Project FY24-FR-002: Early Detection of Fusarium Damaged Kernels Resilient to Barley and Wheat Malting

---

1. **What are the major goals and objectives of the research project?**
   The primary goal of this project is to develop rapid and non-destructive methods to identify Fusarium damaged kernels (FDKs) existing in wheat and barley and developed during malting. These FDKs are the major contributor to mycotoxin occurrence and the increased levels during malting. Specific objectives include: (i) Developing a wet chemical method to analyze *Fusarium* DNA and mycotoxin levels in single kernel, (ii)Developing hyperspectral imaging (HSI)-based models for rapid detection and quantification of FDKs in barley and wheat at the single-kernel level, (iii) Building predictive models using machine learning to forecast DON and Fusarium DNA levels in wheat, barley and during malting.

2. **What was accomplished under these goals or objectives?** *(For each major goal/objective, address these three items below.)*

   What were the major activities?
   (1) Obtaining samples. FHB infected barley and wheat samples were obtained from multiple sources, including commercial samples, variety trial nurseries, and inoculated nurseries.
   (2) Micro-Malting. Two FHB-infected barley samples were malted in the lab.
   (3) Evaluating the effects of kernel orientation, placement angle, and penetration depth on moisture prediction in barley using hyperspectral imaging. To accurately quantify DON content in individual kernels, variation in kernel moisture content must be taken into account. Since moisture levels can range from 5% to 45% during malting, a moisture prediction model was built for normalizing DON content on a dry weight basis.
   (4) Quantifying barley FDKs using wet chemistry to serve as standard samples for model development. One hundreds of individual kernels (N=100) were first scanned by HSI, followed by quantification of DON content using GC-MS and *Fusarium* DNA levels using quantitative PCR.
   (5) Developing a predictive model by combining HSI with machine learning to identify FDKs in barley. **Figure 1** shows the workflow of building the FDK prediction model. In this study, a predictive model for DON concentration in individual barley kernels was developed using raw hyperspectral data as input.

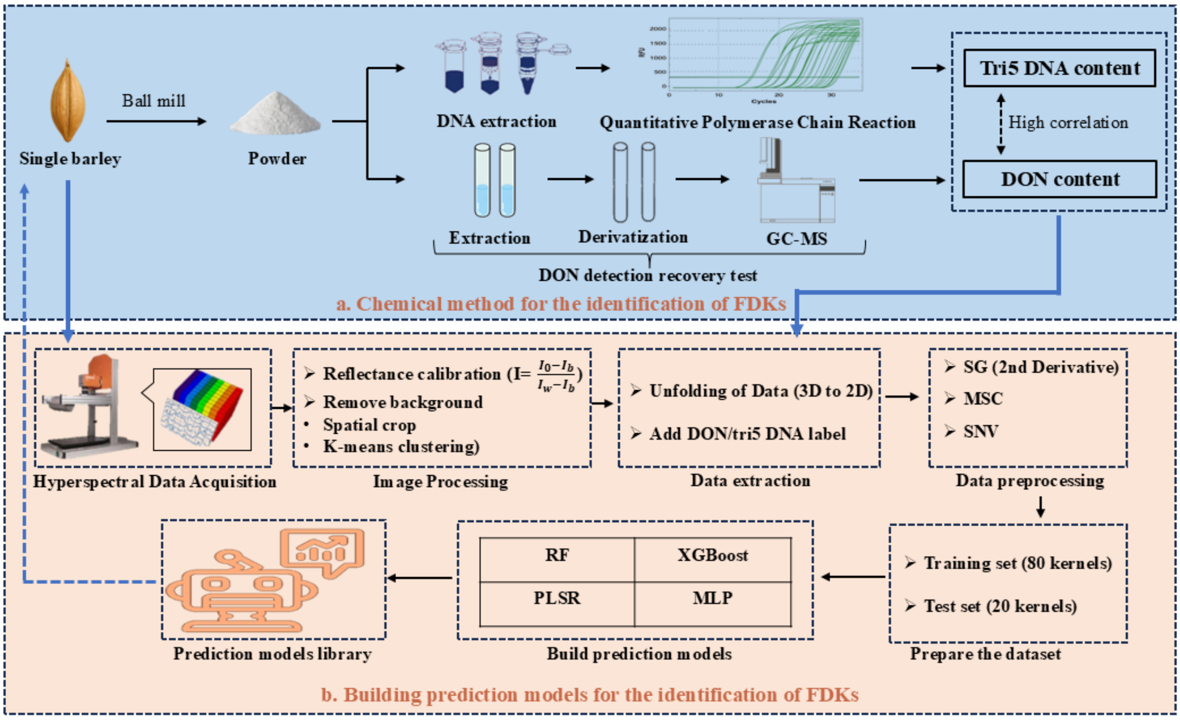

Figure 1. The workflow of building a model to predict FDKs using HIS and machine learning

What were the significant results?

1. The model was built to predict the effects of kernel orientation, placement angle, and penetration depth on moisture prediction in barley using hyperspectral imaging. 3D CNN is well-suited for HSI analysis due to their ability to simultaneously extract spatial and spectral features from volumetric data. In this study, we designed and evaluated four independent 3D CNN models to predict moisture content from hyperspectral data, considering additional contextual factors such as grain orientation and placement angle, which can influence reflectance and thus predictive performance. Models B-D had more dispersed predictions, especially at high moisture levels, confirming the reduced accuracy compared to Model A. When grain angle was added in Model B, model performance slightly declined ($R^2 = 0.9007$, MAE = 1.58%). A similar decrease appeared in Model C, which used orientation alone ($R^2 = 0.9166$, MAE = 1.49%). The combined Model D, which incorporated both angle and orientation, performed worst ($R^2 = 0.8906$, MAE = 1.59%, RMSE = 2.68%). This suggests that adding angle and orientation as scalar inputs, without deeper feature integration (e.g., embedding or attention layers), weakened performance.

Table 1. Performance of four 3D CNN models with different input features for moisture prediction

Model	Input feature	R^2	MAE (%)	RMSE (%)
A	HSI only	0.9227	1.40	2.11
B	HSI + Angle	0.9007	1.58	2.46
C	HSI + Orientation	0.9166	1.49	2.31
D	HSI + Angle + Orientation	0.8906	1.59	2.68

2. The correlation analysis between DON and *Fusarium* tri5DNA in 100 individual barley kernels revealed a strong positive relationship, with a Pearson correlation coefficient (r) of 0.823 and a highly significant *p*-value ($p = 8.79 \times 10^{-25}$), as shown in **Figure 2**. This strong correlation demonstrates that it is feasible to evaluate the extent of Fusarium infection at the single-kernel level by combining chemical (DON) and molecular (tri5DNA) measurements.

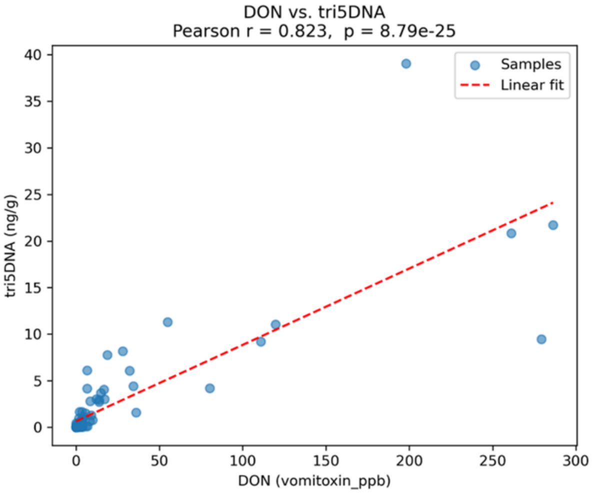

Figure 2. Correlation between DON and *Fusarium* Tri5 DNA in single barley kernels (N=100)

3. A t-distributed Stochastic Neighbor Embedding (t-SNE) plot, as a visualization technique, is used to reduce high-dimensional data to two dimensions, making it easier to observe patterns, clusters, or separations to group samples based on DON levels. As shown in **Figure 3**, the t-SNE plot clearly illustrates the separation of samples with varying DON concentrations, indicating that the raw spectral data contain meaningful structure associated with toxin levels.

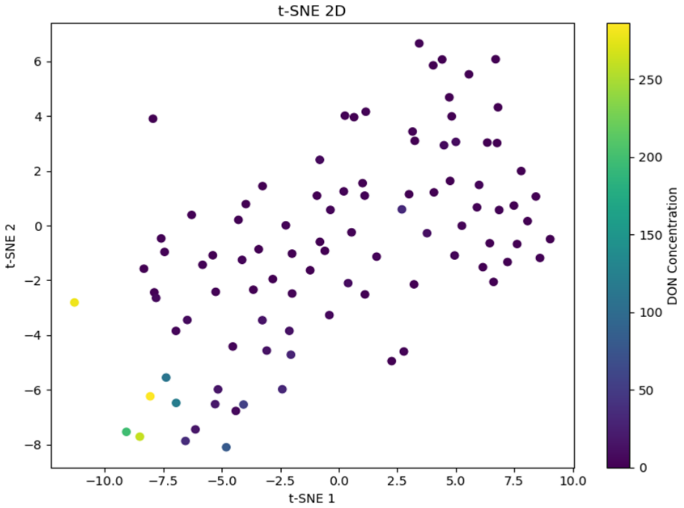

Figure 3. The t-SNE mapping of DON levels in barley kernels (HSI → DON Model)

4. The model selection process involved comparing several machine learning algorithms, and performance metrics including the R^2 and MAE were summarized in **Table 2**. Among all tested models, the multilayer perceptron (MLP) model using raw spectral data achieved the best performance, with an R^2 of 0.9877 and MAE of 3.07% on the training set, and an R^2 of 0.8506 and MAE of 12.31% on the test set. These results suggest that the MLP model effectively captured the nonlinear relationship between hyperspectral features and DON concentration, with high predictive accuracy on the training set and reasonable generalization on the unseen test set. Further visualization of the model performance is shown in **Figure 4**. In the scatter plots for the MLP model, predictions on the training set aligned closely along the 1:1 line, demonstrating excellent fit. On the test set, the model still captured the general trend of DON variation, although slight underestimation was observed for some high-concentration samples. These findings confirm that raw hyperspectral data, even without preprocessing, can provide sufficient information for DON prediction, and that the MLP model can serve as a reliable and practical tool for non-destructive mycotoxin assessment in single barley kernels.

Table 2. Model performance for predicting DON and *Fusarium* Tri5 DNA from HIS

Model type	Best algorithm	Training		Testing	
		R^2	MAE	R^2	MAE
HSI → DON	MLP	0.9877	3.0686	0.8506	12.3102
HSI → Tri5 DNA	PLSR	0.8357	1.0457	0.5667	3.4259
HSI → DON + Tri5 DNA	RF	0.9673 (DON)	3.7500 (DON)	0.7575 (DON)	15.3617 (DON)
		0.8977 (Tri5 DNA)	0.5342 (Tri5 DNA)	0,4769 (Tri5 DNA)	2.6547 (Tri5 DNA)

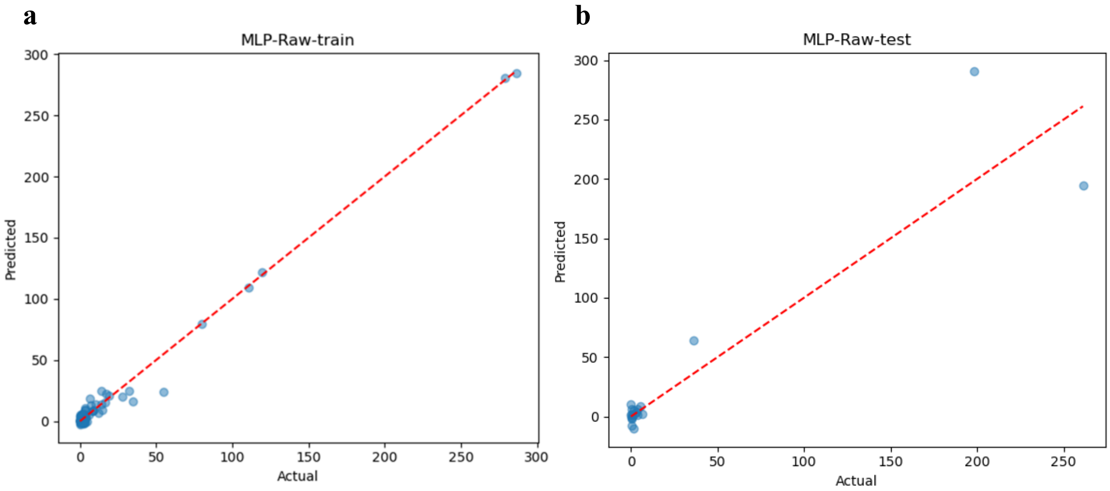

Figure 4. Performance of MLP of using HSI to predict DON. (a) Training set; (b)Test set.

List key outcomes or other achievements.

This study demonstrated that barley kernel orientation, angular variation, and light penetration significantly affect the performance of hyperspectral imaging (HSI) for moisture prediction. Using SWIR HSI and PbS quantum dots, groove-up kernels exhibited consistently higher reflectance in moisture-sensitive regions (1400-2200 nm), while groove-down orientations introduced greater spectral variability during rotation. Light penetration tests further confirmed that barley hulls transmitted more light at 1200 and 1800 nm, whereas intact kernels blocked most light, underscoring the impact of physical structure on spectral behavior.

Four 3D CNN models were developed based on XGBoost-selected wavelengths: Model A (HSI only), Model B (HSI + angle), Model C (HSI + orientation), and Model D (HSI + both). While Models B-D showed slight improvements in validation accuracy, Model A proved to be the most stable and practical ($R^2 = 0.9714$), offering high performance without requiring external metadata-making it well-suited for high-throughput, automated applications.

A strong correlation ($r = 0.823$) was also observed between DON content and *Fusarium* Tri5 DNA at the single-kernel level, validating the use of chemical markers for identifying infected kernels. Recovery tests confirmed reliable quantification across contamination levels. The multilayer perceptron (MLP) model using raw HSI data achieved the best training performance for DON prediction ($R^2 = 0.99$, MAE = 3.07%), though test accuracy was limited by the underrepresentation of highly contaminated samples. Overall, the findings support the integration of HSI and deep learning for rapid, non-destructive assessment of moisture and DON in individual barley kernels.

3. **What opportunities for training and professional development has the project provided?**

Provided hands-on training in hyperspectral imaging operation and data analysis to graduate students and researchers. They gained experience in wet chemistry methods, including GC-MS and qPCR techniques and advanced machine learning tools and modeling strategies. It also provided opportunities to present preliminary findings at scientific conferences and gain professional feedback.

4. **How have the results been disseminated to communities of interest?**

(1) Manuscripts have been submitted and also in the preparation for the submission to peer-reviewed journals focusing on cereal science and food safety areas.
(2) Research findings have been presented at National FHB Forum during December 8-10, 2024 at Austin, TX.

These publications and presentations help share results with academic researchers, barley breeders, and malting industry stakeholders.

5. What do you plan to do during the next reporting period to accomplish the goals and objectives?

(1) Expand sample collection to include additional barley and wheat kernels, particularly targeting more high-DON and highly infected samples.

(2) Build the models for predict FDKs in FHB infected wheat samples.

(3) Conduct further micro-malting trials to monitor DON and *Fusarium* growth through the entire malting process.

(3) Refine hyperspectral data processing methods to improve kernel classification accuracy.

(4) Retrain machine learning models with larger, more diverse datasets to enhance model robustness and prediction performance.

(5) Continue presenting results at scientific conferences and engage industry partners for technology transfer.