African Journal of Rural Development, Vol. 10 (2): April - June, 2025: pp 141- 156 ISSN 2415-2838

Optimizing Deep Learning Models for Aflatoxin Detection in Agricultural Products: A Case Study of Groundnuts

TAMALE, L.[a,d,*], SSEBUGGWAWO, D.[a], MIREMBE, D.P.[b], ALEX MIRUGWE, A.[c] and LUBEGA, J.T.[d]

[a]School of Computing and Information Science, Kyambogo University, P.O. Box 1, Kampala, Uganda
[b]College of Computing and Information Sciences, Makerere University, P.O. Box 7062, Kampala, Uganda
[c]School of Public Health, Makerere University, P.O. Box 7062, Kampala, Uganda
[d]School of Computing and Informatics, Nkumba University, P.O. Box, 237, Entebbe, Uganda
*Correspondence author: ann.lillian14@gmail.com

ABSTRACT
This study presents an automated deep learning-based classification model for aflatoxin detection in groundnuts, addressing the limitations of conventional manual inspection methods, which are often time-intensive and error-prone. Leveraging the Inception-ResNet-V2 deep learning architecture, the model classifies groundnuts into four distinct categories: healthy, moldy, pest-infested, and those exhibiting physiological disorders. A comprehensive dataset comprising 226 healthy, 236 moldy, 191 pest-infested, and 160 physiological disorder samples was utilized for training, validation, and testing. Model performance was evaluated using multiple metrics, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic (ROC) curve (AUC). The proposed model achieved an overall accuracy of 99.29%, with precision and recall values of 100% and 98.44%, respectively. Notably, the moldy category exhibited an AUC of 1.00, underscoring the model's exceptional capability in distinguishing visual patterns and automating classification tasks. Despite these good results, the study highlights the need for future research to incorporate a broader range of agricultural products to enhance model generalizability. The deep learning model developed improves aflatoxin detection, reducing reliance on subjective manual inspections and enhancing food safety practices. This research offers a novel AI-driven solutions in agricultural quality assessment and food safety management.

Keywords: Aflatoxin, Deep learning, Food safety, Groundnuts, Inception-ResNet-V2

Cite as: Tamale, L., Ssebuggwawo, D., mirembe, D.P., Mirugwe, A. and Lubega, J.T. 2025. Optimizing Deep Learning Models for Aflatoxin Detection in Agricultural Products: A Case Study of Groundnuts. *African Journal of Rural Development* 10 (2):141-156.

RÉSUMÉ

Cette étude propose un modèle automatisé de classification fondé sur le Deep-Learning pour la détection des aflatoxines dans les arachides, afin de passer les limites des inspections manuelles conventionnelles, souvent longues et sujettes à l'erreur. En s'appuyant sur l'architecture Inception-ResNet-V2, le modèle classe les arachides en quatre catégories : saines, moisies, infestées par des ravageurs et présentant des désordres physiologiques. Un ensemble de données conséquent, comprenant 226 échantillons sains, 236 moisies, 191 infestés et 160 présentant des désordres physiologiques, a servi à l'entraînement, à la validation et au test. La performance a été évaluée à l'aide de plusieurs métriques : exactitude, précision, rappel, score F1 et aire sous la courbe ROC (AUC). Le modèle atteint une exactitude globale de 99,29 %, avec une précision de 100 % et un rappel de 98,44 %. Notamment, la catégorie « moisie » présente une AUC de 1,00, soulignant la capacité du modèle à distinguer des motifs visuels et à automatiser la classification. Malgré ces résultats probants, l'étude appelle à des travaux futurs intégrant un spectre plus large de produits agricoles pour améliorer la généralisabilité. Le modèle proposé renforce la détection des aflatoxines, réduit la dépendance aux inspections subjectives et améliore les pratiques de sécurité sanitaire des aliments ; il constitue une solution innovante pilotée par l'IA pour l'évaluation de la qualité et la gestion de la sécurité des denrées.

Mots clés: aflatoxine, apprentissage profond, sécurité sanitaire des aliments, arachides , Inception-ResNet-V2.

INTRODUCTION

Aflatoxins are highly toxic secondary metabolites produced primarily by molds such as Aspergillus flavus and Aspergillus parasiticus. These toxins pose a serious threat to food security and public health globally, particularly in tropical and subtropical regions where high humidity and temperatures foster fungal proliferation. Susceptible crops include maize, sorghum, groundnuts (peanuts), and various cereals and nuts (Schrenk *et al.*, 2020). Recognized for their hepatotoxicity, aflatoxins are classified as Group 1 carcinogens by the International Agency for Research on Cancer (IARC), marking them as significant contributors to liver cancer in both humans and animals.

The global burden of aflatoxin contamination extends beyond public health to economic development and food safety. According to the Food and Agriculture Organization (FAO), approximately 25% of the world's food crops are annually compromised by aflatoxins, leading to substantial economic losses and persistent food insecurity (Eskola *et al.*, 2020). For low- and middle-income countries (LMICs), including Uganda, weak food safety frameworks and limited enforcement mechanisms amplify the crisis (Meneely *et al.*, 2023). These challenges mirror findings from ICT-enabled agricultural initiatives that emphasized the importance of context-specific technological models to support rural communities (Mirembe *et al.*, 2016).

In Uganda, groundnuts (Arachis hypogaea) are both a dietary staple and a critical source of income for smallholder farmers (Okello *et al.*, 2010, 2013). However, they remain highly vulnerable to aflatoxin contamination. Recent studies show that groundnut samples from different Ugandan regions often exceed safe aflatoxin levels (Mwesige *et al.*, 2023).

Poor postharvest practices, inadequate storage, and favorable climatic conditions contribute to this high contamination risk (Pandey *et al*., 2020). Addressing these issues calls for contextually relevant technological models, such as those piloted to enhance farmer knowledge-sharing and risk reduction (Mwesigwa *et al*., 2016).

Given the magnitude of health risks and economic implications, early detection of aflatoxins is vital. Traditional methods such as enzyme-linked immunosorbent assays (ELISA) and thin-layer chromatography, while accurate, are resource-intensive, demanding specialized equipment and trained personnel (Akullo *et al*., 2023). Recent innovations show increasing interest in automated systems that leverage artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), as more accessible, scalable, and cost-effective alternatives (Mohanty *et al*., 2016; Balaji *et al*., 2023) which demonstrate the feasibility of adapting high-tech frameworks in constrained settings.

Deep learning, a subset of machine learning, has achieved significant success in image classification across various fields, including agriculture (Sharma *et al*., 2021). Convolutional neural networks (CNNs), in particular, have shown effectiveness in detecting patterns and anomalies in images, which are often imperceptible to the human eye, achieving high accuracy in detecting fungal infections in crops using image data.

"This study represents a pioneering effort in using deep learning to categorise groundnut quality into distinct classes healthy, mouldy, pest-infested, and physiological disorder aiming to offer a rapid and scalable solution for aflatoxin detection."

By training a CNN on labelled images of these categories, we aimed to develop a reliable and scalable model capable of detecting aflatoxins and other defects in groundnuts (Sadimantara *et al*., 2024). This approach provides a faster and more accessible alternative to traditional methods, with the potential for implementation in resource-constrained settings, where rapid and accurate aflatoxin detection is vital to ensuring food security and public health.

MATERIALS and METHODS

Groundnuts were selected as the primary crop for this study due to their nutritional importance, economic relevance, and high susceptibility to aflatoxin contamination, particularly in Uganda. As a staple legume grown extensively by smallholder farmers, groundnuts are frequently subjected to poor post-harvest handling and inadequate storage conditions, which facilitate mold growth and aflatoxin production (Okello *et al*., 2015; Akullo *et al*., 2023). According to Mirembe *et al.* (2016), the high risk of aflatoxin exposure associated with groundnuts, coupled with the crop's substantial economic importance, positions groundnuts as an ideal candidate for AI-based quality detection interventions particularly within low-resource agricultural settings.

To address the image classification task, this study employed the Inception-ResNet-V2 deep learning architecture, which synergistically combines inception modules designed to capture multiscale features with residual connections that ensure stable training across very deep networks (Szegedy *et al*., 2017). This hybrid architecture provides enhanced learning capacity for distinguishing subtle visual features, making it particularly appropriate for detecting nuanced differences in agricultural products (Kumar *et al*., 2021; Pai *et al*., 2024). The model's application in postharvest analysis is consistent with the growing use of intelligent systems in rural development and food safety surveillance (Mirembe *et al*., 2016). This trend has been highlighted in recent studies, emphasizing the relevance of AI-driven approaches in enhancing agricultural outcomes in low-resource settings.

Groundnut images were categorized into four quality classes based on expert agronomic annotation and recognized postharvest inspection guidelines: healthy, moldy, pest-infested, and physiological disorders (Akullo *et al*., 2023). Healthy kernels showed no defects; moldy ones exhibited fungal growth; pest-infested samples presented insect-related damage, while physiological disorders reflected abiotic issues such as shriveling or internal discoloration. These classification categories align with conventional postharvest quality control standards and provide a robust basis for automated detection systems.

The proposed model is intended for deployment at the post-harvest stage, specifically during sorting and grading processes at collection centers, cooperatives, and quality inspection units. Integrating AI at this stage can significantly improve inspection accuracy and processing speed, replacing subjective manual evaluations with objective, scalable, and technology-enabled assessments (Mirembe *et al*., 2019). This aligns with broader efforts to digitize agricultural services and reduce postharvest losses through ICT innovations.

Implementation Framework Figure 1 shows the overall workflow implemented in this study, starting with the acquisition of groundnut images and ending with evaluation of the Inception ResNet V2 model. Each step of the process of image collection, pre-processing, augmentation, model training, and evaluation is explained in detail in the following subsections. This schematic highlights the methodological approach used to address the challenges of aflatoxin detection in groundnuts using the Inception ResNet V2 model to classify groundnut quality into four categories: healthy, moldy, pest-infested, and physiological disorders. The progression through these stages demonstrates the thoroughness of the methodology used, thereby ensuring that the developed model is accurate and reliable for practical use.

Data Collection. This study utilized a comprehensive data collection method to obtain high-resolution images of groundnuts, categorizing them into four primary categories: healthy, moldy, test-infested, and physiological disorders. These categories represent critical factors associated with aflatoxin contamination and other indicators of groundwater quality (Figure 2). A digital camera with a fixed configuration, incorporating standardized lighting and regulated environmental conditions, was employed to guarantee uniformity in all image capture. Each groundnut sample was isolated and photographed repeatedly to document all potential visual attributes, thereby reducing the interference from external factors. The dataset was partitioned using a stratified sampling approach to ensure a balanced representation of all four classes across training, validation, and test sets. This method was selected to minimize sampling bias and ensure the model performance could be reliably assessed across all categories. The final distribution in the different sets was: Training Set: Healthy (226 samples), moldy (236 samples), Pest-infested (191 samples), Physiological Disorders (160 samples).

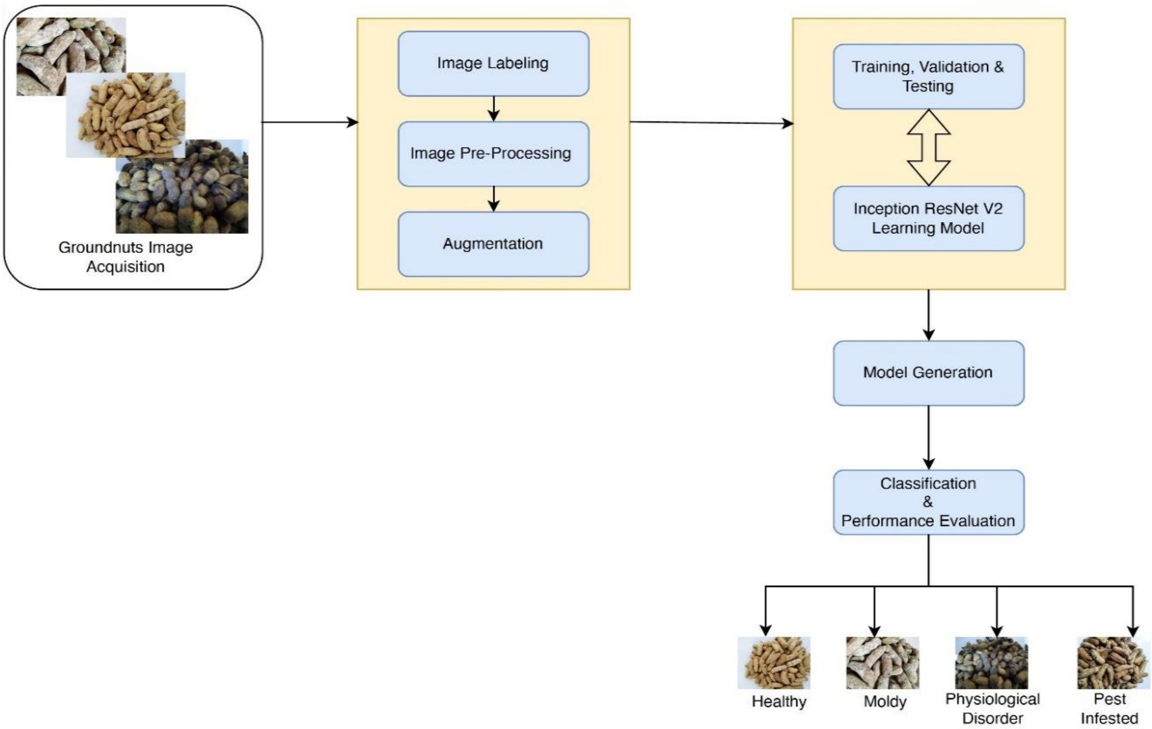

Figure 1. Schematic implementation flowchart

Validation Set: Healthy (50 samples), Moldy (62 samples), Pest-Infested (46 samples), Physiological Disorders (46 samples); Testing Set: Healthy (88 samples), Moldy (65 samples), Pest-Infested (54 samples), Physiological Disorders (48 samples). The class balance was preserved to guarantee equitable representation among categories by employing data augmentation techniques to enhance the dataset diversity, especially for minority categories.

Data Preprocessing. Data preprocessing is a critical step in ensuring consistency and enhancing the model's ability to generalize under diverse conditions. Each groundnut image was resized to a standardized dimension of 256 × 256 pixels to align it with the requirements of the Inception-ResNet-V2 architecture. To further improve the robustness of the model and expand the dataset, a series of data augmentation techniques were applied. These included random rotations within a range of 0 °to 60 °, width and height shifts of up to 20%, zooming of up to 20%, and horizontal and vertical flipping, with each image having a 50% probability of being flipped in either direction or remaining unchanged. These transformations introduced controlled variability and diversity into the dataset. They were designed to reflect conditions commonly seen during ground inspection, such as changes in orientation, scale, and minor occlusions. This helped the model focus on consistent visual features that do not depend on exact positioning or lighting. Simulating real-world conditions and improving the model's ability to perform effectively in practical scenarios. These transformations simulate real-world variability in groundnut images, enabling the model to learn invariant features and improve its performance on unseen data. Mathematically, the image transformations can be expressed as

$$(\boldsymbol{I}'(\boldsymbol{x}', \boldsymbol{y}' = \boldsymbol{I}(\boldsymbol{x} \cdot \boldsymbol{cos}\,\boldsymbol{\theta} - \boldsymbol{y} \cdot \boldsymbol{sin}\,\boldsymbol{\theta}, \boldsymbol{x} \cdot \boldsymbol{sin}\,\boldsymbol{\theta} + \boldsymbol{y} \cdot \boldsymbol{cos}\,\boldsymbol{\theta}) \qquad (\boldsymbol{1})$$

Where I (x', y) represents the original image, I ′ (x,' y) represents the augmented image, and 0 is the rotation angle.

Figure 2. Sample classified images

This transformation ensures that the model is exposed to images that vary in orientation, scale, and position, thereby minimizing the risk of overfitting and enhancing its generalization capability. Different augmentation percentages were applied to each category to address the class imbalances in the dataset. Healthy groundnut images were augmented by 5%, whereas moldy and physiological disorder images were augmented by 30% each. Pest-infested images were augmented by 20%. These percentages were carefully selected to increase the representation of the underrepresented categories and ensure sufficient variability in more complex categories, such as moldy and pest-infested groundnuts. The higher augmentation rates for moldy and physiological disorder images reflect the need for increased representation of these categories in the dataset owing to their visual complexity and lower sample counts. This augmentation strategy exposed the model to a wide range of transformations, thereby improving its ability to classify groundnut images into four categories: healthy, moldy, test-infested, and physiological disorders. By incorporating these augmented images into the training set, the model learned to generalize more effectively, leading to an improved classification performance across all categories.

Model Architecture. The Inception-ResNet-V2 architecture was chosen for this study because of its superior performance in outperforming other state-of-the-art architectures for complex image classification tasks, particularly in agricultural applications. This hybrid design merges the feature extraction power of inception networks with the stability and efficient gradient propagation of residual connections, thereby offering a unique advantage for handling large, complex datasets. These capabilities are especially beneficial for tasks in which multiscale feature extraction and maintenance of gradient flow are critical to model performance (Szegedy *et al.*, 2017). In agricultural contexts, where datasets often include diverse images of plant species, disease symptoms, or crop conditions, Inception-ResNet-V2 has been proven to be highly effective. For example, in a study focused on plant disease detection from leaf images, Inception-ResNet-V2 outperformed architectures, such as ResNet-50 and VGG-16, in terms of classification accuracy and computational efficiency (Kumar *et al*., 2021). This makes it an ideal architecture for tasks that involve handling a wide variety of conditions, such as the classification of groundnut images into categories of healthy, moldy, test-infested, and physiological disorders.

Additional studies in the agricultural domain reinforce the capabilities of the model (Yang *et al.*, 2019***;*** Pai *et al.***,** 2024) For instance, Inception-ResNet-V2 achieved a higher accuracy than Inception-V3 and DenseNet in a weed identification task, where the ability to differentiate subtle morphological differences in plants is crucial (Pai *et al.*, 2024). Similarly, research on rice crop classification using UAV imagery has highlighted the superior performance of Inception-ResNet-V2 over MobileNet and EfficientNet, showing its higher accuracy and resilience to environmental variability (Yang *et al.*, 2019). This emphasizes the potential of the architecture for applications requiring robust image analysis in dynamic agricultural settings.

Transfer Learning. The transfer learning capabilities of Inception-ResNet-V2 were another critical reason for its selection in this study. Agricultural datasets, particularly those with annotated labels, are often limited in size. By leveraging pre-trained weights from large-scale datasets, such as ImageNet, the model can utilize general features learned from millions of images and adapt them to specific tasks, such as groundnut classification. This transfer-learning approach has been shown to improve the classification performance while significantly reducing the time required for training (Kumar *et al.*, 2021; Okaron *et al.*, 2024). In this study, we fine-tuned the Inception-ResNet-V2 model by freezing the initial layers and unfreezing the last ten layers, allowing the model to retain general features while learning domain-specific characteristics relevant to groundnut classification. This method enables the model to focus on fine-grained details, such as mold patches or pest damage, without losing the ability to generalize across different groundnut categories.

Model Design and Training. The Inception-ResNet-V2 architecture begins with a stem block that performs initial convolutional and pooling operations to reduce the input image dimensions. Following this, Inception-ResNet blocks are used, where parallel convolutional filters of varying sizes (1 × 1, 3 × 3, 5 × 5) extract multi-scale features, and residual connections improve gradient flow and stabilize training. The residual connections are mathematically represented as:

$$\boldsymbol{y} = \boldsymbol{F}(\boldsymbol{x}, \{\boldsymbol{Wi}\}) + \boldsymbol{x} \tag{2}$$

allows the model to train efficiently, even with very deep networks, by preventing the vanishing gradient problem. Reduction blocks are placed at strategic points to decrease the spatial resolution and increase the feature depth, thereby enabling the network to handle complex patterns more efficiently. The model includes a fully connected layer and a softmax activation function that generates the final classification probabilities (Szegedy *et al.*, 2017). We extend the model by adding custom dense layers to enhance the learning capacity of the network. The first dense layer consisted of 256 neurons with ReLU activation, followed by batch normalization and a dropout rate of 0.4 for regularization. s dense layer with 256 neurons and a dropout rate of 0.35 was added, followed by a final dense layer with 512 neurons and a dropout rate of 0.3. The model's output layer, a softmax function, comprises four neurons to output the class probabilities for four distinct categories: healthy, moldy, test-infested, and physiological disorders.

Hyperparameter Optimization. The training of the Inception-ResNet-V2 model was optimized using the Adam optimizer with an initial learning rate of 0.0001, chosen for its adaptive properties and ability to handle sparse gradients in high-dimensional datasets effectively (Szegedy *et al.*, 2017). The categorical cross-entropy loss function was used to minimize the difference between the predicted probabilities ˆy and true labels y, as defined by

$$\boldsymbol{L}(\boldsymbol{y}, \boldsymbol{y}\hat{}) = -\sum_{i=1}^{N} \boldsymbol{yi}\, \boldsymbol{log}(\boldsymbol{y}\hat{}) \tag{3}$$

This loss function is well suited for the multiclass classification task addressed in this study, as it handles multiple output classes and ensures a smooth optimization process. Hyperparameter optimization was performed using a grid search across multiple parameter values to fine-tune the model. Specifically, we searched for learning rates α in the range of

$$\boldsymbol{\alpha} \in \mathbf{10-4}\,, \mathbf{10-5}\,, \mathbf{10-6} \tag{4}$$

Batch sizes of 8, 16, and 32, as well as dropout rates ranging from 0.2 to 0.5, were tested. This systematic search allowed us to identify the optimal combination of hyperparameters, with the best results obtained using a learning rate of α = 0.0001, batch size of 16, and dropout rate of 0.3. These values were selected based on the performance of the validation set, ensuring an effective balance between convergence speed and generalization. To address the class imbalance inherent in the dataset, class weights were computed and applied during training. This approach helped prevent the model from becoming biased towards the more frequent classes, such as Healthy, while maintaining accurate classification for minority classes, such as Pest-Infested and Physiological Disorders.

Additionally, a learning rate scheduler, specifically the ReduceLROnPlateau callback, was employed to dynamically adjust the learning rate when the validation loss plateaued, thereby improving model generalization and preventing overfitting. The model was trained over 50 epochs with continuous monitoring of performance on a separate validation set to ensure stability and prevent overfitting. This hyperparameter-tuning process allowed the model to achieve optimal performance in groundnut classification.

RESULTS

Evaluation Metrics. To evaluate the performance of the model, we used accuracy, precision, recall, F1-score, and area under the curve (AUC) metrics (Yacouby and Axman, 2020). These metrics provide a comprehensive assessment of the model's ability to classify groundnut images across four categories. Precision measures the proportion of true positive predictions relative to the total predicted positives and is defined as

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FP+FN}} \dots \quad \dots\dots(\mathbf{5})$$

where TP, TN, FN, and FP refer to true positives, true negatives, false negatives, and false positives, respectively. Accuracy is the most commonly used metric to evaluate the performance of binary classifiers. This is defined as the ratio of correct predictions to the total number of predictions made by the model.

$$\boldsymbol{Precisio} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \dots\dots\dots\dots\dots\dots\dots..\dots\dots\dots(\mathbf{6})$$

where T P refers to true positives and F P refers to false positives. Recall is defined as the proportion of true positive predictions out of the actual positives.

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \dots\dots\dots\dots\dots\dots..\dots\dots(\mathbf{7})$$

where N refers to the false negatives. The F1-score, the harmonic mean of precision and recall, is calculated as follows:

$$\boldsymbol{F1-Score} = \frac{\boldsymbol{Precision\ x\ Recall}}{\boldsymbol{Precision+Recall}} \dots\dots.\dots(\mathbf{8})$$

In addition, the Area under the Curve (AUC) was computed based on the Receiver Operating Characteristic (ROC) curve, which provided 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.

Computational Setup and Resources. The computational setup for this study involved the use of Keras 3.3.3 and TensorFlow 2.16.1 frameworks, implemented on an MSI GL75 Leopard 10SFR laptop equipped with 32 GB of RAM and an 8 GB NVIDIA GeForce RTX 2070 GDDR6 GPU. To enhance deep learning model training efficiency, the system utilized NVIDIA CUDA Toolkit 12.1 and cuDNN SDK 8.7.0 for GPU acceleration. Python 3.10.12 served as the programming environment, supplemented by libraries such as NumPy, Pandas, and Matplotlib for data preprocessing and visualization. A Canon EOS 90D DSLR camera was used for image acquisition due to its high resolution (32.5 MP), APS-C sensor, and adjustable EF-S lens, which provided superior image quality compared to equipment used in similar agricultural studies (Bernacki and Scherer, 2023). This combination of computational and imaging resources enabled effective processing and training of the Inception-ResNet-V2 model, particularly when applying data augmentation techniques for improved aflatoxin detection in groundnuts.

The performance of the Inception-ResNet-V2 model was evaluated using several key metrics, demonstrating its effectiveness in classifying groundnut images into four categories: healthy, moldy, test-infested, and physiologically disordered.

The confusion matrix shown in Figure 3 provides a comprehensive view of the classification ability of the model.

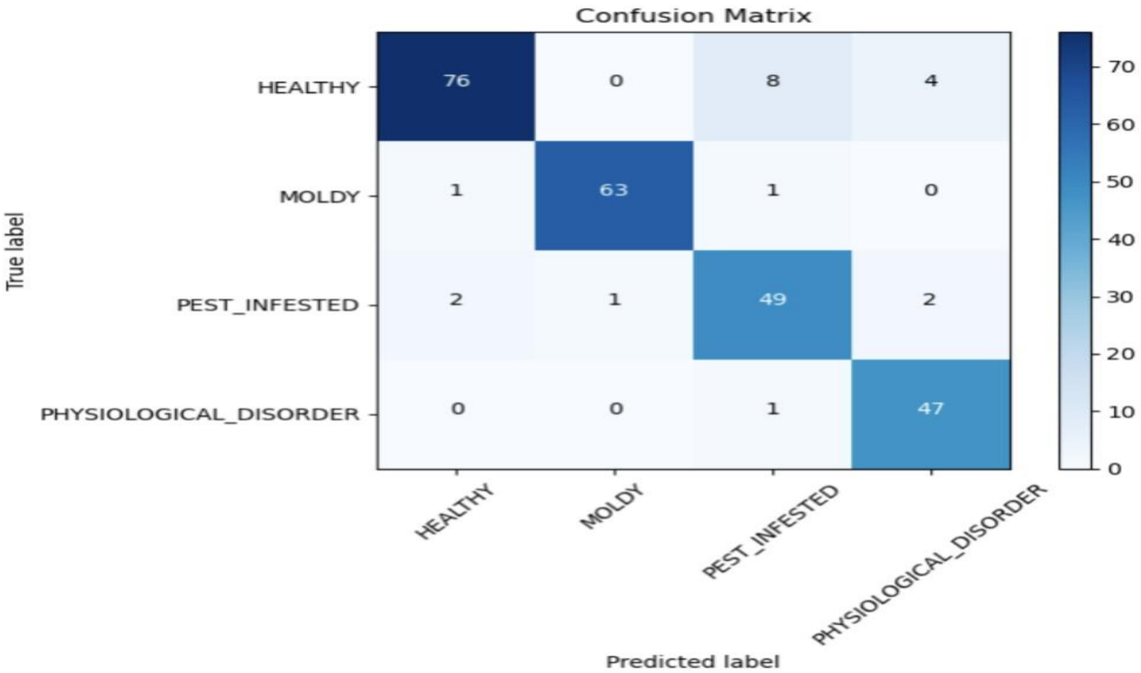

Figure 3. Confusion matrix showing model classification results for the four groundnut categories.

For the Healthy category, the model correctly classified 76 out of 88 samples, with minor misclassifications: eight samples were mistakenly labelled as Pest-Infested, and four as Physiological Disorder. The highest performance was observed in the Moldy category, where the model correctly identified 63 out of 65 samples with only two misclassifications. Similarly, the Pest-Infested category showed strong performance, with 49 out of 54 samples correctly classified, although two samples were misclassified as healthy and two as physiological disorders. In the Physiological Disorder category, the model demonstrated accuracy, misclassifying only one sample as Pest-Infested while correctly identifying 47 samples. These results indicate that despite some minor classification errors, the model is highly effective across all categories.

The model achieved high performance across all metrics, as shown in Figure 4, with an accuracy of 99.29%, precision of 100%, recall of 98.44%, and F1-score of 99.21% on the test dataset. These results highlight the ability of the model to generalize well to unseen data, confirming that it does not overfit the training data. To further assess this, we calculated the bias as the difference between the mean validation accuracy and test accuracy, which was 0.84%. The performance deviation, measured as the standard deviation of the test accuracy across five stratified folds, was 0.71%, indicating stable performance across the different data splits.

The precision, recall, and learning rate curves shown in Figure 5 illustrate the model training process over 50 epochs. Precision improved rapidly in the early epochs, reaching nearly 100% on the training set and stabilizing at around 90% on the validation set. To further test the model's validity, we performed a permutation test using 1000 label shuffles. The resulting p-value was less than 0.01, confirming that the model's performance is statistically significant and unlikely to have occurred by chance. These findings, together with the alignment of training and validation curves, provide strong evidence that the model generalizes effectively and is not overfitted.

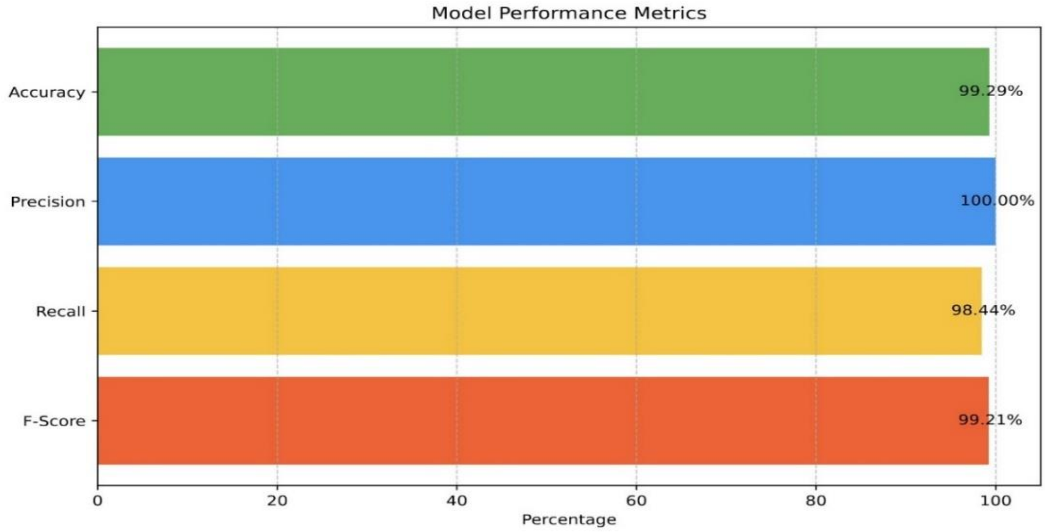

Figure 4. Overall model performance metrics showing accuracy, precision, recall, and F1-score

This indicates that the model was highly accurate in correctly identifying true-positive instances while minimizing false-positives. Similarly, recall followed a steep upward trajectory in the initial epochs, with both training and validation recall exceeding 98% by epoch 30, demonstrating that the model successfully identified most of the relevant instances (true positives).

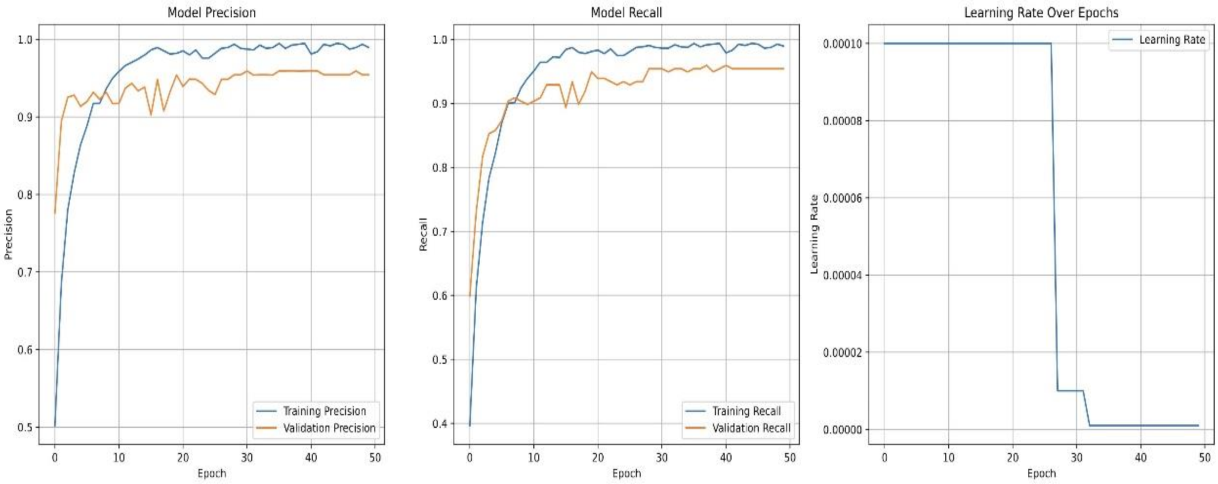

Figure 5. Precision, recall, and learning rate trends across 50 training epochs for both training and validation sets

The learning rate curve reflects the fine-tuning of the model during training, with significant reductions in the learning rate occurring around epoch 20, and again around epoch 30. These reductions helped the model stabilize its performance and prevent overfitting, as evidenced by the steady precision and recall values in later epochs. The final learning rate drop coincided with the plateau in both precision and recall, indicating that the model had reached optimal convergence at this point. The model loss and accuracy curves shown in Figure 6 provide a detailed view of the training and validation performance over 50 epochs. The training loss decreased consistently, showing a smooth downward trend, indicating that the model effectively learned and minimized errors during training. The validation loss followed a similar trend, although it began to stabilize around epoch 20, with minimal fluctuations thereafter. This suggests that the model's performance on the validation set is stable, and no signs of overfitting are evident, as the validation loss remains relatively low and consistent. The accuracy curves further supported these findings. The training accuracy improved rapidly during the first 20 epochs, reached over 90%, and continued to climb, leveling off at approximately 98% by the end of the training process. The validation accuracy closely followed the training accuracy during the early epochs, reaching 90% by epoch 10 and stabilizing between 90% and 93% for the remainder of the training process. This alignment between the training and validation accuracies indicates that the model generalizes well and is capable of maintaining high performance on unseen data.

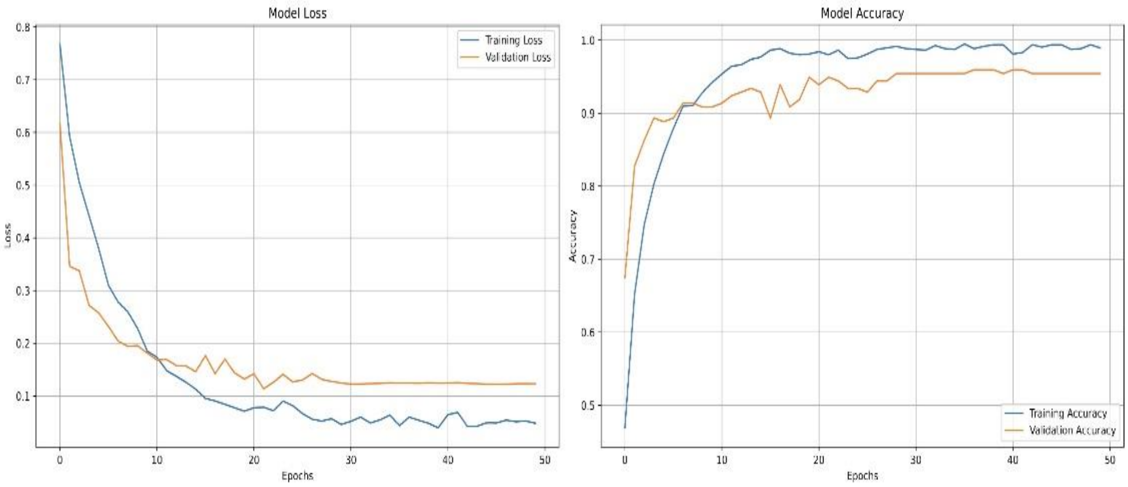

Figure 6. Training and validation loss and accuracy over 50 epochs

The ROC curves shown in Figure 7 provide additional confirmation of the model's strong discriminative power. The AUC values for each class were high, and 0.98) for Healthy, 1.00 for Moldy, 0.97 for Pest-Infested, and 0.99 for Physiological

Disorders, respectively. These high AUC values demonstrate the ability of the model to distinguish between different classes, even at varying decision thresholds. The AUC of 1.00 for the Moldy category indicates perfect classification performance for this class, while the slightly lower AUC for PestInfested suggests that the model faced more challenges distinguishing between pest damage and other conditions, a pattern consistent with the confusion matrix findings.

The qualitative results, shown in Figure 8, offer insight into the model's decision-making by showing incorrectly classified samples. Misclassifications occurred mainly between PestInfested and Physiological Disorder samples, with several Pest-Infested samples being labelled as healthy or physiological disorders. This suggests that the model struggles to distinguish subtle visual differences between these categories, which is likely due to overlapping features.

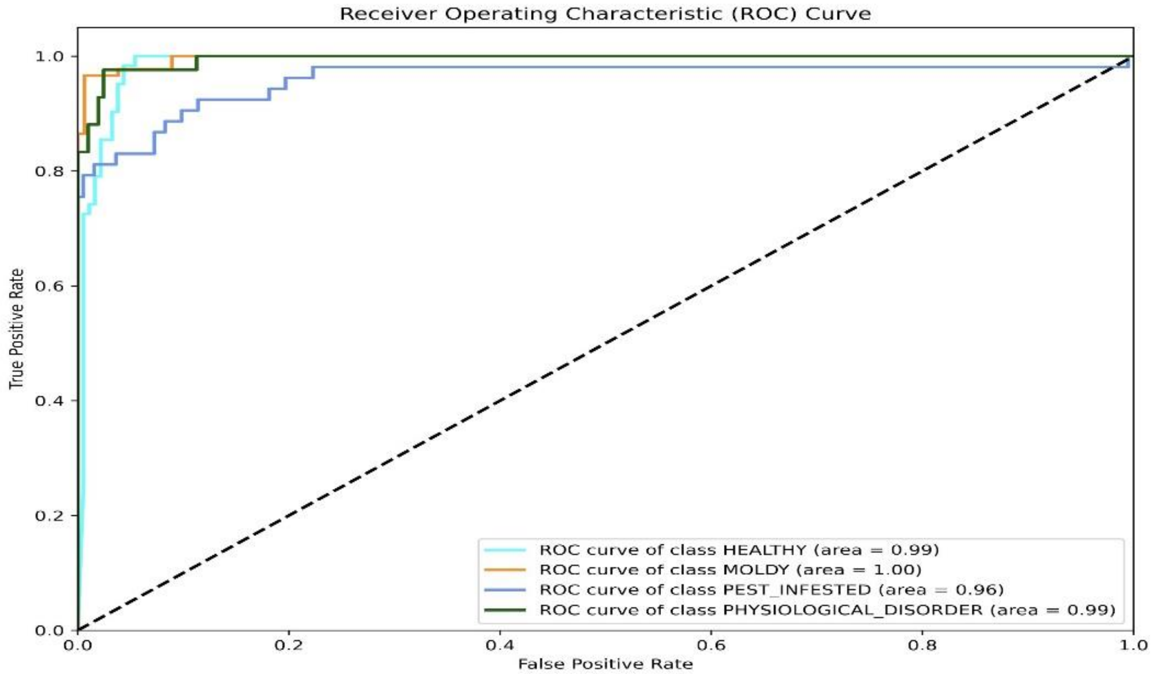

Figure 7. ROC curves with AUC values for the classification of Healthy, Moldy, Pest-Infested, and Physiological Disorder groundnut categories

DISCUSSION

The results of this study demonstrated the efficacy of using the Inception-ResNet-V2 deep learning model for the classification of groundnuts into four categories: healthy, moldy, test-infested, and physiological disorders. The model achieved excellent performance across multiple metrics, including accuracy, precision, recall, and F1-score, with an overall accuracy of 99.29%. These findings highlight the potential of advanced deep learning architectures to automate quality assessment in agricultural products such as groundnuts, which are traditionally labor-intensive processes.

One of the most notable outcomes was the high performance achieved in the Moldy category, where the model reached an AUC of 1.00. This likely reflects the distinct visual patterns associated with mold contamination, such as discoloration and fungal textures, which may have made the class easier to separate during training and testing. This result is consistent with previous studies that showed high accuracy in detecting mold and other diseases using deep learning techniques (Kumar *et al.*, 2021; Pai *et al.*, 2024*)*. The distinct visual features of the mold likely make it easier for the model to generalize, especially when trained on sufficient samples. However, we acknowledge that such a perfect score may not generalize to all real-world settings and should be interpreted with caution. Further testing on more diverse and independent datasets would help validate the model's reliability for this class. However, some misclassifications were observed between the Healthy and Physiological Disorder categories as well as between the Pest-Infested and Healthy samples. The confusion matrix revealed that some healthy samples were classified as either Pest-Infested or Physiological Disorders. This misclassification can be attributed to subtle visual differences between these categories, where Healthy may appear as pest damage or physiological abnormalities. Other studies on agricultural image classification have also noted that visually similar conditions can lead to confusion in model predictions, particularly in the presence of overlapping features (Kamilaris and Prenafeta-Boldú, 2018).

It is important to note that these misclassifications were relatively few and occurred primarily in borderline cases. The model maintained high performance across all metrics, suggesting that its predictions

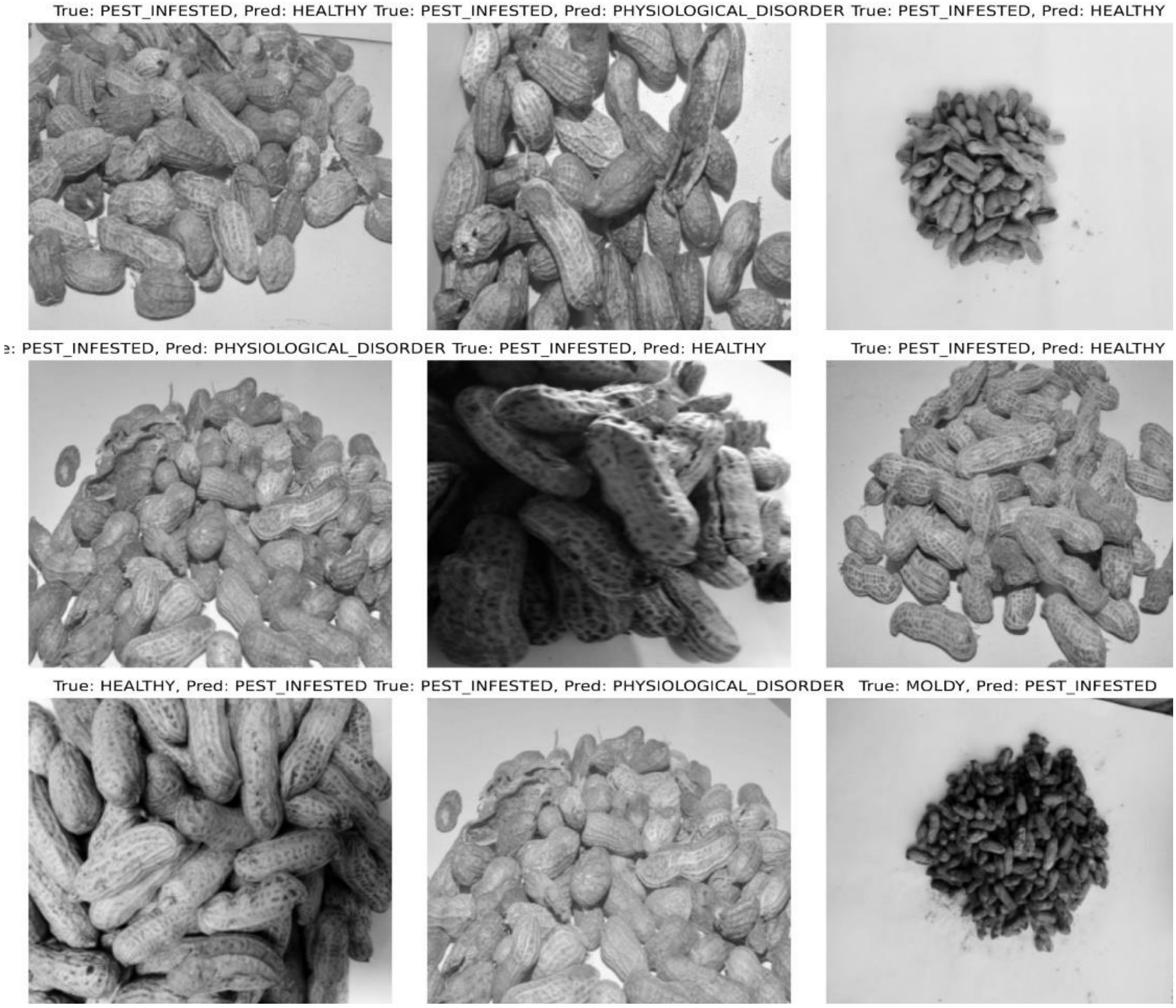

Figure 8. Examples of misclassified groundnut images with true and predicted labels

The overall high precision (100%) indicates that the model has effectively minimized false positives, meaning that when it predicts a sample as moldy or test-infested, it is highly likely to be correct. This is an important factor in agricultural quality control, where incorrectly identifying a healthy product as contaminated can lead to unnecessary food waste and economic losses. Similarly, the high recall (98.44%) demonstrates that the model successfully identified most of the contaminated samples, which is a critical requirement for food safety applications where missed detections could result in significant health risks. The ROC curves and high AUC values further validated the model’s ability to discriminate between different groundnut categories. An AUC value approaching 1.00 for most categories indicates that the model is highly effective across a range of decision thresholds, with particularly strong performance in distinguishing Moldy groundnuts from other categories.

CONCLUSION

This study is the first to classify groundnuts into four distinct categories: using deep learning techniques healthy, moldy, test-infested, and physiological disorders. By applying the Inception-ResNet-V2 model, we address the critical challenge of automating groundnut quality assessment, a task traditionally performed manually, which is both labor-intensive and prone to errors. The model demonstrates the potential of AI to transform agricultural quality control, offering an objective, scalable, and efficient alternative to manual inspection.

A key achievement of this study is the model's exceptional performance, achieving AUC scores of 0.99 for Healthy, 1.00 for Moldy, 0.96 for Pest-Infested, and 0.99 for Physiological Disorder. Despite this overall success, the model struggled to distinguish healthy groundnuts from pests and physiological disorders, with some misclassifications. This indicates that although the model can accurately classify the majority of samples, subtle visual similarities between these categories can still pose challenges, as is often the case in manual inspections.

Overall, this study paves the way for more advanced AI-driven systems in agriculture, in which precision and consistency are critical for ensuring food safety and quality. Future work should focus on refining the ability of the model to handle complex differences between visually similar categories, further enhancing its robustness and reliability. Nonetheless, these promising results highlight the potential of integrating AI technologies into agricultural systems, significantly improving food safety, reducing manual labor, and ensuring more reliable quality control across the supply chain.

ACKNOWLEDGEMENT

This research received funding from The Regional Universities Forum for Capacity Building in Agriculture (RUFORUM) through a Carnegie Corporation of New York grant.

STATEMENT OF CONFLICT OF INTEREST

The Authors declare no conflict of interest in the paper.

REFERENCES

Akullo, J. O., Amayo, R., Okello, D. K., Mohammed, A., Muyinda, R., Magumba, D., Gidoi, R. and Mweetwa, A. M. 2023. Aflatoxin contamination in groundnut and maize food products in Eastern and Northern Uganda. *Cogent Food and Agriculture* 9 (1) : 1–13. https://doi.org/10.1080/23311932.2023.2221015

Balaji, B., Satyanarayana Murthy, T. and Kuchipudi, R. 2023. A Comparative Study on Plant Disease Detection and Classification Using Deep Learning Approaches. *International Journal of Image, Graphics and Signal Processing* 15 (3) :48–59. https://doi.org/10.5815/ijigsp.2023.03.04

Bernacki, J. and Scherer, R. 2023. IMAGINE Dataset: Digital Camera Identification Image BenchmarkinDataset. Proceedings of the International Conference on Security and Cryptography 1(Secrypt) 799–804 pp. https://doi.org/10.5220/0012130300003555

Eskola, M., Kos, G., Elliott, C. T., Hajšlová, J., Mayar, S. and Krska, R. 2020. Worldwide contamination of food-crops with mycotoxins: Validity of the widely cited 'FAO estimate' of 25%. *Critical Reviews in Food Science and Nutrition* 60 (16): 2773–2789. https://doi.org/10.1080/10408398.2019.1658570

Kamilaris, A. and Prenafeta-Boldú, F. X. 2018. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture* 147: 70–90. https://doi.org/10.1016/j.compag.2018.02.016

Kumar, A., Pathak, H., Bhadauria, S. and Sudan, J. 2021. Aflatoxin contamination in food crops: causes, detection, and management: a review. *Food Production, Processing and Nutrition* 3 (1). https://doi.org/10.1186/s43014-021-00064-y

Kumar, R. S., Babu, V. S. and Sunder, R. 2021.

A comparative study on disease detection of plants using machine learning techniques. In: 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). *IEEE* 1: 1937–1941. https://doi.org/10.1109/AIIoT58432.2024.10574617

Meneely, J. P., Kolawole, O., Haughey, S. A., Miller, S. J., Krska, R. and Elliott, C. T. 2023. The Challenge of Global Aflatoxins Legislation with a Focus on Peanuts and Peanut Products: A Systematic Review. *Exposure and Health* 15 (2): 467–487. https://doi.org/10.1007/s12403-022-00499-9

Mirembe, D. P., Lubega, J. T. and Kibukamusoke, M. 2019. Using ICTs to enhance duty bearer accountability and transparency to citizens in Eastern and Northern Uganda. *International Journal of Information Technology, Communications and Convergence* 3 (3): 228. https://doi.org/10.1504/ijitcc.2019.106559

Mirembe, D. P., Obaa, B. B. and Ebanyat, P. 2016. Developing and piloting a multi-channel ICT-Enabled Model to enhance University engagement with smallholder farming communities in Uganda. *African Journal of Rural Development* 1 (1): 13–21

Mohanty, S. P., Hughes, D. P. and Salathé, M. 2016. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science* 7: 1–10. https://doi.org/10.3389/fpls.2016.01419

Mwesige, S., Tushabe, F., Okoth, T., Kasamba, I. and Areu, D. 2023. Levels of total aflatoxins in maize and groundnuts across food value chains, gender and Agro-ecological zones of Uganda. *International Journal of Life Science Research Archive* 5 (1): 090–097. https://doi.org/10.53771/ijlsra.2023.5.1.0081

Mwesigwa, E., Tulinayo, F. P. and Mirembe, D. P. 2016. Enhancing agricultural knowledge sharing among smallholder farmers in Uganda: An evaluation of mobile and web technologies. *RUFORUM Working Document Series* 14 (2) : 247–251. http://repository.ruforum.org

Okaron, V., Mwololo, J., Gimode, D. M., Okello, D. K., Avosa, M., Clevenger, J., Korani, W., Ssemakula, M. O., Odong, T. L. and Odeny, D. A. 2024. Using cross-country datasets for association mapping in Arachis hypogaea L. *Plant Genome, February* 1–18pp. https://doi.org/10.1002/tpg2.20515

Okello, D. K., Biruma, M. and Deom, C. M. 2010. Overview of groundnuts research in Uganda: Past, present and future. *African Journal of Biotechnology* 9 (39): 6448–6459.

Okello, D. K., Monyo, E., Deom, C. M., Ininda, J. and Oloka, H. K. 2013. *Groundnut Production Guide for Uganda :* GROUNDNUT PRODUCTION GUIDE FOR UGANDA *: Recommended practices for farmers*. National Agricultural Research Organisation, Entebbe, Uganda.

Okello, D., Okori, P., Puppala, N., Bravo-Ureta, B., Deom, C. M., Ininda, J., Anguria, P., Biruma, M. and Asekenye, C. 2015. Groundnut seed production manual for Uganda. National Agricultural Research Organisation,Entebbe, Uganda https://www.researchgate.net/publication/277558622_Groundnut_Seed_Production_Manual_for_Uganda_2015

Pai, D. G., Kamath, R. and Balachandra, M. 2024. Deep Learning Techniques for Weed Detection in Agricultural Environments : A Comprehensive Review. *IEEE Access*, 1pp. https://doi.org/10.1109/ACCESS.2024.3418454

Pandey, P., Pandey, N. S. and Chaturvedi, R. 2020. Bio-management of Postharvest Diseases and Mycotoxigenic Fungi. pp. 223–234. In: Prevention and Control Strategies of Aflatoxin Contamination (Issue 1st Edition,. CRC Press. file:///E:/Documents/dosen/buku Metodologi/[John_W._Creswell]_Research_Design_Qualitative,_Q(Bookos.org).pdf

Sadimantara, M. S., Argo, B. D., Sucipto, S., Riza, D. F. Al, and Hendrawan, Y. 2024. The Classification of Aflatoxin Contamination Level in Cocoa Beans using Fluorescence Imaging and Deep learning. *Journal of Robotics and Control (JRC)* 5 (1): 82–91. https://doi.org/10.18196/jrc.v5i1.19081

Schrenk, D., Bignami, M., Bodin, L., Chipman, J. K., del Mazo, J., Grasl-Kraupp, B., Hogstrand, C., Hoogenboom, L., Leblanc, J. C., Nebbia, C. S., Nielsen, E., Ntzani, E., Petersen, A., Sand, S., Schwerdtle, T., Vleminckx, C., Marko, D., Oswald, I. P., Piersma, A. and Wallace, H. 2020. Risk assessment of aflatoxins in food. *EFSA Journal* 18 (3). https://doi.org/10.2903/j.efsa.2020.6040

Sharma, A., Jain, A., Gupta, P. and Chowdary, V. 2021. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access* 9 : 4843–4873. https://doi.org/10.1109/ACCESS.2020.3048415

Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A. 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 31 (1). http://arxiv.org/abs/1512.00567

Yacouby, R. and Axman, D. 2020. Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models. 79–91. https://doi.org/10.18653/v1/2020.eval4nlp-1.9

Yang, Q., Shi, L., Han, J., Zha, Y. and Zhu, P. 2019. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Research* 235*:* 142–153. https://doi.org/10.1016/j.fcr.2019.02.022