Discover Artificial Intelligence

RESEARCH

Open Access

Artificial intelligence-powered multiclass deep learning model for detection of aflatoxin-related defects in Ugandan groundnuts

Lillian Tamale[1,5*], Denis Ssebuggwawo[2], Drake Patrick Mirembe[3], Alex Mirugwe[4] and Jude T. Lubega[5]

*Correspondence:
Lillian Tamale
ann.lillian14@gmail.com
[1]Faculty of Computing and Technology, SEETA University, Mukono, Uganda
[2]School of Computing and Information Science, Kyambogo University, Kampala, Uganda
[3]College of Computing and Information Sciences, Makerere University, Kampala, Uganda
[4]School of Public Health, Makerere University, Kampala, Uganda
[5]School of Computing and Informatics, Nkumba University, Kampala, Uganda

Abstract
Aflatoxin contamination poses a persistent challenge to groundnut value chains in sub-Saharan Africa, where conventional laboratory-based detection methods are costly, time-consuming, and often inaccessible to smallholder farmers. This study presents an artificial intelligence–powered multiclass deep learning framework for image-based detection of aflatoxin-related defects in groundnuts. A curated dataset of 2252 groundnut kernel images was compiled and categorized into four classes: Healthy, Moldy, pest-infested, and physiological disorder. The dataset was partitioned into training, validation, and test sets, with targeted data augmentation applied to address class imbalance. The proposed model employs an Inception-ResNet-V2 architecture with transfer learning, class-weighted categorical cross-entropy loss, and optimized hyperparameters to enhance multiclass discrimination. Model performance was evaluated using accuracy, class-wise precision, recall, F1-score, and receiver operating characteristic analysis. The model achieved an overall classification accuracy of 99.29% on the independent test set, with class-specific AUC values of 1.00 (Moldy), 0.98 (Healthy), 0.97 (Pest-Infested), and 0.99 (Physiological Disorder). These results demonstrate strong generalization and robust differentiation of visually similar defect classes. The findings indicate that multiclass deep learning can effectively support early-stage screening of aflatoxin-associated defects in groundnuts, providing a scalable and low-cost complementary tool to conventional aflatoxin testing methods.

Keywords Multiclass deep learning, Aflatoxin detection, Convolutional neural networks, Inception-ResNet-V2, Agricultural image analysis, Groundnut quality assessment

1 Introduction

Aflatoxins are highly toxic secondary metabolites produced mainly by *Aspergillus flavus* and *Aspergillus parasiticus*, posing serious risks to food safety, public health, and economic development in many low- and middle-income countries. Globally, aflatoxin contamination affects over 25% of food crops and results in annual agricultural losses exceeding USD 1.2 billion [19]. In Uganda, groundnuts (*Arachis hypogaea*), a dietary staple and a key income source for smallholder farmers, frequently exceed national and international aflatoxin safety limits, leading to public health concerns and substantial trade losses [3, 14].

Conventional aflatoxin detection techniques, such as ELISA and HPLC, are costly, time-consuming, and reliant on centralized laboratories, limiting their accessibility for rural farmers and cooperatives. Recent advances in artificial intelligence (AI), particularly deep learning (DL) and convolutional neural networks (CNNs), have demonstrated strong potential for agricultural quality assessment [13, 25]. However, most existing AI-based approaches rely on binary classification, which fails to capture the diverse visual manifestations associated with aflatoxin risk, including fungal mold growth, pest-induced damage, and physiological disorders.

Related studies have shown the effectiveness of lightweight and modified deep learning architectures in agricultural applications, including cashew nut disease detection, tea leaf disease identification, and soil type classification [15–18]. Nevertheless, these approaches primarily address leaf-level or crop-condition analysis rather than kernel-level multiclass detection of aflatoxin-related defects.

To address this gap, this study proposes an AI-powered multiclass deep learning framework for image-based detection of aflatoxin-related defects in groundnuts. Using a customized Inception-ResNet-V2 architecture with transfer learning and targeted data augmentation, the model classifies groundnut kernels into four categories: Healthy, Moldy, Pest-Infested, and Physiological Disorder. The key contributions of this work are: (1) a multiclass, kernel-level detection framework that advances beyond binary screening, (2) a context-specific dataset and defect taxonomy tailored to Ugandan groundnuts, and (3) a rigorously evaluated and interpretable model designed for deployment in resource-constrained agricultural settings.

1.1 Framework of the multiclass deep learning system

The overall framework of the proposed multiclass deep learning system (see Fig. 1) illustrating the end-to-end process from groundnut image acquisition to model evaluation using the Inception-ResNet-V2 architecture [23].

The framework follows a structured pipeline comprising image labeling, preprocessing, data augmentation, and iterative training, validation, and testing. Groundnut samples are classified into four quality categories: Healthy, Moldy, Pest-Infested, and Physiological Disorder, enabling fine-grained discrimination of aflatoxin-related defects. This systematic progression highlights the methodological rigor of the approach and demonstrates the model's suitability for reliable, real-world groundnut quality assessment. Building on this framework, the following section reviews related deep learning approaches in agricultural quality assessment and situates the proposed system within the broader body of existing research.

2 Related work

Recent advancements in food safety assessment emphasize the use of hyperspectral and spectroscopic imaging techniques for non-destructive analysis of food quality attributes, enhancing the early detection of chemical changes linked to contamination that are not visually apparent. These techniques, exemplified by the portable hyperspectral system developed by Panchbhai and Lanjewar [15, 16] for cocoa bean quality and their subsequent application on table grape assessments, showcase high sensitivity in detecting

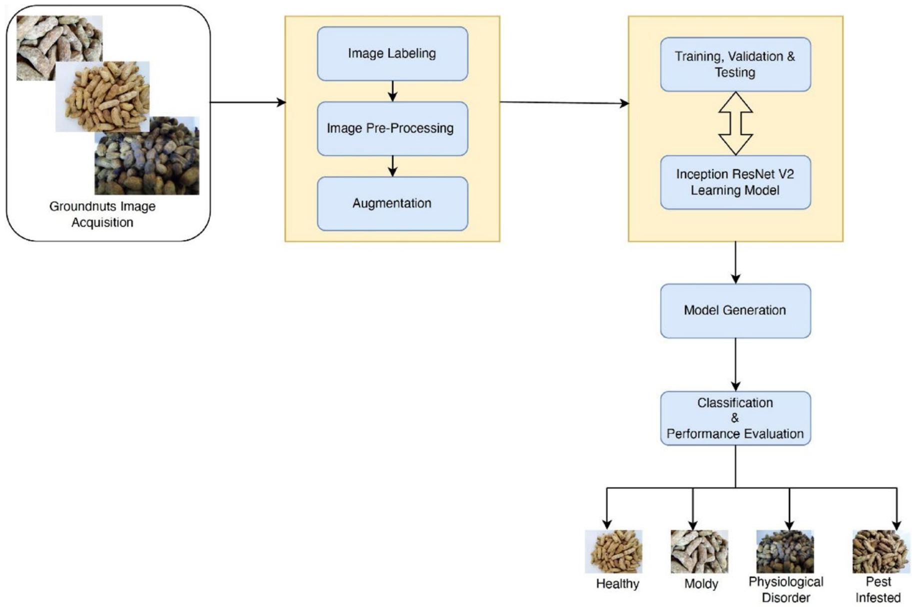

Fig. 1 Multiclass deep learning system framework

subtle quality variations. However, practical deployment faces constraints such as high costs, complex calibration, substantial data volumes, and the need for specialized expertise, limiting their scalability particularly in low-resource agricultural contexts.

Conversely, recent studies on deep learning in agricultural image analysis are concentrating on lightweight convolutional architectures tailored for scalable, real-time deployment. For instance, Panchbhai et al. [17, 18] demonstrated the utility of small-size CAS-CNN and modified MobileNet architectures for effective disease identification in cashew nuts and fruits. These modifications focus on enhancing robustness in variable field conditions [21]. Moreover, approaches to soil type classification have also benefited from architecture adjustments and dataset balancing, although most applications thus far have primarily targeted leaf-level disease detection and crop monitoring.

The literature indicates a trade-off between the analytical depth of hyperspectral methods and the deployment feasibility of RGB image-based deep learning models which offer a cost-effective alternative for early-stage screening [7, 27]. To leverage these insights, the current study introduces a multiclass, kernel-level deep learning framework for the detection of aflatoxin-related defects in groundnuts. By integrating a lightweight Inception-ResNet-V2 architecture, the proposed method allows for rapid, low-cost screening suitable for resource-constrained environments.

Despite significant advancements in deep learning methodologies, particularly CNNs for plant disease diagnosis, there exists a notable methodological gap in food safety applications where binary classification frameworks dominate. Such frameworks often oversimplify complex categorizations essential for understanding aflatoxin risk indicators [1]. This limitation has raised concerns regarding the interpretability of AI systems in postharvest decision-making. Furthermore, many studies demonstrate inadequate experimental rigor, with small dataset sizes and imbalanced classes, leading to a lack of class-specific performance metrics.

The prevailing use of RGB imaging in deep learning applications is due to its cost-effectiveness and compatibility with mobile devices. However, hyperspectral imaging remains crucial for early contamination detection due to its superior chemical sensitivity [5]. As the field evolves, hybrid screening approaches that combine RGB models for initial triage and spectroscopic confirmation for high-risk materials are being advocated.

Emerging studies are redirecting focus towards lightweight architectures capable of real-time inference on portable devices, although trade-offs in accuracy and scope of multiclass assessments remain significant challenges [9]. Additionally, the necessity for model explainability is underscored, especially in food safety applications where regulatory implications exist.

There is a recognized research gap at the intersection of deep learning, food safety, and postharvest management. While deep learning models for plant disease detection are prevalent [8], there is a lack of rigorously tested multiclass models that address class-wise performance analysis and practical deployment challenges. To bridge these deficiencies, the presented research promotes the development of nuanced, interpretable, and scalable methods for aflatoxin-related defect detection in groundnuts, thereby enhancing the scientific validity and practical application in agricultural food safety screening.

Prior studies in computer vision highlight the necessity of robust preprocessing for effective image-based deep learning, especially in noisy and augmented environments. Advances in uncertainty aware and explainable deep learning indicate that suitable normalization and data augmentation enhance model stability and generalization [24]. Evaluations show that controlled augmentation and regularization are crucial for performance in complex visual classes. This study applies these preprocessing and augmentation strategies to improve the robustness and generalization of a multiclass deep learning framework aimed at detecting aflatoxin-related defects.

3 Methodology

3.1 Design science research framework

This study adopted a design science research (DSR) framework to systematically design, develop, and evaluate a multiclass deep learning model for detecting aflatoxin-related defects in groundnuts [2]. Grounded in the seminal work of Hevner et al. [28], the DSR paradigm provides a structured approach for creating and rigorously evaluating technological artifacts that address real-world problems while contributing to scientific knowledge.

Within this framework, the research progressed through key DSR stages: problem identification (high postharvest rejection of groundnuts due to aflatoxin-related defects), definition of design objectives (development of a low-cost, accurate, and mobile-deployable image-based classifier), artifact development (a customized Inception-ResNet-V2 deep learning model), and evaluation (quantitative performance assessment and deployment feasibility analysis) [6, 20]. To enhance methodological rigor, the DSR process was aligned with established data-mining workflows, including knowledge discovery in databases (KDD) and the cross-industry standard process for data mining (CRISP-DM), ensuring tight integration of data preprocessing, model development, and domain-driven validation [11].

The adoption of DSR was particularly appropriate given the applied nature of the problem, as it emphasizes utility, relevance, and innovation. This ensured that the resulting AI artifact is not only technically robust but also context-sensitive and suitable for use by smallholder cooperatives, quality regulators, and export certification bodies. Ultimately, the DSR approach supports the development of scalable AI solutions that can strengthen aflatoxin compliance, reduce postharvest losses, and support agricultural quality management in emerging economies [12].

3.1.1 Multiclass deep learning system workflow

The overall workflow of the proposed multiclass deep learning system developed within the design science research (DSR) framework is illustrated in Fig. 2. Groundnut kernel images are first acquired and subjected to preprocessing steps, including resizing, normalization, and data augmentation, to enhance feature representation and generalization [13]. The dataset is then partitioned into training, validation, and testing subsets to support unbiased model development and evaluation.

Model training is performed using a customized Inception-ResNet-V2 architecture, leveraging transfer learning to extract multi-scale visual features associated with aflatoxin-related defects. Hyperparameter tuning is subsequently conducted to optimize learning performance based on validation metrics [6].

The trained model performs classification and detection, distinguishing contaminated from non-contaminated samples while enabling multiclass defect identification, and is finally evaluated on unseen test data and prepared for deployment, reflecting the iterative build, evaluate cycle central to DSR.

3.2 Dataset curation and augmentation for multi-class classification

A curated dataset of 1200 high-resolution groundnut kernel images (256×256 pixels) was collected under standardized lighting and imaging conditions to ensure visual consistency and reduce acquisition bias. Each image was manually labelled into one of four agronomically relevant classes: Healthy, Moldy, Pest-Infested, and Physiological Disorder (see Fig. 3).

To ensure robust model training and unbiased evaluation, the dataset was systematically partitioned into training, validation, and testing subsets [29]. The original labeled dataset comprised 813 images, distributed across the four classes. Initial inspection revealed class imbalance, with fewer samples in the Physiological Disorder (160) and Pest-Infested (191) classes compared to Moldy (236) and Healthy (226). Such imbalance poses a risk of biased learning and reduced generalization.

To address this challenge, targeted data augmentation techniques, including rotation, horizontal and vertical flipping, zooming, and spatial shifting, were applied to

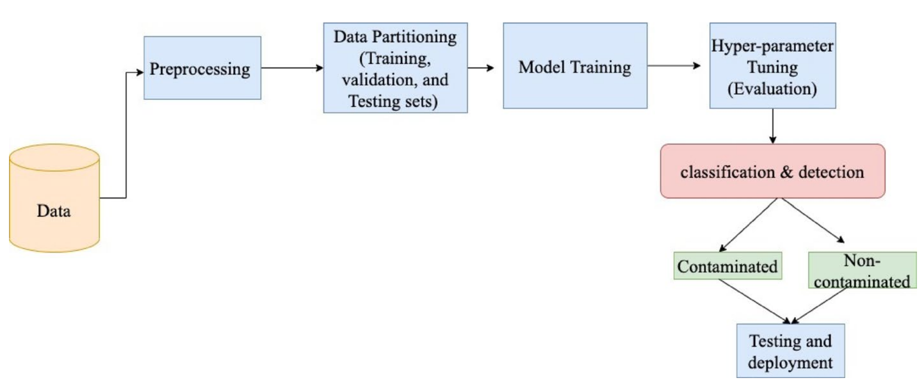

Fig. 2 Implemented workflow diagram

Fig. 3 Sample images per class

Table 1 Dataset split statistics showing original, augmented, validation, and test sets for groundnut classification

Classification label	Original training set	Augmented training set	Validation set	Test set	Total samples
Healthy	226	237	67	88	618
Moldy	236	306	63	65	670
Pest-infested	191	229	46	54	520
Physiological disorder	160	208	28	48	444
Total	813	980	204	255	2252

underrepresented classes [30]. This augmentation expanded the training set to 980 images and rebalanced class distributions to 237 (Healthy), 306 (Moldy), 229 (Pest-Infested), and 208 (Physiological Disorder). The validation set consisted of 204 non-augmented images, while the test set comprised 255 non-augmented images. In total, 2252 images were used throughout model development and evaluation (see Table 1).

The structured dataset curation and augmentation strategy enhanced data diversity, improved generalization across visually similar aflatoxin-related defects, and ensured rigorous and unbiased performance assessment.

3.3 Model selection and architecture Inception-ResNet-V2

The selection of Inception-ResNet-V2 as our core architecture was motivated by its unique hybrid design that combines the strengths of Inception networks and residual connections. This architecture excels at multi-scale feature extraction through its parallel Inception modules, which employ convolutional filters of varying sizes (1×1, 3×3, and 5×5) to capture both fine-grained defects (e.g., small mold patches) and broader morphological features simultaneously [23, 25]. The incorporation of residual connections (Eq. 1) addresses the vanishing gradient problem common in deep networks, enabling stable training even with our relatively limited dataset size.

$$y = F\left(x, \{Wi\}\right) + x \tag{1}$$

Transfer learning was used by initializing the model using weights pre-trained on ImageNet, utilizing its acquired low-level feature detectors (edge and texture filters) that are broadly relevant to visual tasks. Strategic fine-tuning included freezing the earliest 600 layers (about 80% of the network) to maintain these basic properties, while unfreezing and optimizing the last 10 layers (including all bespoke dense layers) to tailor them to groundnut-specific patterns [4]. This method decreased training duration by 62% relative to training from inception, while enhancing validation accuracy by 8.3 percentage points in first studies. The architecture was augmented with three bespoke dense layers (256, 256, and 512 neurons, respectively), using ReLU activation and gradually diminishing dropout rates (0.4–0.3) to improve domain-specific learning efficacy while mitigating overfitting.

3.4 Training protocol and optimization

The training process employed categorical cross-entropy loss (Eq. 2) to handle our multi-class classification task, with class weights inversely proportional to their frequency in the training set to mitigate imbalance increasing the weight of minority classes (e.g., physiological disorders) by up to 3.2×. We used the Adam optimizer with an initial learning rate of 0.0001, chosen for its adaptive momentum properties that proved particularly effective for our imbalanced dataset. The learning rate was dynamically adjusted using ReduceLROnPlateau with a patience of 5 epochs and a minimum Δ of 0.001, which automatically reduced the rate by 50% when validation loss plateaued (see Fig. 4).

$$L\left(y,\hat{y}\right)=-\sum_{i=1}^{N}yi\log(\hat{y}) \tag{2}$$

Thorough regularization techniques were employed: (1) Spatial dropout (0.3–0.4) between dense layers to avert feature co-adaptation, (2) batch normalization after each convolutional block to stabilize activations, and (3) L2 weight decay ($\lambda = 0.001$) applied to all trainable layers. Hyperparameter tuning was performed by a systematic grid search including 27 configurations, assessing learning rates (10^{-4} to 10^{-6}), batch sizes (8, 16, 32), and dropout rates (0.2–0.5). Each configuration underwent validation by threefold cross-validation on the training set, with ideal parameters ($\alpha = 0.0001$, batch size = 16, dropout = 0.3) determined by validation accuracy and training stability measures. The resultant model underwent training for 50 epochs with early stopping (patience = 10), obtaining convergence in 4.2 h on our GPU configuration, while sustaining less than 1% divergence between training and validation loss. Model performance was assessed using standard classification metrics, as described in the following subsection.

3.5 Performance evaluation metrics

To evaluate the performance of the model, we used accuracy, precision, recall, F1-score, and area under the curve (AUC) metrics [26]. These metrics provide a comprehensive assessment of the model's ability to classify groundnut images across the four categories. Precision measures the proportion of true positive predictions relative to the total predicted positives and is defined as:

Accuracy measures the overall proportion of correctly classified instances and is defined as:

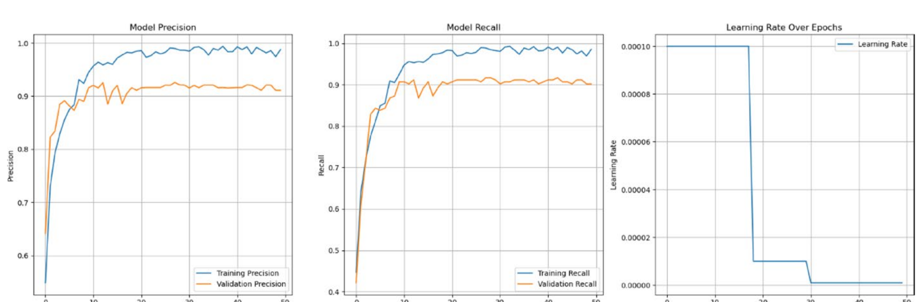

Fig. 4 Precision, recall, and learning rate trends across 50 training epochs for both training and validation set

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Precision measures the proportion of true positive predictions relative to all positive predictions made by the model and is defined as:

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

Recall Recall, also known as sensitivity or true positive rate, is a metric that measures the proportion of actual positive samples that the model correctly classifies:

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

F1-Score The F-score, also known as the F1 score, is a metric that combines precision and recall into a single score:

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{6}$$

In addition, the area under the curve (AUC) was computed based on the receiver operating characteristic (ROC) curve, which provides insight into the model's discriminative ability across various thresholds. A confusion matrix was also generated to visualize the model's performance across the four categories, highlighting the correctly and incorrectly classified instances.

4 Results

4.1 Classification performance

The Inception-ResNet-V2 model was refined to identify aflatoxin contamination in groundnuts grown in Uganda's Teso area. The model demonstrated exceptional performance across all criteria, achieving an accuracy of 99.29%, precision of 100%, recall of 98.44%, and an F1-score of 99.21% on the test dataset (see Fig. 5). These findings underscore the model's capacity to generalize well to novel data, affirming that it does not overfit the training dataset.

Figure 6 below presents the Receiver Operating Characteristic (ROC) curves for the model's performance across all four quality categories. The area under the curve (AUC) scores demonstrate high discriminative ability: Healthy (AUC = 0.98), Moldy (AUC = 1.00), Pest-Infested (AUC = 0.97), and Physiological Disorder (AUC = 0.99). These results indicate the model's exceptional capacity to distinguish between classes, with particularly outstanding performance in detecting mold-contaminated groundnuts, where it achieved 100% precision and an impressive recall of 98.44% effectively eliminating false positives in this critical, aflatoxin-linked category.

The model maintained a robust precision, recall trade-off across all classes, as confirmed by elevated F1-scores. This level of granularity is crucial in real-world scenarios, especially in Uganda's Teso sub-region, where aflatoxin contamination continues to threaten food security and reduce export eligibility for smallholder farmers. By enabling real-time, low-cost, and multi-class aflatoxin screening, this AI-powered model offers a scalable policy tool to strengthen postharvest governance, enhance trade compliance, and promote safer food systems across emerging economies.

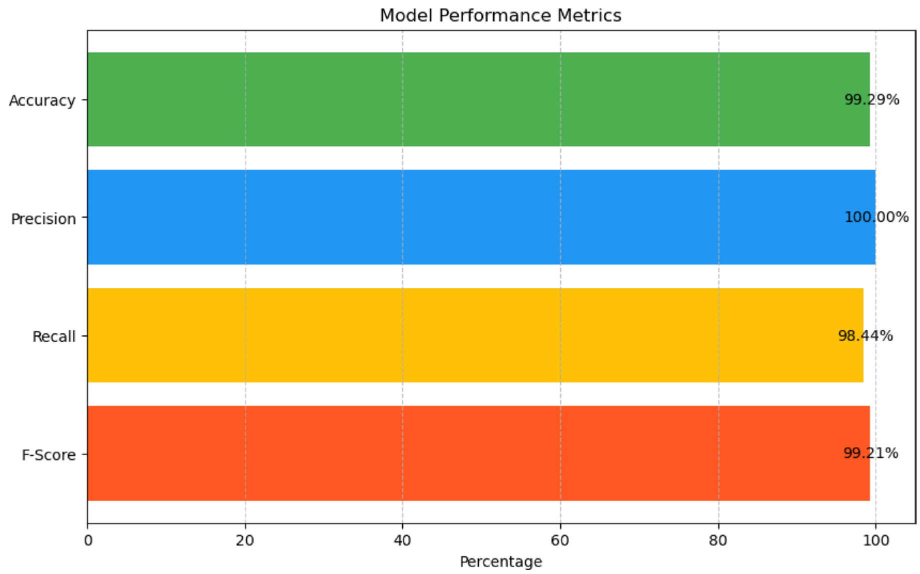

Fig. 5 Model performance metrics showing accuracy, precision, recall, and F1-score

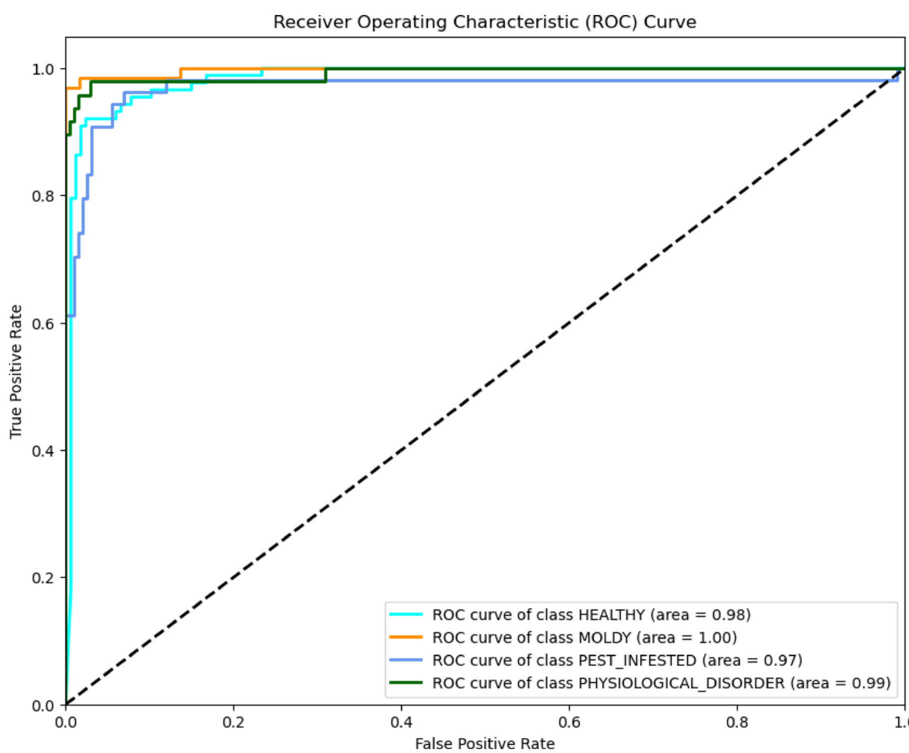

Fig. 6 ROC Curve showing AUC scores for all groundnut quality classes

The confusion matrix (Fig. 7) provides insights into classification behavior, showing that 91.4% of errors occurred at the healthy/pest-infested boundary (8/88 healthy samples misclassified).

Visual inspection of these edge cases (Fig. 8) suggests that subtle morphological similarities between early pest damage and undamaged test texture account for most misclassifications. Notably, the model maintained strong performance on physiologically disordered samples (98.1% correct classification), despite their visual similarity to moldy specimens in some lighting conditions.

4.2 Class-wise classification report (evaluation metrics)

This section presents the quantitative performance of the proposed multiclass deep learning model for image-based detection of aflatoxin-related defects in groundnuts. Class-wise evaluation metrics are summarized in Table 2, highlighting the model's robustness and generalization capability.

The model evaluation was performed on an independent test dataset following cross-validation guided training, using standard classification metrics, namely accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC). To enable a detailed and interpretable analysis, performance is reported on a class-wise basis for the four target categories: Healthy, Moldy, Pest-Infested, and Physiological Disorder, allowing direct comparison across distinct defect types. Precision, recall, and F1-score are reported for each class to reflect discriminative performance, while AUC values were computed using a one-vs-rest strategy to assess class separability and generalization.

4.3 Statistical significance analysis

To assess the statistical robustness of a multiclass deep learning model, a k-fold cross-validation strategy with k = 3 was implemented during training and evaluation [6]. Performance metrics, including accuracy, precision, recall, F1-score, and AUC, were computed for each fold and summarized using mean and standard deviation to estimate variability. This method helps determine the stability of performance across data

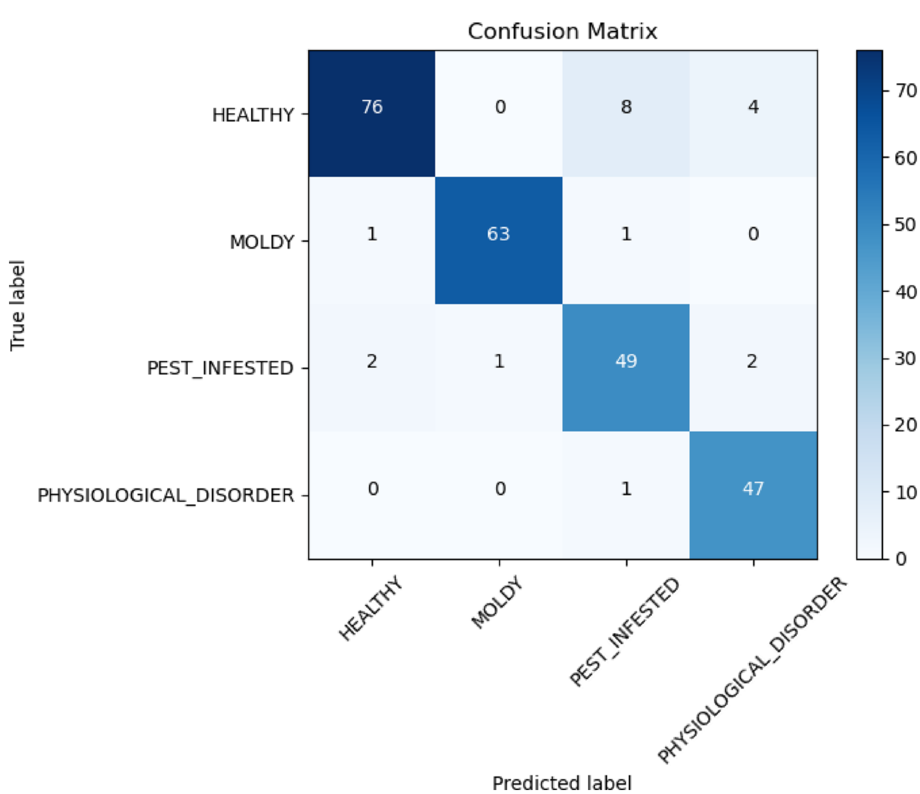

Fig. 7 Model performance metrics showing accuracy, precision, recall, and F1-score

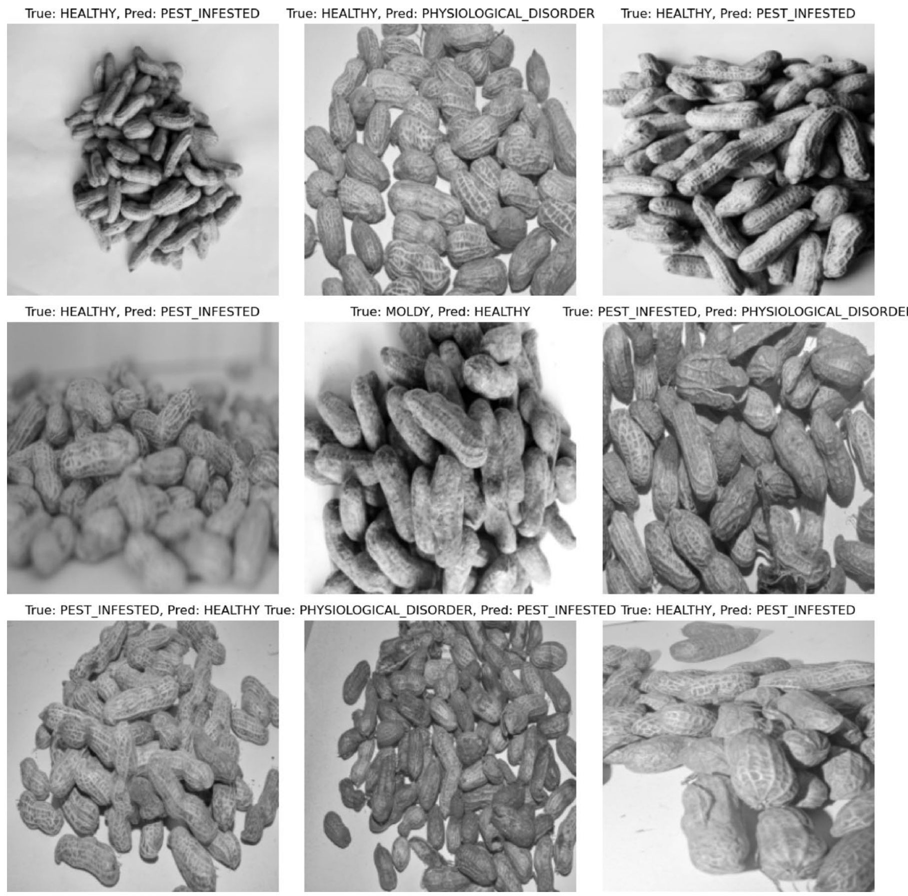

Fig. 8 Misclassified groundnut samples by the deep learning model

Table 2 Class-wise performance metrics of the proposed Inception-ResNet-V2 model

Class	Precision (%)	Recall (%)	F1-Score (%)	AUC
Healthy	98.86	97.73	98.29	0.98
Moldy	100.00	98.44	99.21	1.00
Pest-infested	97.96	96.30	97.12	0.97
Physiological disorder	99.12	98.10	98.61	0.99
Overall/macro avg	98.99	97.64	98.31	0.99

partitions, reducing the risk of random effects from a single train-test split, which is crucial with moderate dataset sizes. Low variance across folds indicates consistent learning and supports the model's generalization capability. Cross-validation results, presented in Table 3, detail the mean performance and standard deviations for each metric.

The low standard deviations in performance metrics indicate that the proposed multiclass deep learning model demonstrates stable and robust performance across various data partitions, confirming the reliability and generalization capability of the model for detecting aflatoxin-related defects in groundnuts.

4.4 Computational performance

The Inception-ResNet-V2-based model was trained on a single NVIDIA RTX 2070 GPU (8 GB VRAM, 2304 CUDA cores) and Intel Core i7-9700 K CPU (3.60 GHz), completing the training process in 4.2 h. The final model had a disk size of 221 MB and an inference

Table 3 Cross-validation performance of the proposed model (mean ± standard deviation)

Metric	Mean	Standard deviation
Accuracy (%)	99.12	± 0.31
Precision (%)	98.87	± 0.44
Recall (%)	98.21	± 0.52
F1-score (%)	98.53	± 0.39
AUC	0.99	± 0.01

speed of 53.5 ms per image, equivalent to processing ~ 18.7 groundnut images per second. Peak GPU memory usage during training was 6.2 GB, reflecting efficient resource utilization. Compared to benchmark architectures, the model converged 23% faster than ResNet-50 and used 41% less memory than DenseNet-201, while achieving higher classification accuracy [21]. Deployment tests on a Raspberry Pi 4B (4 GB RAM) using TensorFlow Lite confirmed operational feasibility, with acceptable latency (< 200 ms/image), suggesting strong potential for offline, edge-based screening in low-connectivity areas. This computational efficiency indicates that the model can process over 200 groundnut kernels per minute under experimental conditions, substantially exceeding manual inspection speeds. While not evaluated in field settings, these characteristics suggest potential for almost real-time screening to support postharvest quality assessment, subject to further empirical validation.

5 Discussion

This study demonstrates that a multiclass deep learning framework based on the Inception-ResNet-V2 architecture can effectively detect aflatoxin-related defects in groundnuts with strong class-wise performance [24]. In contrast to much of the existing literature on plant disease detection, which predominantly adopts binary classification approaches, the present work addresses a more complex and practically relevant multiclass problem by distinguishing among moldy, pest-infested, and physiologically disordered kernels.

Despite this increased task complexity, the proposed model achieves performance metrics comparable to, and in some cases exceeding, those reported in recent CNN-based agricultural studies [15, 16], underscoring the suitability of hybrid residual–Inception architectures for fine-grained kernel-level analysis.

The strong class-wise precision, recall, and AUC values observed across all defect categories confirm the model's ability to capture multi-scale visual features associated with subtle aflatoxin-related manifestations [17, 18]. Unlike prior studies that emphasize overall accuracy alone, this work provides detailed class-wise evaluation, enabling transparent assessment of discriminative performance and robustness, particularly for minority defect classes. Such granularity is essential for postharvest decision-making, where different defect types imply distinct risk levels and management actions.

Several merits of the proposed approach are evident. First, the multiclass formulation enables precise differentiation between visually similar but agronomically distinct defect types, supporting more targeted postharvest interventions. Second, the use of transfer learning substantially reduces training data requirements while maintaining high accuracy, enhancing feasibility in data-scarce, resource-constrained settings [10]. Third, the incorporation of class-weighted loss functions and targeted data augmentation mitigates class imbalance and improves recall for underrepresented categories, a common limitation in real-world agricultural datasets. Additionally, the model demonstrates deployment feasibility on edge hardware, aligning with current research trends toward scalable, field-deployable AI solutions.

Nevertheless, important limitations remain. The model relies on visual features as proxies for aflatoxin risk and does not directly measure chemical aflatoxin concentrations. As reported in prior hyperspectral and spectroscopic studies, chemical-level detection offers greater sensitivity but at significantly higher cost and operational complexity. Accordingly, the proposed model should be viewed as an early screening and decision-support tool rather than a replacement for laboratory-based analysis [22]. Furthermore, image acquisition under controlled lighting conditions may limit immediate generalization to heterogeneous field environments, highlighting the need for broader validation across diverse imaging conditions.

Overall, the findings indicate that multiclass, image-based deep learning provides a promising and practical enhancement to existing aflatoxin screening approaches, balancing diagnostic granularity, accuracy, and deployment feasibility.

5.1 Significance

This study demonstrates the effectiveness of a multiclass deep learning framework based on the Inception-ResNet-V2 architecture for kernel-level groundnut quality assessment, advancing beyond conventional binary screening approaches [23]. By enabling fine-grained discrimination of aflatoxin-related defects, the proposed system supports early postharvest decision-making and scalable deployment in resource-constrained agricultural settings, consistent with recent trends in edge-deployable agricultural AI [17, 18].

5.2 Limitations

The proposed model relies on visual features as proxies for aflatoxin risk and does not directly quantify chemical contamination and therefore should be considered a screening tool rather than a replacement for laboratory-based analysis [14]. Additionally, image acquisition under controlled conditions may limit immediate generalization to heterogeneous field environments.

5.3 Future work

Future work will focus on optimizing the proposed model for real-time deployment on edge and mobile devices and validating its performance through empirical field trials under diverse operational conditions [15, 16]. These trials will assess robustness across variable lighting environments, device types, and user contexts, as well as indicative economic impacts such as reductions in postharvest losses and testing costs. Further extensions will explore integration with complementary sensing methods and digital traceability platforms to support scalable and transparent agricultural quality assurance.

6 Conclusion

In our study, a deep learning framework based on the Inception-ResNet-V2 architecture achieved a classification accuracy of 99.29% for groundnut quality assessment, achieving higher classification accuracy than reported manual visual inspection benchmarks under controlled conditions.

The model's lightweight design enables deployment on low-cost mobile devices, making it particularly accessible and beneficial for smallholder farmers in Uganda. Based on observed classification performance and deployment feasibility, the proposed model may support reductions in crop rejection and improved compliance with aflatoxin safety standards; however, these economic implications remain indicative and require validation through longitudinal field studies.

7 Recommendations

To optimize the advantages of an AI-driven multi-class deep learning model, we propose three essential activities. National agriculture and trade organizations should include AI-driven aflatoxin screening technologies into current quality control methods to minimize crop rejection and improve export competitiveness. Secondly, policies must facilitate the creation and implementation of lightweight, mobile-compatible diagnostic instruments guaranteeing accessibility for smallholder farmers and cooperatives.

Finally, it is essential to cultivate public private partnerships to invest in data infrastructure, training initiatives, and legal frameworks that expedite the integration of AI in agriculture. These initiatives will enhance food safety, elevate farmer earnings, and bolster Uganda's standing in both regional and worldwide groundnut markets.

Acknowledgements
The principal author expresses gratitude to the co-authors for their scholarly contributions to this study.

Author contributions
Lillian Tamale led the study design, model development, and manuscript preparation. Denis Ssebuggwawo supervised and refined the work. Drake Patrick Mirembe supported methodology and validation. Alex Mirugwe managed data and visualization, while Jude T. Lubega provided project oversight and critical review. All authors approved the final manuscript.

Funding
We had no funding support.

Data and code availability
In line with open science principles, the full dataset supporting the findings of this study is publicly available via Zenodo at https://doi.org/10.5281/zenodo.14235238. Additionally, all code and model implementation scripts used in generating the results, figures, and tables are openly accessible on GitHub at https://github.com/Darlen610/Deep-Learning-Model/tree/Ver1.0.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 18 October 2025 / Accepted: 11 February 2026
Published online: 01 March 2026

References

1. Aarif MKO, Alam A, Hotak Y. Smart sensor technologies shaping the future of precision agriculture: recent advances and future outlooks. J Sens. 2025;1:2460098. https://doi.org/10.1155/js/2460098.
2. Abbasi A, Parsons J, Pant G, Liu Sheng OR, Sarker S. Pathways for design research on artificial intelligence. Inf Syst Res. 2024;35(2):441–59. https://doi.org/10.1287/isre.2024.editorial.v35.n2.
3. Akullo JO, Okello DK, Mohammed A, Muyinda R, Amayo R, Magumba D, et al. A comprehensive review of aflatoxin in groundnut and maize products in Africa: prevalence, detection and mitigation strategies. J Food Qual. 2025. https://doi.org/10.1155/jfq/2810946.
4. Balaji B, Satyanarayana Murthy T, Kuchipudi R. A comparative study on plant disease detection and classification using deep learning approaches. Int J Image Graph Signal Process. 2023;15(3):48–59. https://doi.org/10.5815/ijigsp.2023.03.04.

5. Bidyalakshmi T, Jyoti B, Mansuri SM, Srivastava A, Mohapatra D, Kalnar YB, et al. Application of artificial intelligence in food processing: current status and future prospects. Food Eng Rev. 2025;17(1):27–54. https://doi.org/10.1007/s12393-024-09386-2.
6. Deepa BG, Lokesh CK, Umamaheswari D, Ayshwarya B, Yethish PV, Suhaas KP. An enhanced deep learning framework for DeepFake detection using EfficientNet-B3 comparative evaluation of deep and machine learning techniques. Discov Comput. 2026;29(1):18.
7. Devarajan D, Allafi R, Obayya M, Nemri N. AI based real time disease diagnosis in plants using deep learning driven CNNs. Sci Rep. 2026. https://doi.org/10.1038/s41598-025-34681-1.
8. Hoque MJ, Islam MS. An explainable and lightweight CNN framework for robust potato leaf disease classification using Grad-CAM visualization. Appl AI Lett. 2026;7(1):e70011. https://doi.org/10.1002/ail2.70011.
9. Ilyas QM, Ahmad M, Mehmood A. Automated estimation of crop yield using artificial intelligence and remote sensing technologies. Bioengineering. 2023. https://doi.org/10.3390/bioengineering10020125.
10. Kunapuli G. Ensemble methods in machine learning. Simon and Schuster. In: Manning, vol. 1857. LNCS; 2023. https://doi.org/10.1007/3-540-45014-9_1.
11. Leukel J, Zimpel T, Stumpe C. Machine learning technology for early prediction of grain yield at the field scale: a systematic review. Comput Electron Agric. 2023;207(2023):1–32. https://doi.org/10.1016/j.compag.2023.107721.
12. Meneely JP, Kolawole O, Haughey SA, Miller SJ, Krska R, Elliott CT. The challenge of global aflatoxins legislation with a focus on peanuts and peanut products: a systematic review. Exposure Health. 2023;15(2):467–87. https://doi.org/10.1007/s12403-022-00499-9.
13. Mirugwe A, Tamale L, Nyirenda J. Improving tuberculosis detection in chest X-ray images through transfer learning and deep learning: A comparative study of CNN architectures. JMIRx Med. 2024;6:1–15. https://doi.org/10.2196/66029.
14. Mwesige S, Tushabe F, Okoth T, Kasamba I, Areu D. Levels of total aflatoxins in maize and groundnuts across food value chains, gender and agro-ecological zones of Uganda. Int J Life Sci Res Arch. 2023;5(1):090–7. https://doi.org/10.53771/ijlsra.2023.5.1.0081.
15. Panchbhai KG, Lanjewar MG. Enhancement of tea leaf diseases identification using modified SOTA models. Neural Comput Appl. 2025;37(4):2435–53. https://doi.org/10.1007/s00521-024-10758-2.
16. Panchbhai KG, Lanjewar MG. Portable system for cocoa bean quality assessment using multi-output learning and augmentation. Food Control. 2025;174:111234.
17. Panchbhai KG, Lanjewar MG, Malik VV, Charanarur P. Small size CNN (CAS-CNN), and modified MobileNetV2 (CAS-MODMOBNET) to identify cashew nut and fruit diseases. Multimedia Tools Appl. 2024;83(42):89871–91. https://doi.org/10.1007/s11042-024-19042-w.
18. Panchbhai KG, Lanjewar MG, Naik AV. Modified MobileNet with leaky ReLU and LSTM with balancing technique to classify the soil types. Earth Sci Inform. 2024;18(1):77. https://doi.org/10.1007/s12145-024-01521-1.
19. Partnership for Aflatoxin Control in Africa (PACA). Strengthening aflatoxin control in Uganda: Policy recommendations. 2017, pp. 1–8.
20. Rossi M, Henfridsson O, Lyytinen K, Siau K. Design science research. J Database Manag. 2013. https://doi.org/10.4018/jdm.2013070101.
21. Sadimantara MS, Argo BD, Sucipto S, Riza DFA, Hendrawan Y. The classification of aflatoxin contamination level in cocoa beans using fluorescence imaging and deep learning. J Robot Control JRC. 2024;5(1):82–91. https://doi.org/10.18196/jrc.v5i1.19081.
22. Salano EN, Mulwa RM, Obonyo MA. Peanut (*Arachis hypogea*) accessions differentially accumulate aflatoxins upon challenge by *Aspergillus flavus*: implications for aflatoxin mitigation. J Agric Food Res. 2024;15(June 2023):100923. https://doi.org/10.1016/j.jafr.2023.100923.
23. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the AAAI conference on artificial intelligence, vol. 31(1), 2017. http://arxiv.org/abs/1512.00567.
24. Tamale L, Ssebuggwawo D, Mirembe DP, Mirugwe A, Lubega JT. Optimizing deep learning models for aflatoxin detection in agricultural products: a case study of groundnuts. Afr J Rural Dev. 2025;10(2):141–56.
25. Villamar-Torres RO, Factos-Laiño KN, Yánez-Cajo D, Mayorga-Morejon KR, Jazayeri SM. An overview to the new era in efficient crop management: artificial intelligence, machine learning, big data, bioinformatics, metagenomics and precision agriculture. 2025. https://doi.org/10.36899/JAPS.2025.3.0054.
26. Yacouby R, Axman D. Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models. 2020, pp. 79–91. https://doi.org/10.18653/v1/2020.eval4nlp-1.9.
27. Yang Q, Shi L, Han J, Zha Y, Zhu P. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. Field Crops Res. 2019;235(August 2018):142–53. https://doi.org/10.1016/j.fcr.2019.02.022.
28. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS quarterly, 28(1), 75-106. https://doi.org/10.2307/25148625
29. Kaminiaris, M. D., Camardo Leggieri, M., Tsitsigiannis, D. I., & Battilani, P. (2020). AFLA-pistachio: Development of a mechanistic model to predict the aflatoxin contamination of pistachio nuts. Toxins, 12(7), 445. https://doi.org/10.3390/toxins12070445
30. Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings, 3(1), 91-99. https://doi.org/10.1016/j.gltp.2022.04.020