Vomitoxin Prediction in Corn using Hyperspectral Imaging

Dishant Bothra,IIT Kharagpur

March 14, 2025

1 Introduction

Vomitoxin contamination in corn is a major concern in agriculture. This project leverages hyperspectral imaging data and machine learning to predict vomitoxin levels in parts per billion (ppb). The model is deployed as a Streamlit application allowing CSV/Excel uploads or manual spectral value input.

2 Approach

The following steps were taken:

- Data Preprocessing: Principal Component Analysis (PCA) was applied for dimensionality reduction, followed by MinMax scaling.
- Model Selection: Several models were tested, including:
  - Deep Learning(Simple NN): Provided better performance but was computationally expensive.
  - Decision Tree: Can be a good approach but need to studied upon.
  - Gradient Boosting: Showed strong performance but required fine-tuning.
  - Final Model (Random Forest): Selected due to its balance of accuracy, speed, and robustness.
- Pipeline Development: The final model, PCA, and scaler were saved in `model_pipeline.pkl` for deployment.
- Web Application: A Streamlit app allows users to make predictions interactively.

3 Results and Discussion

The model was trained and evaluated on hyperspectral data with 448 bands. The final model achieved competitive accuracy, balancing between performance and computational efficiency. The Streamlit app enables real-time vomitoxin prediction.

Model	R2-Score	MSE
Decision Tree	0.9428	15981057
Random Forest	**0.9496**	**14076686**
XGBoost	0.9348	18216874
Deep Learning	0.9502	13904483

Table 1: Model performance comparison

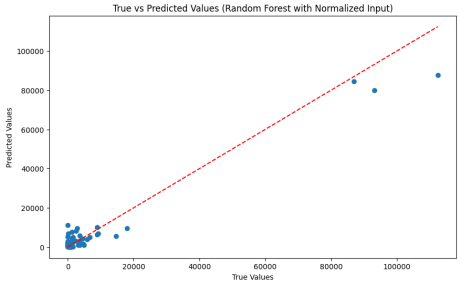

Figure 1: Test prediction.

4 Challenges and Trade-offs

- Dimensionality Reduction: PCA helped in handling high-dimensional data but may cause minor information loss.
- Model Complexity: A balance was struck between model accuracy and real-time prediction speed.
- Input Constraints: The system strictly requires exactly 448 spectral bands to function correctly.

5 Conclusion

This project successfully predicts vomitoxin levels using hyperspectral imaging. The interactive web application makes it accessible for real-world agricultural use. Future improvements could explore deep learning techniques for enhanced accuracy.