

Report for 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.

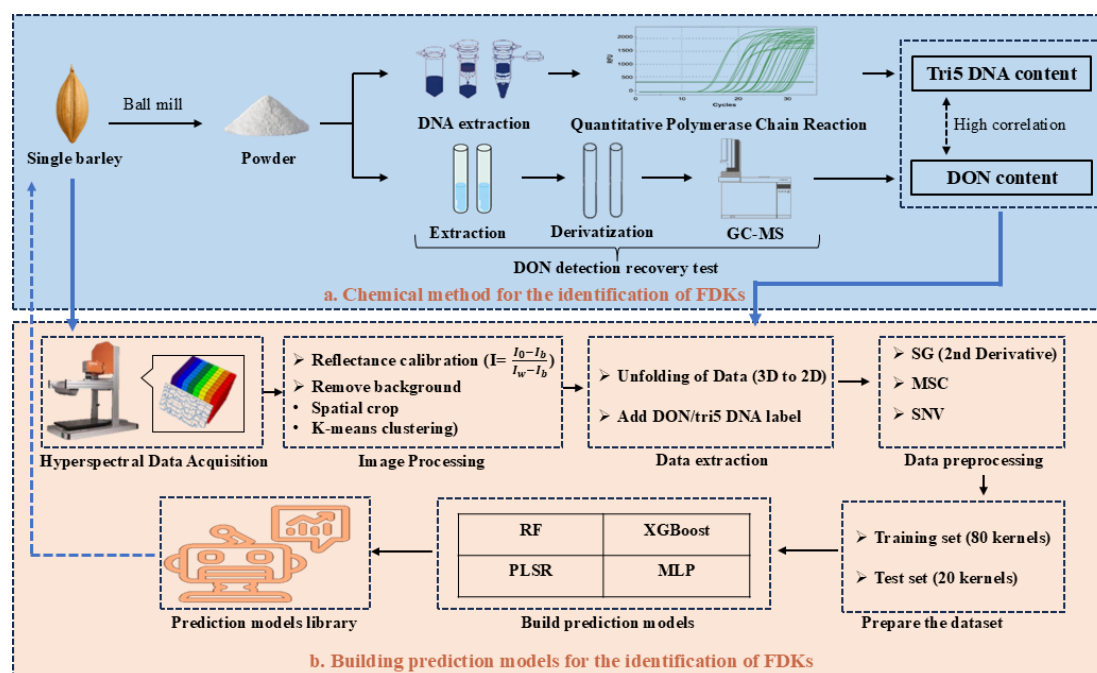


Figure 1. The workflow of building a model to predict FDKs using HIS and machine learning.

What were the significant results?

DON contamination in barley poses a serious food safety concern because toxin accumulation is highly heterogeneous among kernels, with a small portion of fusarium damaged kernels potentially driving the risk of an entire grain lot. Conventional DON detection methods are accurate but destructive, time-consuming, and costly, limiting their suitability for rapid single kernel screening. This study developed a non-destructive framework combining short wave infrared hyperspectral imaging (SWIR HSI) with machine learning to predict DON content in individual barley kernels.

Figure 2. shows that kernel-DON concentration exhibits a highly skewed, right-tailed distribution, with most kernels containing low levels and only a small fraction showing extremely high concentrations. This pronounced long-tail pattern reflects substantial kernel-to-kernel variation in fungal colonization and toxin biosynthesis, consistent with the heterogeneous nature of FHB infection within a single grain lot. Such heterogeneity has been reported before, it showed that DON distribution in cereals is uneven due to localized fungal colonization and variable toxin biosynthesis across kernels.

In this study, four regression models, PLSR, MLP, RF, and XGBoost, were constructed and evaluated for the prediction of DON contents in individual barley kernels based on HSI data. The performance metrics for all models on both training and test sets are summarized in **Table 1**. On the test set, RF achieved the most stable generalization ($R^2 = 0.869$, $\Delta R^2 \approx -0.001$), followed by XGBoost ($R^2 = 0.854$, $\Delta R^2 = 0.145$), PLSR ($R^2 = 0.713$), and MLP ($R^2 = 0.666$, $\Delta R^2 = 0.328$).

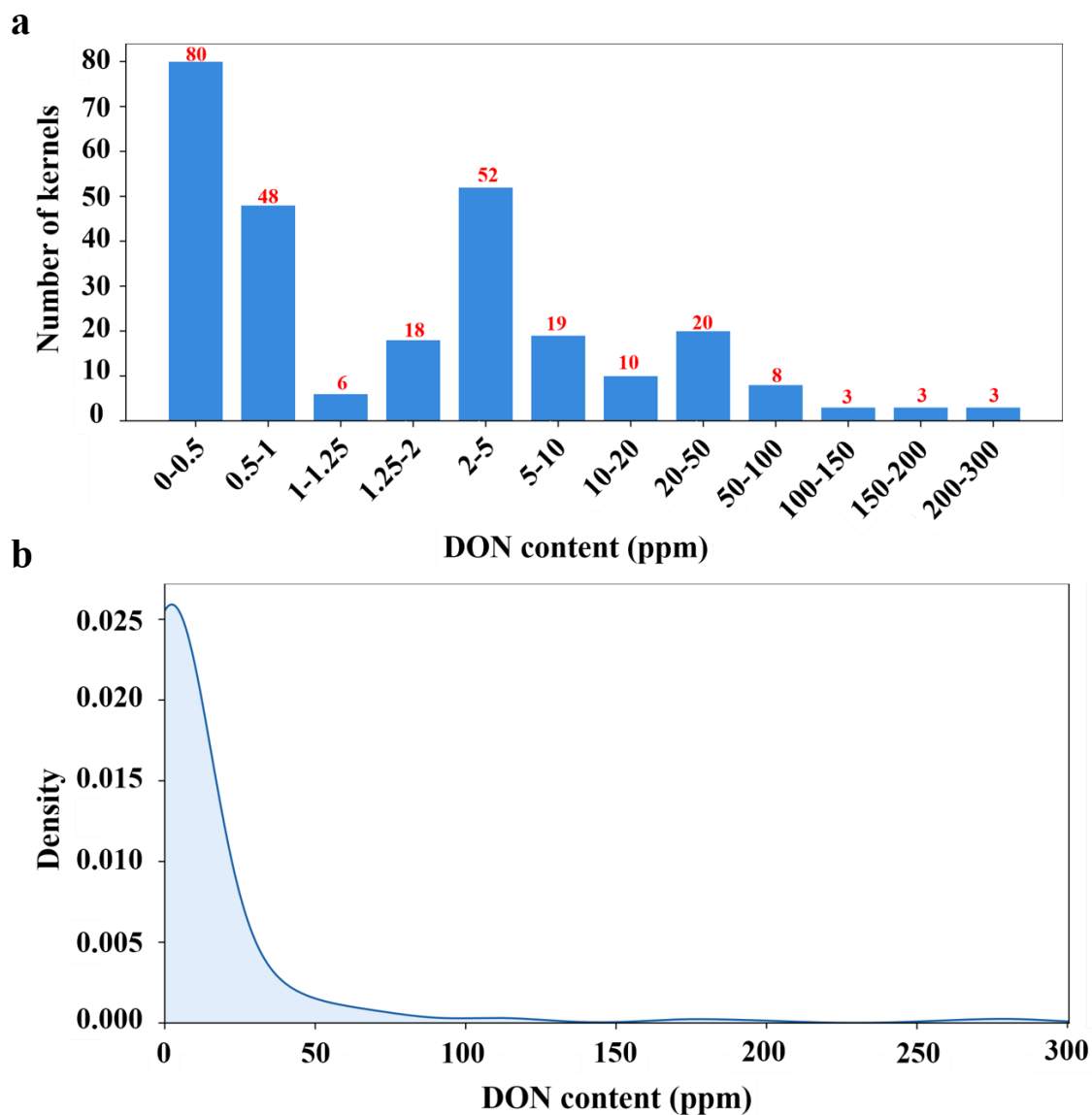


Figure 2. The distribution of kernel-DON concentration among individual barley kernels. (a) Histogram showing the frequency distribution of DON concentration; (b) Kernel density curve illustrating the skewed distribution of DON concentration.

Table 1. The performance of the kernel-DON content predicting regression models.

Models	Data set	R ²	RMSE	MAE	Slope
PLSR	train	0.447	28.885	16.725	0.447
	test	0.713	17.314	13.724	0.857
MLP	train	0.994	3.058	1.546	0.994
	test	0.666	18.669	8.098	1.372
RF	train	0.868	14.130	4.792	0.795
	test	0.869	11.672	5.692	1.096
XGBoost	train	0.999	1.084	0.108	0.999
	test	0.854	12.352	6.010	1.166

To select feature wavelength for the kernel DON detection, original 272 bands were reduced to 8 wavelengths (**Table 2**). Representative HSI data are illustrated in **Figure 3**. The raw push broom scan (Fig. 3a) captures all kernels under SWIR illumination prior to any processing, while the segmented image (Fig. 3b) confirms that background removal successfully isolated individual kernel regions without distorting kernel morphology. The false-color reflectance map at 1309.07 nm (Fig. 3c) reveals spatially heterogeneous reflectance distribution across and within kernels, with values ranging approximately from 0.2 to 0.8. This spatial variation indicates that meaningful spectral differences exist at the single-kernel level at this wavelength, consistent with its top-ranked importance across all three feature selection methods.

Table 2. Selected feature wavelengths for kernel-DON content prediction with corresponding Pearson correlation coefficients ($|r|$), RF feature importance, and SHAP importance values.

No.	Wavelength (nm)	$ r $	RF Importance	SHAP Importance
1	1309.07	0.638	0.101	1.392
2	1192.31	0.629	0.029	0.903
3	1208.99	0.627	0.020	0.846
4	1114.47	0.590	0.069	1.034
5	1131.15	0.620	0.018	1.140
6	1164.51	0.575	0.062	0.920
7	1259.03	0.545	0.082	1.940
8	1097.79	0.506	0.011	1.001

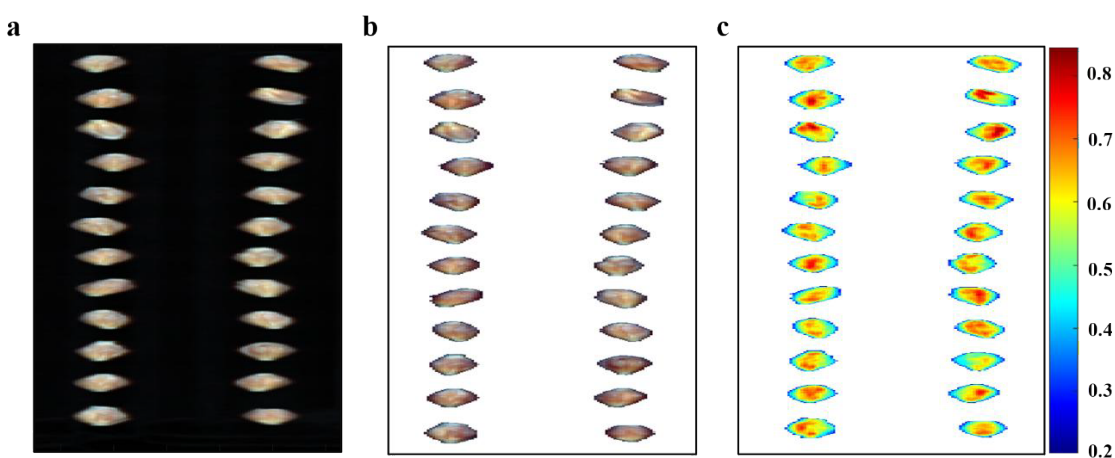


Fig. 3. HSI of individual barley kernels and spatial reflectance distribution at the consensus wavelength. (a) Raw scan image; (b) segmented kernel image after background removal; (c) false-color reflectance map at 1309.07 nm, a consensus wavelength identified by all three feature selection methods. Warmer colors (red) indicate higher reflectance and cooler colors (blue) indicate stronger absorption.

Following the identification of eight consensus wavelengths, RF and XGBoost models were retrained using these bands and compared with full-spectrum (272-band) and algorithm-specific top-20 models (Table 1 and Fig. 4). RF achieved its best performance with the top-20 wavelengths ($R^2 = 0.882$, RMSE = 11.056), slightly outperforming the full-spectrum model, whereas XGBoost performed best with the full spectrum ($R^2 = 0.854$, RMSE = 12.352). Reducing

the models to eight bands decreased test performance ($R^2 = 0.757$ for RF and 0.715 for XGBoost) and produced slopes greater than 1, indicating overestimation at high DON concentrations. Therefore, the eight-band models are not intended to replace the more accurate models but represent an engineering trade-off: a 97% reduction in spectral dimensionality enables lower-cost, less computationally demanding multispectral systems while retaining chemically interpretable features and sufficient performance for rapid, batch-level DON risk screening followed by confirmatory analysis of high-risk samples.

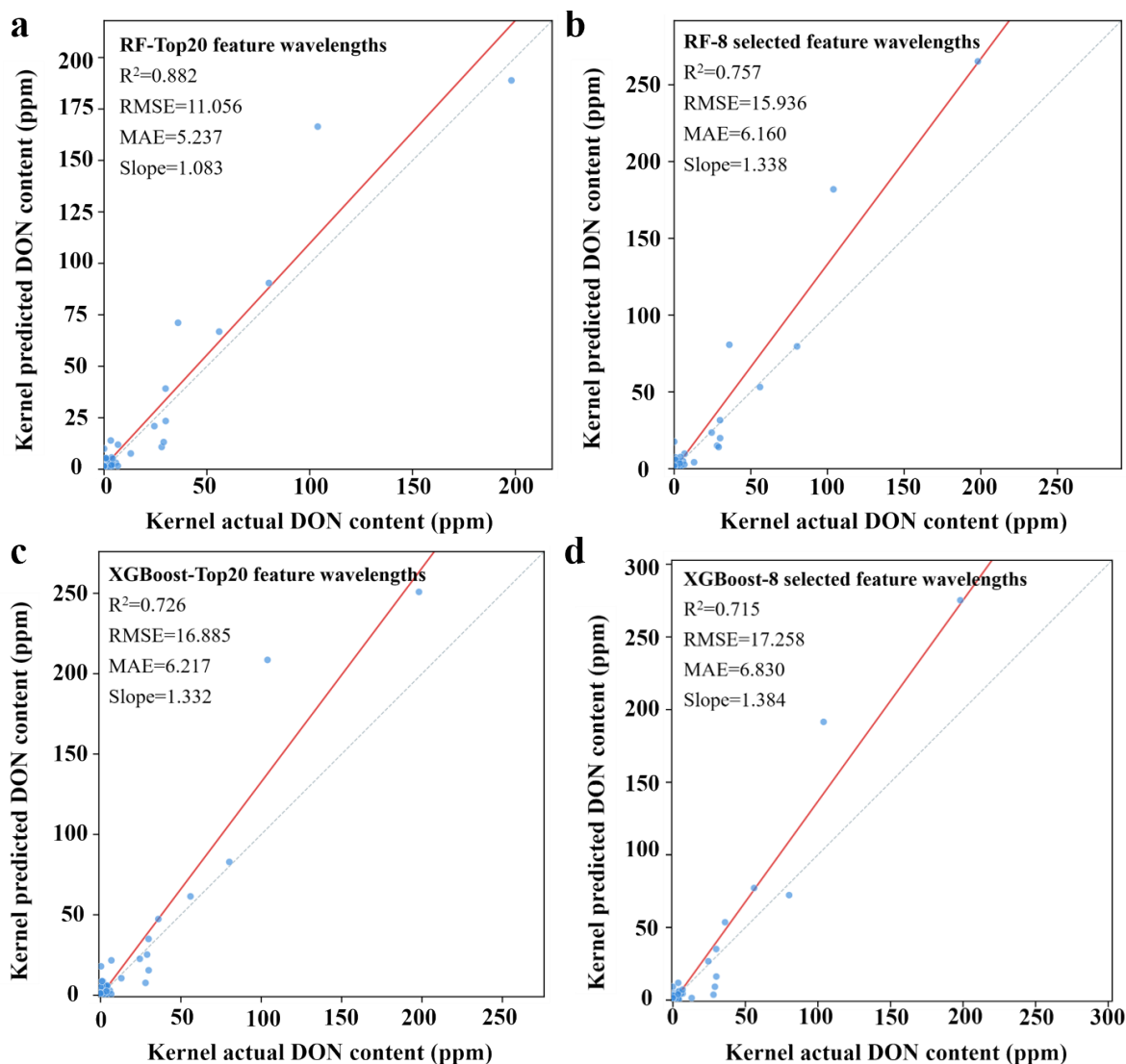


Figure 4. Comparison of kernel-DON content prediction performance between RF and XGBoost models based on different feature wavelength subsets.

List key outcomes or other achievements.

- Established and validated a rapid, non-destructive framework combining SWIR hyperspectral imaging and machine learning to predict DON concentrations in individual barley kernels.

- Identified Random Forest as the most accurate and stable regression model, achieving a test R^2 of 0.869 with minimal train-test divergence.
- Reduced spectral dimensionality by 97%, from 272 bands to eight chemically interpretable wavelengths associated with Fusarium-induced changes in lipids, proteins, polysaccharides, and moisture.
- Achieved 92.6% accuracy and a Macro-F1 score of 0.904 for binary food-safety classification at the 5-ppm DON threshold using RBF-SVM.
- Demonstrated that larger, more balanced datasets are needed to reliably distinguish intermediate feed-grade DON classes.
- Independently validated model predictions using *tri5* DNA content ($R^2 = 0.892$), confirming that spectral predictions reflected Fusarium infection-related biochemical changes.
- Provided a foundation for developing lower-cost multispectral systems for rapid DON screening in grain-handling and food-safety applications.

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) A manuscript has been submitted to the journal of Food Control and currently is under the review (FOODCONT-D-26-04738).
- (2) Research findings have been presented at National FHB Forum during December 8-10, 2025 at Denver, Colorado.

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?

Enlarging the sampling size to include additional barley and wheat kernels, particularly targeting more high-DON and highly infected samples.