifier more stable but potentially losing local patterns and reducing accuracy.

A.2.1.5 Decision Trees (DT) And Random Forests (RF)

Decision Trees (DT) and their ensemble extensions, Random Forests (RF), are machine learning-based classification methods that rely on recursive partitioning of the feature space. Unlike probabilistic models such as LDA and QDA, or margin-based classifiers like SVM, tree-based methods provide an interpretable, rule-based approach to classification (opposite of a black box approach). These models are particularly useful in cases where model transparency and feature importance analysis are desired. This transparency make these classifiers ideally suited for applications affecting human health and safety as one can extensively study and understand their functioning.

A Decision Tree (DT) is a nonlinear classifier that uses a flowchart-like structure of *if-then-else* rules to predict class labels. The algorithm recursively partitions the dataset based on feature values, creating a hierarchical structure of decision nodes. Each internal node represents a decision rule—typically comparing one or more features against a threshold—while each leaf

corresponds to a final predicted class label. At each node, the dataset of N training samples is split into two subsets based on a threshold t for a selected feature j . The goal is to find the optimal (j, t) pair that minimizes impurity. The two most common impurity metrics used in DT construction are the Gini impurity defined as

$$G = \sum_{k=1}^K p_k(1 - p_k), \quad (\text{A.23})$$

where p_k is the proportion of samples belonging to class k in the node, and Entropy defined as

$$H = - \sum_{k=1}^K p_k \log_2 p_k, \quad (\text{A.24})$$

where a lower entropy value indicates a more homogeneous class distribution. A greedy search is performed over all possible splits, selecting the feature-threshold pair that minimizes impurity. The process is repeated recursively until a stopping criterion (e.g., maximum depth, minimum samples per node) is met.

To address the overfitting problem of DT, RF introduce an ensemble learning approach based on bootstrap aggregation (bagging). Instead of relying on a single tree, RF constructs multiple decision trees, each trained on a randomly drawn subset of the training data (bootstrapped samples). The final classification decision is made by majority voting across all trees.

A.2.1.6 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) and their extended form, Deep Neural Networks (DNN), are machine learning-based models inspired by the structure and functionality of biological neurons. These models consist of multiple layers of interconnected artificial neurons that transform input features into high-level representations for classification and regression tasks. Unlike traditional statistical models such as LDA and LSVM, neural networks can learn complex, nonlinear relationships in spectral data, making them well-suited for high-dimensional problems like spectroscopic data analyses. The strength of NN is evident from the fact that they can approximate any function, as stated by the universal approximation theorem.

These methods do however require sufficient training data and can, especially DNN, lead to overfitting, where the model learns the training data by heart and no longer generalizes to unseen data. Model performance assessment using validation sets during training is therefore essential to avoid overfitting. Additionally, the black-box nature of NN makes it difficult to interpret the decisions made, making it hard to really know what is happening and how decisions are made.

An ANN consists of three main types of layers, an input layer that accepts the feature vector $x_i \in \mathbb{R}^n$ (e.g., spectral intensities at different wavelengths), a (number of) hidden layer(s) that implement a nonlinear transformations to the input using activation functions and a final output layer, that yields the classification or regression decision. Mathematically, each neuron in layer p processes its input as follows:

$$z^{(p)} = W^{(p)}x^{(p-1)} + b^{(p)}, \quad (\text{A.25})$$

$$x^{(p)} = g(z^{(p)}), \quad (\text{A.26})$$

where: $W^{(p)}$ is the weight matrix connecting layer $p - 1$ to layer p , $b^{(p)}$ is the bias vector, $g(\dots)$ is the activation function, commonly chosen as sigmoid, ReLU, softmax, or Swish, and $x^{(p)}$ is

the transformed output, passed to the next layer. The loss function for classification is typically the categorical cross-entropy:

$$L = - \sum_{i=1}^N \sum_{k=1}^K y_{i,k} \log \hat{y}_{i,k}, \quad (\text{A.27})$$

where $y_{i,k}$ is the true label of sample i and $\hat{y}_{i,k}$ is the predicted probability.

Training is performed using gradient descent-based optimization, commonly stochastic gradient descent, with backpropagation to update weights:

$$W^{(l)} = W^{(l)} - \eta \frac{\partial L}{\partial W^{(l)}}, \quad (\text{A.28})$$

where η is the learning rate.

A Deep Neural Network (DNN) is simply an ANN with multiple hidden layers. Increasing the depth of the network allows for the learning of hierarchical features, making it better suited for complex, high-dimensional datasets. But these networks are more prone to lead to overfitting due to their increased complexity, especially for training on limited datasets.

A.2.1.7 Partial Least Squared Discriminant Analyses (PLS-DA)

Partial Least Squares Discriminant Analysis (PLS-DA) is another supervised classification technique that merges the dimensionality reduction capabilities of Partial Least Squares Regression (PLSR) (discussed in Section A.2.2.5) with the discriminative objectives of a classification. In essence, PLS-DA projects high-dimensional spectral data X into a lower-dimensional latent space where the separation between classes—defined by Y —is maximized.

Mathematically, the method decomposes the predictor matrix X into latent scores T and loadings P such that

$$X = TP^T + E, \quad (\text{A.29})$$

while simultaneously modeling the relationship with the response (Y) as

$$Y = TQ^T + F, \quad (\text{A.30})$$

where Q represents the loadings for Y and E and F are the corresponding residual matrices. The latent scores T are computed by maximizing the covariance between X and Y , thereby ensuring that the extracted components capture the most relevant spectral variations associated with class membership.

This projection not only reduces the dimensionality of the spectral data but also enhances class discrimination by emphasizing features that are most correlated with the response variable. The resulting PLS-DA model assigns class labels to new samples by projecting their spectral profiles onto the latent space and employing a decision rule—often based on distance metrics or thresholding—to determine their class membership. As such, PLS-DA is particularly advantageous when dealing with highly collinear and noisy data, common challenges in spectroscopic analyses, where it provides a robust framework for classification.

A.2.2 Regression Models

In addition to classification models, regression models can be used for spectroscopic mycotoxin detection. Unlike classification, which assigns samples to discrete categories, regression quantifies contamination levels, offering a more granular assessment. From a practical perspective, this

facilitates targeted interventions and supports risk-based decision-making. This capability makes regression models particularly valuable for in-field monitoring devices, enabling the continuous evaluation and tracking of mycotoxin proliferation in agricultural settings.

As summarized in Table 2.10, regression models have been widely applied for mycotoxin detection in both corn and wheat. The identified studies focus on the detection of single mycotoxins, such as AFLA [60], DON [115], and FUM [119], and they have utilized reflectance, transmittance, and Fourier-transformed reflectance data. The literature indicates that regression models have typically been successful only when applied to sample series with a broad contamination range. For instance, in studies focusing on DON detection, the contamination levels extended beyond 10mg/kg far exceeding the legal limits [31, 121, 122]. Furthermore, the large standard errors generally reported and the high Limits Of Detection (LODs) indicate that most attempts to develop quantification models result in insufficiently precise predictions that fail to meet legal thresholds [31].

In this study, following the approach used for classification models, we will perform a comprehensive comparative analysis to determine the best-performing regression models for our specific application. We will evaluate a range of models, including Support Vector Regression (SVR), Random Forest Regressors (RFR), Artificial Neural Network Multi-Layer Perceptron Regressors (ANN-MLPR), Lasso and Elastic Net, and finally, Partial Least Squares Regression (PLSR). Notably, previous comparative studies have consistently shown that PLSR outperforms alternative models, with all identified studies ultimately employing PLSR as their final regressor model [31].

The following sections¹ will provide a detailed discussion on these regression models, explaining their methodologies while drawing parallels to the classification techniques outlined earlier and highlighting their specific advantages and limitations.

A.2.2.1 Support Vector Regression (SVR)

Support Vector Regression (SVR) extends the principles of Support Vector Machines (SVM) from classification to regression tasks. While SVM focuses on separating classes (see Section A.2.1.3), SVR aims to find a function that models the relationship between the input data X (spectra) and a continuous response Y (mycotoxin concentration). This is done by determining a regression function $f(x_i)$ that deviates from the true target, y_i , by no more than a specified margin, ϵ , for all data elements, while keeping the function as flat as possible to avoid overfitting to noise. Mathematically, SVR seeks a function of the form

$$f(x_i) = \langle w, x_i \rangle + b, \quad (\text{A.31})$$

where w is the weight vector and b is the bias term. The model is trained by minimizing the norm $\|w\|^2$, while adhering to the constraint that each sample falls within the ϵ -tube:

$$|y_i - f(x_i)| \leq \epsilon, \quad \forall i. \quad (\text{A.32})$$

Since real-world data will rarely perfectly adhere to this constraint, slack variables ξ_i and ξ_i^* are introduced to handle data points outside the ϵ -tube, leading to the optimization problem:

$$\min_{w, b, \xi, \xi^*} \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n (\xi_i + \xi_i^*), \quad (\text{A.33})$$

¹This theoretical discussion is based on the course work; "AI for chemical Engineering", by Jaewook Nam thought at Seoul National University.

subject to

$$\begin{aligned} y_i - \langle w, x_i \rangle - b &\leq \epsilon + \xi_i, \\ \langle w, x_i \rangle + b - y_i &\leq \epsilon + \xi_i^*, \\ \forall i, \quad \xi_i, \xi_i^* &\geq 0, \end{aligned} \tag{A.34}$$

where c is a regularization parameter that balances model complexity and the degree to which deviations larger than ϵ are tolerated.

In summary, SVR inherits the robust optimization framework of SVM while adapting it to continuous outcomes. Its capacity to control the precision of regression through the ϵ -insensitive loss function provides a powerful tool for spectroscopic data processing.

A.2.2.2 Random Forest Regressors (RFR)

Random Forest Regressors (RFR) extend the ensemble learning framework of Random Forest classifiers, discussed in Section A.2.1.5, to the domain of regression. Much like their classification counterparts, RFR builds an ensemble of decision trees; however, instead of outputting class votes, each regression tree predicts a continuous value. The final prediction is obtained by aggregating the outputs of all individual trees—typically by computing the mean of their predictions.

In practice, each decision tree in the forest is constructed using a bootstrap sample drawn from the original dataset, and at each split, a random subset of features is considered. This randomness not only reduces the variance of the overall model but also helps mitigate overfitting, as individual trees are less correlated.

Mathematically, if the prediction from the t -th tree for a given input x_i is denoted as $f_t(x_i)$, then the Random Forest Regressor estimates the target value as:

$$\hat{y}_i = \frac{1}{T} \sum_{t=1}^T f_t(x_i), \tag{A.35}$$

where T is the total number of trees in the ensemble and $\hat{y}_i$ is the averaged, predicted concentration corresponding to x_i . Alternatively, some implementations may use the median of predictions to further enhance robustness against outliers.

In summary, Random Forest Regressors adapt the principles of Random Forest classification to continuous prediction tasks by averaging the outputs of multiple decision trees. This approach leverages the strength of ensemble learning to manage the complex, non-linear relationships inherent in spectroscopic data, while retaining the advantageous, white-box nature of the model.

A.2.2.3 Artificial Neural Network Regressor (ANN-MLPR)

Artificial Neural Network Multi-Layer Perceptron Regressors (ANN-MLPR) extend the neural network architecture—previously discussed for classification in Section A.2.1.6—to continuous prediction tasks. Like the ANN-classifier, the ANN-MLPR consists of interconnected layers of neurons that process input data using non-linear activation functions. However, while the classifier produces discrete class probabilities, the regressor is designed to generate continuous outputs corresponding to estimated mycotoxin concentrations.

In ANN-MLPR, the network architecture typically consists of an input layer, one or more hidden

layers, and a final output layer with a linear activation function to yield unbounded continuous values. Training the network involves minimizing a loss function—commonly the Mean Squared Error (MSE)—which quantifies the difference between the predicted output and the actual target values. Mathematically, the MSE is defined as:

$$\text{MSE} = \frac{1}{l} \sum_{i=1}^l (y_i - \hat{y}_i)^2, \quad (\text{A.36})$$

where y_i represents the true contamination level for the i -th sample, $\hat{y}_i$ is the network's prediction, and l is the number of training samples.

To optimize the networks weights and biases gradient descent is typically used. The inherent flexibility of ANN-MLPR allows it to capture complex, non-linear relationships between spectral features and toxin concentrations, making it particularly well suited for variable and noisy data.

A.2.2.4 Lasso And Elastic Net

Lasso and Elastic Net are powerful extensions of the traditional least squared regression techniques that improve predictive accuracy while performing feature reduction. These methods are especially useful when dealing with high-dimensional data, such as spectroscopic measurements, where many features (wavelengths) may be redundant or irrelevant.

Lasso, short for Least Absolute Shrinkage and Selection Operator, incorporates an L_1 penalty into the least squares objective. This penalty encourages sparsity in the model coefficients, effectively shrinking some to zero, which results in automatic feature reduction. Mathematically, Lasso minimizes the following objective function:

$$\min_{\beta} \frac{1}{2l} \sum_{i=1}^l (y_i - x_i^\top \beta)^2 + \lambda \|\beta\|_1, \quad (\text{A.37})$$

where β represents the vector of regression coefficients, x_i the feature vector (spectral data), and y_i the target variable (concentration) corresponding to the i -th sample. l is the number of observations/data points, and λ is a regularization parameter controlling the magnitude of the L_1 penalty ($\|\beta\|_1$). By tuning λ , Lasso balances the trade-off between minimizing the residual sum of squares and enforcing sparsity in the model. A higher λ results in more coefficients being set to zero, thus simplifying the model and hindering overfitting.

Elastic Net extends Lasso by combining the L_1 penalty with an L_2 penalty. This hybrid approach is especially useful when dealing with highly correlated features—a common scenario in spectroscopic analyses. The Elastic Net objective function is defined as:

$$\min_{\beta} \frac{1}{2l} \sum_{i=1}^l (y_i - x_i^\top \beta)^2 + \lambda \left(\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right), \quad (\text{A.38})$$

where $\alpha \in [0, 1]$ controls the balance between the L_1 and L_2 penalties, with $\alpha = 1$ yielding Lasso as a special case. The L_2 term, $\|\beta\|_2^2$, helps stabilize the solution when predictors are highly correlated, while the L_1 term, $\|\beta\|_1$, retains the benefit of automatic feature reduction, which lowers the risk of overfitting and enhance the model's ability to generalize to new, unseen data. This is critical when the data comes from complex food matrices where spectral signatures can be noisy and highly correlated.

In summary, Lasso and Elastic Net improve upon traditional regression methods by incorporating regularization terms that simultaneously perform feature reduction and dimensionality shrinkage. Their mathematical framework is well-suited for handling the high-dimensional, collinear nature of spectroscopic data, making them robust tools of potential interest for our application.

A.2.2.5 Partial Least Squares Regression (PLSR)

A final regression model of particular interest, based on our literature review, is Partial Least Squares Regression (PLSR). PLSR extends the dimensionality reduction strategy employed in PLS-DA (see Section A.2.1.7) to regression tasks, where the objective is to predict continuous outcomes rather than assign discrete class labels. In PLSR, the model function is formulated in a manner analogous to PLS-DA; the key distinction is that the response variable, Y , is continuous, enabling the model to generate quantitative predictions of mycotoxin concentrations.

In summary, PLSR utilizes the same mathematical framework as PLS-DA but adapts it for regression. Its capability to condense high-dimensional spectral data into informative latent variables renders it a robust tool for predicting continuous outcomes in spectroscopic mycotoxin detection. This approach provides detailed quantitative insights that are essential for informed, risk-based decision-making in food safety monitoring.

A.3 Feature Selection Methods

A crucial preprocessing step that has not yet been elaborated-on is feature selection, which involves identifying and selecting the most informative features (wavelengths) from the spectral data. Irrelevant or redundant information can then be removed, thereby reducing the dimensionality, retaining only a limited number of key intensity values at carefully chosen wavelengths. Such reduction not only enhances the interpretability of the model—allowing one to discern which spectral features underpin the classification or regression decisions—but also decreases system complexity. This simplification is beneficial both for model training and use in industrial applications, such as in-line food quality control units or in-field measurement devices, where working with a discrete set of wavelengths is preferred, as discussed in the introduction, in Section 1.2.1.

It is important to note that some classification and regression models, such as PLS-DA, PLSR, Lasso, and Elastic Net (discussed in Sections A.2.1.7, A.2.2.5, and A.2.2.4), inherently perform a form of dimensionality reduction by generating latent variables or linear combinations of wavelengths. However, this form of reduction does not equate to selecting discrete, individual wavelengths which we desire to implement.

Some commonly used methods for selecting only the most informative features (wavelengths) include Principal Component Analysis (PCA), Sequential Feature Search (SFS), and Minimum Redundancy Maximum Relevance (MRMR). In the following sections, each of these techniques will be discussed in some detail.

A.3.1 Principal Component Analysis (PCA)

A first method for feature selection is Principal Component Analysis (PCA), one of the most widely used techniques for dimensionality reduction and feature extraction. PCA transforms the original dataset into a new set of uncorrelated variables—called Principal Components

(PCs)—that retain the highest variance in the data while reducing redundancy. Notably, PCA is an unsupervised method, meaning it only considers the spectral data without incorporating label information.

Mathematically, given a dataset, X , with dimensions $l \times n$ (where l is the number of training samples or spectra and n is the number of features per spectra or wavelengths), PCA seeks a linear transformation that maps X onto a new space Z with uncorrelated principal components, following the functional form:

$$Z = XW, \quad (\text{A.39})$$

where W is the eigenvector matrix of the covariance matrix $\Sigma = X^T X$, with its columns representing the principal components, ranked in decreasing order of variance. By retaining only the first k principal components (where $k < n$), PCA effectively reduces the dimensionality of the data while preserving most of the variance.

This method is computationally simple and efficient. However, for highly complex data—such as spectra where the relevant information is obscured by a highly variable background—PCA may not always be the optimal feature selection strategy, as it can select for variance that is uninformative with respect to the task at hand [125].

Moreover, while PCA reduces the dimensionality of the dataset by projecting the data onto a smaller set of principal components, these lower-dimensional PCA scores are still linear combinations of the original spectral variables. This means that, although the data is represented in fewer dimensions, the transformation does not eliminate or isolate specific spectral features; instead, it re-expresses the entire original spectrum in a new basis. As a result, PCA does not truly reduce the number of spectral features in the sense of selecting individual wavelengths or bands, which may limit its interpretability and usefulness in applications that require direct feature selection or wavelength interpretation.

A.3.2 Sequential Feature Search (SFS)

An alternative approach is to use a wrapper-based feature selection method, such as Forward or Backward Sequential Feature Search (SFS or SBS). These methods iteratively select the best subset of features based on the performance of a given model architecture on these features—whether it be a classification or regression model. SFS is of particular interest for our application, as it has already been successfully applied for spectral feature selection by Magnus et al. in the detection of aflatoxin on pistachios [30, 125].

Sequential Feature Search begins with an empty feature set, $S = \emptyset$, and progressively builds an optimal subset of features. First, a classifier or regressor—such as those described in Sections A.2.1 and A.2.2—is trained using each individual feature, x_{i_j} , from the available pool $\{x_{i_1}, x_{i_2}, \dots, x_{i_n}\}$. The feature that results in the best model performance, $x_{i_s} \in \{x_{i_1}, x_{i_2}, \dots, x_{i_n}\}$, is then selected and added to S . Next, the algorithm evaluates every remaining feature by combining it with the features already in S , training a new model for each combination. The feature that yields the greatest improvement in performance is added to S . This iterative process continues until either the model performance plateaus or a predetermined maximum number of features is reached.

As a supervised feature selection method, SFS is effective in ensuring that the most important features are chosen for model training. However, it is computationally expensive because it requires iteratively training models for many possible feature combinations. In contrast, SBS

starts with the full set of features and iteratively removes the feature whose exclusion has the least impact on performance. For our application—where the objective is to identify a small subset of features from a high-dimensional dataset—SFS is preferred due to its computational efficiency and the reduced number of iterations required.

A.3.3 Minimum Redundancy Maximum Relevance (MRMR)

A third alternative for feature selection is to use Minimum Redundancy Maximum Relevance (MRMR). This technique selects features based on two criteria, maximum relevance (selected features should be highly correlated with the target variable) and minimum redundancy (selected features should be minimally correlated with each other to avoid redundancy). This makes MRMR much more computationally efficient than iterative processes like SFS.

Given a dataset with n features (wavelengths), the objective is to maximize the relevance D and minimize the redundancy R , defined as:

$$D = \frac{1}{|S|} \sum_{x_{i_a} \in S} I(x_{i_a}, y_i), \quad (\text{A.40})$$

$$R = \frac{1}{|S|^2} \sum_{x_{i_a}, x_{i_b} \in S} I(x_{i_a}, x_{i_b}), \quad (\text{A.41})$$

where x_{i_j} represents the j -th feature in the spectrum and y_i is the corresponding label. S is the feature subset and $|S|$ the number of features therein, $I(a, b)$ represents the mutual information between two variables a and b , often defined as

$$I(a, b) = \sum_{a \in A} \sum_{b \in B} p(a, b) \log \left(\frac{p(a, b)}{p(a)p(b)} \right), \quad (\text{A.42})$$

where $p(a, b)$ is the joint probability distribution of a and b , and $p(a)$ and $p(b)$ are their marginal probability distributions. A and B represent the sets of all possible values (the supports) for the random variables a and b , respectively. Signifying that the summations run over every possible value that a and b can assume.

The final feature set is obtained by solving:

$$\max_S (D - R). \quad (\text{A.43})$$

A.4 Model Validation & Performance Metrics

Evaluating the performance of a trained model—whether a regressor or a classifier—is a critical final step in developing a spectroscopic platform for mycotoxin detection in foodstuff. A robust evaluation framework must comprehensively capture model performance, ensuring that the model avoids overfitting and generalizes well to unseen data. This rigorous assessment provides a solid foundation for determining the model's suitability and reliability in real-world applications.

There are numerous validation strategies available, each with their own advantages and limitations. Choosing the right validation scheme is therefore essential. In this section, we will

explore various validation approaches commonly used with spectroscopic data, including external hold-out sets, cross-validation (CV), and leave-one-sample-out cross-validation (LOSO-CV). We will also discuss key performance metrics for both classification and regression models. For classification, metrics such as accuracy and False Positive Rate (FPR) will be introduced, while for regression, we will detail measures like the correlation coefficient (R), the coefficient of determination (R^2), and error metrics like the Root Mean Squared Error (RMSE). This discussion aims to provide clarity on how these evaluation tools contribute to a comprehensive understanding of model performance.

A.4.1 Validation Strategies

In an ideal scenario, the final model evaluation is performed on an external hold-out set, comprising samples that were not used during training, to provide an unbiased estimate of model performance on completely unseen data. This approach ensures that the evaluation fully reflects the model's ability to generalize across a wide range of independent samples and spectra.

However, in our application, the in-situ detection of mycotoxins, obtaining contaminated samples with known toxin compositions and capturing their respective spectra is both challenging and costly. Therefore researchers are typically limited to only a few tens to a few hundred samples at best [34]. Because both classification and regression models require large amounts of data to learn the complex relationships between spectral features and mycotoxin levels, reserving a large external hold-out set would leave insufficient data for training.

To overcome this limitation, various forms of Cross-Validation (CV) are employed. CV is a resampling technique that systematically partitions the data into training and validation sets multiple times, ensuring that each sample is used for both purposes in different iterations. This approach allows for a more efficient use of the available data while providing a robust estimate of model performance.

One popular method is k-fold cross-validation. In k-fold CV, the dataset is divided into k equally sized subsets (folds). The model is then trained k times, each time using $k - 1$ folds for training and the remaining fold for validation. The overall performance metric is computed by averaging the results over all k iterations. Mathematically, the overall cross-validation accuracy is given by:

$$A_{CV} = \frac{1}{k} \sum_{i=1}^k A_i, \quad (\text{A.44})$$

where A_i represents the accuracy obtained on the i -th fold.

Alternatively, Leave-One-Sample-Out Cross-Validation (LOSO-CV) is a special case where k is equal to the total number of unique samples. In each iteration, the model is trained on the spectral data from all samples except one, which is used for validation. This approach maximizes the training data available in each iteration while ensuring that every sample is used for external validation exactly once. LOSO-CV, in particular, has been successfully applied in similar contexts. Dr. Francesca Venturini and colleagues demonstrated its effectiveness for multi-mycotoxin detection on corn, making it an especially promising validation scheme.

Furthermore, LOSO-CV is particularly advantageous in situations where multiple spectra are acquired per sample—such as when the same physical sample is measured several times under varying conditions—because it ensures that no spectra from the same sample appear in both the training and validation sets. This distinction between "*spectra*" and "*sample*" is crucial, as it

prevents data leakage and more accurately reflects the model’s generalizability to new, unseen samples.

A.4.2 Key Performance Metrics

Evaluating model performance is essential for validating spectroscopic mycotoxin detection systems. In the following sections, we introduce key metrics for classification—such as accuracy, sensitivity, specificity, precision, and the F1-score—and for regression, including R^2 , R , RMSE, LOD, and LOQ. These metrics provide a robust framework for assessing model reliability and generalizability in real-world applications.

A.4.2.1 Classification Models

For classification tasks, a variety of performance metrics are used to quantify a model’s ability to distinguish between classes. These metrics are derived from the confusion matrix—a fundamental tool in evaluating classification outcomes. For a binary classification problem (e.g., contaminated vs. non-contaminated, safe vs. unsafe), the confusion matrix is structured as follows:

	Predicted Positive	Predicted Negative
Actual Positive (contaminated)	True Positive (TP)	False Negative (FN)
Actual Negative (safe)	False Positive (FP)	True Negative (TN)

Table A.1: Structure of a two-class confusion matrix

In this table, True Positives (TP) represent samples correctly identified as contaminated, while False Positives (FP) are safe samples incorrectly classified as contaminated. False Negatives (FN) occur when contaminated samples are misclassified as safe, and True Negatives (TN) correspond to safe samples correctly identified as such.

From the confusion matrix, several key performance metrics can be defined:

- **Accuracy**, a measure of the overall correctness of the model:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (\text{A.45})$$

usually expressed as a percentage.

- **Sensitivity** (or Recall), the proportion of actual positive samples correctly identified:

$$Sensitivity = \frac{TP}{TP + FN}. \quad (\text{A.46})$$

- **Specificity**, the proportion of actual negative samples correctly identified:

$$Specificity = \frac{TN}{TN + FP}, \quad (\text{A.47})$$

- **Precision**, the fraction of samples predicted as positive that are actually positive:

$$Precision = \frac{TP}{TP + FP}, \quad (\text{A.48})$$

Additional metrics include:

- **Negative Predictive Value (NPV):**

$$NPV = \frac{TN}{TN + FN}, \quad (\text{A.49})$$

- **False Positive Rate (FPR):**

$$FPR = \frac{FP}{FP + TN}, \quad (\text{A.50})$$

- **False Negative Rate (FNR):**

$$FNR = \frac{FN}{FN + TP}. \quad (\text{A.51})$$

Finally, the **F1-Score**—the harmonic mean of precision and sensitivity—is particularly useful for imbalanced datasets:

$$F1 = 2 \cdot \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}, \quad (\text{A.52})$$

In the context of mycotoxin detection and food safety, both sensitivity and specificity are critical. Low sensitivity (high FNR) means that contaminated samples may be misclassified as safe, posing significant health risks by allowing unsafe food into the supply chain. Conversely, low specificity (high FPR) can lead to unnecessary food waste, as safe samples are erroneously discarded. Thus, achieving an optimal balance between these metrics is essential for ensuring consumer safety while minimizing food loss.

A.4.2.2 Regression Models

Performance metrics for regression models quantify how effectively the model predicts continuous outcomes, based on the differences between the predicted values, $\hat{y}_i$, and the actual targets, y_i . One key metric is the unit-less Coefficient of Determination, R^2 , which measures the proportion of variance in the target variable that is explained by the model. It is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^l (y_i - \hat{y}_i)^2}{\sum_{i=1}^l (y_i - \bar{y})^2}, \quad (\text{A.53})$$

where $\bar{y}$ is the mean of the observed values (y_i 's). A higher R^2 indicates a better linear fit between the model and the data, with an $R^2 = 1$ presenting a perfect fit.

The Pearson correlation coefficient, R , evaluates the linear relationship between the observed and predicted values. While similar in concept to R^2 , R provides a direct measure of correlation:

$$R = \frac{\sum_{i=1}^l (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^l (y_i - \bar{y})^2 \sum_{i=1}^l (\hat{y}_i - \bar{\hat{y}})^2}}, \quad (\text{A.54})$$

where $\bar{\hat{y}}$ is the mean of the predicted values ($\hat{y}_i$'s).

Another important metric is the Root Mean Squared Error (RMSE), which quantifies the average magnitude of the prediction error:

$$RMSE = \sqrt{\frac{1}{v} \sum_{i=1}^v (y_i - \hat{y}_i)^2}, \quad (\text{A.55})$$

where v is the number of spectra used for validation or testing. RMSE is particularly useful for evaluating the model's predictive accuracy and assessing its compliance with legal requirements. In the specific case of cross-validation, the RMSE is referred to as the Standard Error of Cross-Validation (SECV), whereas when evaluated on an external hold-out set, it is called the Standard Error of Prediction (SEP).

From the error we can define two metrics, the Limit of Detection (LOD) and the Limit of Quantification (LOQ). The LOD is the lowest concentration of a toxin that can be reliably distinguished from background noise, whereas the LOQ is the lowest concentration that can be quantitatively measured with acceptable precision and accuracy. In analytical chemistry these are defined as

$$LOD = 3.3 \frac{\sigma}{S}, \quad (\text{A.56})$$

and

$$LOQ = 10 \frac{\sigma}{S}, \quad (\text{A.57})$$

with σ the standard deviation of the data and S the slope of the linear relation. For σ we commonly substitute the RMSE of a blank sample.

In summary, these performance metrics— R , R^2 , RMSE, SECV, SEP, LOD, and LOQ—provide a comprehensive evaluation framework for regression models in spectroscopic mycotoxin detection. They enable researchers to assess model accuracy, precision, and sensitivity, ensuring that the developed models meet regulatory limits and are robust and reliable for real-world applications.

Bibliography

- [1] LB Bullerman. Significance of mycotoxins to food safety and human health. *Journal of Food Protection*, 42(1):65–86, 1979.
- [2] Sofia Agriopoulou, Eygenia Stamatelopoulou, and Theodoros Varzakas. Advances in analysis and detection of major mycotoxins in foods. *Foods*, 9(4):518, 2020.
- [3] JI Pitt. Toxigenic fungi and mycotoxins. *British medical bulletin*, 56(1):184–192, 2000.
- [4] S Marin, AJ Ramos, Germán Cano-Sancho, and V Sanchis. Mycotoxins: Occurrence, toxicology, and exposure assessment. *Food and chemical toxicology*, 60:218–237, 2013.
- [5] Massimo Reverberi, Alessandra Ricelli, Slaven Zjalic, Anna A Fabbri, and Corrado Fanelli. Natural functions of mycotoxins and control of their biosynthesis in fungi. *Applied microbiology and biotechnology*, 87:899–911, 2010.
- [6] Aycan Cinar and Elif Onbaşı. Mycotoxins: The hidden danger in foods. *Mycotoxins and food safety*, pages 1–21, 2019.
- [7] Kajal Karsauliya, C Yahavi, Anushka Pandey, Manisha Bhateria, Ashish Kumar Sonker, Harshita Pandey, Manu Sharma, and Sheelendra Pratap Singh. Co-occurrence of mycotoxins: A review on bioanalytical methods for simultaneous analysis in human biological samples, mixture toxicity and risk assessment strategies. *Toxicon*, 218:25–39, 2022.
- [8] PMCP Kirk, Paul Cannon, David Minter, and Joost Stalpers. Ainsworth and bisby’s dictionary of the fungi., 2008.
- [9] Gary P Munkvold, Silvina Arias, Ines Taschl, and Christiane Gruber-Dorninger. Mycotoxins in corn: Occurrence, impacts, and management. In *Corn*, pages 235–287. Elsevier, 2019.
- [10] Zsolt Ráduly, László Szabó, Anett Madar, István Pócsi, and László Csernoch. Toxicological and medical aspects of aspergillus-derived mycotoxins entering the feed and food chain. *Frontiers in microbiology*, 10:2908, 2020.
- [11] Inger Völkel, Eva Schröer-Merker, and Claus-Peter Czerny. The carry-over of mycotoxins in products of animal origin with special regard to its implications for the european food safety legislation, 2011.
- [12] Michael J Sweeney and Alan DW Dobson. Mycotoxin production by aspergillus, fusarium and penicillium species. *International journal of food microbiology*, 43(3):141–158, 1998.
- [13] Ákos Mesterházy, Judit Oláh, and József Popp. Losses in the grain supply chain: Causes and solutions. *Sustainability*, 12(6):2342, 2020.

- [14] Rahim Khan, Farooq Anwar, and Farinazleen Mohamad Ghazali. A comprehensive review of mycotoxins: Toxicology, detection, and effective mitigation approaches. *Heliyon*, 2024.
- [15] Shabir Ahmad Mir, BN Dar, Manzoor Ahmad Shah, Sajad Ahmad Sofi, Afshan Mumtaz Hamdani, Carlos AF Oliveira, Motahareh Hashemi Moosavi, Amin Mousavi Khaneghah, and Anderson S Sant’Ana. Application of new technologies in decontamination of mycotoxins in cereal grains: Challenges, and perspectives. *Food and Chemical Toxicology*, 148:111976, 2021.
- [16] Lien Smeesters, Thomas Kuntzel, Hugo Thienpont, and Ludovic Guilbert. Handheld fluorescence spectrometer enabling sensitive aflatoxin detection in maize. *Toxins*, 15(6):361, 2023.
- [17] Wenhua Wang, Qi Zhang, Fei Ma, and Peiwu Li. Simultaneous determination of aflatoxins, fumonisin b1, t-2 and cyclopiazonic acid in agri-products by immunomagnetic solid-phase extraction coupled with uhplc-ms/ms. *Food Chemistry*, 378:132020, 2022.
- [18] Dakai Guo, Qingwen Huang, Rui Zhao, Wenbo Guo, Kai Fan, Zheng Han, Zhihui Zhao, and Dongxia Nie. Mil-101 (cr)@ fe₃o₄ nanocomposites as magnetic solid-phase extraction adsorbent for the determination of multiple mycotoxins in agricultural products by ultra-high-performance liquid chromatography tandem mass spectrometry. *Food Control*, 146:109540, 2023.
- [19] Chencheng Wei, Handong Wang, Gaozheng Li, Jianhua Li, Fang Zhang, Yuezhong Wu, and Zuquan Weng. Multiplex detection methods for mycotoxins in agricultural products: a systematic review. *Food Control*, 158:110207, 2024.
- [20] Ke-Yu Xing, Juan Peng, Shan Shan, Dao-Feng Liu, Yi-Na Huang, and Wei-Hua Lai. Green enzyme-linked immunosorbent assay based on the single-stranded binding protein-assisted aptamer for the detection of mycotoxin. *Analytical Chemistry*, 92(12):8422–8426, 2020.
- [21] Xiaoqian Tang, Xin Li, Peiwu Li, Qi Zhang, Ran Li, Wen Zhang, Xiaoxia Ding, Jiawen Lei, and Zhaowei Zhang. Development and application of an immunoaffinity column enzyme immunoassay for mycotoxin zearalenone in complicated samples. *PloS one*, 9(1):e85606, 2014.
- [22] Jyoti Singh and Alka Mehta. Rapid and sensitive detection of mycotoxins by advanced and emerging analytical methods: A review. *Food science & nutrition*, 8(5):2183–2204, 2020.
- [23] Jie Li, Xiuli Xu, Xiujuan Wang, Chen Li, Xuesong Feng, Yuan Zhang, and Feng Zhang. Construction of a magnetic covalent organic framework for magnetic solid-phase extraction of aflatoxin B₁ and aflatoxin G₁ in milk prior to quantification by lc-ms/ms. *Microchimica Acta*, 189(4):149, 2022.
- [24] Yan Yang, Zhuoqun Su, Di Wu, Jianghua Liu, Xianlong Zhang, Yongning Wu, and Guoliang Li. Low background interference sers aptasensor for highly sensitive multiplex mycotoxin detection based on polystyrene microspheres-mediated controlled release of raman reporters. *Analytica Chimica Acta*, 1218:340000, 2022.
- [25] Mengxiang Qiao, Yong Liu, and Min Wei. Dual-signal output fluorescent aptasensor based on dna programmability and gold nanoflowers for multiple mycotoxins detection. *Analytical and Bioanalytical Chemistry*, 415(2):277–288, 2023.

- [26] Lien Smeesters, Wendy Meulebroeck, S Raeymaekers, and Hugo Thienpont. Non-destructive detection of mycotoxins in maize kernels using diffuse reflectance spectroscopy. *Food Control*, 70:48–57, 2016.
- [27] B Jia, W Wang, XZ Ni, X Chu, SC Yoon, and KC Lawrence. Detection of mycotoxins and toxigenic fungi in cereal grains using vibrational spectroscopic techniques: A review. *World Mycotoxin Journal*, 13(2):163–178, 2020.
- [28] Yong-Kyoung Kim, Insuck Baek, Kyung-Min Lee, Geonwoo Kim, Seyeon Kim, Sung-Youn Kim, Diane Chan, Timothy J Herrman, Namkuk Kim, and Moon S Kim. Rapid detection of single-and co-contaminant aflatoxins and fumonisins in ground maize using hyperspectral imaging techniques. *Toxins*, 15(7):472, 2023.
- [29] David J Griffiths and Darrell F Schroeter. *Introduction to Quantum Mechanics, third edition*. Cambridge University Press, 2020.
- [30] Indy Magnus, Fatemeh Abbasi, Hugo Thienpont, and Lien Smeesters. Laser-induced fluorescence spectroscopy enhancing pistachio nut quality screening. *Food Control*, 158:110192, 2024.
- [31] Cecile Levasseur-Garcia. Updated overview of infrared spectroscopy methods for detecting mycotoxins on cereals (corn, wheat, and barley). *Toxins*, 10(1):38, 2018.
- [32] Wen-Hao Su, Hong-Ju He, and Da-Wen Sun. Non-destructive and rapid evaluation of staple foods quality by using spectroscopic techniques: a review. *Critical reviews in food science and nutrition*, 57(5):1039–1051, 2017.
- [33] Lien Smeesters, Indy Magnus, Martin Virte, Hugo Thienpont, and Wendy Meulebroeck. Potato quality assessment by monitoring the acrylamide precursors using reflection spectroscopy and machine learning. *Journal of Food Engineering*, 311:110699, 2021.
- [34] Lien Smeesters, Wendy Meulebroeck, S Raeymaekers, and Hugo Thienpont. Optical detection of aflatoxins in maize using one-and two-photon induced fluorescence spectroscopy. *Food Control*, 51:408–416, 2015.
- [35] Balkis Aouadi, John-Lewis Zinia Zaukuu, Flora Vitális, Zsanett Bodor, Orsolya Fehér, Zoltan Gillay, George Bazar, and Zoltan Kovacs. Historical evolution and food control achievements of near infrared spectroscopy, electronic nose, and electronic tongue—critical overview. *Sensors*, 20(19):5479, 2020.
- [36] Dushani L Palliyaguru and Felicia Wu. Global geographical overlap of aflatoxin and hepatitis c: controlling risk factors for liver cancer worldwide. *Food Additives & Contaminants: Part A*, 30(3):534–540, 2013.
- [37] Xiaohu Luo, Ren Wang, Li Wang, Yong Wang, and Zhengxing Chen. Structure elucidation and toxicity analyses of the degradation products of aflatoxin b1 by aqueous ozone. *Food Control*, 31(2):331–336, 2013.
- [38] LGC Standards GmbH. Certified reference material for aflatoxin b1. *Certificate of Analysis*, DRE-C10047100:1–6, 2022.
- [39] LGC Standards GmbH. Certified reference material for aflatoxin b2. *Certificate of Analysis*, DRE-C10047200:1–6, 2025.

- [40] AA El-Banna, JI Pitt, and L Leistner. Production of mycotoxins by penicillium species. *Systematic and Applied Microbiology*, 10(1):42–46, 1987.
- [41] Sherif O Sherif, Emad E Salama, and Mosaad A Abdel-Wahhab. Mycotoxins and child health: The need for health risk assessment. *International journal of hygiene and environmental health*, 212(4):347–368, 2009.
- [42] Annie Pfohl-Leszkowicz and Richard A Manderville. Ochratoxin a: An overview on toxicity and carcinogenicity in animals and humans. *Molecular nutrition & food research*, 51(1):61–99, 2007.
- [43] T. Kőszegi and M. Po’or. Ochratoxin a: Molecular interactions, mechanisms of toxicity and prevention at the molecular level. *Toxins*, 8(4):111, 2016.
- [44] Javad Hashemi and Naader Alizadeh. Investigation of solvent effect and cyclodextrins on fluorescence properties of ochratoxin a. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 73(1):121–126, 2009.
- [45] Iris Cagnasso, Glauco Tonachini, Silvia Berto, Agnese Giacomino, Luisa Mandrile, Andrea Maranzana, and Francesca Durbiano. Comprehensive study on the degradation of ochratoxin a in water by spectroscopic techniques and dft calculations. *RSC advances*, 9(34):19844–19854, 2019.
- [46] CM Placinta, JP Felix D’Mello, and AMC Macdonald. A review of worldwide contamination of cereal grains and animal feed with fusarium mycotoxins. *Animal feed science and technology*, 78(1-2):21–37, 1999.
- [47] Gary P Munkvold. Fusarium species and their associated mycotoxins. *Mycotoxigenic fungi: methods and protocols*, pages 51–106, 2017.
- [48] Ran Ran, Canhua Wang, Zheng Han, Aibo Wu, Dabing Zhang, and Jianxin Shi. Determination of deoxynivalenol (don) and its derivatives: Current status of analytical methods. *Food control*, 34(1):138–148, 2013.
- [49] Chengzhi Wang, Xiaping Fu, Ying Zhou, and Feng Fu. Deoxynivalenol detection beyond the limit in wheat flour based on the fluorescence hyperspectral imaging technique. *Foods*, 13(6):897, 2024.
- [50] Karolina Ropejko and Magdalena Twarużek. Zearalenone and its metabolites—general overview, occurrence, and toxicity. *Toxins*, 13(1):35, 2021.
- [51] Del Zotto Stefania Della Riccia Giacomo. A multivariate regression model for detection of fumonisins content in maize from near infrared spectra. *Food Chem*, 141:4289–4294, 2013.
- [52] Venkataramana Mudili, Chandra Nayaka Siddaih, Madhukar Nagesh, Phanikumar Garapati, Kalagatur Naveen Kumar, Harishchandra Sreepathi Murali, Tapani Yli Mattila, and Harsh Vardan Batra. Mould incidence and mycotoxin contamination in freshly harvested maize kernels originated from india. *Journal of the Science of Food and Agriculture*, 94(13):2674–2683, 2014.
- [53] M Hove, Christof Van Poucke, Emmanuel Njumbe-Ediage, LK Nyanga, and Sarah De Saeger. Review on the natural co-occurrence of aflb1 and fb1 in maize and the combined toxicity of aflb1 and fb1. *Food Control*, 59:675–682, 2016.

- [54] Trevor K Smith, Ewen G McMillan, and Jorge B Castillo. Effect of feeding blends of fusarium mycotoxin-contaminated grains containing deoxynivalenol and fusaric acid on growth and feed consumption of immature swine. *Journal of animal science*, 75(8):2184–2191, 1997.
- [55] LF Kubena, TS Edrington, RB Harvey, SA Buckley, TD Phillips, GE Rottinghaus, and HH Casper. Individual and combined effects of fumonisin b1 present in fusarium moniliforme culture material and t-2 toxin or deoxynivalenol in broiler chicks. *Poultry Science*, 76(9):1239–1247, 1997.
- [56] Panagiotis D Tassis, Ioannis A Tsakmakidis, Veronika Nagl, Nicole Reisinger, Eleni Tzika, Christiane Gruber-Dorninger, Ilias Michos, Nikolaos Mittas, Athina Basioura, and Dian Schatzmayr. Individual and combined in vitro effects of deoxynivalenol and zearalenone on boar semen. *Toxins*, 12(8):495, 2020.
- [57] AS Ficheux, Y Sibiril, and D Parent-Massin. Co-exposure of fusarium mycotoxins: in vitro myelotoxicity assessment on human hematopoietic progenitors. *Toxicon*, 60(6):1171–1179, 2012.
- [58] European Commission. Mycotoxins. https://food.ec.europa.eu/food-safety/chemical-safety/contaminants/catalogue/mycotoxins_en, 2024. Accessed: 2025-04-29.
- [59] U.S. Food and Drug Administration. Mycotoxins. <https://www.fda.gov/food/natural-toxins-food/mycotoxins>, 2024. Accessed: 2025-04-29.
- [60] V Fernández-Ibañez, A Soldado, A Martínez-Fernández, and B De la Roza-Delgado. Application of near infrared spectroscopy for rapid detection of aflatoxin b1 in maize and barley as analytical quality assessment. *Food Chemistry*, 113(2):629–634, 2009.
- [61] Eloisa D Caldas and Angela CS Silva. Mycotoxins in corn-based food products consumed in brazil: an exposure assessment for fumonisins. *Journal of Agricultural and Food Chemistry*, 55(19):7974–7980, 2007.
- [62] Khayria M Abdel-Gawad and AA Zohri. Fungal flora and mycotoxins of six kinds of nut seeds for human consumption in saudi arabia. *Mycopathologia*, 124:55–64, 1993.
- [63] PD Andrade, M Homem de Mello, JA França, and ED Caldas. Aflatoxins in food products consumed in brazil: a preliminary dietary risk assessment. *Food Additives & Contaminants: Part A*, 30(1):127–136, 2013.
- [64] Shahzad Z Iqbal. Mycotoxins in food, recent development in food analysis and future challenges; a review. *Current Opinion in Food Science*, 42:237–247, 2021.
- [65] Susan J MacDonald, Sharron Anderson, Paul Brereton, Roger Wood, and Andrew Damant. Determination of zearalenone in barley, maize and wheat flour, polenta, and maize-based baby food by immunoaffinity column cleanup with liquid chromatography: interlaboratory study. *Journal of AOAC International*, 88(6):1733–1740, 2005.
- [66] Carlo Brera, Francesca Debegnach, Valentina Minardi, Elena Pannunzi, Barbara De Santis, and Marina Miraglia. Immunoaffinity column cleanup with liquid chromatography for determination of aflatoxin b1 in corn samples: interlaboratory study. *Journal of AOAC International*, 90(3):765–772, 2007.

- [67] Antonello Santini, Rosalia Ferracane, Maria Carmela Somma, Alejandro Aragón, and Alberto Ritieni. Multitoxin extraction and detection of trichothecenes in cereals: an improved lc-ms/ms approach. *Journal of the Science of Food and Agriculture*, 89(7):1145–1153, 2009.
- [68] Barbara De Santis, Francesca Debegnach, Emanuela Gregori, Simona Russo, Francesca Marchegiani, Gabriele Moracci, and Carlo Brera. Development of a lc-ms/ms method for the multi-mycotoxin determination in composite cereal-based samples. *Toxins*, 9(5):169, 2017.
- [69] Dong-Ho Kim, Sung-Yong Hong, Jea Woo Kang, Sung Min Cho, Kyu Ri Lee, Tae Kyung An, Chan Lee, and Soo Hyun Chung. Simultaneous determination of multi-mycotoxins in cereal grains collected from south korea by lc/ms/ms. *Toxins*, 9(3):106, 2017.
- [70] Xian Zhang, Ke He, Yun Fang, Tong Cao, Narayan Paudyal, Xiao-feng Zhang, Hou-hui Song, Xiao-liang Li, and Wei-huan Fang. Dual flow immunochromatographic assay for rapid and simultaneous quantitative detection of ochratoxin a and zearalenone in corn, wheat, and feed samples. *Journal of Zhejiang University. Science. B*, 19(11):871, 2018.
- [71] Birgitta Maria Kunz, Felicitas Wanko, Sabine Kemmlein, Arnold Bahlmann, Sascha Rohn, and Ronald Maul. Development of a rapid multi-mycotoxin lc-ms/ms stable isotope dilution analysis for grain legumes and its application on 66 market samples. *Food Control*, 109:106949, 2020.
- [72] Ihuoma E Chibuzor-Onyema, Obinna T Ezeokoli, Michael Sulyok, Iviwe Notununu, Awanwee Petchkongkaew, Christopher T Elliott, Rasheed A Adeleke, Rudolf Krska, and Chibundu N Ezekiel. Metataxonomic analysis of bacterial communities and mycotoxin reduction during processing of three millet varieties into ogi, a fermented cereal beverage. *Food Research International*, 143:110241, 2021.
- [73] Laura Anfossi, Cristina Giovannoli, and Claudio Baggiani. Mycotoxin detection. *Current opinion in biotechnology*, 37:120–126, 2016.
- [74] Jiayi Chen, Pengjie Luo, Zhiwei Liu, Zixian He, Yimeng Pang, Hongtao Lei, Zhenlin Xu, Hong Wang, and Xiangmei Li. Rainbow latex microspheres lateral flow immunoassay with smartphone-based device for simultaneous detection of three mycotoxins in cereals. *Analytica Chimica Acta*, 1221:340138, 2022.
- [75] Qing Yu, Heng Li, Chenglong Li, Suxia Zhang, Jianzhong Shen, and Zhanhui Wang. Gold nanoparticles-based lateral flow immunoassay with silver staining for simultaneous detection of fumonisin b1 and deoxynivalenol. *Food Control*, 54:347–352, 2015.
- [76] Suquan Song, Na Liu, Zhiyong Zhao, Emmanuel Njumbe Ediage, Songling Wu, Changpo Sun, Sarah De Saeger, and Aibo Wu. Multiplex lateral flow immunoassay for mycotoxin determination. *Analytical chemistry*, 86(10):4995–5001, 2014.
- [77] Wanjun Zhang, Shusheng Tang, Yongpeng Jin, Chunjiang Yang, Lidong He, Jiayi Wang, and Yiqiang Chen. Multiplex sers-based lateral flow immunosensor for the detection of major mycotoxins in maize utilizing dual raman labels and triple test lines. *Journal of Hazardous Materials*, 393:122348, 2020.
- [78] Mohamad Kamal Abdul Kadir and Ibtisam E Tothill. Development of an electrochemical immunosensor for fumonisins detection in foods. *Toxins*, 2(4):382–398, 2010.

- [79] Zhiwei Liu, Qicheng Hua, Jin Wang, Zaoqing Liang, Jiahao Li, Jinxiao Wu, Xing Shen, Hongtao Lei, and Xiangmei Li. A smartphone-based dual detection mode device integrated with two lateral flow immunoassays for multiplex mycotoxins in cereals. *Biosensors and Bioelectronics*, 158:112178, 2020.
- [80] Yuhao Wu, Yaofeng Zhou, Hai Huang, Xirui Chen, Yuankui Leng, Weihua Lai, Xiaolin Huang, and Yonghua Xiong. Engineered gold nanoparticles as multicolor labels for simultaneous multi-mycotoxin detection on the immunochromatographic test strip nanosensor. *Sensors and Actuators B: Chemical*, 316:128107, 2020.
- [81] Ruipeng Chen, Shuang Li, Yunfeng Sun, Bingyang Huo, Yating Xia, Yingkai Qin, Shuning Li, Baodi Shi, Defu He, Jun Liang, et al. Surface-enhanced raman spectroscopy aptasensor for simultaneous determination of ochratoxin a and zearalenone using au@ ag core-shell nanoparticles and gold nanorods. *Microchimica Acta*, 188:1–9, 2021.
- [82] Ajay Kumar Yagati, Sachin Ganpat Chavan, Changyoon Baek, Min-Ho Lee, and Junhong Min. Label-free impedance sensing of aflatoxin b1 with polyaniline nanofibers/au nanoparticle electrode array. *Sensors*, 18(5):1320, 2018.
- [83] Zakir Hossain, Mark Busman, and Chris M Maragos. Immunoassay utilizing imaging surface plasmon resonance for the detection of cyclopiazonic acid (cpa) in maize and cheese. *Analytical and bioanalytical chemistry*, 411:3543–3552, 2019.
- [84] Xian Zhang, Chao-Rui Li, Wei-Cheng Wang, Jian Xue, Ya-Ling Huang, Xian-Xian Yang, Bin Tan, Xi-Peng Zhou, Chuang Shao, Shi-Jia Ding, et al. A novel electrochemical immunosensor for highly sensitive detection of aflatoxin b1 in corn using single-walled carbon nanotubes/chitosan. *Food chemistry*, 192:197–202, 2016.
- [85] Jessica MD Machado, Ruben RG Soares, Virginia Chu, and João P Conde. Multiplexed capillary microfluidic immunoassay with smartphone data acquisition for parallel mycotoxin detection. *Biosensors and Bioelectronics*, 99:40–46, 2018.
- [86] Baoshan He and Xiaohai Yan. An amperometric zearalenone aptasensor based on signal amplification by using a composite prepared from porous platinum nanotubes, gold nanoparticles and thionine-labelled graphene oxide. *Microchimica Acta*, 186:1–10, 2019.
- [87] Baoshan He and Xiaohai Yan. A “signal-on” voltammetric aptasensor fabricated by hcpt@aunfs/pei-rgo and fe3o4nrs/rgo for the detection of zearalenone. *Sensors and Actuators B: Chemical*, 290:477–483, 2019.
- [88] Lin Lu and Sundaram Gunasekaran. Dual-channel ito-microfluidic electrochemical immunosensor for simultaneous detection of two mycotoxins. *Talanta*, 194:709–716, 2019.
- [89] Kundan Kumar Mishra, Vikram Narayanan Dhamu, Abhinav Kokala, Sriram Muthukumar, and Shalini Prasad. Advancing food safety: Two-plex electrochemical biosensor for mycotoxin detection in food matrices. *Biosensors and Bioelectronics: X*, page 100626, 2025.
- [90] European Commission. Alert and cooperation network annual report 2022. https://food.ec.europa.eu/safety/rasff_en, 2023. Accessed: 2025-04-29.
- [91] Herbert Fraenkel, Patrick B Gough, and William S Maughan. Method and apparatus for sorting agricultural products, May 20 1980. US Patent 4,203,522.

- [92] Maksim N Moskovskiy, Aleksey V Sibirev, Anatoly A Gulyaev, Stanislav A Gerasimenko, Sergey I Borzenko, Maria M Godyaeva, Oleg V Noy, Egor I Nagaev, Tatiana A Matveeva, Ruslan M Sarimov, et al. Raman spectroscopy enables non-invasive identification of mycotoxins p. fusarium of winter wheat seeds. In *Photonics*, volume 8, page 587. MDPI, 2021.
- [93] Yong-Kyoung Kim, Insuck Baek, Kyung-Min Lee, Jianwei Qin, Geonwoo Kim, Byeung Kon Shin, Diane E Chan, Timothy J Herrman, Soon-kil Cho, and Moon S Kim. Investigation of reflectance, fluorescence, and raman hyperspectral imaging techniques for rapid detection of aflatoxins in ground maize. *Food Control*, 132:108479, 2022.
- [94] M Aoun, C Siegel, GL Windham, WP Williams, and RJ Nelson. Application of reflectance spectroscopy to identify maize genotypes and aflatoxin levels in single kernels. *World Mycotoxin Journal*, 15(4):327–342, 2022.
- [95] Xianbin Cheng, Andrea Vella, and Matthew J Stasiewicz. Classification of aflatoxin contaminated single corn kernels by ultraviolet to near infrared spectroscopy. *Food Control*, 98:253–261, 2019.
- [96] Tom C Pearson, Donald T Wicklow, Elizabeth B Maghirang, Feng Xie, and Floyd E Dowell. Detecting aflatoxin in single corn kernels by transmittance and reflectance spectroscopy. *Transactions of the ASAE*, 44(5):1247, 2001.
- [97] TC Pearson, DT Wicklow, and MC Pasikatan. Reduction of aflatoxin and fumonisin contamination in yellow corn by high-speed dual-wavelength sorting. *Cereal Chemistry*, 81(4):490–498, 2004.
- [98] Matthew J Stasiewicz, Titilayo DO Falade, Murithi Mutuma, Samuel K Mutiga, Jagger JW Harvey, Glen Fox, Tom C Pearson, James W Muthomi, and Rebecca J Nelson. Multi-spectral kernel sorting to reduce aflatoxins and fumonisins in kenyan maize. *Food Control*, 78:203–214, 2017.
- [99] Gregor Kos, Markus Sieger, David McMullin, Celine Zahradnik, Michael Sulyok, Tuba Öner, Boris Mizaikoff, and Rudolf Krska. A novel chemometric classification for ftir spectra of mycotoxin-contaminated maize and peanuts at regulatory limits. *Food Additives & Contaminants: Part A*, 33(10):1596–1607, 2016.
- [100] Kyung-Min Lee, Jessica Davis, Timothy J Herrman, Seth C Murray, and Youjun Deng. An empirical evaluation of three vibrational spectroscopic methods for detection of aflatoxins in maize. *Food Chemistry*, 173:629–639, 2015.
- [101] Brigitta Gaspardo, Stefania Del Zotto, Emanuela Torelli, SR Cividino, Giuseppe Firrao, Giacomo Della Riccia, and Bruno Stefanon. A rapid method for detection of fumonisins b1 and b2 in corn meal using fourier transform near infrared (ft-nir) spectroscopy implemented with integrating sphere. *Food Chemistry*, 135(3):1608–1612, 2012.
- [102] Bei Wang, Fei Shen, Xueming He, Yong Fang, Qiuhui Hu, and Xingquan Liu. Simultaneous detection of aspergillus moulds and aflatoxin b1 contamination in rice by laser induced fluorescence spectroscopy. *Food Control*, 145:109485, 2023.
- [103] Haibo Yao, Zuzana Hruska, Russell Kincaid, Robert L Brown, Deepak Bhatnagar, and Thomas E Cleveland. Detecting maize inoculated with toxigenic and atoxigenic fungal strains with fluorescence hyperspectral imagery. *Biosystems Engineering*, 115(2):125–135, 2013.

- [104] Fengle Zhu, Haibo Yao, Zuzana Hruska, Russell Kincaid, Robert L Brown, Deepak Bhatnagar, and Thomas E Cleveland. Integration of fluorescence and reflectance visible near-infrared (vnir) hyperspectral images for detection of aflatoxins in corn kernels. *Transactions of the ASABE*, 59(3):785–794, 2016.
- [105] Haibo Yao, Zuzana Hruska, Russell Kincaid, Robert L Brown, Deepak Bhatnagar, and Thomas E Cleveland. Svm-based feature extraction and classification of aflatoxin contaminated corn using fluorescence hyperspectral data. In *2012 4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4. IEEE, 2012.
- [106] Haibo Yao, Zuzana Hruska, Russell Kincaid, R Brown, T Cleveland, and D Bhatnagar. Correlation and classification of single kernel fluorescence hyperspectral data with aflatoxin concentration in corn kernels inoculated with aspergillus flavus spores. *Food additives and contaminants*, 27(5):701–709, 2010.
- [107] Wei Liu, Haiyang Deng, Yule Shi, Changhong Liu, and Lei Zheng. Application of multi-spectral imaging combined with machine learning for rapid and non-destructive detection of zearalenone (zen) in maize. *Available at SSRN 4116066*, 2023.
- [108] Hyunjung Min and Byoung-Kwan Cho. Spectroscopic techniques for nondestructive detection of fungi and mycotoxins in agricultural materials: A review. *Journal of Biosystems Engineering*, 40(1):67–77, 2015.
- [109] Pauline Ong, I-Chun Tung, Ching-Feng Chiu, I-Lin Tsai, Hsi-Chang Shih, Suming Chen, and Yung-Kun Chuang. Determination of aflatoxin b1 level in rice (oryza sativa l.) through near-infrared spectroscopy and an improved simulated annealing variable selection method. *Food Control*, 136:108886, 2022.
- [110] Han Song, Feng Li, Peiwen Guang, Xinhao Yang, Huanyu Pan, and Furong Huang. Detection of aflatoxin b1 in peanut oil using attenuated total reflection fourier transform infrared spectroscopy combined with partial least squares discriminant analysis and support vector machine models. *Journal of Food Protection*, 84(8):1315–1320, 2021.
- [111] C-L Lai, Y-M Fuh, and DY-C Shih. Detection of mycotoxin patulin in apple juice. *Journal of Food and Drug Analysis*, 8(2):5, 2000.
- [112] Pranita Jaiswal, Shyam Narayan Jha, Jaspreet Kaur, Anjan Borah, and HG Ramya. Detection of aflatoxin m1 in milk using spectroscopy and multivariate analyses. *Food Chemistry*, 238:209–214, 2018.
- [113] Haibo Yao, Zuzana Hruska, Russell Kincaid, Ambrose Ononye, Robert L Brown, and Thomas E Cleveland. Spectral angle mapper classification of fluorescence hyperspectral image for aflatoxin contaminated corn. In *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, pages 1–4. IEEE, 2010.
- [114] Haibo Yao, Zuzana Hruska, Russell Kincaid, Robert L Brown, Deepak Bhatnagar, and Thomas E Cleveland. Hyperspectral image classification and development of fluorescence index for single corn kernels infected with aspergillus flavus. *Transactions of the ASABE*, 56(5):1977–1988, 2013.
- [115] Hans Pettersson and Lena Åberg. Near infrared spectroscopy for determination of mycotoxins in cereals. *Food Control*, 14(4):229–232, 2003.

- [116] Marco Beyer, Friederike Pogoda, Franz K Ronellenfitsch, Lucien Hoffmann, and Thomas Udelhoven. Estimating deoxynivalenol contents of wheat samples containing different levels of fusarium-damaged kernels by diffuse reflectance spectrometry and partial least square regression. *International Journal of Food Microbiology*, 142(3):370–374, 2010.
- [117] Feng Jin, Guihua Bai, Dadong Zhang, Yanhong Dong, Lingjian Ma, William Bockus, and Floyd Dowell. Fusarium-damaged kernels and deoxynivalenol in fusarium-infected us winter wheat. *Phytopathology*, 104(5):472–478, 2014.
- [118] CS Tibola, JMC Fernandes, and R Delanora. Predicting wheat mycotoxin content using near-infrared reflectance spectroscopy. In *Proceedings of the 2010 National Fusarium Head Blight Forum, December*, pages 7–9, 2010.
- [119] Denize Tyska, Adriano Olnei Mallmann, Juliano Kobs Vidal, Carlos Alberto Araújo de Almeida, Luciane Tourem Gressler, and Carlos Augusto Mallmann. Multivariate method for prediction of fumonisins b1 and b2 and zearalenone in brazilian maize using near infrared spectroscopy (nir). *PloS one*, 16(1):e0244957, 2021.
- [120] William Appaw, John-Lewis Zinia Zaukuu, Balkis Aouadi, Eric Tetteh Mensah, Ibok Nsa Oduro, and Zoltan Kovacs. Predicting aflatoxin contamination in white and yellow maize using vis/nir spectroscopy combined with pca-lda and plsr models through aquaphotomics approaches. *Applied Food Research*, page 100841, 2025.
- [121] Annalisa De Girolamo, Salvatore Cervellieri, Angelo Visconti, and Michelangelo Pascale. Rapid analysis of deoxynivalenol in durum wheat by ft-nir spectroscopy. *Toxins*, 6(11):3129–3143, 2014.
- [122] Marta Bradaáčová, Kristýna Bezděková, and Pavlína Smutíná. Použití ft-nir spectroscopy for the determination of mycotoxins in barley grain by ft-nir spectroscopy, 2020.
- [123] Wei Liu, Haiyang Deng, Yule Shi, Changhong Liu, and Lei Zheng. Application of multi-spectral imaging combined with machine learning methods for rapid and non-destructive detection of zearalenone (zen) in maize. *Measurement*, 203:111944, 2022.
- [124] Joseph R Lakowicz. *Introduction to fluorescence*. Springer, 1999.
- [125] Indy Magnus. Combining spectroscopy and machine learning processing for precise sample classification. *VUB*, 2020.
- [126] Robert K Neely, Jochem Deen, and Johan Hofkens. Optical mapping of dna: Single-molecule-based methods for mapping genomes. *Biopolymers*, 95(5):298–311, 2011.
- [127] Feifei Tao, Haibo Yao, Zuzana Hruska, Loren W Burger, Kanniah Rajasekaran, and Deepak Bhatnagar. Recent development of optical methods in rapid and non-destructive detection of aflatoxin and fungal contamination in agricultural products. *TrAC Trends in Analytical Chemistry*, 100:65–81, 2018.
- [128] Eugene Hecht. *Optics, Fifth Edition*. Pearson, 2017.
- [129] Claudia Rasch, Maike Böttcher, and Michael Kumke. Determination of aflatoxin b 1 in alcoholic beverages: comparison of one-and two-photon-induced fluorescence. *Analytical and Bioanalytical Chemistry*, 397:87–92, 2010.
- [130] Romer Labs. Innovative diagnostic solutions for food and feed safety. <https://www.romerlabs.com/en/>. Accessed: Feb 9th 2025.

- [131] Romer Labs. Product name: Aflatoxin in corn, low level. <https://plm.erber.group/WDD/?partNumber=10007113&lotNumber=122294>, 2024. Part Number:10007113, lot: 122294.
- [132] Romer Labs. Product name: Deoxynivalenol in corn, high level. <https://plm.erber.group/WDD/?partNumber=10007115&lotNumber=123359>, 2024. Part Number:10007115, lot: 123359.
- [133] Romer Labs. Product name: Zearalenone in corn, low level. https://www.romerlabs.com/at_en/shop/zearalenone-in-corn-low-level/#theme.catalog.product.additional.information.documents.detailed, 2024. Part Number:10003624, lot: M21055Z.
- [134] Romer Labs. Product name: Afla, don, zea and fum in corn, mid level. https://www.romerlabs.com/at_en/shop/zearalenone-in-corn-low-level/#theme.catalog.product.additional.information.documents.detailed, 2024. Part Number:10007114, lot: 123368.
- [135] Romer Labs. Product name: Don, t-2, ht-2 and ota in wheat, low level. <https://plm.erber.group/WDD/?partNumber=10007116&lotNumber=123359>, 2024. Part Number:10007116, lot: 123359.
- [136] Romer Labs. Wheat, blank. https://www.romerlabs.com/us_en/shop/wheat-blank/, 2025. Part Number:10003611, lot: 123359.
- [137] Joseph A Curcio and Charles C Petty. The near infrared absorption spectrum of liquid water. *Journal of the optical society of America*, 41(5):302–304, 1951.
- [138] Elias Alisaac, Jan Behmann, Anna Rathgeb, Petr Karlovsky, Heinz-Wilhelm Dehne, and Anne-Katrin Mahlein. Assessment of fusarium infection and mycotoxin contamination of wheat kernels and flour using hyperspectral imaging. *Toxins*, 11(10):556, 2019.
- [139] Wannes Luts de Marelalere. Github repository with the full code (classification/regression) modeling pipeline. <https://github.com/wannes122/MasterThesis.git>, 2025.
- [140] Trilogy Analytical Laboratory. Certified reference standards. <https://trilogylab.com/collections/certified-reference-standards>, 2025. Accessed: 2025-05-26.
- [141] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, volume 2. Springer, 2009.

Generative AI, specifically gpt-4, was used throughout the development of this thesis work. It provided support in various aspects, including coding, data processing, and enhancing the efficiency and quality of the literature review by summarizing scientific papers, organizing information into tables, and identifying relevant new sources of interest. In addition, it contributed to the writing process by assisting with text structuring, improving clarity and grammar, and ensuring consistent formatting. All scientific results, analyses, and conclusions presented in this thesis were independently conducted and critically evaluated by the author, whom is responsible for their accuracy and validity.

Acknowledgments

I would like to express my deepest gratitude to **Prof. Dr. Lien Smeesters** for her invaluable support, expert guidance, and consistent, constructive feedback throughout this thesis. I truly appreciated how involved she remained despite her many responsibilities, and I am sincerely grateful for the time and care she dedicated to supporting me.

I would also like to thank **Ir. Indy Magnus**, my supervisor, for his practical help, insightful suggestions on coding challenges, and valuable ideas for improving this work. His input was instrumental in shaping the technical depth of the thesis. A warm thank you goes to **Prof. Dr. Francesca Venturini** for her helpful comments and thoughtful suggestions, which greatly contributed to refining this work.

I am also immensely grateful to my **parents**, whose support throughout my studies has made this journey possible. Special thanks go to my girlfriend, **Hai Anh Bui**, for her encouragement, patience, and for helping me stay focused during the long hours of writing—even when motivation was low, *quac quac*.

Lastly, I want to thank the **Vrije Universiteit Brussel** and all the professors who have taught and mentored me over the years. Their dedication and expertise have helped shape me into the engineer I am today, thank you.

Give thanks to G-d