Journal of Hazardous Materials 498 (2025) 139883

Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

A novel approach for predicting aflatoxin B_1 production using regression models and whole-cell biosensors in moldy maize and peanut kernels

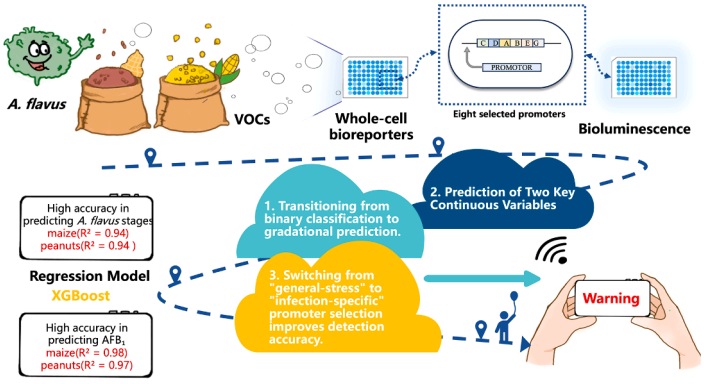
Lu Sun [a,b,1], Junning Ma [a,1], Yongping Jiang [a], Giorgia Purcaro [b], Yuanyuan Tian [a], Gang Wang [a], Weizhao Li [a], Bowen Tai [a], Fuguo Xing [a,*]

[a] 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, PR China
[b] Gembloux Agro-Bio Tech, University of Liège, Passage des Déportés 2, Gembloux 5030, Belgium

HIGHLIGHTS

- Transcriptome-guided whole-cell biosensor enables early *A. flavus* detection.
- Eight novel infection-induced promoters integrated into alginate-immobilized reporters.
- Bioluminescence signals analyzed with ensemble regressors for infection stage and AFB_1 risk.
- Regression models enable quantitative prediction of AFB_1 levels from mere binary classification.
- XGBoost achieved R^2 up to 0.98 in maize and 0.97 in peanuts with strong validation.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Aspergillus flavus
Whole-cell biosensor array
Post-harvest storage
Volatile organic compound
CatBoost
XGBoost

ABSTRACT

Aspergillus flavus is a major cause of post-harvest losses in maize and peanuts through aflatoxin B_1 (AFB_1) contamination, highlighting the urgent need for sensitive and scalable early detection strategies. In this study, we developed a transcriptome-guided whole-cell biosensor array by integrating eight infection-induced promoters, identified from *E. coli* transcriptomic responses to volatile organic compounds, into calcium alginate–immobilized bioreporters coupled with machine learning regression models. Time-resolved bioluminescence signals were used to train ensemble regressors for the first time, including XGBoost, CatBoost, and RandomForest, for quantitative prediction of infection stages and AFB_1 levels. XGBoost consistently achieved superior performance with R^2 values of 0.94 and 0.98 in internal validation for maize infection staging and AFB_1 quantification and maintained strong generalization in external validation with independent *A. flavus* isolates ($R^2 = 0.92$ and 0.91). Comparable results were observed in peanuts, where XGBoost achieved $R^2 = 0.94$ and 0.97 internally and 0.92 and 0.86 externally, confirming robustness across different substrates and fungal strains. Direct comparison with

* Correspondence to: Institute of Food Science and Technology, Chinese Academy of Agricultural Sciences, 2 Yuanmingyuan West Road, Haidian District, Beijing 100193, PR China.
E-mail address: xingfuguo@caas.cn (F. Xing).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jhazmat.2025.139883
Received 14 January 2025; Received in revised form 11 September 2025; Accepted 14 September 2025
Available online 15 September 2025

our previous biosensors constructed from 14 general stress-responsive promoters revealed that the novel biosensors based on 8 new transcriptome-guided promoters yielded markedly higher predictive accuracy, particularly under external validation conditions. 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. Together, these findings establish a biologically informed, non-invasive, and cost-efficient biosensing platform that integrates promoter-level transcriptomic insights with ensemble learning, offering a versatile approach for real-time aflatoxin risk assessment and scalable food safety monitoring across diverse agroecosystems.

1. Introduction

Mold contamination, particularly the presence of mycotoxins, poses a significant threat to the food safety and quality of grain and oil products, impacting both exporters and manufacturers worldwide [1]. Considering the vast scale of global cereal production, estimated at 2854 million tonnes with stocks of 894 million tonnes in 2024 [2], safeguarding this supply is critical. However, this production is challenged by the pervasive issue of mycotoxins, which affects approximately 25 % of food crops worldwide, highlighting the urgent need for effective monitoring and management strategies. Among the various mycotoxins, aflatoxin B_1 (AFB_1), produced by the pathogen *Aspergillus flavus,* stands out as a particularly concerning human carcinogen [3]. This fungus threatens crops like peanuts and maize and proliferates under diverse environmental conditions, causing substantial economic losses. The ability of *A. flavus* to produce AFB_1 under various conditions makes contamination pervasive and difficult to prevent, further underscoring the need for robust strategies against mold contamination to safeguard the global food supply.

The stark economic consequences of mycotoxin contamination, such as the high rejection rates of agricultural crop shipments and exacerbated food and feed loss, motivate the present research. Between January 2021 and March 2022, the Rapid Alert System for Food and Feed (RASFF) reported 290 serious cases of mycotoxin risks in food and feed in Asia, with the shipments predominantly rejected for exceeding the EU limit of 10 μg/kg for AFB_1 in feed [4]. Consequently, the development of early detection approaches and risk prediction models to identify effective intervention strategies has garnered significant attention [5]. Accurate, rapid, and early detection of mold infection during storage is crucial for implementing timely measures to reduce food losses and establish predictive infection models, thereby mitigating the economic and health impacts of mycotoxins.

Conventional detection methods, such as visual examination, culturing and plating methods, and isozyme analysis, which rely on morphological, microbiological, and biochemical identifications, have long been the foundation of mold detection [6]. However, these methods are often slow and lack specificity. More advanced laboratory techniques, including DNA-based methods like PCR [7–9] and antigen-based assays like enzyme-linked immunosorbent assay (ELISA) [10], offer high sensitivity and specificity. Nevertheless, they are labor-intensive, time-consuming, and not suitable for real-time or large-scale monitoring [11]. This has driven interest in non-invasive approaches. In contrast, these alternatives that leverage spectroscopic-based methods, electronic noses (E-noses), and colorimetric sensors, that are integrated with machine learning models, have emerged as promising alternatives. Machine-learning models can analyze the data collected by these sensors, enabling continuous monitoring of crops. This integration effectively overcomes the labor-intensive and time-consuming issues of lab-based methods and the lack of real-time monitoring ability of traditional methods [12,13]. These innovative strategies address the key limitations of traditional and lab-based methods, offering more efficient and scalable solutions crucial for timely intervention and safeguarding the food supply chain.

Recent advancements have highlighted the potential of volatile organic compounds (VOCs) emitted from infected crops as valuable indicators of mold presence, with specific VOCs identified as biomarkers for *A. flavus* infection in crops such as maize and peanuts. These VOCs enable the development of sensors for early detection, thus offering a promising approach for preventing crop spoilage. However, conventional detection methods, including gas chromatography and E-noses, often face challenges in terms of portability, high costs, and the need for specialized technical expertise [14]. Given the limitations of conventional detection methods, researchers have turned their attention to alternative solutions, among which whole-cell biosensors show great promise. To overcome these limitations, the integration of whole-cell biosensors presents a cost-effective and adaptable solution for monitoring *A. flavus* infections. These biosensors, which can be engineered to selectively detect specific VOCs, offer the advantage of being incorporated into portable devices and integrated into smart monitoring systems, making them suitable for on-site, real-time monitoring and decision-making [15].

Recent studies have shown that whole-cell biosensors can predict early spoilage events in fungi-infected postharvest commodities, such as maize, peanuts, potatoes, and grape berries, by sensing headspace VOCs, demonstrating their viability for practical agricultural applications [16–18]. These biosensors are developed by identifying key sensing promoters through transcriptomic analysis of liquid cultures of free-living *E. coli* cells exposed to infection-related VOC signatures, followed by immobilization of the cells in calcium alginate microbeads, which promotes biofilm formation [19]. Coupling these biosensors with various machine-learning classification models, including Neural Networks (NN), Support Vector Machines (SVM), and Random Forests (RF), have shown high prediction accuracy [16,17]. However, the characteristic VOCs associated with spoilage are often context-dependent, varying with factors such as substrates, storage conditions, and microbial species. This variation in complex VOC profiles limits the ability of biosensors to reflect the spoilage process [20]. Furthermore, using liquid cultures of free-living *E. coli* cells to identify potential promoters does not replicate the real conditions of immobilized cells in a biofilm state. Additionally, machine-learning classification models provide a binary "yes" or "no" signal, which fails to capture the degree of mold contamination or AFB_1 production.

To overcome the limitations of our previous binary classification approach—specifically, its reliance on non-specific promoters and lack of quantitative detail—we propose a transcriptome-guided RNA-seq strategy. This method will identify highly responsive promoters in sodium calcium-immobilized *E. coli*, mimicking their state in real applications, when triggered by VOCs from *A. flavus*-infected agricultural substrates. We will integrate these promoters into a whole-cell biosensor array, utilizing calcium alginate immobilization to enhance cell stability and gas diffusion for improved VOC sensing. This array will be paired with advanced ensemble regression and classification machine learning models, such as CatBoost, RandomForest, Extreme Gradient Boosting (XGBoost), SVM, Sparse Partial Least Squares Discriminant Analysis (SPLSDA) and Stochastic Gradient Boosting (SGB) [21–23], to enable dynamic, time-resolved monitoring of fungal contamination across multiple time points. The prediction accuracy will show strong correlation with total mold count and AFB_1 levels. This integrated biosensor-machine learning system offers a promising strategy for reliable, early fungal contamination monitoring in agricultural diagnostics

and food safety.

2. Materials and methods

2.1. Strains and materials

Two aflatoxigenic strains of *A. flavus* were used in this study: the reference strain NRRL 3357 was stored in our lab, and a second strain was isolated from peanut soil and maintained in our laboratory. Both strains were cultured on potato dextrose agar (PDA) for routine propagation. For plasmid propagation and whole-cell bioreporter construction, *E. coli* DH5α and BL21 strains, procured from Ruibio Biotech in Beijing, China, were employed and cultivated in Luria-Bertani (LB) medium supplemented with 100 μg/mL ampicillin from Solarbio in Beijing, China, at 37 °C with agitation at 180 rpm. The parental plasmid pGEN-luxCDABE, essential for the construction of bioluminescent bacteria, was generously provided by Harry Mobley (Addgene plasmid # 44918). Shelled peanuts and maize kernels were purchased from a local supermarket in Beijing, China. Restriction enzymes *Pme*I and *Sna*BI for plasmid double digestion and T4 DNA ligase for ligation were obtained from New England Biolabs (Beijing, China). The 2 × Tag PCR Mix was purchased from Tiangen Biotech (Beijing, China), and primers for promoter cloning were synthesized by Sangon Biotech (Beijing, China). For DNA purification, the HiPure Gel Pure DNA Mini Kit was sourced from Magen Biotechnology (Guangzhou, China). Sodium alginate powder was procured from Solarbio Life Sciences (Beijing, China). Lastly, flat bottom 96-well white and transparent polystyrene microplates were acquired from Corning (Shanghai, China) for experimental assays.

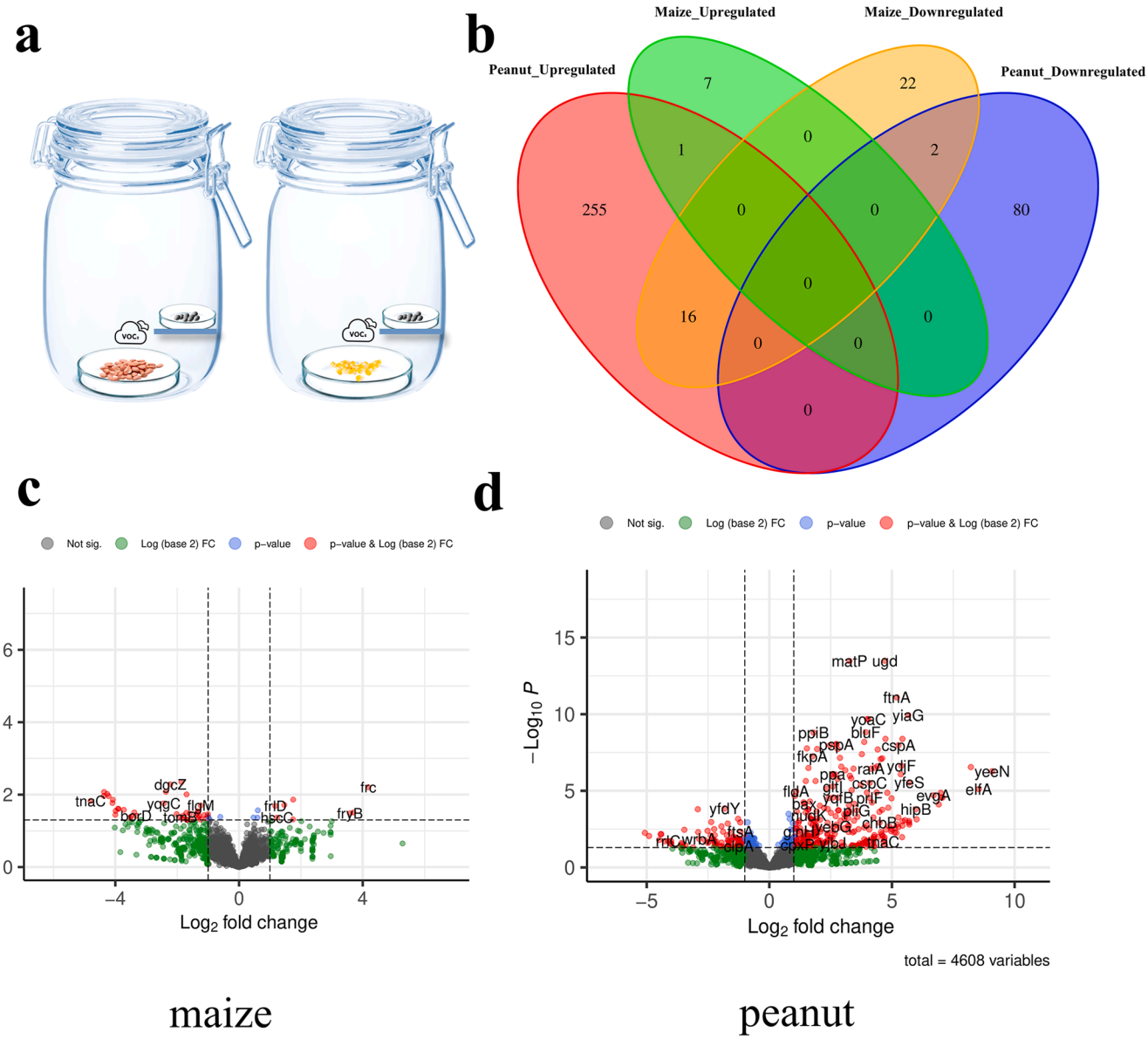


Fig. 1. Analysis of gene expression of immobilized *E. coli* in calcium alginate beads in response to the VOCs from *A. flavus* infected maize and peanuts. (a) Schematic representation of immobilized *E. coli* within calcium alginate beads co-cultured with *A. flavus* -infected maize and peanuts within a sealed environment. (b) Venn diagram illustrating the overlap and uniqueness of differentially expressed genes in *E. coli* in response to VOCs emitted by *A. flavus*-infected maize and peanuts. The sets are defined as follows: Maize_Upregulated (upregulated genes in response to VOCs from infected maize), Maize_Downregulated (downregulated genes in response to VOCs from infected maize), Peanut_Upregulated (upregulated genes in response to VOCs from infected peanut), and Peanut_Downregulated (downregulated genes in response to VOCs from infected peanut). (c) Volcano plot depicting the differential expression of genes in maize. (d) Volcano plot illustrating the differential expression of genes in peanuts. Note: the criteria for DEGs are defined by an absolute $\log_2$FC greater than 1 and an adjusted p-value less than 0.05.

2.2. Infection of peanuts and maize by A. flavus

Before *A. flavus* infection, the water activity of peanut and maize kernels was adjusted to ~0.90. Five grams of kernels were surface disinfected by immersion in 80 % (v/v) ethanol for 10 s, rinsed with distilled water, soaked in 5 % (v/v) sodium hypochlorite for 15 min, rinsed three times, and air-dried for 30 min. The kernels were then inoculated with 500 μL of 1×10^6 conidia/mL suspension, placed on Petri dishes, and incubated at 28 °C for symptom monitoring or VOCs analysis with the whole-cell biosensor array.

2.3. Bacterial cultivation, immobilization, and exposure to volatile compounds

The *E. coli* strain BL21 was initially cultivated in LB medium overnight, followed by a subculture into fresh LB medium for a further 3 h at 37 °C with constant agitation at 220 rpm. The cells were encapsulated in 1.5 % calcium alginate by mixing the suspension with alginate and extruding the mixture dropwise into 0.1 M $CaCl_2$, as previously described [17]. The resulting beads were placed in a controlled chamber and exposed for 2 h to VOCs emitted from corn and peanut kernels inoculated for 2 days (Fig. 1a). After exposure, the beads were flash-frozen in liquid nitrogen to preserve cell integrity.

2.4. Liberating immobilized E. coli cells from calcium alginate beads for RNA extraction

E. coli cells immobilized within calcium alginate beads were first ground to a fine powder using a mortar and pestle, with the aid of liquid nitrogen. Following grinding, the powdered beads containing *E. coli* cells were liberated using a chemical dissolution method. Specifically, the beads were treated with a solution of 0.1 M ethylene diamine tetraacetic acid (EDTA) [24] and 0.2 M sodium citrate [25] for a duration of 10 min to dissolve the calcium alginate matrix. This treatment facilitated the release of the bacterial cells while maintaining their integrity. Following the dissolution, the mixture was filtered through sterilized filter paper to remove any undissolved alginate pellets, ensuring a clean separation of the bacterial cells from the matrix debris. The filtrate, enriched with liberated *E. coli* cells, was subjected to centrifugation to pellet the bacterial cells. The supernatant was discarded, and the bacterial cell pellet was carefully collected for subsequent RNA extraction procedures.

2.5. mRNA sequencing and bioinformatic analysis

Total RNA was extracted, according to the protocol provided in SV Total RNA Isolation System (Promega, Madison, Wisconsin, USA), followed by ribosomal RNA depletion using MicroExpress kit (Ambion, Thermo Fisher Scientific) to acquire purified mRNA. Subsequently, fragmentation buffer was added to randomly break the obtained mRNA into short fragments, and a library was constructed according to the strand-specific library preparation method [26]. After the library quality check, different libraries were pooled based on the required effective concentration and target sequencing data volume and then proceeded to Illumina sequencing by Novogene (Beijing, China).

The raw sequencing data were subjected to stringent quality control using FastQC v0.11.9 and Fastp v0.23.4 to remove contaminants and low-quality reads, ensuring that only high-confidence data proceeded to downstream analyses. The filtered reads were aligned to the *E. coli* K-12 MG1655 reference genome (BioProject: PRJEB37847) using the Bowtie2 v2.5.1 software with default alignment parameters, which offers ultrafast and memory-efficient alignment with support for gapped, local, and paired-end alignment modes. This makes Bowtie2 particularly suitable for aligning reads to long reference sequences and provides superior performance in terms of speed and memory usage compared to other tools like BWA. Post-alignment, gene-level quantification of mapped reads was conducted using the featureCounts function within the Subread package v2.0.8. FeatureCounts is highly accurate and efficient in summarizing mapped reads at the gene or transcript level, and it provides flexibility in handling multi-mapping reads and various annotation formats. It is also significantly faster and more memory-efficient than alternatives like htseq-count. For differential expression analysis, DESeq2 v1.34.0 was employed to identify differentially expressed genes (DEGs) between the experimental and control conditions, applying a threshold of adjusted p-value < 0.05 and absolute value of $\log_2$ fold change being greater than 1 as criteria for significance. To further explore the functional implications of the DEGs, gene ontology (GO) enrichment analysis was performed using the clusterProfiler package v4.4.0.

2.6. Construction of promoter-reporter plasmids

Promoters of target genes were obtained from the BioCyc database (https://biocyc.org/). Each promoter region, defined as approximately 300 bp flanking either side of the core promoter sequence, was amplified via PCR using specific primers designed for each sequence (Table S1, Supplementary Information). The amplified promoter fragments were inserted upstream of the luciferase operon in the pGEN-luxCDABE plasmid to create recombinant reporter plasmids. Plasmid construction involved the double digestion of both the purified promoter amplicons and the pGEN-luxCDABE vector with *Pme*I and *Sna*BI restriction enzymes. Digested products were purified and subsequently ligated using T4 DNA ligase. The resulting recombinant plasmids were transformed into *E. coli* DH5α competent cells following the transformation protocol previously established [16]. Transformants were selected on LB agar plates supplemented with 100 μg/mL ampicillin. Positive transformants were confirmed by colony PCR and sequencing, and bioluminescent activity was assessed using a microplate reader to ensure proper expression of the luciferase operon.

2.7. Immobilization of bioluminescent cells and assembly of whole-cell biosensor array

Each confirmed bioluminescent *E. coli* strain carrying a distinct promoter–reporter construct, was immobilized in calcium alginate as described in Section 2.3. The immobilized cells were then dispensed into the wells of a 96-well microplate, producing a biosensor array for simultaneous multiplex detection.

2.8. Exposure to VOCs and bioluminescence measurement

The whole-cell biosensor array monitored volatile emissions from *A. flavus*–infected peanut and maize kernels at 0, 1, 2, 3, 4, 5, 6, and 7 days post-inoculation, as described in Section 2.3. The 96-well plates were arranged in a sealed chamber to facilitate headspace exposure to VOCs without direct contact between the samples and biosensor cells.

Bioluminescence signals were recorded at 0, 1, 2, 3, 4, and 5 h after exposure using a microplate reader. These measurements captured the dynamic response of each promoter over time. Multiple independent experiments were conducted, resulting in 258 valid observations per *A. flavus* strain.

2.9. Calculation of induction factors

To identify the most responsive luminescent whole-cell biosensors to VOCs emitted by *A. flavus*-infected peanut and maize kernels, the induction factor (IF) was employed as a key metric for comparative analysis.

The induction factor was calculated in two main steps for each whole-cell biosensor across all tested time points: The bioluminescence value obtained from the biosensors exposed to VOCs from infected samples was divided by the bioluminescence value from the same biosensors exposed to control (uninfected) samples. This ratio provided a

measure of the relative response of the biosensor to infection-induced VOCs at each time point.

(1) Ratio= Bioluminescence (infected sample) / Bioluminescence (control sample)

The induction factor (IF) was defined as the highest ratio recorded for each biosensor across all measurement times, reflecting the peak bioluminescent response to VOC exposure. IF was calculated as follows:

(2) IF = max {Ratio at 0, 1, 2, 3, 4, 5 h}.

These IF values were then used as input features for the regression models and to compare biosensor performance across treatments. To select the most responsive luminescent cells, the calculated induction factors for all biosensors were compared. Biosensors with higher induction factors were considered more sensitive and effective in detecting VOCs from *A. flavus*-infected samples. This approach facilitated the identification of optimal whole-cell biosensors for subsequent analysis and potential applications in the detection of fungal contamination in food products.

2.10. Analysis of AFB_1 production

AFB_1 in peanuts and maize samples was extracted and purified with the ToxinFast immunoaffinity column according to the protocol (Huaan Magnech, Beijing, China). Waters Alliance 2690 high-performance liquid chromatography (Waters, United States) with a reversed-phase C18 column (4.6 mm × 250 mm, 5 μm) was used to analyze the concentration of AFB_1. A photochemical derivator was added for the detection of AFB_1. The mobile phase of AFB_1 was 70 % methanol. The excitation and emission wavelengths of AFB_1 determination were 360 nm and 440 nm. The mobile phase flow rate was 1 mL/min.

2.11. Development of machine-learning regression models for infection staging and AFB_1 estimation

In this study, data for constructing and validating predictive regression models were generated using two distinct aflatoxigenic *A. flavus* strains, cultured under controlled laboratory conditions. A total of 258 valid observations were collected from multiple independent experimental replicates for each *A. flavus* strain, with each observation comprising bioluminescence intensity measurements from a panel of whole-cell biosensors and corresponding biological parameters. For model development, data obtained from the first strain were randomly partitioned into a training set (60 %) and an internal testing set (40 %) using a stratified sampling approach. The entire dataset from the second strain was subsequently employed as an independent external validation set to rigorously assess model generalizability across diverse biological backgrounds. Two continuous response variables were defined for prediction: (i) infection stage, quantified as days post-inoculation (dpi; range: 0–7), and (ii) AFB_1 concentration. The primary input feature set consisted of bioluminescence intensities derived from 8 promoter-responsive elements (*dkgB*, *yiaG*, *hipB*, *yeeN*, *ygcF*, *frC*, *elfA*, and *evgA*), which were initially identified through comprehensive transcriptomic screening. Additionally, for comparative analysis and to evaluate the predictive capacity of established biosensor components, bioluminescence intensities from 14 previously characterized promoters (*pspA*, *soxS*, *grpE*, *dnaK*, *fabA*, *glgS*, *katG*, *leuA*, *ompF*, *oxyR*, *recA*, *spy*, *uvrA*, *yibT*), originally utilized in a whole-cell biosensor array for classification of early mold contamination in maize and peanut kernels [17], were also incorporated as input features for constructing parallel regression models.

All raw signals were z-score-standardized prior to modelling. To enlarge the training space while minimizing class imbalance, a hybrid data-augmentation pipeline was implemented: (i) interpolation-based synthesis inspired by SMOTE, in which new samples were generated along the line connecting two random neighbours using interpolation factors of 0.2–0.8; and (ii) additive Gaussian noise with a standard deviation equal to 1–2 % of the original feature or target deviation. Aberrant observations were discarded according to the three-sigma (3 σ) criterion. Feature engineering followed a three-step strategy: (i) removal of highly collinear descriptors (Pearson $r > 0.90$), (ii) principal-component analysis (PCA) retaining components that jointly explained $\geq$ 95 % of cumulative variance, and (iii) ranking of the remaining variables by model-specific importance scores to identify the most influential predictors.

Six supervised regressors were benchmarked: RandomForest, XGBoost, CatBoost, SGB, SVM, and SPLSDA. Hyper-parameter optimization was performed with Bayesian optimization wrapped in a 10-fold cross-validation framework and was driven by maximization of the coefficient of determination (R^2), while mean absolute error (MAE) and root-mean-square error (RMSE) were monitored as secondary objectives. The tuned search spaces included: RandomForest—number of variables tried at each split (mtry), minimum terminal-node size, and splitting rule; XGBoost—learning rate (eta), maximum tree depth, minimum child weight, subsampling ratios (subsample, colsample_bytree), and regularization term (gamma); CatBoost—learning rate, tree depth, L2 leaf regularization, and random-strength parameter; SGB—number of trees, interaction depth, shrinkage rate, subsampling fraction, and minimum observations per terminal node; SPLSDA—number of latent components (K) and sparsity parameter (eta); SVM—penalty constant (C), kernel width (γ), and insensitivity parameter (ε).

Model performance was first assessed on the held-out internal test subset, and subsequently on the external strain dataset, using R^2, MAE and RMSE. Variable importance was quantified by each algorithm's relative-influence metric and visualized as horizontal bar charts of the top 20 contributors (scaled 0–100 %). All resampling and external-validation statistics were summarized across ten independent runs and displayed as box-and-whisker plots. The entire workflow was executed in R v4.4.1 with the following core packages: ranger (RandomForest), xgboost (XGBoost), catboost (CatBoost), gbm (SGB), kernlab (SVM), spls (SPLSDA), rBayesianOptimization (hyper-parameter tuning), caret (resampling orchestration), ggplot2 (visualization), and dplyr (data wrangling).

3. Results and discussion

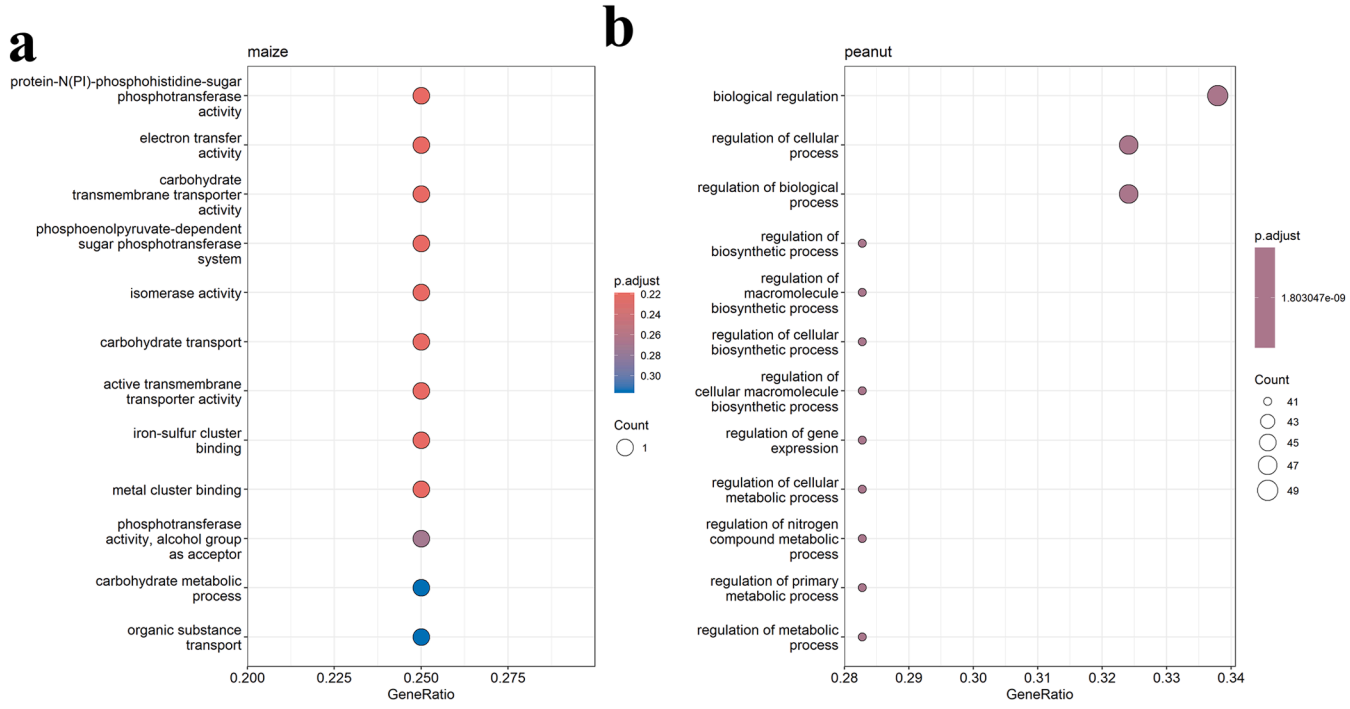
3.1. Differential gene expression analysis of calcium alginate-immobilized *E. coli* cells in response to headspace VOCs of *A. flavus*-infected maize and peanuts

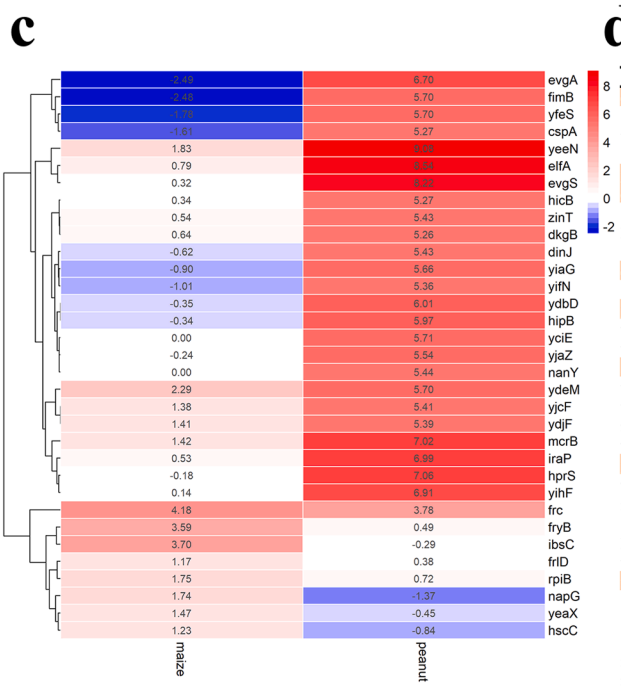
To identify the promoters that serve as responsive elements in whole-cell bioreporters, we subjected calcium-alginate immobilized *E. coli* cells to the VOCs emitted by *A. flavus*-infected peanut and maize kernels (Fig. 1a). Calcium alginate offers significant advantages over agar and other hydrogels for immobilizing bacterial cells as gas sensing elements, including mechanically robust, high stability, tunability for gas diffusion, biocompatibility, and practical ease of use [27,28]. Its selection was informed by its suitability for our needs, as evidenced by our optimization work [29], where we fine-tuned its parameters to enhance VOCs sensing capability. These attributes make it a preferred choice for applications requiring efficient gas exchange and cell functionality, supported by extensive research into its properties and applications [27, 30]. The porous structure of the calcium alginate microbeads facilitated VOCs diffusion, allowing them to act as signaling molecules or stressors, thereby inducing changes in gene expression in encapsulated *E. coli* cells [18].

Differential gene expression analysis revealed distinct transcriptional profiles in calcium alginate-immobilized *E. coli* cells exposed to VOCs from the infected substrates (Fig. 1a). Exposure to VOCs from *A. flavus*-infected peanuts upregulated 272 genes, whereas only 8 genes were

upregulated in response to VOCs from infected maize. Similarly, the number of downregulated genes differed significantly, with 82 genes affected in the peanut condition versus 60 genes in the maize condition (Fig. 1b). These results suggest that the VOCs mixtures emitted by infected peanuts elicit a stronger transcriptional response in *E. coli* compared to those from maize. This disparity could be attributed to differences in the composition, concentration, or bioactivity of VOCs produced during *A. flavus* infection of the two substrates [31].

Volcano plot analyses provided further insights into the transcriptional changes induced by the VOCs (Figs. 1c and 1d). In the maize condition, the most significantly upregulated genes included *frc* (involved in the cellular response to acidic pH), *fryB* and *frlD* (both





d

Gene	GO
evgA	GO:0000160 - phosphorelay signal transduction system
fimB	GO:0006974 - DNA damage response
yfeS	GO:0005829 - cytosol
cspA	GO:0009409 - response to cold
yeeN	GO:0006355 - regulation of DNA-templated transcription
elfA	GO:0043709 - cell adhesion involved in single-species biofilm formation
evgS	GO:0010447 - response to acidic pH
hicB	GO:0006950 - response to stress
zinT	GO:0070301 - cellular response to hydrogen peroxide;
dkgB	GO:0051596 - methylglyoxal catabolic process; GO:0019853 - L-ascorbic acid biosynthetic process
dinJ	GO:0015643 - toxic substance binding; GO:0044010 - single-species biofilm formation
yiaG	GO:0003677 - DNA binding
yifN	GO:0003677 - DNA binding
ydbD	GO:0051595 - response to methylglyoxal
hipB	GO:0040008 - regulation of growth; GO:0006355 - regulation of DNA-templated transcription
yciE	GO:0005829 - cytosol
yjaZ	GO:0009408 - response to heat; GO:0005829 - cytosol
nanY	GO:0006974 - DNA damage response
ydeM	GO:0006974 - DNA damage response
yjcF	GO:0005829 - cytosol
ydjF	GO:0006355 - regulation of DNA-templated transcription
mcrB	GO:0006308 - DNA catabolic process
iraP	GO:0009267 - cellular response to starvation; GO:0016036 - cellular response to phosphate starvation
hprS	GO:0006974 - DNA damage response; GO:0051716 - cellular response to stimulus
yihF	GO:0005886 - plasma membrane
frc	GO:0071468 - cellular response to acidic pH; GO:0033611 - oxalate catabolic process
fryB	GO:0016310 - phosphorylation; GO:0071702 - organic substance transport
ibsC	GO:0012501 - programmed cell death; GO:0005886 - plasma membrane
frlD	GO:0016310 - phosphorylation
rpiB	GO:0005975 - carbohydrate metabolic process
napG	GO:0051536 - iron-sulfur cluster binding
yeaX	GO:0016491 - oxidoreductase activity
hscC	GO:0006950 - response to stress

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.

involved in phosphorylation), and *hscC* (associated with stress response). Downregulated genes included *tnaC*, *borD*, *yqgC*, *dgcZ*, *tomB*, and *flgM*. In contrast, VOCs from infected peanuts resulted in a significantly higher number of upregulated genes, including *yeeN* (regulation of DNA-templated transcription), *elfA* (biofilm formation), and *evgA* (phosphorelay signal transduction). Fewer downregulated genes were observed, notably *rrlC* and *wrbA*.

The variability in the number of DEGs between conditions underscores the distinct impacts of VOCs from the two substrates. While prior studies reported greater diversity of VOCs emitted by *A. flavus*-infected maize (55 VOCs) [32] compared to peanuts (31 VOCs) [17], these findings are highly dependent on experimental conditions, including different strains and VOCs detection methodology (e.g., GC-MS or GC-IMS). Moreover, VOCs from infected peanuts may possess unique signaling molecules or stress-inducing compounds that provoke a stronger transcriptional response in *E. coli*. More importantly, the distinct transcriptional profiles observed in *E. coli* exposed to VOCs from *A. flavus*-infected peanuts and maize highlight the potential of using the upregulated promoters in *E. coli* as sensing elements for detecting spoilage in stored agricultural products. Previous studies have demonstrated that specific promoters in *E. coli* can respond dynamically to environmental conditions, including VOCs, and their activity can be harnessed for biosensor development [33]. Calcium alginate immobilization provides a protective niche for *E. coli*, enhancing its stability, longevity and gas diffusion, which is critical for practical applications in monitoring stored agricultural products [18]. Our approach leverages the natural transcriptional responses of calcium alginate-immobilized *E. coli* cells to VOCs, providing a promising avenue for developing biosensors that can non-invasively monitor spoilage in stored peanuts and maize kernels.

3.2. Identification and characterization of upregulated genes in response to VOCs

Following differential gene expression analysis of *E. coli* exposed to the headspace VOCs (HVOCs) emitted from *A. flavus*-infected peanut and maize kernels, we performed GO term enrichment analysis on the significantly upregulated genes. This allowed us to identify potentially responsive elements for constructing whole-cell bioreporters. The GO enrichment analysis revealed distinct biological processes that were significantly enriched in response to the HVOCs from infected maize (Fig. 2a). The enriched terms predominantly involved metabolic processes related to carbohydrates, transport activities, electron transfer, and specific enzymatic functions. The enrichment of these GO classifications indicates that *E. coli* is adjusting its metabolic pathways and transport systems to respond to the HVOCs from moldy maize. The bacterium may be utilizing these VOCs as a nutrient source [34] or coping with potential stress induced by the environmental conditions. This adaptive response reflects *E. coli*'s ability to sense and respond to changes in its environment, optimizing its survival and metabolic efficiency under challenging conditions.

In contrast, the GO enrichment analysis for *E. coli* exposed to HVOCs from moldy peanuts showed a different set of enriched biological processes (Fig. 2b). The enriched processes included regulation of metabolic processes, primary metabolic processes, nitrogen compound metabolic processes, cellular metabolic processes, gene expression, cellular macromolecule biosynthetic processes, and broader biological processes. These findings suggest that the VOCs from moldy peanuts may act as signaling molecules that influence bacterial metabolism, gene expression, and macromolecule synthesis. The low *p*-value adjustment (<0.01) for these GO terms indicated that these processes were significantly regulated in response to the HVOCs, highlighting *E. coli*'s ability to sense and adapt to the environmental cues provided by the volatiles. This adaptive mechanism is crucial for the bacterium's survival and underscores the dynamic interaction between *E. coli* and its environment.

To further investigate the transcriptional response of *E. coli* to the VOCs emitted from *A. flavus*-infected peanut and maize kernels, we focused on the significantly upregulated genes. From the list of upregulated genes, we selected key candidates for functional analysis. Specifically, we examined the top 8 upregulated genes in response to maize-derived VOCs and the top 25 upregulated genes in response to peanut-derived VOCs (Fig. 2c). Most of these genes were associated with stress response mechanisms, including *fimB* (DNA damage response), *hicB* (general stress response), and *nanY* (DNA damage response), or signal transduction processes such as *yeeN* (regulation of DNA-templated transcription), *fryB* (phosphorylation), and *ydjF* (regulation of DNA-templated transcription) in response to environmental stimuli (Fig. 2d). These findings align with previous studies demonstrating the role of VOCs as signaling molecules that influence bacterial behavior [33].

In the case of maize-derived volatiles, the most significantly upregulated genes were *frc*, *fryB*, and *ibsC*, with $\log_2$ fold change ($\log_2$FC) values of 4.1, 3.6, and 3.7, respectively (Fig. 2c). The *frc* gene, known for its role in acidic pH stress response and oxalate catabolism, likely aids in mitigating the toxic effects associated with such environmental conditions [35]. The upregulation of *fryB* and *frlD*, both implicated in phosphorylation, suggests their involvement in global signaling systems within *E. coli*, enabling the bacterium to sense environmental cues and activate stress-related pathways [36]. The gene *ibsC*, linked to programmed cell death, further underscores the severe stress experienced by *E. coli* upon exposure to VOCs from infected maize [37]. These findings align with the notion that VOCs emitted from infected plants can induce oxidative stress, DNA damage, and membrane disruption, leading to broad cellular stress responses [33].

In contrast, *E. coli* exposed to VOCs from infected peanut samples exhibited a different pattern of gene upregulation. The most significantly upregulated genes included *yeeN* ($\log_2$FC = 9.0), *elfA* ($\log_2$FC = 8.5), and *evgS* ($\log_2$FC = 8.2) (Fig. 2c). *YeeN* is involved in the regulation of DNA-templated transcription, *elfA* [38] is associated with biofilm formation, and *evgS* is known to respond to acidic pH [35].

Overall, the high expression levels of these 33 genes indicate that *E. coli* mounts a robust transcriptional response to the complex and potentially harmful VOCs produced during the interaction between *A. flavus* and food substrates. The differential gene expression patterns observed in *E. coli* exposed to VOCs from *A. flavus*-infected peanut and maize kernels suggest that specific genes and their associated promoters could be harnessed to develop whole-cell bioreporters for spoilage detection. By integrating these promoters into genetic circuits, it may be possible to develop sensitive and specific bioreporters capable of detecting spoilage VOCs in real-time.

3.3. Construction and functional validation of a whole-cell biosensor array for time-dependent VOC response analysis

Subsequently, from the initial pool of 33 highly upregulated genes identified through RNA-seq analysis, a subset of 16 candidate promoters (Fig. S1, Supplementary Information) was selected based on a dual criterion: exhibiting substantial relative gene expression increases and successful cloning into luciferase reporter plasmids suitable for whole-cell biosensor construction. To further prioritize these candidates and streamline downstream efforts towards an effective biosensor array, we conducted a preliminary functional validation using calcium alginate-immobilized whole-cell biosensors harboring these 16 distinct promoter-luciferase fusions as described in Section 2.6. Following exposure to headspace VOCs from *A. flavus*-infected kernels, bioluminescence signals were collected and analyzed for relative fold change and Z-score normalization (detailed in Fig. S1, Supplementary Information). This screening revealed that eight specific promoters (namely those associated with the genes *evgA, frC, elfA, yiaG, ygcF, yeeN, dkgB,* and *hipB*) demonstrated particularly prominent and significant bioluminescence responses upon VOC exposure. Consequently, these eight

highly responsive promoters were chosen as the optimal candidates for constructing the final multi-channel whole-cell biosensor array, providing a robust and focused set for subsequent machine-learning model development and predictive analysis.

To enable the dynamic and real-time monitoring of VOCs released from *A. flavus*-infected maize and peanut kernels, we engineered a whole-cell biosensor array comprising eight *A. flavus* reporter strains. Each strain incorporated a distinct, highly VOC-responsive promoter, previously identified via a high-throughput functional screening pipeline, driving bioluminescence. The biosensor array was systematically exposed to VOCs emanating from both infected and uninfected control kernels under sealed incubation conditions. Bioluminescence signals were quantitatively recorded in triplicate at eight discrete time points over a seven-day period post-inoculation (0–7 dpi), generating a comprehensive dataset for temporal analysis. To accurately quantify promoter induction in response to VOCs, an induction factor (IF) was calculated for each biosensor at every time point, using signals from non-infected samples as the baseline reference (detailed in Section 2.8). All IF values were subsequently $\log_2$-transformed (Log_2IF) to facilitate comprehensive visualization and comparative analysis, and presented as a heatmap (Fig. 3a) illustrating the temporal dynamics of the biosensor responses. Analysis of the heatmap revealed a predominant pattern of positive Log_2IF values (indicated by red coloration), signifying a robust upregulation of biosensor activity upon exposure to VOCs emitted from *A. flavus*-infected samples. This finding aligns consistently with our prior transcriptional profiling (Fig. 2c), thereby reinforcing the functional responsiveness and specificity of the selected promoters. Notably, VOCs released from infected maize and peanut kernels elicited distinct response profiles, with the biosensors forming separate temporal clusters based on the host substrate. Specifically, biosensors exposed to maize-derived VOCs consistently exhibited generally higher Log_2IF values throughout the time course compared to those exposed to peanut-derived VOCs, implying a stronger or more sustained transcriptional activation. This observed substrate-dependent difference critically suggests that the nutritional composition and metabolic environment provided by the host material directly influences the metabolic activity and secondary metabolite production of *A. flavus*, consequently altering the specific profile of emitted VOCs that trigger differential biosensor responses.

Further temporal analysis, illustrating Log_2IF changes for individual promoters across the 7-day infection period (0–7 dpi), elucidated distinct substrate-specific response patterns (Fig. 3b,c). In maize, the responses of biosensors (Fig. 3b) exhibited a transient upregulation, generally peaking around 3 dpi, followed by a subsequent decline or plateau. For instance, promoter *dkgB* showed the most pronounced activation, reaching a peak Log_2IF value exceeding 4.8 at 5 dpi, while *yiaG* displayed significant but less pronounced induction kinetics ($Log_2IF \approx 2.7$). Conversely, promoters like *elfA* and *yeeN* showed only marginal upregulation, indicating limited responsiveness to the volatile profile associated with *A. flavus* infection in maize. In stark contrast, peanut samples induced a markedly different biosensor response profile (Fig. 3c), characterized by a more gradual and sustained increase in Log_2IF values over time, with most VOC-responsive promoters reaching maximal activation at 7 dpi. This extended activation window in peanuts suggests either a delayed or prolonged release of specific VOCs, or sustained microbial activity, which maintains biosensor stimulation. These contrasting kinetics highlight significant substrate-dependent variations in VOCs dynamics and biosensor responsiveness, carrying implications for designing time-sensitive detection systems tailored to specific host matrices. To contextualize these biosensor dynamics, we concurrently assessed the temporal production of AFB_1 in maize kernels infected with two different *A. flavus* strains under identical sealed incubation conditions. As shown in Fig. 3d, infection by the NRRL_3357 strain led to a pronounced exponential surge in AFB_1 accumulation starting at 4 dpi, reaching a peak median concentration of approximately 75 μg/kg by 7 dpi, albeit with notable inter-replicate variability reflecting differential colonization or toxin biosynthesis rates. In comparison, the second aflatoxigenic *A. flavus* strain induced a more gradual AFB_1 accumulation, rising steadily from 3 dpi and plateauing around 35 μg/kg by 7 dpi (Fig. 3e). Importantly, despite these quantitative differences, the temporal AFB_1 profiles of both strains exhibited a strong positive correlation (Pearson's $r = 0.986$; Fig. 3f), which supports the potential applicability of predictive models developed from NRRL_3357 data for estimating AFB_1 levels induced by other aflatoxigenic strains. Collectively, these results demonstrate that our constructed multi-channel biosensor array effectively captures both the magnitude and intricate temporal patterns of VOC emissions during *A. flavus* infection with high sensitivity and distinct substrate specificity. The observed divergence in transcriptional response profiles between maize and peanut samples further underscores the profound influence of the host metabolic environment on the precise composition of emitted VOCs, thereby establishing a critical foundation for developing pathogen detection strategies that are specifically tailored to crop-specific biochemical contexts.

3.4. Regression model performance for predicting *A. flavus* infection stages in maize kernels

To develop a reliable, non-destructive pipeline for early detection of *A. flavus* colonization in maize kernels, we benchmarked six supervised learning algorithms—Random Forest, XGBoost, CatBoost, SGB, SVM, and SPLSDA—for their ability to regress infection stage (0–7 dpi) from integrated VOC profiles and bioluminescent biosensor readouts. After comprehensive hyperparameter optimization (Table S2, Supplementary Information), CatBoost and SGB achieved the highest apparent accuracies on internal datasets (training $R^2 \approx 0.98$; test $R^2 \approx 0.95$), but their predictive performance declined markedly when applied to an external dataset derived from a second aflatoxigenic *A. flavus* strain ($R^2 \leq 0.5$), reflecting overfitting and limited strain generalizability (Fig. 4a–c). In contrast, XGBoost maintained consistently high performance across training ($R^2 = 0.95$), test ($R^2 = 0.94$), and external validation datasets ($R^2 = 0.92$), underscoring its robustness against pathogen-specific VOC variability and its capacity to capture stable, non-linear relationships between fungal VOC emissions and biosensor responses. Feature importance analysis further revealed that XGBoost predictions were strongly driven by early infection-associated biosensor outputs, with key contributors including *hipB*_3 h, *elfA*_4 h, *evgA*_0 h, and *dkgB*_3 h, which regulate processes such as growth, biofilm formation, phosphorelay signaling, and methylglyoxal detoxification (Fig. 4d). These findings are mechanistically consistent with the oxidative stress responses triggered by early-stage fungal VOCs in host tissues.

To evaluate novelty relative to our previous approach, which employed 14 literature-reported stress-responsive promoters, we constructed regression models under identical preprocessing and tuning protocols. Although the SVM model achieved perfect fit on training data ($R^2 = 0.94$; Fig. S2a), its predictive accuracy decreased substantially on internal test ($R^2 = 0.84$) and external validation datasets ($R^2 = 0.76$; Fig. 4e), confirming overfitting. By contrast, the transcriptomic-guided selection of infection-specific promoters coupled with XGBoost modeling in the present study achieved superior and more balanced accuracy across all datasets (training $R^2 = 0.95$; test $R^2 = 0.94$; external $R^2 = 0.92$; Fig. 4e), demonstrating enhanced specificity, reduced variance, and improved biological interpretability. These results demonstrate that transcriptomics-guided promoter selection not only enhances biosensor specificity and signal strength but also mitigates model overfitting, thereby improving predictive accuracy in heterogeneous biological contexts. Such integration of molecularly informed biosensor design with machine learning establishes a scalable and transferable framework for early, non-destructive detection of fungal contamination, consistent with previous evidence highlighting transcriptomic analysis as a powerful approach for identifying VOC-responsive promoters in diverse biosensing applications [18,33,39].

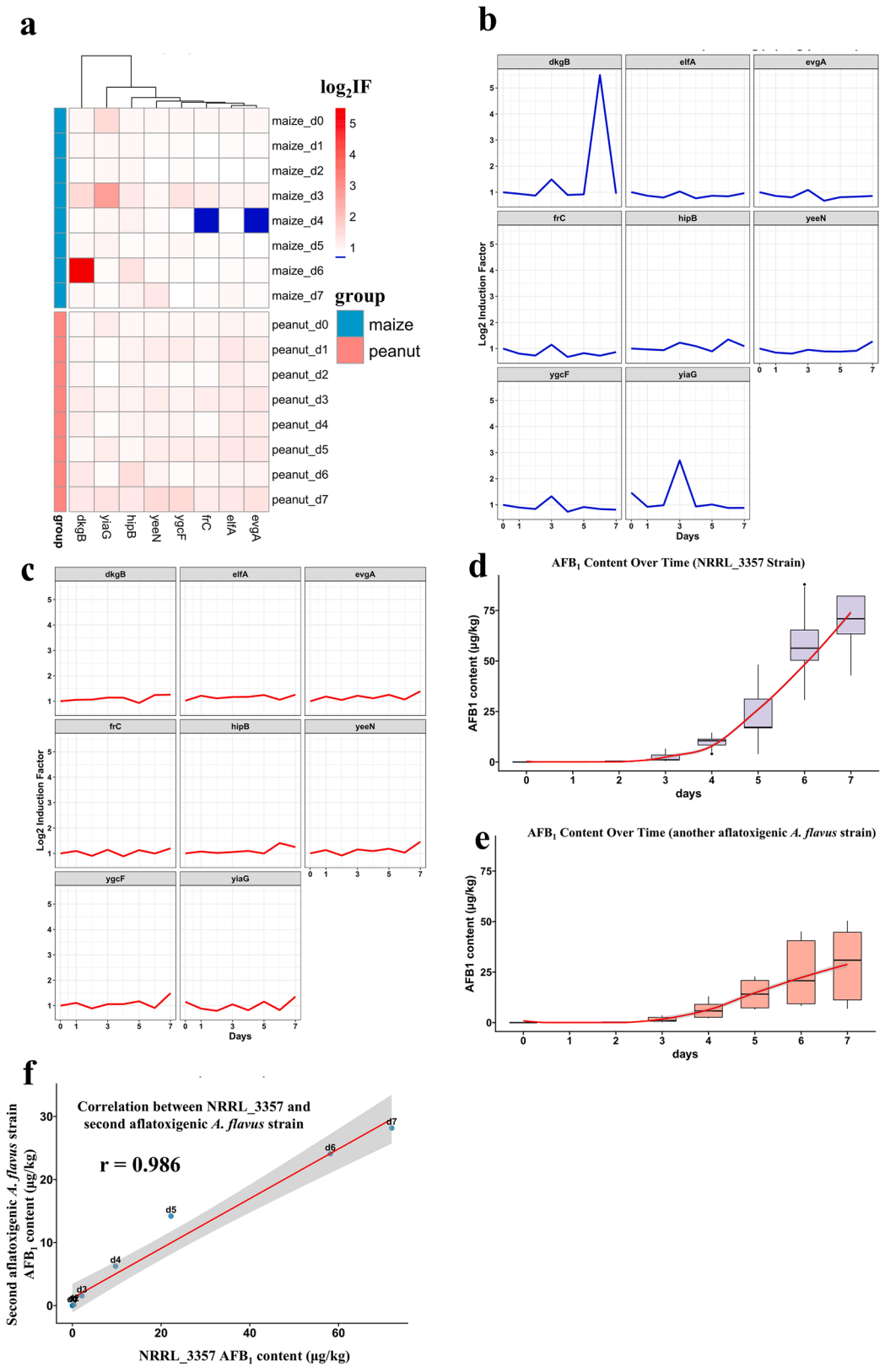


Fig. 3. Temporal dynamics and clustering of whole-cell bioreporter responses to *A. flavus*-induced VOCs in maize and peanut. (a) Hierarchical clustering heatmap of Log_2-transformed induction factor (Log_2IF) values across 8 promoter-driven biosensors exposed to VOCs emitted from *A. flavus*-infected 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.

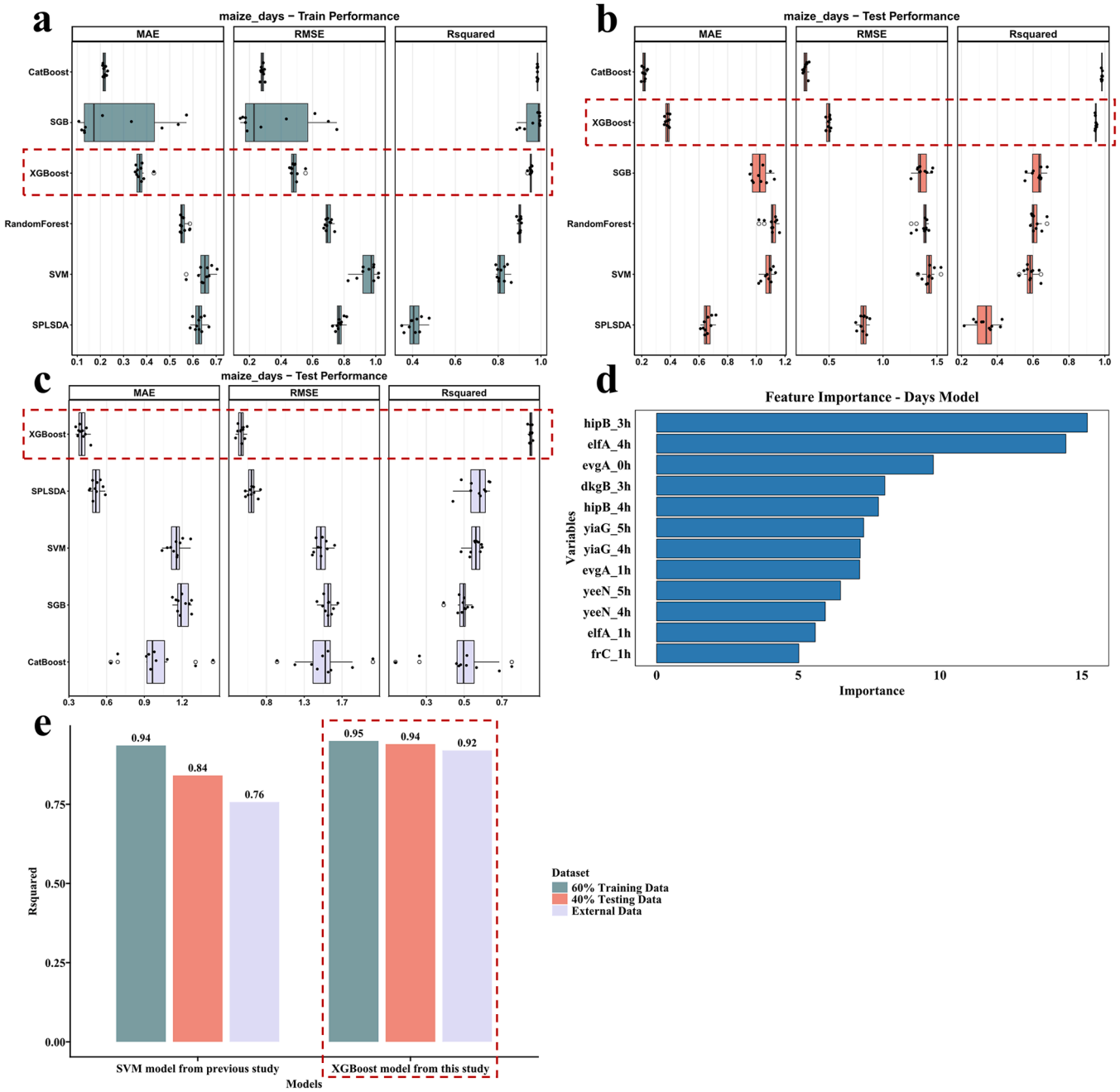


Fig. 4. Comparative performance of regression models for predicting *A. flavus* infection stages in maize kernels. (a–c) Boxplots summarizing the predictive accuracy of six regression algorithms—RF, SGB, CatBoost, XGBoost, SVM, and SPLSDA—evaluated using mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2). Models were trained and tested on: (a) 60 % of maize kernel samples inoculated with *A. flavus* strain NRRL_33557 (training set); (b) 40 % of the same dataset (testing set); and (c) an independent external maize kernel dataset inoculated with a distinct aflatoxigenic *A. flavus* strain. (d) Feature importance plot showing the top 20 variables contributing to infection stage prediction, as identified by the best-performing XGBoost model, with importance scaled from 0 % to 100 %. (e) Parallel comparison of R^2 values between an SVM model which utilized previously-published 14 promoters for whole-cell biosensor array construction, and the current XGBoost model informed by transcriptomic-guided selection of 8 promoters in this study, revealing enhanced predictive accuracy in the latter across testing (40 %), and external datasets. The red dashed rectangles in panels (a–c) highlight the most accurate and robust models across metrics and datasets.

3.5. Modeling AFB_1 accumulation in maize kernels

To assess the practical applicability of the proposed biosensing platform in food safety, we applied predictive modeling to quantify AFB_1 levels in maize kernels, a critical indicator of *A. flavus* colonization. Among the algorithms evaluated, CatBoost showed the highest accuracy on the training (60 %) and internal testing (40 %) datasets ($R^2 \approx 0.98$). However, its predictive performance significantly declined on an external dataset derived from a second aflatoxigenic strain ($R^2 \approx 0.5$), highlighting its susceptibility to strain-specific variations (Fig. 5a–c). In contrast, XGBoost, optimized with the hyperparameters provided in Table S2 (Supplementary Information), demonstrated robust predictive power, achieving R^2 values of 0.98 on both the training and internal testing datasets, and maintained high accuracy ($R^2 = 0.91$) on independent external validation (Fig. 5a–c). Feature importance analysis revealed that model predictions were primarily driven by transcripts

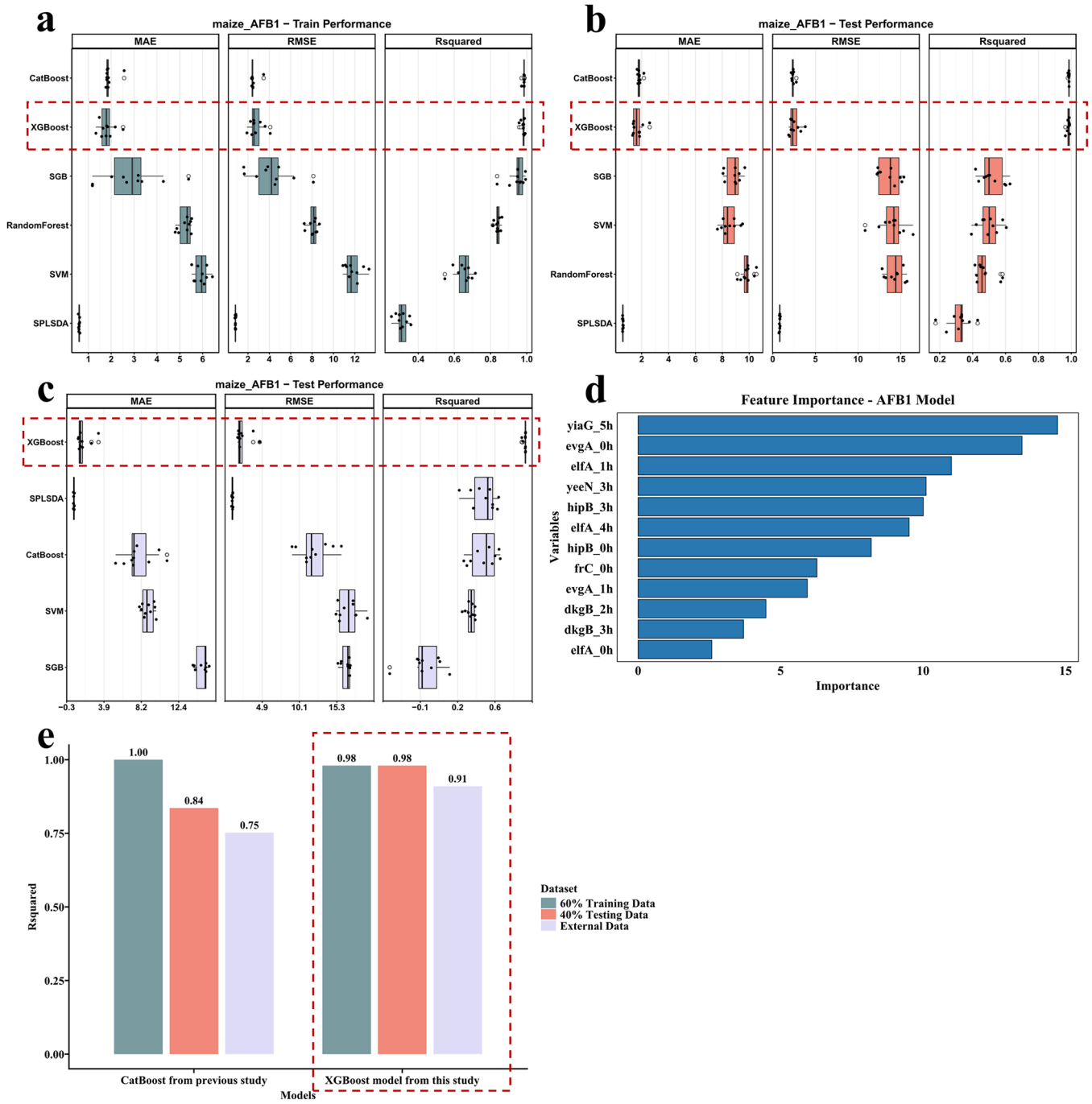


Fig. 5. Predictive performance benchmarking of regression models for estimating AFB₁ production by *A. flavus* in maize kernels. (a–c) Boxplots illustrating the predictive accuracy of six machine learning regression algorithms—RF, SGB, CatBoost, XGBoost, SVM, and SPLSDA—in forecasting AFB₁ accumulation. 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 AFB₁ estimation, as determined by the best-performing XGBoost model. (e) Parallel comparison of R^2 values between a CatBoost model which utilized previously-published 14 promoters for whole-cell biosensor array construction, and the current XGBoost model informed by transcriptomic-guided selection of 8 promoters in this study, revealing enhanced predictive accuracy in the latter across testing (40 %), and external datasets. The red dashed rectangles in panels (a–c) highlight the most accurate and robust models across metrics and datasets.

associated with stress adaptation and fungal developmental regulation, including *yiaG*_5h (DNA binding), *evgA*_0h (phosphorelay signaling), *elfA*_1h (biofilm formation), *yeeN*_3h (transcriptional regulation), and *hipB*_3h (growth control), which are all involved in early stages of fungal infection (Fig. 5d). These molecular markers corroborate the known role of oxidative stress and transcriptional reprogramming as key triggers for the aflatoxin biosynthetic pathway, providing mechanistic support for the predictive accuracy of our model. In comparison, a previously reported whole-cell biosensor array incorporating 14 promoters and six regression models exhibited lower performance, with CatBoost achieving the best results on the 60 % training dataset but underperforming on the 40 % testing ($R^2 = 0.84$) and external datasets ($R^2 = 0.75$) (Fig. 5e). Notably, our novel method, based on the transcriptomic selection of 8 promoters, demonstrated superior stability and

robustness, achieving R^2 values of 0.98, 0.98, and 0.91 on the 60 % training, 40 % testing, and external datasets, respectively (Fig. 5e). The capacity of XGBoost to integrate biologically relevant signals from biosensor data further highlights its potential for non-invasive, real-time monitoring of mycotoxin contamination across diverse environmental conditions. Taken together, these results establish the XGBoost-based biosensing platform as a scalable, field-deployable early warning system for post-harvest management, advancing food safety monitoring by combining mechanistic insight with high predictive accuracy.

3.6. Regression model performance for predicting *A. flavus* infection stages in peanut kernels

To assess the robustness and generalizability of the proposed modeling framework, we applied it to a distinct agricultural system by predicting infection stages in peanut kernels inoculated with *A. flavus*

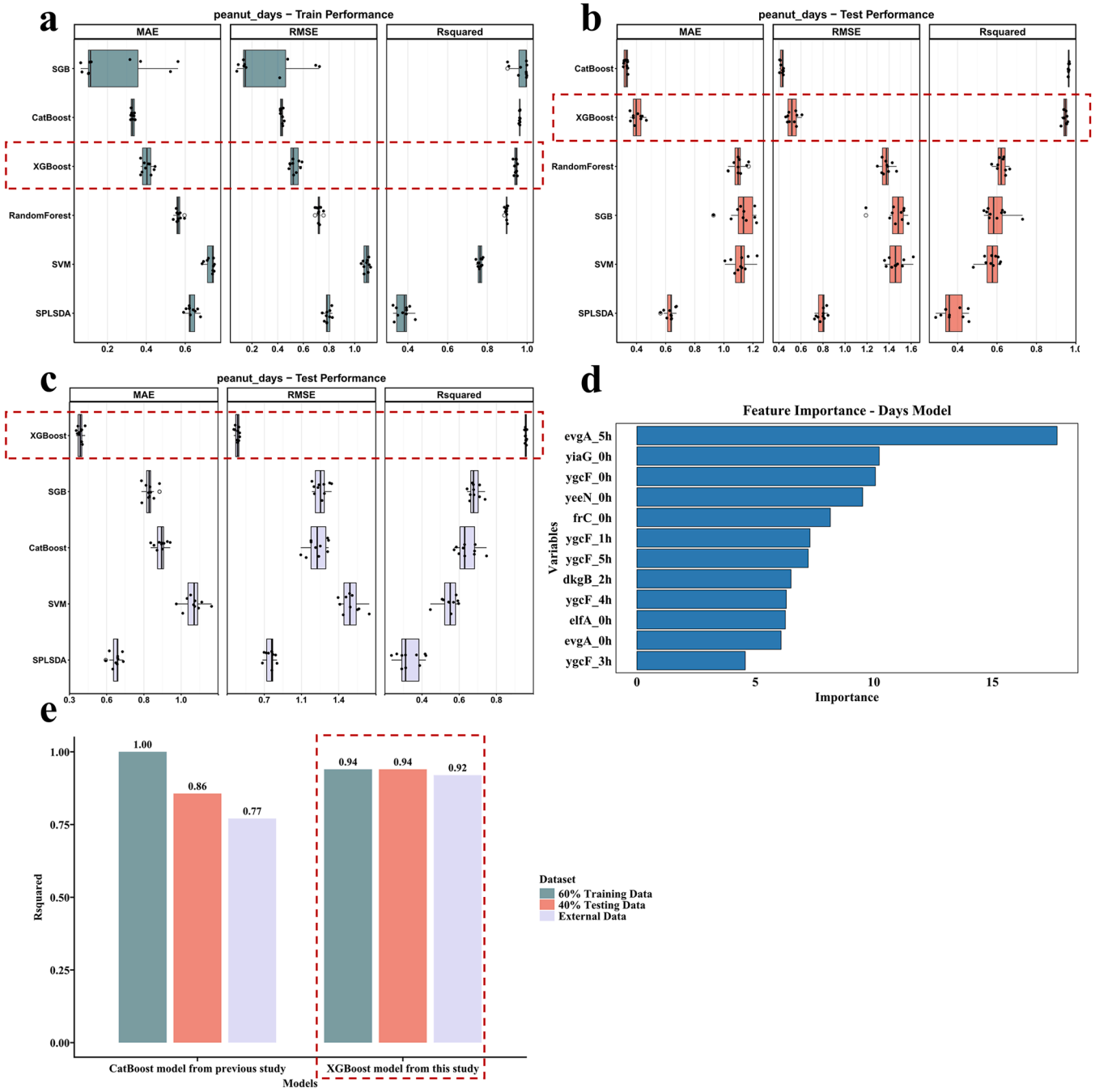


Fig. 6. Comparative performance of six regression models for predicting *A. flavus* infection stages in peanut kernels. (a–c) Boxplots illustrating the predictive performance of RF, SGB, CatBoost, XGBoost, SVM, and SPLSDA across three evaluation metrics: MAE, RMSE, and R^2. Models were trained and validated 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 of R^2 values between a CatBoost model which utilized previously-published 14 promoters for whole-cell biosensor array construction, and the current XGBoost model informed by transcriptomic-guided selection of 8 promoters in this study, revealing enhanced predictive accuracy in the latter across testing (40 %), and external datasets. The red dashed rectangles in panels (a–c) highlight the most accurate and robust models across metrics and datasets.

using six regression models trained on eight transcriptomics-guided promoters. Among them, the SGB model achieved the lowest MAE and RMSE with an R^2 of 0.97 on the 60 % training set, yet its performance declined markedly on the 40 % test ($R^2 = 0.60$) and external datasets ($R^2 < 0.70$), indicating overfitting (Fig. 6a–c). The CatBoost model displayed better balance, yielding R^2 values of 0.96 on both training and test sets, but its external validation ($R^2 = 0.68$) similarly revealed limited generalization. In contrast, XGBoost consistently outperformed the other models, maintaining high predictive accuracy across training ($R^2 = 0.94$), test ($R^2 = 0.94$), and external datasets ($R^2 = 0.92$), thereby underscoring its superior capacity to capture infection dynamics in heterogeneous biological contexts (Fig. 6a–c). Feature importance analysis further revealed that promoters such as *evgA*_5h, *yiaG*_0h, *ygcF*_0h and *ygcF*_1h, *yeeN*_0h, and *frc*_0h were critical determinants, suggesting that early-stage fungal VOCs rapidly activate cellular signaling pathways in whole-cell biosensors. For comparison, models constructed using 14 literature-reported stress-responsive promoters exhibited strong training fits (CatBoost, $R^2 = 1.0$) but showed poor

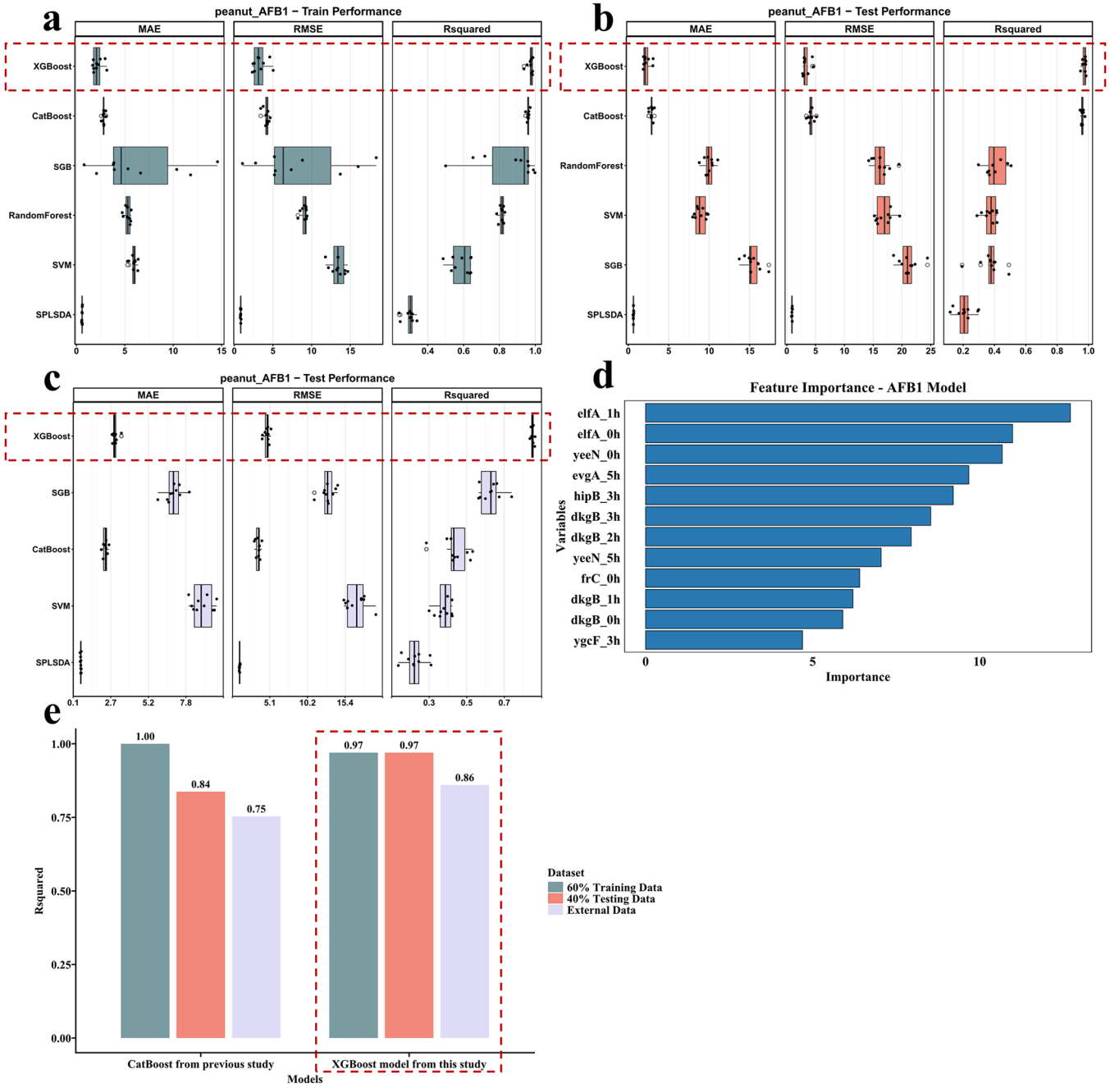


Fig. 7. Predictive performance benchmarking of regression models for estimating AFB₁ production by *A. flavus* in peanut kernels. (a–c) Boxplots illustrating the predictive performance of RF, SGB, CatBoost, XGBoost, SVM, and SPLSDA across three evaluation metrics: MAE, RMSE, and R^2. Models were trained 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 of R^2 values between a CatBoost model which utilized previously-published 14 promoters for whole-cell biosensor array construction, and the current XGBoost model informed by transcriptomic-guided selection of 8 promoters in this study, revealing enhanced predictive accuracy in the latter across testing (40 %), and external datasets. The red dashed rectangles in panels (a–c) highlight the most accurate and robust models across metrics and datasets.

generalization on internal ($R^2 = 0.86$) and external ($R^2 = 0.77$) datasets (Fig. 6e), highlighting the inadequacy of non-specific promoters. Collectively, these results demonstrate that transcriptomics-guided promoter selection combined with XGBoost modeling achieves stable, biologically interpretable predictions with reduced overfitting, thereby providing a more reliable strategy for biosensor-based infection monitoring across diverse host–pathogen systems.

3.7. Prediction of AFB_1 production in peanut kernels

Our predictive framework was further evaluated for its capacity to quantify AFB_1 accumulation in peanut kernels, a critical parameter for food safety assessment. The XGBoost regression model exhibited strong predictive performance, achieving coefficients of determination (R^2) of 0.97 on both training and internal testing datasets (Fig. 7a, b). Importantly, its robustness was maintained in an independent external validation set containing peanuts infected by a genetically distinct toxigenic *A. flavus* strain, where an R^2 of 0.86 was obtained (Fig. 7c). This consistency across fungal genotypes underscores the model's high generalization capability, a prerequisite for deployment in heterogeneous agricultural environments. Feature importance analysis further revealed that the most influential predictors corresponded to transcriptional markers of stress responses and metabolic regulation (Fig. 7d), including *elfA*_0h and *elfA*_1h, *yeeN*_0h, *evgA*_5h, and *dkgB*_2h/3h. These pathways are tightly coupled with fungal stress adaptation and aflatoxin biosynthesis, supporting the mechanistic relevance of the predictors and highlighting how VOC-based profiles capture the pathogen's physiological state during infection. For comparative benchmarking, we constructed regression models using 14 literature-reported stress-responsive promoters under identical preprocessing and optimization protocols. Although CatBoost achieved a perfect fit on the training dataset ($R^2 = 1.0$; Fig. S2d), its performance dropped significantly on internal test ($R^2 = 0.84$) and external validation datasets ($R^2 = 0.75$; Fig. 7e), indicative of overfitting and limited generalizability. In contrast, transcriptomic-guided promoter selection coupled with XGBoost modeling produced consistently higher and more balanced performance across all datasets (training $R^2 = 0.97$; test $R^2 = 0.97$; external $R^2 = 0.86$; Fig. 7e), confirming enhanced specificity, reduced variance, and improved biological interpretability. Collectively, these results demonstrate that integrating transcriptomics-driven feature selection with robust machine learning architectures not only improves prediction accuracy but also establishes VOCs fingerprints as reliable, non-invasive indicators of *in situ* toxigenesis. The ability to estimate both fungal infection and toxin accumulation with high fidelity, consistent with previous findings in maize systems, suggests broad applicability across crops and provides a valuable framework for pre-harvest and early post-harvest risk assessment in real-world agricultural settings.

3.8. Comparative assessment: Our whole-cell biosensor array versus established techniques for mold monitoring in postharvest agricultural products

Benchmarking our transcriptome-guided whole-cell biosensor array against existing non-invasive detection platforms revealed notable differences in predictive performance and operational feasibility. Near-infrared (NIR) spectroscopy, which has been extensively applied for rapid AFB_1 quantification in maize and barley, typically reports coefficients of determination (R^2) ranging from 0.82 to 0.88 under controlled conditions [40]. Similarly, colorimetric sensor arrays utilizing chemically responsive dyes for detecting VOCs from fungal contamination in wheat have demonstrated classification accuracies exceeding 97 % [41,42]. Although these optical and chemical platforms offer quick, minimally invasive measurements, they often require high-precision instrumentation and specialized chemistries, which contribute to substantial capital and maintenance costs, limiting their scalability in decentralized agricultural settings.

In contrast, our biosensor platform integrates genetically encoded luminescent reporters within Ca-alginate–immobilized *E. coli* and employs ensemble learning regression models, specifically XGBoost, trained on time-resolved bioluminescence profiles. This design achieved consistently high internal cross-validation performance, with R^2 values ≥ 0.94 for both infection staging and AFB_1 quantification in *A. flavus*–infected maize and peanut kernels (Figs. 4–7). More critically, the framework maintained robust generalization to independent external validation datasets, achieving R^2 values of 0.92, 0.91, 0.92, and 0.86 for infection staging and toxin prediction across both maize and peanut substrates (Figs. 4e–7e). These metrics not only surpass the upper range of NIR spectroscopy but also do so with a much lower material and instrumentation overhead. Bioluminescence readouts can be captured using inexpensive CMOS sensors or even smartphone cameras [29,43], making it a cost-effective solution suitable for field deployment in agricultural environments.

Beyond cost-effectiveness, our biosensor demonstrates superior specificity compared to our earlier 14-promoter RandomForest classifier, which achieved only 80 % substrate distinction accuracy [17]. By using transcriptomic data to select eight promoters specifically induced during *A. flavus* infection and aflatoxin production—rather than relying on broadly stress-responsive promoters—we established a more biologically relevant predictor set. This refinement, combined with XGBoost regression, resulted in balanced and reproducible performance across internal ($R^2 > 0.94$) and external ($R^2 > 0.86$) datasets. In contrast, models trained on the previous 14-promoter set, such as CatBoost (the best-performing model among six trained regressors; Fig. S2, supplementary information), suffered from overfitting and considerably lower external accuracy. Thus, transcriptome-informed promoter selection not only improves predictive fidelity but also enhances the biological interpretability of the biosensor by linking output signals directly to pathogen-specific processes.

From a mechanistic perspective, early predictive features linked to oxidative stress, signal regulation, and biofilm formation emerged as key markers, highlighting transcriptional programs associated with *A. flavus* infection and aflatoxin production. Such mechanistic resolution, which provides insight into the biological basis of toxigenesis, is not readily achievable using NIR or colorimetric assays, which infer contamination indirectly. Nevertheless, established optical platforms retain certain advantages: NIR spectroscopy offers analyte-agnostic detection, making it less susceptible to biological variability, while colorimetric arrays are adaptable for multiplex VOCs profiling. Nonetheless, our model's performance ceiling is currently constrained by the diversity of training datasets, emphasizing the need for expansion to include a broader range of geographic isolates, environmental conditions, and substrate types. Future developments in synthetic biology, such as promoter engineering [44] and directed evolution [45], could further enhance biosensor robustness and sensitivity, particularly in varying environmental contexts.

In conclusion, while NIR and colorimetric assays remain valuable for centralized, high-throughput screening, our transcriptome-guided whole-cell biosensor array offers a highly competitive alternative, combining strong internal validation, robust external generalization, low instrumentation demands, and mechanistic interpretability. The integration of promoter-level transcriptomic signals with ensemble learning creates a versatile, non-invasive platform that bridges laboratory precision with real-world agricultural applicability. Future efforts to expand datasets and optimize biosensor performance will further position this approach as a leading tool for pre- and post-harvest aflatoxin risk management in diverse agricultural systems.

4. Conclusion

This study establishes a transcriptome-guided whole-cell biosensor array as a robust platform for early detection of *A. flavus* contamination and AFB_1 production in maize and peanuts. By selecting infection-

induced promoters identified through transcriptomic analysis and embedding them in calcium alginate–immobilized *E. coli*, the biosensor achieved substantial improvements over previous designs that relied on general stress-responsive promoters. When integrated with machine learning regression models, particularly XGBoost, the system consistently demonstrated high predictive accuracy ($R^2 > 0.94$ in internal validation) while maintaining strong generalizability in external validation datasets ($R^2 = 0.86$), surpassing the performance of conventional non-invasive detection methods such as NIR spectroscopy. Importantly, feature importance analysis revealed that transcriptional programs associated with DNA-templated signal transduction, cytosolic regulation, growth regulation, methylglyoxal catabolism, and biofilm formation served as critical early predictors of fungal infection and toxin accumulation, thereby providing mechanistic interpretability alongside predictive performance.

The validation of this approach across two distinct crop substrates underscores its versatility and scalability for broader agricultural contexts. Coupling promoter-level biological specificity with low-cost luminescence readouts, which can be captured using CMOS sensors or smartphone cameras, further highlights the feasibility of this platform for decentralized, field-based monitoring. Nonetheless, practical deployment will require systematic evaluation of biosensor stability under variable storage and environmental conditions, as well as adaptation to diverse crop types and fungal isolates. Future optimization through synthetic biology strategies, such as promoter engineering or directed evolution, combined with expansion of the training dataset to include geographically and genetically diverse strains, is expected to enhance sensitivity, specificity, and robustness. Collectively, these results position the transcriptome-guided biosensor as a promising tool for scalable, real-time aflatoxin risk assessment, offering significant potential to strengthen food safety monitoring systems and reduce postharvest losses in global supply chains.

Environmental Implication

Aspergillus flavus, a widespread saprotrophic and pathogenic fungus, produces aflatoxins, potent carcinogenic toxins that contaminate crops such as maize and peanuts. The environmental implications are significant, as these toxins not only threaten agricultural productivity but also pose serious risks to human and animal health through food contamination. Aflatoxins persist in the environment, especially in warm and humid climates, and can accumulate in food storage systems. On-site and real-time early detection of *A. flavus* contamination is crucial to mitigate aflatoxin production and reduce its environmental and health impacts.

CRediT authorship contribution statement

Lu Sun: Data curation, Formal analysis, Investigation, Visualization, Writing–original draft. **Junning Ma:** Methodology, Formal analysis, Conceptualization, Visualization, Writing–review & editing. **Yongping Jiang:** Investigation. Giorgia Purcaro: Validation, Writing–review & editing, Supervision. **Yuanyuan Tian:** Validation, Visualization. **Gang Wang:** Validation , Investigation. **Weizhao Li:** Validation, Visualization. **Bowen Tai:** Project administration. **Fuguo Xing:** Supervision, Writing–review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the Research and Development of Common Key Technologies and Equipment for Grain Postharvest Loss Reduction (CAAS-ZDRW202414, CAAS-ASTIP-G2025-IFST-09), and National Natural Science Foundation of China (32372458). The authors are grateful for the support provided by the foundation.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jhazmat.2025.139883.

Data availability

The authors do not have permission to share data.

References

[1] Jiarpinijnun, A., Osako, K., Siripatrawan, U., 2020. Visualization of volatomic profiles for early detection of fungal infection on storage jasmine brown rice using electronic nose coupled with chemometrics. Measurement 157, 107561. https://doi.org/10.1016/j.measurement.2020.107561.
[2] FAO, 2024. World food situation. Food and Agriculture Organization of the United Nations.
[3] Eskola, M., Kos, G., Elliott, C.T., Hajšlová, J., Mayar, S., Krska, R., 2020. Worldwide contamination of food-crops with mycotoxins: validity of the widely cited 'FAO estimate'of 25%. Crit Rev Food Sci Nutr 60 (16), 2773–2789. https://doi.org/10.1080/10408398.2019.1658570.
[4] Sun, L., Li, R., Tai, B., Hussain, S., Wang, G., Liu, X., Xing, F., 2023. Current status of major mycotoxins contamination in food and feed in Asia— a review. Food sci technol, 3 (2), 231–244. https://doi.org/10.1021/acsfoodscitech.2c00331.
[5] Stutt, R.O., Castle, M.D., Markwell, P., Baker, R., Gilligan, C.A., 2023. An integrated model for pre-and post-harvest aflatoxin contamination in maize. NPJ Sci Food 7 (1), 60. https://doi.org/10.1038/s41538-023-00238-7.
[6] Ray, M., Ray, A., Dash, S., Mishra, A., Achary, K.G., Nayak, S., Singh, S.J.B., Bioelectronics, 2017. Fungal disease detection in plants: traditional assays, novel diagnostic techniques and biosensors. Biosens Bioelectron 87, 708–723. https://doi.org/10.1016/j.bios.2016.09.032.
[7] Abd El-Aziz, A.R., Shehata, S.M., Hisham, S.M., Alobathani, A.A., 2021. Molecular profile of aflatoxigenic and non-aflatoxigenic isolates of *aspergillus flavus* isolated from stored maize. Saudi J Biol Sci 28 (2), 1383–1391. https://doi.org/10.1016/j.sjbs.2020.11.073.
[8] Garcia-Lopez, M.T., Luo, Y., Ortega-Beltran, A., Jaime, R., Moral, J., Michailides, T. J., 2021. Quantification of the aflatoxin biocontrol strain *aspergillus flavus* AF36 in soil and in nuts and leaves of pistachio by real-time PCR. Plant Dis 105 (6), 1657–1665. https://doi.org/10.1094/PDIS-05-20-1097-RE.
[9] Li, J., Macdonald, J., Von Stetten, F., 2019. A comprehensive summary of a decade development of the recombinase polymerase amplification. Analyst 144 (1), 31–67. https://doi.org/10.1039/C8AN01621F.
[10] Wang, T., Li, P., Zhang, Q., Zhang, W., Zhang, Z., Wang, T., He, T., 2017. Determination of *aspergillus* pathogens in agricultural products by a specific nanobody-polyclonal antibody sandwich ELISA. Sci Rep 7 (1), 4348. https://doi.org/10.1038/s41598-017-04195-6.
[11] Gong, Y., Yang, T., Liang, Y., Ge, H., Shen, E., 2021. Comparative assessments between conventional and promising technologies for wheat aging or mold detection. Cereal Res Commun 1–9. https://doi.org/10.1007/s42976-021-00137-1.
[12] Guo, Z., Guo, C., Chen, Q., Ouyang, Q., Shi, J., El-Seedi, H.R., Zou, X., 2020. Classification for *penicillium* expansum spoilage and defect in apples by electronic nose combined with chemometrics. Sens (Basel) 20 (7), 2130. https://doi.org/10.3390/s20072130.
[13] Lin, H., Kang, W., Kutsanedzie, F.Y., Chen, Q., 2019. A novel nanoscaled chemo dye–based sensor for the identification of volatile organic compounds during the mildewing process of stored wheat. Food Anal Methods 12, 2895–2907. https://doi.org/10.1007/s12161-019-01617-1.
[14] Ali, A., Mansol, A.S., Khan, A.A., Muthoosamy, K., Siddiqui, Y., 2023. Electronic nose as a tool for early detection of diseases and quality monitoring in fresh postharvest produce: a comprehensive review. Compr Rev Food Sci Food Saf 22 (3), 2408–2432. https://doi.org/10.1111/1541-4337.13151.
[15] Chalupowicz, D., Veltman, B., Droby, S., Eltzov, E., 2020. Evaluating the use of biosensors for monitoring of *penicillium digitatum* infection in citrus fruit. Sens Actuat BChem 311, 127896. https://doi.org/10.1016/j.snb.2020.127896.
[16] Ma, J., Guan, Y., Liu, Y., Wang, G., Tai, B., Xing, F., 2023. Assessment of *escherichia coli* bioreporters for early detection of fungal spoilage in postharvest grape berries. Postharvest Biol Technol 204, 112481. https://doi.org/10.1016/j.postharvbio.2023.112481.
[17] Ma, J., Guan, Y., Xing, F., Eltzov, E., Wang, Y., Li, X., Tai, B., 2023. Accurate and non-destructive monitoring of mold contamination in foodstuffs based on whole-cell biosensor array coupling with machine-learning prediction models. J Hazard Mater 449, 131030. https://doi.org/10.1016/j.jhazmat.2023.131030.
[18] Veltman, B., Ma, J., Harpaz, D., Xing, F., Eltzov, E., 2023. Genetically engineered bacterial strains constructed as a whole-cell biosensor for specific volatiles

identification of infected potato tubers with a soft rot disease. Sens Actuat BChem 387, 133788. https://doi.org/10.1016/j.snb.2023.133788.
[19] Smidsrød, O., Skjåk-Br1k, G., 1990. Alginate as immobilization matrix for cells. Trends Biotechnol 8, 71–78. https://doi.org/10.1016/0167-7799(90)90139-O.
[20] Fan, M., Rakotondrabe, T.F., Chen, G., Guo, M., 2023. Advances in microbial analysis: based on volatile organic compounds of microorganisms in food. Food Chem 418, 135950. https://doi.org/10.1016/j.foodchem.2023.135950.
[21] Dorogush, A.V., Ershov, V., Gulin, A., 2018. CatBoost: gradient boosting with categorical features support. arXiv: 1810.11363. https://doi.org/10.48550/arXiv.1810.11363.
[22] Lawrence, R., Bunn, A., Powell, S., Zambon, M., 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. Remote Sens Environ 90 (3), 331–336. https://doi.org/10.1016/j.rse.2004.01.007.
[23] Nielsen, D., 2016. Tree boosting with xgboost-why does xgboost win" every" machine learning competition? NTNU. ⟨https://api.semanticscholar.org/CorpuSID:114191144⟩.
[24] Chueh, B.H., Zheng, Y., Torisawa, Y.S., Hsiao, A.Y., Ge, C., Hsiong, S., Huebsch, N., Franceschi, R., Mooney, D.J., Takayama, S., 2010. Patterning alginate hydrogels using light-directed release of caged calcium in a microfluidic device. Biomed Micro 12 (1), 145–151. https://doi.org/10.1007/s10544-009-9369-6.
[25] Wang, Z., Feng, Z., Yang, L., Wang, M., 2021. Effective removal of calcium and magnesium ions from water by a novel alginate-citrate composite aerogel. Gels (Basel) 7 (3), 125. https://doi.org/10.3390/gels7030125.
[26] Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., Soldatov, A., 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res 37, e123. https://doi.org/10.1093/nar/gkp596.
[27] Duarte, J.C., Rodrigues, J.A., Moran, P.J., Valença, G.P., Nunhez, J.R., 2013. Effect of immobilized cells in calcium alginate beads in alcoholic fermentation. AMB Express 3 (1), 31. https://doi.org/10.1186/2191-0855-3-31.
[28] Voo, W.-P., Ooi, C.-W., Islam, A., Tey, B.-T., Chan, E.-S., 2016. Calcium alginate hydrogel beads with high stiffness and extended dissolution behaviour. Eur Polym J 75, 343–353. https://doi.org/10.1016/j.eurpolymj.2015.12.029.
[29] Ma, J., Veltman, B., Tietel, Z., Tsror, L., Liu, Y., Eltzov, E., 2020. Monitoring of infection volatile markers using CMOS-based luminescent bioreporters. -121347 Talanta 219, 121333. https://doi.org/10.1016/j.talanta.2020.121333.
[30] López-Menchero, J.R., Ogawa, M., Mauricio, J.C., Moreno, J., Moreno-García, J., 2021. Effect of calcium alginate coating on the cell retention and fermentation of a fungus-yeast immobilization system. LWT 144, 111250. https://doi.org/10.1016/j.lwt.2021.111250.
[31] Josselin, L., De Clerck, C., De Boevre, M., Moretti, A., Jijakli, M.H., Soyeurt, H., Fauconnier, M.L., 2021. Volatile organic compounds emitted by *aspergillus* flavus strains producing or not aflatoxin B1. Toxins 13 (10). https://doi.org/10.3390/toxins13100705.
[32] Li, H., Kang, X., Wang, S., Mo, H., Xu, D., Zhou, W., Hu, L., 2021. Early detection and monitoring for *aspergillus flavus* contamination in maize kernels. Food Control 121, 107636. https://doi.org/10.1016/j.foodcont.2020.107636.
[33] Yung, P.Y., Lo Grasso, L., Mohidin, A.F., Acerbi, E., Hinks, J., Seviour, T., Marsili, E., Lauro, F.M., 2016. Global transcriptomic responses of *escherichia coli* K-12 to volatile organic compounds. Sci Rep 6 (1), 19899. https://doi.org/10.1038/srep19899.
[34] Abis, L., Loubet, B., Ciuraru, R., Lafouge, F., Houot, S., Nowak, V., Tripied, J., Dequiedt, S., Maron, P.A., Sadet-Bourgeteau, S., 2020. Reduced microbial diversity induces larger volatile organic compound emissions from soils. Sci Rep 10 (1), 6104. https://doi.org/10.1038/s41598-020-63091-8.
[35] Masuda, N., Church, G.M., 2002. *escherichia coli* gene expression responsive to levels of the response regulator EvgA. J Bacteriol 184 (22), 6225–6234. https://doi.org/10.1128/jb.184.22.6225-6234.2002.
[36] Tchieu, J.H., Norris, V., Edwards, J.S., Saier Jr., M.H., 2001. The complete phosphotransferase system in *escherichia coli*. J Mol Microbiol Biotechnol 3 (3), 329–346.
[37] Mok, W.W., Navani, N.K., Barker, C., Sawchyn, B.L., Gu, J., Pathania, R., Zhu, R.D., Brown, E.D., Li, Y., 2009. Identification of a toxic peptide through bidirectional expression of small RNAs. Chembiochem 10 (2), 238–241. https://doi.org/10.1002/cbic.200800591.
[38] Serina, S., Nozza, F., Nicastro, G., Faggioni, F., Mottl, H., Dehò, G., Polissi, A., 2004. Scanning the *escherichia coli* chromosome by random transposon mutagenesis and multiple phenotypic screening. Res Microbiol 155 (8), 692–701. https://doi.org/10.1016/j.resmic.2004.05.006.
[39] Tan, J., Kan, N., Wang, W., Ling, J., Qu, G., Jin, J., Shao, Y., Liu, G., Chen, H., 2015. Construction of 2,4,6-trinitrotoluene biosensors with novel sensing elements from *escherichia coli* K-12 MG1655. Cell Biochem Biophys 72 (2), 417–428. https://doi.org/10.1007/s12013-014-0481-8.
[40] Fernández-Ibañez, V., Soldado, A., Martínez-Fernández, A., de la Roza-Delgado, B., 2009. Application of near infrared spectroscopy for rapid detection of AFB_1 in maize and barley as analytical quality assessment. Food Chem 113, 629–634. https://doi.org/10.1016/j.foodchem.2008.07.049.
[41] Jiang, H., Wang, J., Chen, Q., 2021. Comparison of wavelength selected methods for improving of prediction performance of PLS model to determine aflatoxin B_1 (AFB_1) in wheat samples during storage. Microchem J 170, 106642. https://doi.org/10.1016/j.microc.2021.106642.
[42] Jiang, H., Wang, J., Mao, W., Chen, Q., 2022. Determination of aflatoxin B_1 in wheat based on colourimetric sensor array technology: optimization of sensor features and model parameters to improve the model generalization performance. Microchem J 175, 107173. https://doi.org/10.1016/j.microc.2022.107173.
[43] Ma, J., Harpaz, D., Liu, Y., Eltzov, E., 2020. Smartphone-based whole-cell biosensor platform utilizing an immobilization approach on a filter membrane disk for the monitoring of water toxicants. Sensors 20 (19), 5486–5490. https://doi.org/10.3390/s20195486.
[44] Liu, C., Yu, H., Zhang, B., Liu, S., Liu, C., Li, F., Song, H., 2022. Engineering whole-cell microbial biosensors: design principles and applications in monitoring and treatment of heavy metals and organic pollutants. Biotechnol Adv 60, 108019. https://doi.org/10.1016/j.biotechadv.2022.108019.
[45] Chen, C., Liu, J., Yao, G., Bao, S., Wan, X., Wang, F., Wang, K., Song, T., Han, P., Liu, T., Jiang, H., 2023. A novel, genetically encoded whole-cell biosensor for directed evolution of myrcene synthase in *escherichia coli*. Biosens Bioelectron 228, 115176. https://doi.org/10.1016/j.bios.2023.115176.