A Novel Approach for *Aspergillus flavus* Contamination Monitoring: Whole-Cell Biosensor Array Coupled with Regression Models in Agricultural Commodities

Lu Sun[1,2], Giorgia Purcaro[1], Fuguo Xing[2]

1 Analytical Chemistry Lab, Gembloux Agro-Bio Tech, University of Liège, Passage de Déportés 2, 5030 Gembloux, Belgium
2 Institute of Food Science and Technology, Chinese Academy of Agricultural Sciences /Key Laboratory of Agro-Products Quality and Safety Control in Storage and Transport Process, Ministry of Agriculture and Rural Affairs, Beijing, 100193, P.R. China

1

I.Introduction

Aspergillus flavus is a major cause of post-harvest losses in maize and peanuts due to production of aflatoxin B_1 (AFB_1) contamination. Current detection methods are often slow, invasive, and unsuitable for real-time monitoring, highlighting the urgent need for sensitive and scalable early detection strategies.

2

II.Objectives

This study aims to develop a non-invasive, transcriptome-guided whole-cell biosensor array integrated with machine learning models for early and quantitative prediction of *A. flavus* infection stages and AFB_1 levels in maize and peanut kernels.

4

IV.Results and Expectation

We identified eight infection-induced promoters *in E. coli* by analyzing its transcriptomic response to HVOCs in infected maize and peanuts.

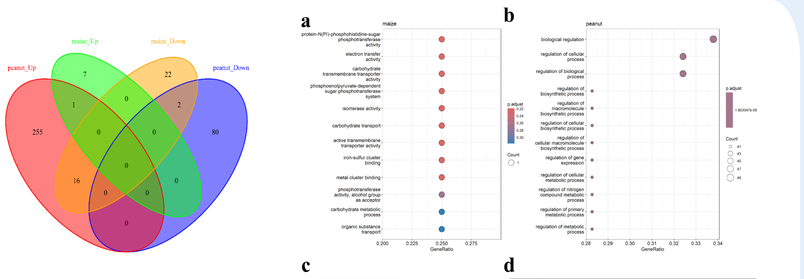


Fig. 2. GO-term-based enrichment analysis of significantly upregulated genes in *E. coli* exposed to VOCs from *A. flavus*-infected maize and peanut samples. (a) GO-term enrichment analysis of upregulated genes in *E. coli* in response to head-space VOCs from infected maize samples. (b) GO-term enrichment analysis of upregulated genes in *E. coli* in response to head-space VOCs from infected peanut samples. (c) Heatmap representation of the top 8 upregulated genes in response to VOCs from infected maize samples (left) and the top 25 upregulated genes in response to VOCs from infected peanut samples (right). (d) GO categories associated with the 33 selected upregulated genes. Highlighted in orange are the 8 genes (*evgA*, *yeeN*, *elfA*, *dkgB*, *yiaG*, *hipB*, *yjcF*, and *frc*) whose promoter regions were selected for downstream whole-cell biosensor construction.

Feature importance analysis revealed that early host responses, including transcriptional regulation and biofilm formation, served as key predictive features, thereby providing mechanistic interpretability not attainable with conventional optical or chemical assays.

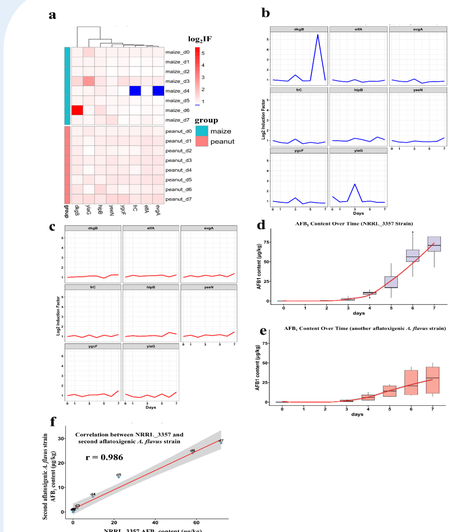


Fig. 3. Temporal dynamics and clustering of whole-cell bioreporter responses to *A. flavus*-induced VOCs in maize and peanut. (a) Hierarchical clustering heatmap of Log_2-transformed induction factor (Log_2IF) values across 8 promoter-driven biosensors exposed to VOCs emitted from *A. flavus*-infected kernels, illustrating distinct temporal activation profiles. (b) Time-course plots of Log_2IF values in maize samples (blue), measured at 0, 1, 2, 3, 4, 5, 6, and 7 days post-inoculation (dpi). Each panel corresponds to a specific promoter. (c) Corresponding Log_2IF trajectories in peanut samples over the same infection timeline. (d,e) AFB_1 accumulation dynamics in maize kernels infected with *A. flavus* strain NRRL_3357 (d) and a second aflatoxigenic strain (e), quantified across identical dpi stages. (f) Pearson correlation coefficient of AFB_1 levels between the two strains across all time points, indicating concordance in temporal AFB_1 accumulation patterns.

3

III.Methods

A transcriptome-guided whole-cell biosensor array was developed by integrating eight infection-induced promoters, identified from *E. coli* transcriptomic responses to the headspace volatile organic compounds (HVOCs), into calcium alginate-immobilized bioreporters coupled with machine learning regression models, as summarized in Figure 1. Time-resolved bioluminescence signals were used to train ensemble regressors, including XGBoost, CatBoost, and RandomForest, for quantitative prediction of infection stages and AFB_1 levels.

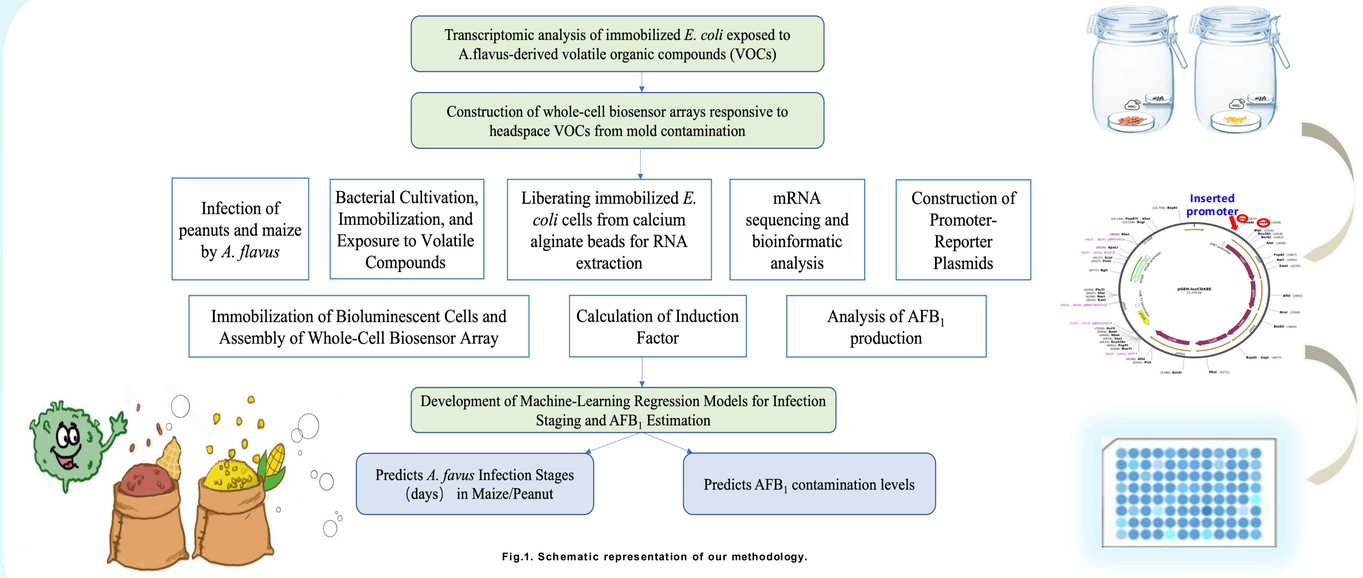
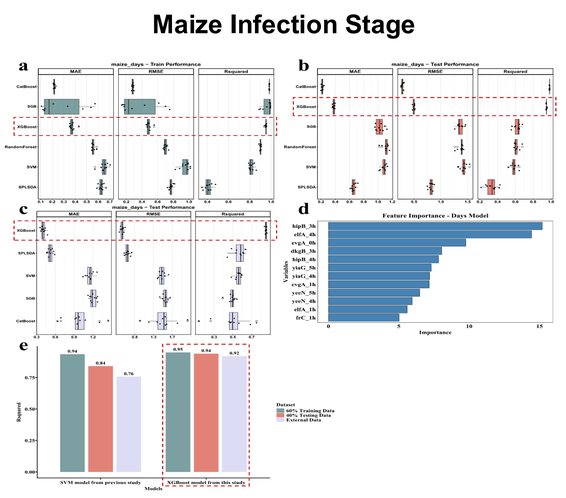


Fig.1. Schematic representation of our methodology.

5

For maize: XGBoost consistently achieved superior performance in internal validation, with R^2 values of 0.94 and 0.98 for maize infection staging and AFB_1 quantification, respectively.



Infected Maize AFB_1 Production

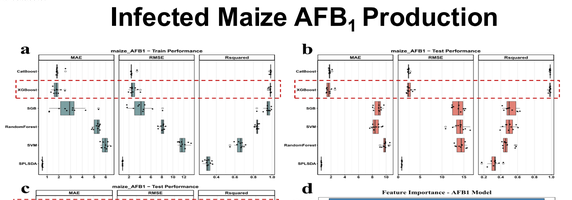


Fig. 5 Predictive performance benchmarking of regression models for estimating AFB_1 production by *A. flavus* in maize kernels. (a–c) Boxplots illustrating the predictive accuracy of six machine learning regression algorithms—RF, SGB, CatBoost, XGBoost, SVM, and SPLSDA—in forecasting AFB_1 accumulation. Model performance was assessed using MAE, RMSE, and R^2 on: (a) 60 % of maize kernel samples inoculated with *A. flavus* strain NRRL_33557 (training set), (b) the remaining 40 % of the same dataset (testing set), and (c) an independent external maize kernel dataset inoculated with a genetically distinct aflatoxigenic *A. flavus* strain. (d) Relative importance (scaled 0–100 %) of the top 20 predictive variables for AFB_1 estimation, as determined by the best-performing XGBoost model. (e) Parallel comparison.

The model also maintained strong generalization in external validation using independent *A. flavus* isolates, achieving R^2 values of 0.92 and 0.91.

6

Peanut Infection Stage

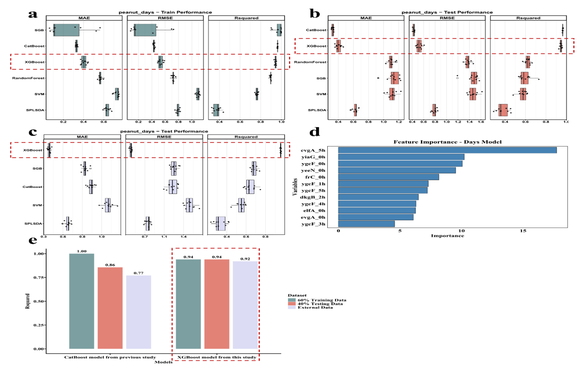


Fig. 6. Comparative performance of six regression models for predicting *A. flavus* infection stages in peanut kernels. (a–c) Boxplots illustrating the predictive performance of RF, SGB, CatBoost, XGBoost, SVM, and SPLSDA across three evaluation metrics: MAE, RMSE, and R^2. Models were trained and validated on: (a) 60 % of peanut kernel samples inoculated with *A. flavus* strain NRRL_33557 (training set), (b) 40 % of the same dataset (testing set), and (c) an independent test set comprising kernels infected with a distinct aflatoxigenic *A. flavus* strain. (d) Ranked feature importance plot from the top-performing CatBoost model, highlighting the 20 most influential variables contributing to model predictions, with normalized importance scores (0–100 %). (e) Parallel comparison.

Infected Peanut AFB_1 Production

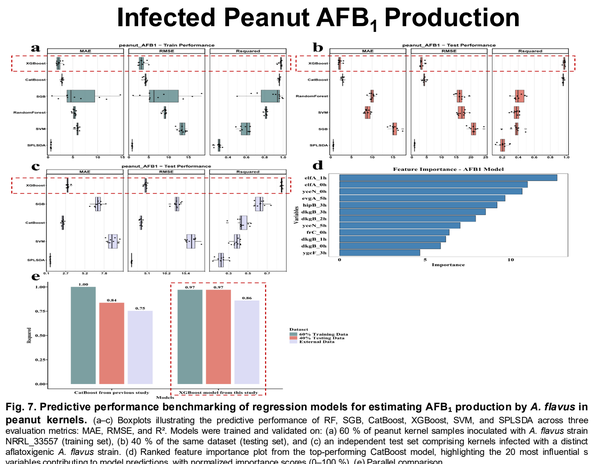


Fig. 7. Predictive performance benchmarking of regression models for estimating AFB_1 production by *A. flavus* in peanut kernels. (a–c) Boxplots illustrating the predictive performance of RF, SGB, CatBoost, XGBoost, SVM, and SPLSDA across three evaluation metrics: MAE, RMSE, and R^2. Models were trained and validated on: (a) 60 % of peanut kernel samples inoculated with *A. flavus* strain NRRL_33557 (training set), (b) 40 % of the same dataset (testing set), and (c) an independent test set comprising kernels infected with a distinct aflatoxigenic *A. flavus* strain. (d) Ranked feature importance plot from the top-performing CatBoost model, highlighting the 20 most influential s variables contributing to model predictions, with normalized importance scores (0–100 %). (e) Parallel comparison.

For peanuts: XGBoost achieved high accuracy in internal validation (R^2 = 0.94 and 0.97) and maintained robust performance in external validation (R^2 = 0.92 and 0.86).

Model comparison

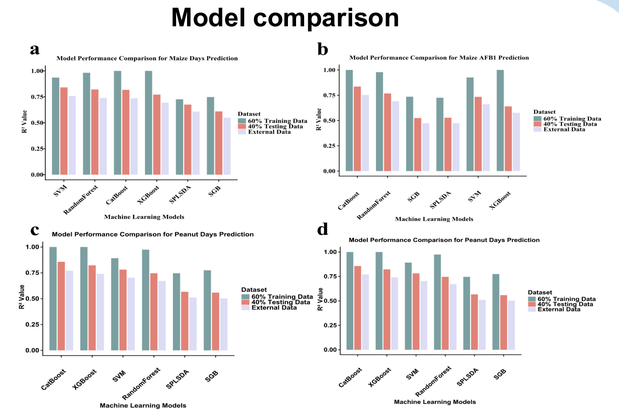


Fig.8. Comparison of six regression models (CatBoost, RandomForest, SGB, SPLSDA, SVM, and XGBoost) for the construction of whole-cell biosensor arrays designed to stage *A. flavus* infection and predict aflatoxin B_1 production in infected maize and peanuts. (a) R^2 metrics for the six regression models in predicting *A. flavus* infection staging in infected maize. (b) R^2 metrics for the six regression models in predicting aflatoxin B1 production in infected maize. (c) R^2 metrics for the six regression models in predicting *A. flavus* infection staging in infected peanuts. (d) R^2 metrics for the six regression models in predicting aflatoxin B1 production in infected peanuts.

7

V.Conclusion

This integrated biosensor-machine learning platform provides a robust, non-invasive approach for real-time monitoring of *A. flavus* contamination, offering significant potential for scalable food safety applications across diverse agroecosystems.

2nd PhD Symposium GEMBLOUX AGRO-BIO TECH 13 November 2025, Gembloux