## A PROPOSED MODEL FOR PREDICTING THE MATURITY OF GROUNDNUT

**\*[1]Suleiman T., [2]Saidu I. R., [2]Musa M. N., [3]Isa M. M., [1]Hassan K. A. and [4]Abdul A. J.**

[1]Department of Computer Science and IT, Federal University Dutsinma, Katsina State, Nigeria
[2]Department of Intelligence and Cyber Security, Nigerian Defence Academy, Kaduna State, Nigeria.
[3]Department of Computer Science, Nigerian Defence Academy, Kaduna State, Nigeria.
[4]Department of Computer Science, Ahmadu Bello University Zaria, Kaduna State, Nigeria

*Corresponding Author's Email: tsuleiman@fudutsinma.edu.ng

**ABSTRACT**

The use of information communication Technology (ICT) has grown exponentially that its monumental application can be seen in almost every aspect of human endeavor of which agriculture is not an exception to the profound benefits provided by this field. Groundnut is used to be one of the most remunerative farming enterprises in Nigeria prior to the discovery of crude oil. In Nigeria groundnut is crushed to produce roasted snacks, groundnut oil or boiled either in the shell or unshelled for direct consumption. This research work aims at mitigating the difficulty associated with manual detection of groundnut maturity using certain features of the leaves. Customary, to examine the maturity of groundnut it requires constant monitoring and observation of changes in the color of groundnut leaves from purely green to purely yellow. This method of maturity assessment in order to harvest the crop without excessive loss is less accurate and consumes an awful lot of time particularly in a large farm. Hence, this approach cannot be fully reliable as color is subjective to our naked eyes and failure to harvest the crop when it reaches optimum maturity stage might cause the seeds pod to decay/ germinate underground due to moisture which might eventually result in quantity reduction of the expected yield. Consequently, design of an automated system is pivotal to farming and becomes necessary in the context of ICT era. This task is achieved by identifying the stages of the leaves of the groundnut plant using a convolutional neural network classifier. An accuracy of 99.1% is achieved.

**Keywords:** Groundnut, Leaves, CNN, ICT, Maturity, RGB, Color.

## INTRODUCTION

According to Vibhute & Bodhe (2012), Image processing has been proved to be an effective tool for analysis in various fields and applications. The application of image processing in agriculture can exponentially augment the quality of decision making for ripeness assessment, measuring severity of plant diseases, fruit sorting and grading, quantity of expected yield, weed detection amongst others. However, one of the strategies employed in improving agricultural activities is image processing. The use of image processing in agriculture has a wide range of applications; from image analysis to computer vision. The application involves the extraction of features which uniquely identifies an object to perform the needed task. These features include: shape, size color and texture (Abdulhamid, Aminu, & Daniel, 2018).According to Mustafa et al., (2008) most fruits/crops, color is one of the essential attributes that serve as a good ripeness indicator, where the percentage of ripeness can be determined by evaluating the individual pixels of the image.

Groundnut is used to be one of the most remunerative farming enterprises in Nigeria before the discovery of crude oil. In Nigeria groundnut is crushed to produce confectionary snack roasted, groundnut oil or boiled either in the shell or unshelled for direct consumption. The classical approach of groundnut ripeness and disease assessment is done by professionals or human inspectors, which is labor intensive, exorbitant high in cost, less accurate and consumes an awful lot of time as it is affected by external menace like fatigue, subjectivity etc. due

to human fallible nature. Consequently, design of an automated system is pivotal to farming and becomes necessary to substitute the traditional method adopted by farmers in the context of Information communication technology era so as to address the aforementioned factors. In order to determine the harvest date, a farmer must constantly scout his crop on a regular basis. The groundnut plant usually gives an indication when the crop reaches optimum ripeness stage. To determine the correct harvest date, the development of the plant must be considered. The groundnut should be harvested when approximately 75% of the pods have reached maturity. When 75% of the selected pods already show the dark discoloration, the harvest process can start. It is strongly advised that the groundnut should be harvested when the leaves turned from green to yellow; this is because delaying the harvest process when the leaves color changed to yellow could make the seedpods under the ground to decay, which will imminently cause an awful loss of the expected yield.

### Related Works

In one of the early works involving computer-aided classification of plants, Mustafa et al., (2008) proposed a novel method of classifying bananas according to shape, size, color, texture and taste of the banana. The process of getting size and ripeness of the banana is done from its image using image processing toolbox in MATLAB. The percentage of ripeness can be determined by evaluating individual pixels of the image and the area, perimeter, length and the thickness of a banana

was determined using MATLAB.

Savakar, (2012) Proposed a different method that classifies and identifies different types of bulk fruits images using a neural network. Features such as color and texture were extracted which represents the 24- bit input images and the artificial neural network is applied to the extracted features to reach a decision regarding the represented type of images. Five different fruits, namely: apple, chikoo, Mango, Orange and Sweet Lemon were considered and a back propagation neural network (BPNN) was used to recognize and classify the fruit images using different types feature sets, viz; color, texture and combination of both color and texture features. Accuracy achieved for the classification based on color feature of mango and sweet lemon is about 87%, orange is 88% and chikoo is 90%. Similarly, an accuracy of 94% is achieved for chikoo, Mango and orange 92% using combination of both color and texture features.

Fadilah, Sale, Halim, Ibrahim and Ali(2012) like Savakar, (2012)proposed a neural network-based intelligent system for ripeness classification of oil palm fresh fruit bunch (FFB). The images of oil palm FFBs of type DXP were collected and analyzed using digital image processing techniques. Then the color features from the acquired images were extracted and used as the inputs for artificial neural networks (ANN) learning. The performance of the ANN for ripeness classification of oil palm FFB was investigated using two methods: training ANN with full features and training ANN with reduced features based on the Principal Component Analysis (PCA) data reduction technique.

Ramakrishna and Sahaya (2015) used similar approach but differed in its use case by detecting plant diseases such as fungi, bacteria and viruses. The upgraded processing pattern comprises of four leading steps. Initially a color renovation constitutes intended for the input RGB image is formed, This RGB is converted into H because RGB is for Color generation and color descriptor. The next step is plane separation. Next performed is the color feature after which back propagation algorithm is used in the detection of leaf disease.

Abdulhamid et al., (2018) used a similar approach by extracting not just the color features as done by Ramakrishnan and Sahaya (2015) but also textures of the leaves serving as the input to the model. The proposed approach was used in predicting the ripeness of soya beans through image processing technique and artificial neural network (ANN) as a classifier for the ripeness detection while MATLAB was used as a simulation tool.  It classifies the leaves of the soya beans as being healthy leaf, ripe leaf, ripe but not ready for harvesting and diseased leaf of the soya beans by using the captured leaves of the plant. The approach adopted involves image acquisition, image pre-processing, feature extraction and classification using neural network.

While Abdulhamid et al., (2018) work involved using color and texture features to determine the ripeness of soya beans, Swetha and ashvant (2016) however used those features in a different method by detecting the severity of soya beans foliar diseases. The disease severity detection is done using 3 parameters (infection per region IPR, disease severity index, DSI and disease level parameter DLP) and classify the disease of soya beans using the leaves of the plant. The system development involves image acquisition, pre-processing and classification based on neural networks. Texture and color features are extracted which serves as an input to the neural network and the output are the different soya beans diseases. An accuracy of more than 90% is achieved in classifying soya beans foliar disease.

Mazen and Nashat (2019) expanded on the previous works by employing development of brown spots, and Tamura statistical texture in addition to color features to classify and grade banana fruit ripening stage. It involved the preparation of a four-class homemade database serving as input data which is then passed to an artificial neural network based framework. Comparing the performance of the proposed system with other techniques such as SVM, Naive Bayes, KNN, decision tree, and discriminant analysis classifiers revealed that the proposed system has the highest overall recognition rate, which is 97.75%. Among other techniques, the major signal processing steps involves: Image Acquisition, Image Pre-Processing, Extracting Banana Fruit from Background & Brown Spots, Extracting Texture Features and ultimately banana ripening classification based on the extracted features. An artificial neural network (ANN) is then used for ripening classification decision, which depends upon color and Tamura statistical texture features. An accuracy of 95.7% in the classification of the various soya beans leaves is achieved.

Sabanci and Aydin (2014) approach however classified olives by introducing sizes as an added feature to be used in classification after the color feature. The color and size serve as the input to the machine learning algorithm (Artificial Neural Network) carrying out the classification task which was simulated in a MATLAB environment.

**Justification**
As can be ascertained from the literatures reported, virtually all the research/experiments conducted in regard to ripeness classification of fruits/crops were done using artificial neural network (ANN) classifier. ANN has no embedded or inbuilt feature to enable it to perform tasks like image pre-processing, segmentation and feature extraction amongst others. Consequently, this makes all the experiments conducted using this classifier to take a colossal amount of time before the desired result is obtained. In view of this, a convolutional neural network classifier is proposed to mitigate the complexity of the approach used in building a model for ripeness classification.

**MATERIALS AND METHODS**
**Image Acquisition:**
During the image acquisition task 1,507 images of groundnut leaves of various categories ranging from matured and non-matured were captured using a digital camera with certain resolution (16.0 Megapixel) under uncontrolled environment. The captured images serve as inputs to the developed model. In this research work convulutional neural networks (ConvNets) as the machine learning technique was used to classify the leaves of groundnuts into matured and non-matured. Thus, this is a binary classification task of maturity detection that is with only two outputs.Training of the model started with few layers which initially showed underfitting. Hyperparametre is then

carried out. This involves tweaking the number of layers or number of units per layers until the developed model shows signs of overfitting. The images of the groundnuts used for this research work were captured from a farm located in a village called Badole, 4 kilometers west of Safana Local Government Area of Katsina state at around 4:45pm. The images were captured under uncontrolled environment, to obtain their

natural color without introducing any variation in intensity/brightness to the images. The research work started during the raining season (i.e 6 weeks after the groundnut is planted)and ended during the dry season, i.e. the time when the groundnut reaches optimum maturity stage(i.e the time when the crop is fully ready for harvest).

Out[184]: &lt;matplotlib.image.AxesImage at 0x1e8553b2c8&gt;
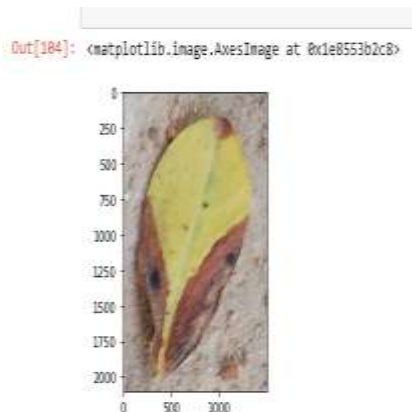


Figure 1: Mature groundnut leaf

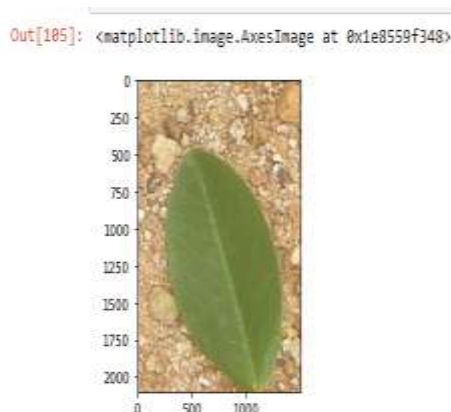Out[185]: &lt;matplotlib.image.AxesImage at 0x1e8559f348&gt;



Figure 2: Immature groundnut leaf

Figure 1 is a snapshot portraying the mature groundnut leaf from the pool of dataset. As can be patently seen, the image is purely yellow in color. Similarly, Figure 2 is a sample image portraying the immature groundnut leaf from the pool of dataset. As can be seen clearly, the image is purely green in color.

I.      **Feature Normalization:**This is carried out during the data preparation by dividing the pixel values (integers ranging from 0-255) by 255. The results in values between 0 and 1, enabling the model to converge faster. Normalization speeds up learning, hence, leading to faster convergence.

II.     **Feature Extraction:** Feature extraction involves the manipulation and transformation of input data into a set of features that uniquely identifies or represents an object (image) [5].

III.    **Image Classification:**In image classification pixels are classified based on certain criteria, that if a pixel certifies the criteria then it is selected, because the pixel represents the pattern we recognize. This task could be achieved with the aid of data trained by the classifier. However, in this research work CNN is chosen in the classification process. That is to classify the extracted features into Matured or non-matured groundnut. The principal objective of image classification is to predict the class of input image based on its features. There are a number of approaches used in classification, some of which are supervised (requiring training) while others are unsupervised which do not require training.

**Data Split**
Before training the model, the dataset was divided into training, validation and testing data. Usually, the model is trained on the training set and evaluated using the validation set. The model is tuned based its performance on the validation set. It is rather advised that the dataset is divided into three; training, validation and test sets (Chollet, 2018). This is to counter the problem of information leak. The dataset in this our project was divided into 80% and 20% for training and validation respectively. While on the other hand, 220 images were used for testing the overall performance of the model.

**Model development**
The model was built starting with two layers with varying number of filters. The number of filters and layers were gradually incremented until the model began to show sign of overfitting.

In the defined architecture, is the use of a sequential model which is a linear stack of layers. Compile method which is used in defining the loss functions, optimizers and metrics is also uniform. Binary_crossentropy is chosen as the loss function. Though it isn't the only available choice: others that could be used are mean_absolute_error, mean_squared_error etc. Binary cross entropy was chosen for this task as it was desired for the model to output probability to which it performs very well. Cross entropy is a quantity from the field of Information theory that measures the distance between probability distributions or, in this case, between the ground-truth distribution and one's predictions (Chollet, 2018). Adam which is an adaptive learning rate optimizer proved to perform better than others tried during the training. The other optimizers considered are stochastic gradient descent (SGD), RMSprop and Adagrad. The metrics being monitored during training was accuracy. This is used to evaluate the performance of the model on the validation

set. The final model built consists of 4 convolutional layers,  each   with   16,   32,   32   and   64   filters   respectively.
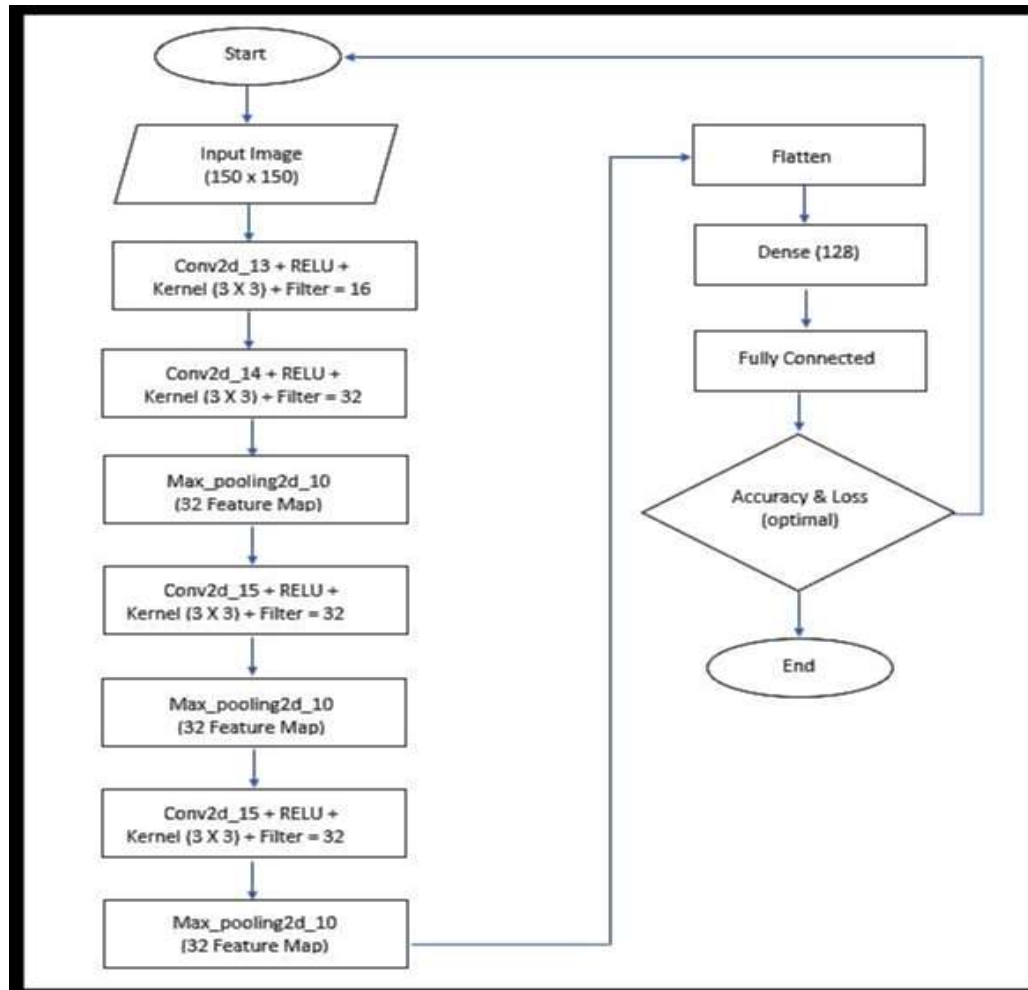


Figure 3: Model architecture

From figure 3 the output from the convolutional and pooling layers represent high-level features of the input image. The output of these layers is passed to a fully connected layer. This is done after a flattening operation is done via a flatten layer. Flattening involves converting the volume of feature map outputs from convolution and pooling operations into 1D vector so it can be served to the fully connected layer. Apart from classification, the fully connected layer also assists in learning non-linear features as activation function such as SoftMax is attached to it (Chollet, 2018). The value 150x150 and 3 represented number of pixels in the image and an image having 3 channels, i.e. Red Green Blue (RGB). Images were processed through convolution and pooling processes at the feature learning stage. This study applied 4 convolution layers.

The first 2D convolutional layer is responsible for capturing low level features, such as: gradient orientation, color, edges etc. but when 3 more layers were added, the network adapts to the high-level features as well, given us a robust network which has a wholesome understanding of the images in our dataset.  Same padding was added to each convolutional layer to keep the input and output feature maps of the same size and the Relu activation function to introduce non-linearity. In the defined architecture as illustrated in the above figure, max-pooling layers were periodically inserted in between successive convolutional blocks. Its basic function is to reduce the spatial size of the convoled feature in order to reduce the computational power required by the network to process the images.

**RESULTS**
**Data Visualization and Statistics**

As it is known beforehand that the dataset consists of two classes; Mature and Immature, it is important to know the number of images for each class. This statistics is important as it helps decides the next cause of action to be taken.
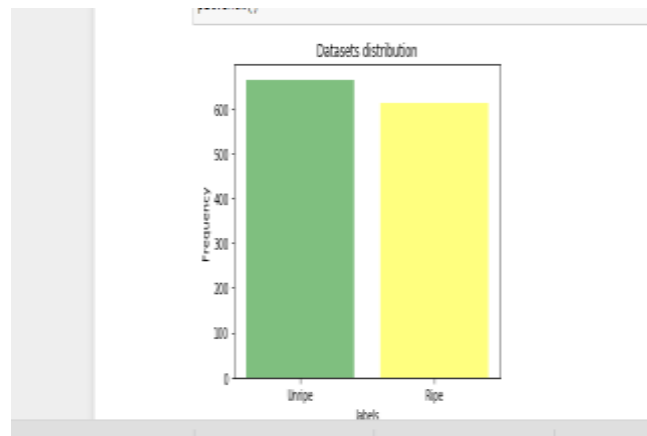


Figure 4: Snapshot of data statistics

As can be seen from figure 4, the there is almost equal number of images for both classes. This implies that, there would be no need to handle the problem of data imbalance.
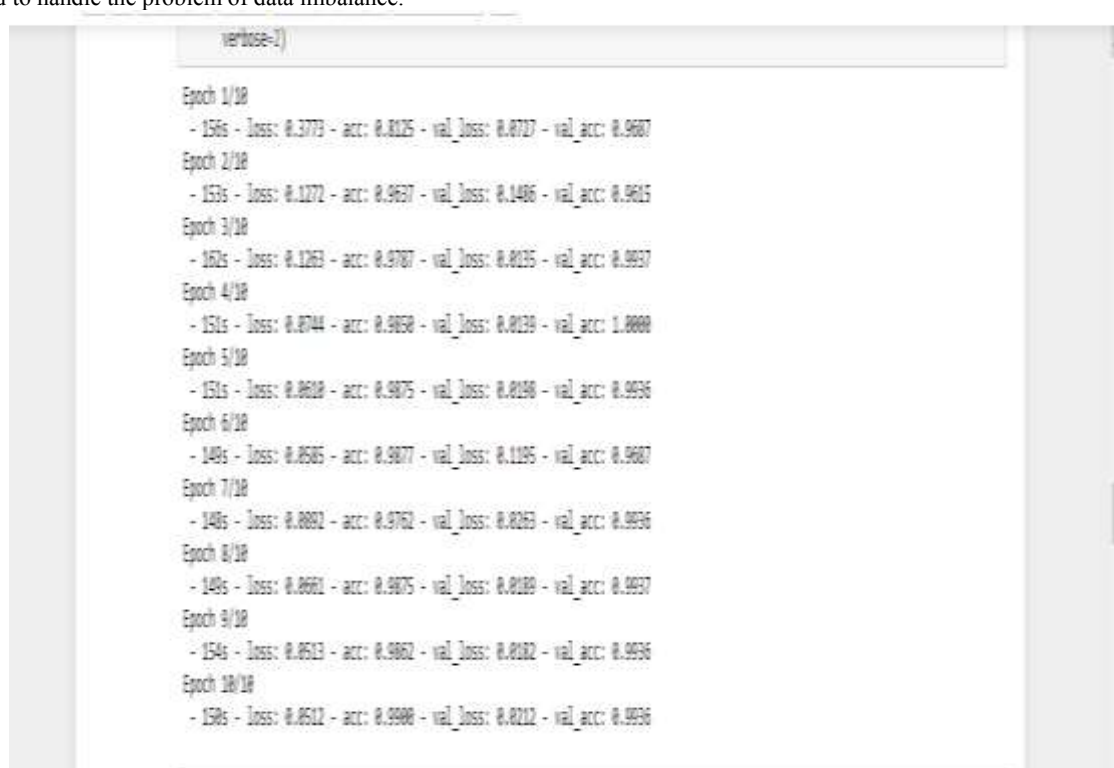


Figure: 5 Model training process

As portrayed in the figure 5 above, two quantities are often displayed during the training process, i.e. the loss of the network over the training data and its accuracy over the same training data. As can be seen in the figure our model was able to learn from the pool of dataset provided.
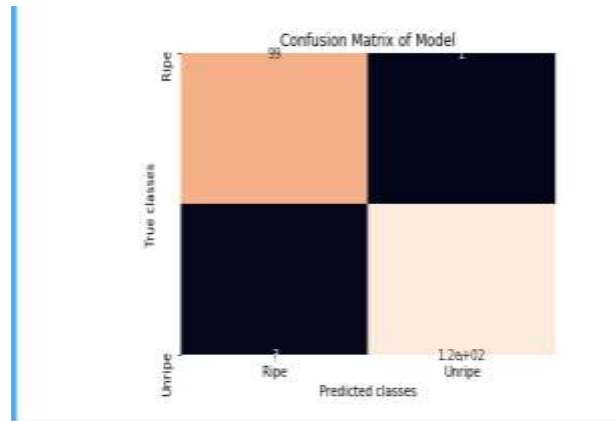
Figure 6:  Confusion Matrix

Figure 6 is a screenshot of the statistics of the images correctly classified and those that were not classified correctly.
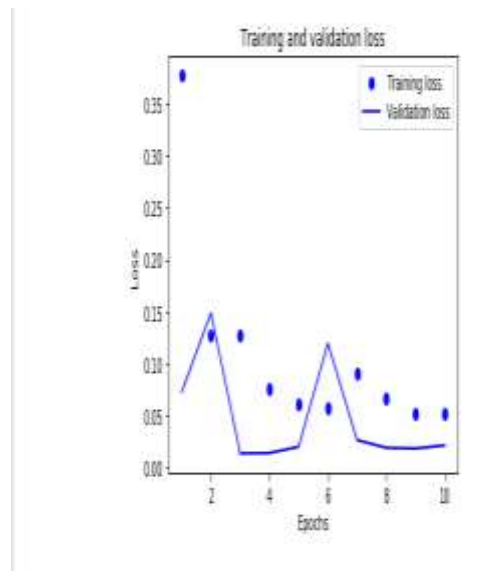


Figure 7: Graph of training against validation loss

Figure 7 is a graph showing the loss experienced during the validation and training of the obtained dataset. This indicates that, our network learns the pattern in the dataset provided. It shows robustness in learning these patterns.
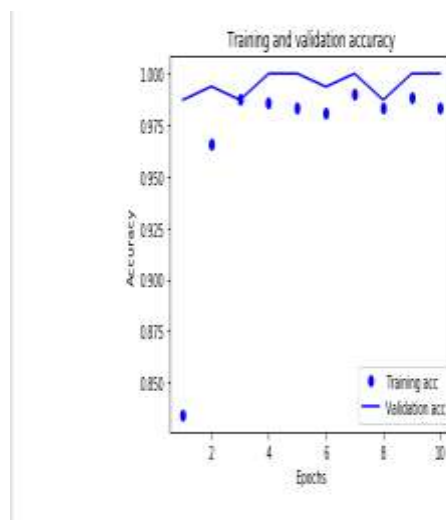
Figure 8: Graph of training against validation accuracy

Figure 8 is a graph depicting the training against validation accuracy.   Geometrically, from the graph it shows how intelligent the model is when faced with new sets of data that it has not seen previously.

**RECOMMENDATION/ FEATURE WORK**

It might be relevant to increase the number of labels being classified by the CNN model considering aspects such as white and grey matter rather than just the binary classification that has been carried out; mature and immature. The image analysis of leaves affected by diseases has not been taken care of. However, a further research could be undertaken to monitor and classify the various diseases that affect groundnut plant and provide suggestions for prompt action.

Finally the CNN model developed in this work could be deployed to web to allow remote access.

**CONCLUSION**

This Research main goal is to develop a CNN model for predicting the maturity of groundnut to high degree of accuracy. The developed CNN model had succeeded in classifying the images of groundnut leaves into mature and immature with 99.1% overall prediction accuracy when the model was tested with different images of groundnut leaves that were not used during the training process. The proposed model indicated that it was efficient and suitable for the prediction of groundnut maturity. This work provides a foundational basis to develop an online monitoring system for maturity detection of groundnut.

**REFERENCES**

Vibhute, A., & K. Bodhe, S. (2012). Applications of Image Processing in Agriculture: A Survey. *International Journal of Computer Applications*, *52*(2), 34–40. https://doi.org/10.5120/8176-1495

Abdulhamid, U., Aminu, M., & Daniel, S. (2018). Detection of Soya Beans Ripeness Using Image Processing Techniques and Artificial Neural Network. *Asian Journal of Physical and Chemical Sciences*, *5*(2), 1–9. https://doi.org/10.9734/ajopacs/2018/39653

Mustafa, N. B. A., Fuad, N. A., Ahmed, S. K., Abidin, A. A. Z.,

Ali, Z., Wong, B. Y., &Sharrif, Z. A. M. (2008). Image processing of an agriculture produce: Determination of size and ripeness of a banana. *Proceedings - International Symposium on Information Technology*. https://doi.org/10.1109/ITSIM.2008.4631636

Cilliers, A. J. Groundnut Production a Concise Guide. *Groundnut Production a Concise Guide*, 1–10.

Garima, T., (2014). Review on Color and Texture Feature Extraction Techniques. *Internal journal of Enhanced Research in Management and Computer Applications,* 3(5), pp. 77-81.

Chollet,F.(2018).Deep Learning with Python. 1-66. http://faculty.neu.edu.cn/yury/AAI/Textbook/Deep%20Learning%20with%20Python.pdf

Savakar, D. (2012). Identification and classification of bulk fruits image using artificial neural networks. *International Journal of Engineering and Innovative Technology*, *1*(3), 36–40.

Swetha, V. & Ram, A.(2016). ACLASSIFICATION AND SEVERITY DETECTION OF SOYABEAN FOLIAR DISEASES USING. (2016). II(Vii). International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE). 1-9. https://ijarmate.com

Ramakrishnan, M. &Sahaya, A.(2015). Groundnut leaf disease detection and classification by using back propagation algorithm. 1-5

Fadilah, N., Mohamad-Saleh, J., Halim, Z. A., Ibrahim, H., & Ali, S. S. S. (2012). Intelligent color vision system for ripeness classification of oil palm fresh fruit bunch. *Sensors (Switzerland)*, *12*(10), 14179–14195.

https://doi.org/10.3390/s121014179

Sabanci, K., & Aydin, C. (2014). *Using Image Processing and Artificial Neural Networks to Determine Classification Parameters of Olives*. *10*(3), 243–246.