Development of a Machine Learning Digital Image Models of Groundnut Pods for Intelligent Threshing Machine Application Using Convolution Neural Network

B. Bello [1*]., J.D. Jiya[1], A.Tokan[2], A. Mohammed[1]
[1]Department of Electrical and Electronics Engineering, Abubakar Tafawa Balewa Unicersity Bauchi, 740272, Nigeria.
[2]Department of Machanical and Production Engineering, Abubakar Tafawa Balewa Unicersity Bauchi, 740272, Nigeria.
[*]corresponding author's email: benburgaper1@gmail.com

Abstract – *Groundnut is an oilseed crop cultivated for food, oil extraction, and industrial applications. Understanding the engineering properties of groundnut varieties is essential for reducing postharvest losses and improving design of efficient threshing machines. Manual methods such as vernier caliper, ruler to measure groundnut pod dimensions are time consuming, prone to errors, and unsuitable for large scale processing. These limitations contribute to inefficiencies in threshing, such as increased groundnut pod breakage, foreign material contamination, and poor machine performance. This research employed machine vision and augmentation techniques to address these challenges by automating groundnut identification, classification, and sizing. Principal Component Analysis is applied to analyze the shape, size, and orientation of groundnut pods, enabling faster, more accurate, and simultaneous measurement of multiple pods far superior to manual linear measurements. To enhance classification, ResNet 101, a deep Convolutional Neural Network (CNN), was employed. This model enabled segmentation, feature extraction, identification and classification of different groundnut varieties such as Ex-Dakar, Jarma, and Samnut26. The trained model achieved a high accuracy of 98.6% and demonstrated strong performance across other evaluation metrics. Such as Ex-Dakar achieved 99% precision, 98% recall, 98% F1-score, and 100% ROC-AUC, Jarma scored 98%, 99%, 98%, and 100% and Samnut26 recorded 99%, 99%, 99%, and 100% respectively. This study establishes an intelligent framework for groundnut geometric analysis and classification using machine learning. The integration of PCA and deep learning not only improves accuracy and reduces human error but also supports the development of a smart, efficient groundnut threshing system that addresses postharvest processing challenges in agriculture.*

Keywords: *Classification, Convolutional Neural Network, Groundnut pods, Principal Component Analysis, ResNet101*

Article History
Received 02 January 2026
Received in revised form 03 March 2026
Accepted 11 April 2026

I. Introduction

Groundnuts are essential oilseed crop widely cultivated for food, oil extraction, and industrial applications [1]. Understanding the engineering properties of different groundnut varieties is crucial for their identification, classification, and for designing and optimizing threshing processes. This research had provides a comparative analysis of engineering properties and its classification of three known groundnut varieties Ex-Dakar, Samnut26 and Jarma. The determination of their physical and mechanical properties of groundnut pods play an important role in the problems associated with design, development of a groundnut threshing machine for processing, separation, cleaning and storage [2]. However, these processes have solved the challenges of lack of precision, reduce high losses, resulting to optimal performance. The research has addressed these challenges of lack of intelligent on the groundnut threshing machine that integrates machine vision-based classification and machine learning techniques for groundnut processing.

The machine vision system was able to identify the groundnuts based on their engineering characteristics and classified them accordantly for threshing performance through the enhancement of machine learning classification algorithms that was developed and deployed to the model that was trained and validated for real-time decision adapting to data capturing, data processing, model training and testing that aimed to minimizing groundnut damage while maximizing the threshing efficiency. This integration will improve the overall quality, yield, economic, time consume, labour intensity of groundnut processing. According to [3] reported that application of Machine Learning (ML), Image Processing (IP) is attributed to the use of farm mechanization, the advancement of higher-yielding crop and livestock varieties, robotics, remote sensing, data analytics, unmanned aerial and terrestrial vehicles, fertilizer and pesticide use, and tractor navigation via the Global Positioning System (GPS).

According to Venkatasaichandrakanth and Ivapparaja[4] there are many techniques used to classify various varieties of groundnut. Traditionally, classification has been performed using the human eye, which is unaided and often insufficient especially when dealing with erroneous or visually similar groundnuts. Modern techniques have become significant for accurate identification. These involve processes such as segmentation, feature extraction, and classification. Machine learning model such as Convolutional Neural Networks (CNN) have proven excellently in performing the tasks.

These models primarily focus on predicting groundnut varieties. They are well suited for image identification and classification, offering improved accuracy and efficiency. The results were achieved by fine-tuning the models parameter, inputting processed images, and embedding them into the models' layers for training, testing, and validation and this process enabled accurate identification of groundnut varieties. The developed model of an intelligent groundnut threshing machine that integrate machine vision and machine learning, hereby enhancing data capture and classification accuracy and will also improve productivity, reduced losses, and better quality control in the groundnut threshing processing

II. Literature Review

The studies of engineering properties of groundnuts for designing a battery-operated decorticator by [5] become necessary to examine the physical engineering properties of groundnuts before developing the decorticator, the properties influence its design. The key factors considered included size (i.e., length, thickness, and width) and shape.

A total of 20 samples were measured from a set of 100 groundnut pods. However, the method used did not incorporate artificial intelligence techniques such as machine learning models, Copilot, or image generators. Instead, it relied on a digital vernier caliper with an accuracy of ±0.02 mm.

Due to structural variations in groundnut pods such as surface roughness, irregular shape, minute deformations, and variations in positioning the caliper's jaws, precision measurement may be affected. Additionally, when measuring a large sample size of groundnut pods, using a digital caliper is time-consuming compared to automated model system.

However, [6] investigated the engineering properties of groundnut pods for advanced pneumatic pod collection systems. The study considered eighteen (18) different groundnut varieties, selecting twenty (20) samples from each for analysis. A digital vernier caliper was used to measure the size of groundnut pods in terms of width (W), length (L), and thickness (T). The results showed that each variety had different average pod length, breadth, thickness, geometric mean, and sphericity.

To evaluate the efficacy and efficiency of the pneumatic groundnut pod collector (GPC), certain engineering properties of groundnuts must be established. However, the digital vernier caliper measures only at specific points, which may not fully capture the true variability in groundnut, pod dimensions, particularly due to their irregular shapes. Another researcher [7] worked on the modification and performance evaluation of a manually operated groundnut decorticator. Their study achieved certain results in terms of threshing efficiency and cleanliness. However, it failed to determine the groundnut properties required for sieve design and did not incorporate automation in the threshing process for improved technology.

Also [2] investigated the engineering properties of groundnut pods and kernels relevant to the design of post-harvest machines. The physical properties of groundnuts were measured using a vernier caliper with an accuracy of 0.05 mm. Twenty (20) samples were selected from a total of one hundred (100) for measurement. However, a device with 0.05 mm accuracy cannot effectively reduce rounding errors, as small variations may not be captured, leading to less precise data results.

Furthermore, [8] conducted a study on the rupture resistance of groundnut (SAMNUT22) kernels using a digital vernier caliper. The three primary pod dimensions length, thickness, and width were measured with an accuracy of 0.01 mm. The kernel size was determined for classification purposes using geometric mean diameter and sphericity. One limitation of this approach is that measuring multiple groundnuts individually with a vernier caliper is slow and labour intensive, making it impractical for large scale sample analysis. Therefore, there is a need to improve the method using artificial intelligence techniques such as principal component analysis. The research conducted by [9] on the physical and engineering properties of three groundnut varieties released by BARI.

A total of 100 groundnut pod samples were randomly selected, and their dimensions breadth, thickness, and length were measured using a vernier caliper with a precision of 0.01 mm to determine their physical and engineering characteristics. However, measuring a large number of groundnut pods using a vernier caliper is time-consuming. Implementing a machine learning model could significantly improve efficiency and accuracy. Meanwhile [10] worked on the classification of peanut images based on multi-features and SVM where he classified the groundnut pod using three handcrafted features with SVM each and they achieved a credible result of accuracy performance. Aspect ratio with SVM had 96.72% accuracy, HOG with SVM had 81.97%, and Hu Invariant Moment with had 81.97%.

However, the research has limitation despite with the good result, the research used limited dataset for the training of the model which around 202 groundnut pod this may affect it generalization to real world task, also using milt-feature such as Aspect ratio, HOG and Hu Invariant Moment which they are handcrafted method may not adapt well with the image variations and SVM worked on smaller data when large number of groundnut is added it may not compute which can lead to overfitting and [11] worked on an intelligent classification model for peanut's varieties by colour and texture features.

The research proposed a testing method according to image processing and computer vision the process is good, fast with high differences in rate. The machine learning method used were Support Vector Machine (SVM), Random Forest (RF), and Multilayer Perceptron (MLP) with an accuracy result of SVM had 86.07%, RF had 82.27% and MLP had 84.9%, thus the research has shortcoming of using only colour and texture for it classification which will not captured other important features that will influence the variety classification. Also colour features based is sensitive to lighting condition during image capturing. The dataset for the training is not justified since only 210 samples were used.

III. Methodology

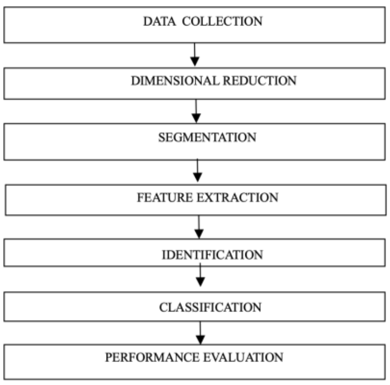


Fig. 1: Methodological approach

A. Data Collection

The groundnut pods such as Ex-dakar, Jarma and Samnut 26, which were obtained from the Muda Lawal market in Bauchi. A camera with a good resolution of a 20 megapixels was used to capture all the groundnut images at equal distance of 20cm under the same light intensities. The research had a dataset of 12,166 images of the groundnut pods. The Ex-dakar variety had three thousand seven hundred and thirty-three (3,733) images, Jarma variety had three thousand nine hundred (3,900) images and Samnut26 variety four thousand three hundred and thirty-three (4,533) images that were captured respectively.

B. Dimensional Reduction using Principal Component Analysis

The basic principle of principal component analysis (PCA) is to reduce the dimensionality of a groundnut dataset that is made up of many interconnected variables or properties while preserving the data's variance or characteristic to a manageable level. The research have used machine learning model to compute the geometric parameter of the groundnut pod such as length, width, through the use of Python (software) PCA library. Through the following processes i.e groundnut image to matrix conversion, constructing the dataset, data centering, covariance matrix and eigen decomposition, dimensionality reduction, and visualization. Equation 1 was used to compute and established different sizes of groundnut pods and was adopted from [12].

$$\text{Geometric Mean Diameter } (D_g) = \sqrt[3]{L \times W \times T} \quad \dots (1)$$

where: L = Length of the seeds
W = Width of the seeds
T = Thickness of the seeds

C. Segmentation

The segmentation was achieved through logical operations on the saturation and value channels of the HSV transformed image (colour based segmentation) to isolate the pod based on its distinctive color, also edge detection like Canny that the groundnut pod was located by bounding box, and morphological operations to refine segmented regions and remove noise. The mathematical model in equation 2 and 3 which were adopted from [13]

$$S = \begin{cases} 1 \text{ if } I\,(x,y) \geq T \\ 0 \text{ if otherwise} \end{cases} \quad \dots (2)$$

$$R = S(I) \quad \dots (3)$$

where I (x,y) = pixel intensity at location (x,y) of grayscale image

T = Threshold value chosen that separate the dark groundnut pod from the light background

S = Binary mask for groundnut pod region

R = Segmented region of the binary mask of the groundnut pod

D. Feature Extraction

Feature extraction is where the groundnut pod input images are transformed into higher level informative representations that are good for image identification and classification. The groundnut pod images were introduced through a convolutional layer and filters. These filters detected features and patterns such as edges, textures, sizes and shapes of the groundnut pods, and the convolutional layer processed the input image in a hierarchical transformation order also, a series of ResNet where the Pooling layers (max-pooling) reduced the spatial dimensions of the feature maps without losing any relevant information. The mathematical model for feature extraction is expressed in equation 15 was adopted from [14]

$$\mathrm{F}_{pod} = \Phi = f(A, P, AR, \mu, M) \quad \ldots(4)$$

where F_{pod} = Segmented groundnut pod

f = feature extraction function that maps raw segmented image

A= Area of the pod

P= Perimeter of the pod

AR= Aspect ratio of width and height

μ = intensity colour value of segmented groundnut pod

M = Shape moment (Hu invariant moment)

Φ = Feature extraction

E. Identification

The identification of groundnut pods images using a machine learning model in Python library that used Keras model, in which the Fully Connected layer in ResNet of Convolutional Neuron Network (CNN) is the central decision maker of the process. Then feature extraction was performed and data features were also obtained and the data was also used and trained the machine learning model which learned the patterns also differentia their individual features from the labeled datasets automatically. After the model is validated using separate test data and fine-tuned the accuracy. Then, the trained model is now used to identify each variety class by the new unseen groundnut pods images and the model were evaluated and maps them as specific output classes as Exdakar, Jarma and Samnut26, by the Fully Connected layer and we can confidently say the pod was identified and even established confidence score of the models by stating the percentage of recognition of the groundnut pod. . The equation 5 was adopted from [15]

$$I(x, y, c) \in \mathbb{R}^{H \times W \times C} \quad (5)$$

where H= is height of image

W = is width of image

C = number colour channel

(x,y,) = pixels coordinate

F. Classification

The classification of groundnut pod images involves using a trained machine learning model, as Convolutional Neural Network (CNN), that identify and classified the groundnut pods based on their visual features automatic. These layers improved the analyses and differentiated all the three different groundnut pods varieties by identifying each based on their characteristic. A labeled image was fed into the model of different groundnut varieties such as Exdakar, Jarma and Samnut26 and was trained. The mathematical model is given in an equation 6 and 7 which were adopted from [14]

$$\dot{y} = C(\mathrm{F_{pod}}) \quad \ldots(6)$$

can also be written as

$$\dot{y} = C(\Phi(S(I(x,y)))) \quad \ldots(7)$$

where

ẏ = Predicted output class label (Exdakar, Jarma, and Samnut26)

C = Classifier (CNN layer)

Φ = Feature extraction

S = Segmentation function

I (x,y) = Input image of the groundnut pod

IV. The Structural layer of Convolutional Neural Network (CNN)

The Convolutional Neural Network (CNN) is a machine learning model designed to extract, learn, recognize features from images and classifier them. Here in this research the Convolutional Neural Network was used in identified patterns, shape, and texture of our groundnut pods and classified them into various classes accurately, these images are digital in nature. It begins with an input layer that receives the groundnut pod image as shown in Fig. 2 and obtained from [16], for the purpose of preprocessed. However, the groundnut pod images were trained on ResNet101 platform of CNN according to the deep learning neural network architecture that consists of 101 layers and is built on the concept of residual learning. The model was pre-trained on ImageNet, which gave it information of the low-level features of the three groundnut varieties such as edges, textures, and object shapes.

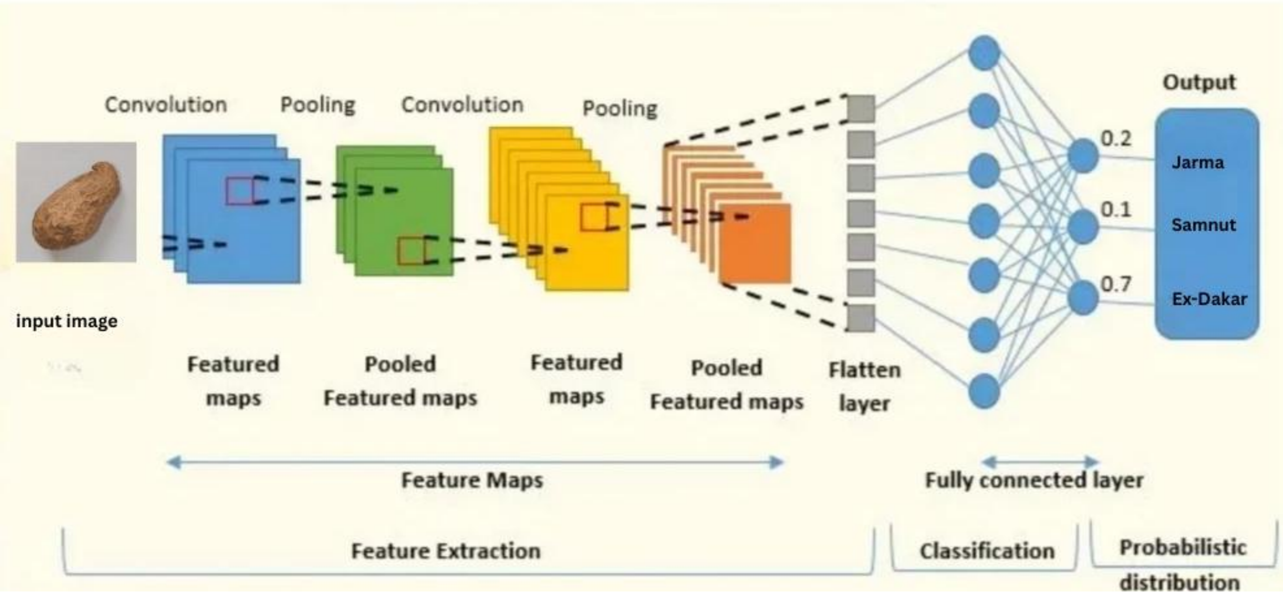


Fig. 2: Convolutional Neural Network Architecture

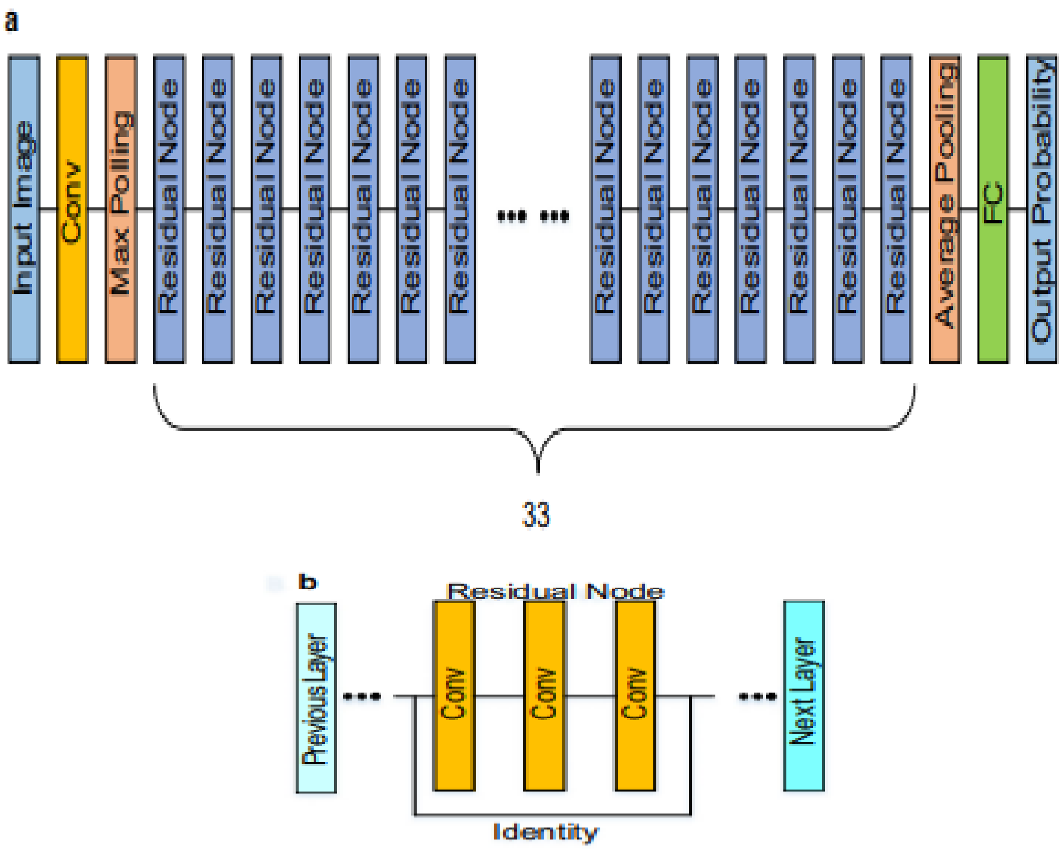


Fig. 3: The ResNet 101 Architecture [17]

When groundnut pod images of Exdakar, Jarma, and Samnut26 were introduced, they are resized and normalized also match the ImageNet dataset preprocessing, that where early introduced to the ResNet101 layers and were fine-tuned and classified. While training the data images, the images are passed through the residual blocks, which consist of convolutional layers, batch normalization, ReLU activation, that helped to preserved important features of each pods, also the ResNet101 consist of the three layers such as convolutional layer, pooling layer and fully connected layers with total 33 residual block (93 layers), 2 initial pooling layers and 2 final fully connected layers as shown in Fig. 3 also gotten from [17].

The convolutional layers filter the entire input image and detected the features such as edges, textures, and shape pod through the use of Rectified Linear Unit (ReLU) to learn complex pattern and non-linear. The Pooling layers, such as max pooling, were then used to reduce the spatial dimensions of the feature maps, making the network more efficient and focused on dominant features. This process of convolution and pooling may repeat several times to form a deep feature hierarchy. Furthermore, the flatten layer transforms the 2D feature maps into a 1D vector, which is passed into fully connected layers that perform high-level operation on the features. The last layer which is SoftMax layer, which indicated their outputs like the input image that belong to each of the groundnut variety class i.e Exdakar, Jarma, and Samnut26.

This architecture had helped the CNN to learn complex patterns directly from image data and classify groundnut pods with high accuracy, even when there is variation in shape, size, but haven the same color. The model continues train, learn and extract deeper and more discriminative features that distinguished between the groundnut pods types with high accuracy. The used of ImageNet-pretrained weights speeds up the convergence and improved the model performance, making it ideal for groundnut pod image classification task. This method ensured accurate classification of groundnut pods varieties and when implemented on the hardware and will contribute to the intelligent system of groundnut threshing processing and other processes.

Note:

a = is a Schematics of the ResNet 101 Architecture that include 33 residual nodes in total

b = the residual node served as building block for the ResNet 101 Architecture

However, the structural flow of the groundnut classification method model like CNN had summarize the entire process as indicated in Fig. 4 and this process was across all three varieties for well and accurate results. The labeled images presented in Fig. 5 validated the groundnut pods dataset was properly prepared for deep learning model training. Each image in the grid labeled with its corresponding class name Jarma, Exdakar, and Samnut26, these confirmed the uniqueness of supervised learning model was effectively applied. The groundnut pod images well aligned and resized uniformly standard of 224×224×3 dimensions, ensuring compatibility with CNN architectures such as ResNet101.

More so, the variation in pod orientation and positioning within each class helps the model generalize better, avoiding overfitting. This organized image confirmed that the dataset is correctly labeled according to its orientation and angle were analyzed at 0^0, 90^0, 180^0 and 270^0 that assessed classes for reliable geometric features, learning, training, classification, and evaluation purposes.

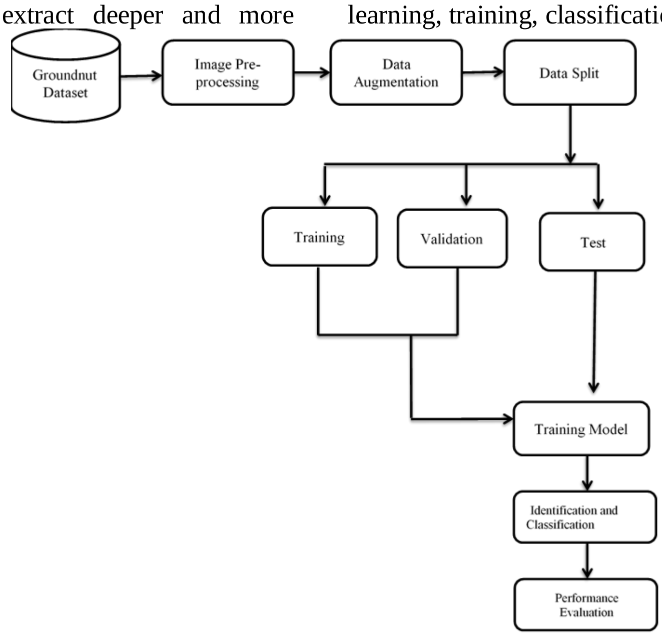


Fig. 4: Block Diagram of Classification Model Flow

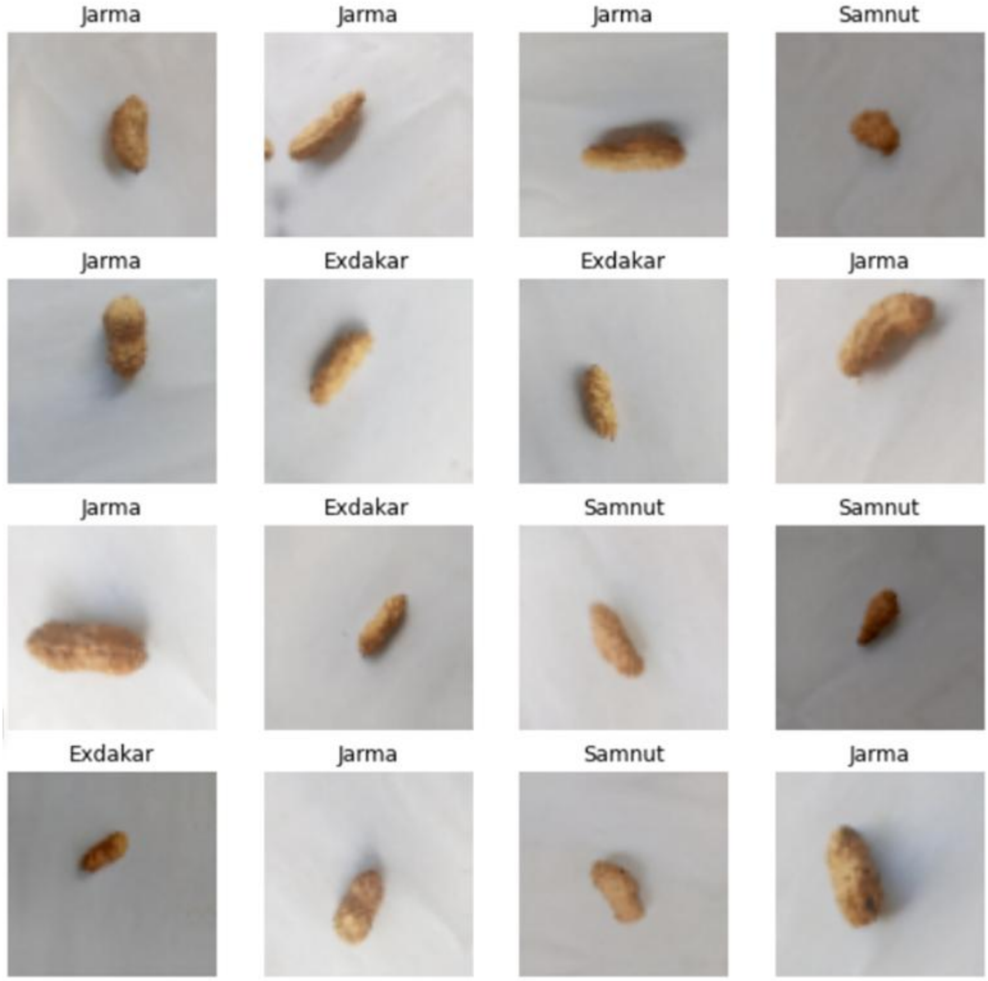


Fig. 5: Groundnut Pod Image Grid of dataset

V. Performance Evaluation of the Machine Learning model

One of the essential techniques for performance evaluation of a classification models is a confusion matrix (decision matrix). This provides insights into how well the model distinguishes between different classes by showing the number of correctly and incorrectly classified instances. According to [18] Accuracy, Precision, Recall, and F-Score were used to assess the classification algorithms' performance. The parameter used for the evaluation are as follows; the True Positives (TP) is the true correctly predicted label as positive, the False Positives (FP) is the actual label as negative because of incorrect prediction, the True Negatives (TN) is the correctly predicted negatively when label, and the False Negatives (FN) is the incorrectly predicted as negative though label true. Below are the parameters used to evaluate the model performance as adapted [18]

i. Accuracy is the ratio of all the correct classified samples and to the total number of samples

$$\text{Accuracy}=\frac{\text{TP}}{\text{TP+TN+FP+FN}} \quad \dots (8)$$

ii. Precision is the proportion of true positives out of the total predicted positives

$$\text{Precision}=\frac{\text{TP}}{\text{TP+FP}} \quad \dots (9)$$

iii. Recall is the proportion of positive samples of classified as true

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad \dots (10)$$

iv. F1-Score is the harmonic mean of recall and precision

$$\text{F1} = 2* \frac{\text{Precision}*\text{Recall}}{\text{Precision+Recall}} \quad \dots (11)$$

VI. Performance Evaluation of the Machine Learning model

These present the performance analysis of PCA, machine learning model using matrices.

A. Properties of the groundnut pods

The Principal Component Analysis (PCA) was effectively captured the groundnut pods for true geometric dimensions, regardless of its orientation of the groundnut images and the results were presented in Fig. 6 to 8.

8x8 Matrix of Exdakar

1.71	1.72	1.72	1.74	1.75	1.71	1.71	1.71
1.71	1.74	1.68	1.42	1.53	1.78	1.7	1.7
1.71	1.75	1.57	1.2	0.71	1.18	1.74	1.68
1.7	1.73	1.53	1.34	0.9	0.52	1.52	1.72
1.69	1.72	1.6	1.35	0.9	0.45	1.22	1.75
1.69	1.69	1.67	1.16	0.41	0.42	0.89	1.74
1.68	1.68	1.69	1.68	1.22	0.78	1.23	1.71
1.67	1.67	1.67	1.64	1.71	1.71	1.67	1.65

Fig. 6: Geometric diameter of Exdakar variety

8x8 Matrix of Jarma

1.92	1.86	1.75	1.5	1.32	1.32	1.52	1.71
1.91	1.86	1.51	0.98	0.9	0.93	1.23	1.64
1.95	1.76	1.02	0.79	0.82	0.79	1.17	1.77
1.97	1.39	0.77	0.86	1.04	0.81	0.93	1.83
1.94	1.54	1.16	1.16	1.16	1.24	1.52	1.88
1.93	1.87	1.65	1.61	1.53	1.73	1.89	1.85
1.94	1.93	1.93	1.91	1.91	1.88	1.85	1.85
1.94	1.93	1.93	1.91	1.88	1.86	1.86	1.87

Fig. 7: Geometric diameter of Jarma variety

8x8 Matrix Samnut26

1.85	1.87	1.86	1.87	1.89	1.86	1.85	1.85
1.82	1.83	1.86	1.76	1.6	1.8	1.86	1.84
1.79	1.82	1.75	1.47	1.34	1.38	1.84	1.81
1.77	1.81	1.57	1.14	1.04	1.15	1.8	1.83
1.74	1.77	1.48	1.03	0.69	1.29	1.71	1.8
1.72	1.75	1.28	1.0	0.92	1.46	1.68	1.78
1.74	1.77	1.47	0.84	1.19	1.61	1.73	1.81
1.75	1.73	1.73	1.5	1.48	1.69	1.77	1.8

Fig. 8: Geometric diameter of Samnut26 variety

Furthermore, Table 1 presents the summary of the various sizes of the groundnut images. The Exdakar has the most variation variety in sizes i.e it has a mixed size of small and large pods, **S**amnut26 is consistent stable in size and evenly distributed and Jarma has the largest size and most uniform distribution making it the most consistent variety in size**.**

TABLE I
SUMMARY OF THE GROUNDNUT GEOMETRIC DIAMETER

Parameter	Exdakar	Samnut26	Jarna
Maximum Size (cm)	1.78	1.89	1.95
Minimum Size (cm)	0.41	0.69	0.77
Size Distribution	Moderate variation (0.41 – 1.78 cm), significant presence of small pods (cm)	Moderate variation (0.69 – 1.89 cm), most sizes between 1.5 – 1.8 cm	Moderate variation (0.77 – 1.95 cm), majority between 1.6 – 1.9 cm
Uniformity	Fairly uniform with some fluctuations	Moderate uniformity with less extreme variations	Relatively uniform with some small outliers

B. Machine Learning models

The groundnut varieties were trained on an open-source of machine learning Python library module where an algorithms model was developed on ResNet 101 of CNN and were tested with the dataset as Ex-Dakar, Samnut26 and Jarma and the predictions results were displayed on the confusion matrix (Decision matrix) table for performance evaluation through training accuracy and training loss, Receiver Operating Characteristic (ROC) and Confidence score. These metrics provided an assessment that is well comprehensive about the model's performance on groundnut classification.

C. Training Dataset

In this research about 12,166 samples of groundnut pods as dataset was used to train the machine learning model, which comprise of three varieties of groundnut pods such as Exdakar, Jarma, and Samnut26, with image counts of 3,733, 3,900, and 4,533 respectively. The dataset was divided into three distinct subsets as 70:15:15 ratios for training, validation, and testing purposes as indicated in Table 2. This means that for each groundnut variety, 70% of the images were allocated for training the model, that enable it to learn the specific groundnut variety, 70% of the images were allocated for training the model, that enable it to learn the specific features of each pod variety. Subsequently, 15% of the images were used for validation to fine-tune the model's parameters and prevent overfitting or underfitting due to the complexity nature of our data are listed in Table 3. While the remaining 15% were reserved and test also ascertain the model's generalization performance during validation.

TABLE II
DATA TRAINING SET

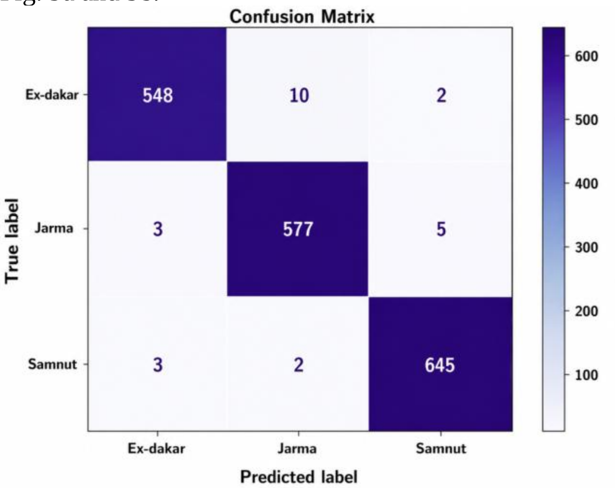
Classification	Training Data	Validation Data	Testing Data
Exdakar	2613	560	560
Jarma	2730	585	585
Samnut26	3233	650	650

TABLE III
TRAINING PARAMETER

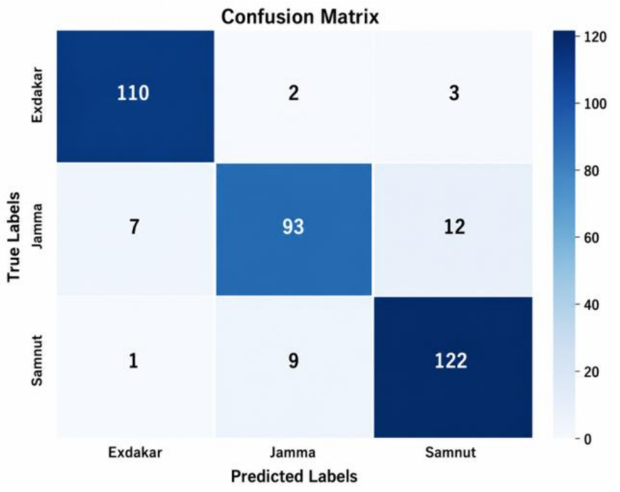
Parameter	Typical Value	Description
Learning Rate	1×0^{-4}	Controls the learning process during training and update the weight
Batch Size	64	Number of samples processed before model updates
Epochs	20 – 100	Training iteration times of dataset
Activation Function	ReLU	Introducing non-linearity after convolutional layer
Input Image Size	224 × 224 × 3	Standard image input size (width x height x RGB channels) to avoid error
Weight Decay (L2 regularization)	1×10^{-4}	Prevents overfitting by penalizing weight

D. Confusion Matrix of CNN and VIT

The groundnut data images were trained on the models and the results were analysed using confusion matrix on how well the model classified the three varieties of groundnuts Exdakar, Jarma, and Samnut26 as shown in Fig. 9a and 9b.



(a)

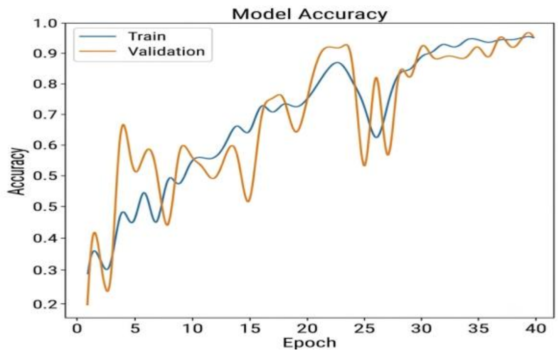


(b)

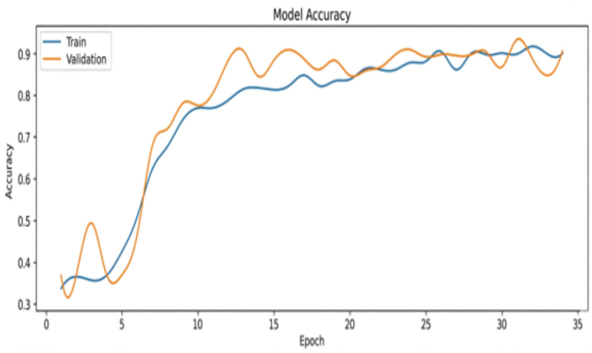
Fig. 9: Confusion matrix of (a) CNN model and (b) VIT model

E. Training and Validation Accuracy of CNN and VIT Model

The graph in Fig. 10 is the accuracy performance of the machine learning model in terms of training and validation accuracy over multiple epochs. The training accuracy is in blue line while the validation accuracy is in orange line.



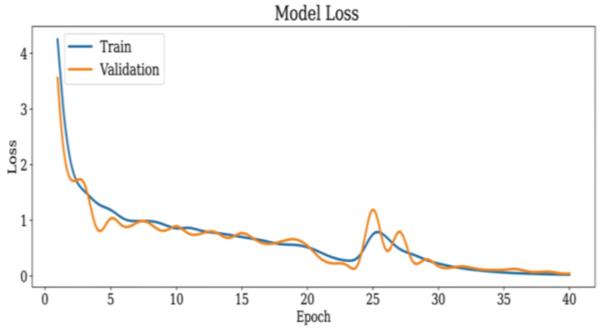
(a)



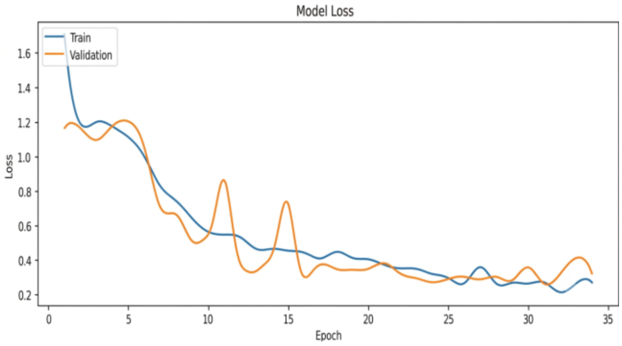
(b)

Fig. 10: Training and Validation accuracy of (a) CNN model and (b) VIT model

The training and the validation loss which are represented both in blue line and orange line as indicated in Fig. 11 which expresses the performance of the machine learning model in terms of training and validation accuracy across multiple epochs.



(a)



(b)

Fig. 11: Training and Validation loss of (a) CNN model and (b) VIT model

F. Receiver Operating Characteristics

The Receiver Operating Characteristic (ROC) curve shown in Fig. 12, presents the performance evaluation of a multi-class classification model distinguishing between the three classes Exdakar, Jarma, and Samnut26.

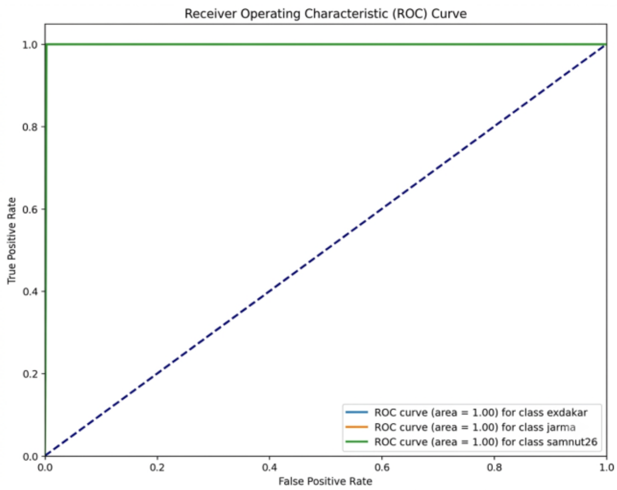


Fig. 12: Receiver Operating Characteristics

G. Confidence Score

The confidence score presents the ability of the model to detect the groundnut pods images from the neural network and assigned the confidence score of each class of the groundnut pods such as Exdakar, Jarma and Samnut26 The model reads pixel patterns based on the groundnut pod image orientation that passed through convolution and pooling layers, which extract features texture, shape, and size unique of the groundnut variety as shown in Fig 13 and the confidence score of the three varieties are shown in Table 4.

TABLE IV
CONFIDENCE SCORE OF EACH GROUNDNUT

Class	Confidence Score of First Variety (%)	Confidence Score of Second Variety (%)	Confidence Score of Third Variety (%)
Exdakar	99.37	95.77	99.73
Jarma	99.55	99.99	99.72
Samnut26	99.90	100	100

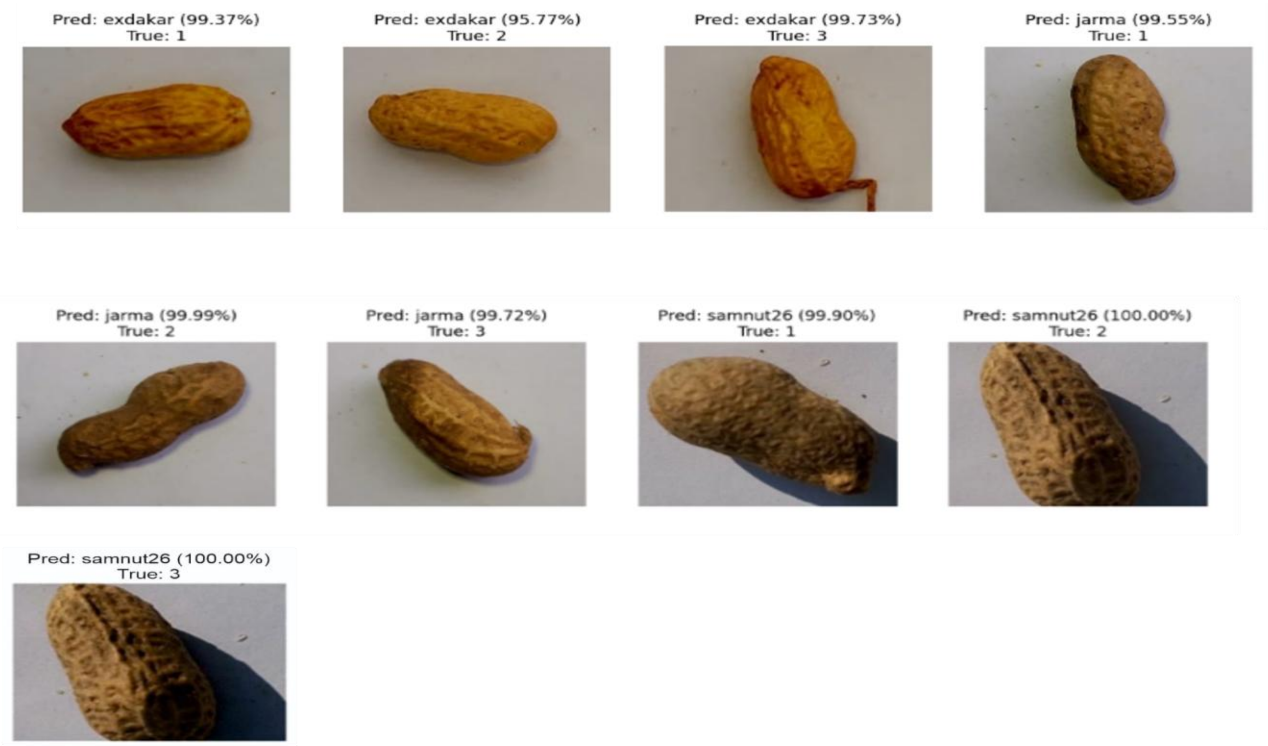


Fig. 13: Confidence Score of Three Varieties of Groundnut

H. Performance Evaluation of CNN

The performances of the model were evaluated using the following Accuracy, Precision, Recall and F1-Score as shown in Table 5.

TABLE V
MACHINE LEARNING MODEL PERFORMANCE (CNN)

Groundnut Pod Varieties	Accuracy	Precision	Recall	F1-Score	ROC
Exdakar		99	98	98	100
Jarma	98.6	98	99	98	100
Samnut26		99	99	99	100

VII. Discussion

The physical and the engineering properties of different varieties of the groundnut such as Exdakar, Jarma, and Samnut26 were achieved using principal component analysis (PCA) from Python modules math. This research targets the dimensional reduction of the groundnut pods to established its different sizes for the purposed of having a quality threshing efficiency of the groundnut seeds. The engineering properties obtained, will now be used to develop a concave (screen) clearance space gap of each variety due to their differences in sizes, this process will improved the performance of the groundnut threshing machine.

The principal component analysis is the automated digital method used in determining groundnut pods sizes within a short time with high precision unlike the used of digital vernier caliper which is time consuming, error prone and fatigue when used [19], also [20] who used ruler to measured groundnut pod parameter with excessive error and low threshing efficiency The PCA method was achieved through the following process as groundnut image to matrix conversion, constructing the dataset, data centering, covariance matrix and eigen The ResNet101 of Convolution Neural Network was the model used for the feature extraction, identification and classification due to its ability and strength of handling complex image for analysis.

During the training, the groundnut pod images undergoes different transformations that improved its generalization such as flipping, scalable, rotation of 00, 900, 1800 and 2700 degrees, resizing, blowing and cropping to a dimension of 224x224 pixels especially when dealing with imbalanced or limited data set for model performance. Traditional machine relied based on operator selection feature of their physical and engineering properties using their domain skill knowledge which is seen as setback engineering practice, time consuming and error prone due to worker fatigue that led to more minor error resulting to misalignment and misclassification during operation [17].

However, the introduction of deep learning approach adopted in this research has addressed such problem [22] stated that the invention of ResNet was a great important in the development of CNN in which ResNet had outshine previous models in image processing, detection and recognition for it variant differ in layers. The research data were divided into 70% for training, 15% for validation and fine-tune the parameters in order to avoid underfitting, and overfitting for better performance and generalization of the model while the remaining 15% was allocated for testing to revalidated the performance of the new, unseen data as presented in Table 2.

The model underwent training for 40 epochs with a batch size of 64 samples. The models utilized a hyper parameter of learning rate of 0.0001, activation function of Rectifier Linear Unit, image input size of 224 × 224 × 3 and to have well performance weight decay (L2) was applied. The model has good performance as shown in Table 3, this is because the ResNet101 has properties of 33 residual note with 93 layers at the convolution layer neck, Initial Conv and Pooling with 2 layer each while Final Pooling and Full Connected also has 2 layer each. The ResNet has a total 101 layers, these made it a successful model in handling complex image, converging easily during training, haven significant high performance.

The model demonstrated high accurate prediction and formed a confusion matrix as shown in Fig. 9, where the correct classified results was presented in a diagonal with dark blue colour while the misclassifications were represented in off-diagonal. Exdakar was classified as 548 varieties and was also misclassified 10 varieties as Jarma and 2 varieties as Samnut26. Jarma was classified as 577 varieties and was also misclassified 3 varieties as Ex-dakar and 5 varieties as Samnut26. Likewise, Samnut26 was classified as 645 varieties and was also misclassified 3 varieties as Exdakar, and 2 varieties as Jarma. The confusion matrix comparatively shows low rate of misclassified varieties which mean the model performance was achieved.

The graph in Fig. 10 explained the accuracy performance of the machine learning model in terms of training accuracy and validation accuracy over multiple epochs. The training accuracy which is the blue line and validation accuracy which is the orange line had both shown a consistent upward trend in learning, starting from approximately 30% to 100% and 20% to 100% respectively across multiple iterations of 40 epochs. This indicated that the model learned effectively from the training data while maintaining a high degree of generalization to unseen validation data, the closed alignment between training and validation accuracy curves which signified a very minimal overfitting also making the model more robust.

However, Fig. 11 shows the loss graphs which present the training loss and validation loss trajectories across 40 epochs, providing the learning dynamics of the deep learning model. Both the training loss and validation loss exhibit a consistent downward trend, indicating successful convergence and effective learning. The validation loss closely tracks the training loss throughout the epochs without significant divergence, and the model generalizes well to unseen data and there was no overfitting. Furthermore, the steady reduction in both losses stabilizes at a very low range approximately 2% these imply that the training and validation processes is very effective and the model minimizes the error function efficiently.

The Receiver Operating Characteristic (ROC) curve shown in Fig. 12, as a graphical representation of a diagnostic model that presents the performance evaluation of a multi-class classification model that distinguished between the three classes Exdakar, Jarma, and Samnut26. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1-specificity) for different classification thresholds. The Area under the curve (AUC) for all three classes of groundnut pod is 1.00 indicating perfect classification performances.

An AUC of 1.00 implies that the model has achieved complete separate between the positive and negative instances for each class, with good classifications. However, for confidence score represents the certainty of the model about the groundnut identification and its prediction on all the three varieties Exdakar, Jarma and Samnut26 images as shown in Fig. 13, the model identified the groundnut pod with the following confidence score Exdakar had 99.37, 95.77, and 99.73 %, Jarma have 99.55, 99.99, 99.72 %, and Samnut26 have 99.90, 100 and 100 % respectively, which indicated a very high certainty in its identification. Therefore for this research the model had attended its threshold. Where the confidence score is below 50% the model is said to be unreliable and uncertainty while Table 4 summarized the confidence score that were assigned by softmax in ResNet 101 of Convolutional Neural Network.

Therefore, Table 5 presents the performance evaluation of machine learning models classification of the three varieties of the groundnut pod with an Accuracy of 98.6%. These was another performance model analysis of Precision performed by the machine learning model ResNet 101 of CNN across three groundnut varieties, which achieved the precision of Exdakar is 99%, Jarma is 98% and Samnut26 is 99% these indicated that the model had strong classification reliability by minimizing the false positive. The Recall model also achieved the performance as follows Exdakar is 98%, Jarma is 99% and Samnut26 is 99%, which means that the model is capable of minimizing the false negatives in groundnut classification. Furthermore, the F1-Score model performed as Exdakar is 98%, Jarma is 98% and Samnut26 is 99% respectively, this means that the model was able to minimize both false positive and false negative of the model.

The comparatives analysis performance of the models accuracy among different models Convolutional Neural Network (CNN), Vision Transformer (ViT) in Fig. 9,10 and 11 where the CNN performance is better than VIT, the CNN has accuracy of 98.6% while VIT is 91%, also CNN training and validation of accuracy and loss model is better than that of VIT.

VIII. Conclusion

In conclusion, the images of the groundnut pods were successful captured, and their sizes were analyzed and determined across different groundnut varieties, as shown in Fig. 11, 12, and 13. Exdakar had a minimum size of 0.41 cm and a maximum of 1.78 cm, Samnut26 ranged from 0.69 cm to 1.89 cm, while Jarma ranged from 0.77 cm to 1.95 cm. These measurements were achieved using Principal Component Analysis (PCA). Also a robust dataset of the groundnut pod was formed.

The groundnut pods images were cropped, image processing algorithms of the groundnut pod was also developed and achieved for image enhancement. An augmentation technique was effectively implemented that improved the quality of the real-time image processing that ensured robustness and consistency in feature extraction.

Furthermore, machine learning model was successful developed, trained, and validated and also the groundnut pods were classified and their validation test were presented in matrix with good performance classification with little misclassification.

Lastly, the performance evaluation of the model was conducted using performance metrics such as accuracy, precision, recall, F1-Score and ROC were used to assess the model, the analysis across each varieties were as followed, Ex-dakar had Precision of 99%,Recall of 98%, F1-Score of 98% and ROC 100%. Jarma had Precision of 98%, Recall of 99%, F1-Score of 98% and ROC 100%. While Samnut26 achieved Precision of 99%, Recall of 99%, F1-Score of 99% and ROC of 100%. Therefore Samnut26 variety performed the best across all varieties and in conclusion the model had attested to the engineering practices through improving technology and human socio-economic benefit if applied to real life with a better efficiency.

Acknowledgements

The authors wish to acknowledge the Abubakar Tafawa Balewa University, Faculty of Engineering and Engineering Technology Bauchi, Nigeria. For her great support and the platform give to us to carry out this novelty research.

Conflict of Interest

The authors declare no conflict of interest in the publication process of the research article.

Author Contribution

Author 1 conceptualized the research, design experiments system and prepared the article paper. Author 2 provides the technical support on machine learning. Author 3 assisted with the development of the research framework and Author 4 contributed to the preparation of dataset, validation of the experiment and fine-tuned the article manuscript.

References

[1] Bhatnaga, S., Chaudhary, R., Sharma, S., Janjhua, Y., Thakur, P., Sharma, P., and Keprate,A.(2024). Exploring the dynamics of climate-smart agricultural practices for sustainable resilience in a changing climate. Environmental and Sustainability Indicators, 24,100535. https://doi.org/10.1016/j.indic.2024.100535

[2] Hassan, M. H. A., and Geasa, M. M. (2021). An investigation of engineering properties of groundnut pods and kernels related to design the post-harvest machines. *Archives of Agriculture Sciences Journal, 4*(3), 144–160

[3] Ngugi,L.C., Abelwahaba,M., Abo-Zahhad,M (2020) Recent advances in image processing techniques for automated leaf pest and disease recognition -A review, Information Processing Intelligent Control of Threshing and Cleaning System Based on Big Data in Agriculture, https://doi.org/10.1016/j.inpa.2020.04.004

[4] Venkatasaichandrakanth,P., Ivapparaja,M. (2024) Asurvey on Pest Detection and Classification in the Field Crops using Artificial Intelligent Technique.International Journal of Intelligent Robotics and Application 8(3), 709-734

[5] Ugargol, M., Gowtham, and Santos. (2024). Engineering properties of groundnut for designing a battery-operated decorticator. *Journal of Scientific Research and Reports, 30*(9), 205216, ISSN: 2320-0227

[6] Verma, K., Kathiria, R. K., Mehta, T. D., Chauhan, P. M.,Yadav, S., and Lakhani, A. L.(2024). Investigation of engineering properties of groundnut pods for advanced pneumatic pod collection systems.

Plant Archives, *ttps://doi.org/10.51470; 24*(2) e-ISSN:2581-6063(online), ISSN:0972-5210

[7] Kadlag, R., Pawase, P., Khurdal, J., and Naraya, G. (2022). Modification and performance evaluation of manually operated groundnut decorticator. The Pharma Innovation Journal, 11(2), SSN (E): 2277- 7695 ISSN (P): 2349-8242

[8] Uguru, H., Akpokodje, O. I., and Altuntaş, E. (2021). A study on rupture resistance of groundnut (cv. SAMNUT 22) kernel. *Turkish Journal of Agricultural Engineering Research, 2*(1),https://doi.org/10.46592; e-ISSN: 2717-8420; 2(1)[9] Hoque, M. A. (2019). Physical and engineering properties of BARI released three groundnut varieties. *Bangladesh Journal of Agricultural Research, 44*(4), ISSN0258-7122 (Print), 2408-8293 (Online)

[10] Li, Z., Niu, B., Peng, F., Li, G., Yang, Z., & Wu, J. (2018). Classification of peanut images based on multi-features and SVM. *IFAC-PapersOnLine,51*(17),726-731.https://doi.org/10.1016/ j.ifacol.2018.08.444

[11] Narendra, V. G and Kini, A.S (2018) An intelligent classification model for peanut's varieties by color and texture features. International Journal of Engineering & Technology Website: www.sciencepubco.com/index.php/IJET.

[12] Verma, K., Kathiria, R. K., Mehta, T. D., Chauhan, P. M., Yadav, S., and Lakhani, A. L.(2024).Investigation of engineering properties of groundnut pods for advanced pneumatic pod collection systems. *Plant Archives, ttps://doi.org/10.51470; 24*(2) e-ISSN:2581-6063 (online), ISSN:0972-5210

[13] Wang, Y., Ye, F., Zeng, J., Cai, J., & Huang, W. (2024). Adherent peanut image segmentation based on multi-modal fusion. *Sensors, 24*(14), 4434. https://doi.org/10.3390/s24144434

[14] Zhang, Y., Wu, L., & Wang, S. (2019). Feature extraction and classification for image-based plant recognition using machine learning techniques. *Computers and Electronics in Agriculture, 162*,112– 121.https://doi.org/10.1016/j.compag.2019.04.018

[15] Gonzalez, R. C. and Woods, R. E. (2018). A Text Book Digital image processing (4th edition) *Printed and bound in Malaysia ISBN 10: 1-292-22304-9 ISBN 13: 978-1 292-22304-9*

[16] Ingoampt (2024) Mastering the Mathematics Behind Convolutional Neural Networks in Deep Learning, *https:// developerbreach.com/convolutional neural network deep learning*

[17] Li, F., Yan, L., Wang, Y., Shi, J., Chen, H., Zhang, X., Jiang, M., Wu, Z., & Zhou, K. (2020). Deep learning-based automated detection of glaucomatous optic neuropathy on color fundus photographs. *Graefe's Archive for Clinical and Experimental Ophthalmology*. https://doi.org/10.1007/s00417-020-04636-8

[18] Taner, A., Mengstu, M. T., Selvi, K. Ç., Duran, H., Gür, İ., and Ungureanu, N. (2024).*Apple varieties classification using deep features and machine learning*.https://doi.org/10.3390/ agriculture 14020252

[19] Ghosal, M. K., and Sarangi, P. (2020). Studies on some physical properties of groundnut of TMV-2 variety. *International Journal of Chemical Studies, 8*(1), 2415–2419. https://doi.org/10.22271/chemi.2020.v8.i1o.8399; P-ISSN: 2349–8528 E-ISSN: 23214902

[20] Ofori, H., Amoah, F., Arah, I. K., Piegu, M.-K. A., Aidoo, I. A., and Commey, E. D. N. A. (2020). Physical properties of selected groundnut (*Arachis hypogea* L.) varieties and its implication to mechanical handling and processing. *African Journal of Food Scienc.* DOI: 10.5897/AJFS2020.1974; ISSN: 1996-0794

[21] Yu,X., Lu,S., Guo ,L., Wang,S.H., Zhang,Y.H (2020) ResGNet-C A graph convolutional neural network for detection of COVID-19. Journal homepage: www.elsevier.com/locate/neucom. https://doi.org/10.1016/j.neucom.2020.07.144 0925-2312