2016

Vision-based model for maize leaf disease identification: a case study in Nyeri County

Maina, C. N.
Faculty of Information Technology (FIT)
Strathmore University

Vision-Based Model for Maize Leaf Disease Identification:

A Case Study in Nyeri County

Maina Njeri Christine

Master of Science in Information Technology [MSc.IT]

2016

Vision-Based Model for Maize Leaf Disease Identification: A Case Study in Nyeri County

Christine Njeri Maina

Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Technology at Strathmore University

Faculty of Information Technology
Strathmore University
Nairobi, Kenya

June, 2016

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Signature …….. Date ………………..

Christine Njeri Maina

047919

The thesis of Maina, Njeri Christine was reviewed and approved by the following:

Dr. Joseph Orero (PhD)
Faculty of Information Technology
Strathmore University

Dr. Joseph Orero (PhD)
Dean of Faculty of Information Technology
Strathmore University

Professor Ruth Kiraka
Dean, School of Graduate Studies
Strathmore University

Abstract

Biotic stress which includes pest and diseases affect crop productivity due to either death of affected crops or reduced yield per crop. Abiotic stress such as water and temperature also contribute to lower yields. Maize is Kenya's staple food with most households having limited choices of other foodstuffs thus increasing their reliance on maize. Diseases affecting maize in Kenya include: Maize Grey Leaf Spot disease, Maize stem borer, Maize Lethal Necrosis Disease, Ear Rot, Stem Borers, and Maize Streak Virus.

Currently, the human visual examination is the most commonly used method for classifying diseases. The method gives room for a lot of errors as the diagnosis is based on the experience of the farmer or the extension worker. The method also takes a great deal of effort and time to identify crop diseases based on the visually observable characteristics. Different experts diagnose the same disease as a different disease due to their varied experiences leading to erroneous identification of diseases. Introduction of artificial intelligence in various aspects of agriculture has gained momentum in today's world. Artificial intelligence has seen its application in predicting soil organic matter based on remote sensing data as well as in prediction of crop yield based on factors of production and in identification of crop diseases.

The research sought to propose use of an artificial intelligence model for identification of maize leaf diseases. In the proposed model, images of maize leaves were acquired and extracted color features used to identify the specific disease. Artificial Neural Network was used to identify the disease by implementing a back propagation learning algorithm. The data obtained was segmented into training and test data for the model. The algorithm was preferred due to its strengths in adaptive learning, its fast processing speed and the accuracy of its output.

The performance evaluation of the model was based on the accuracy of the classification, the precision, recall ratio and the F- Measure. The model was proven to be significantly accurate with an accuracy of 78.94 % while the precision obtained was 0.778. The recall ratio from the neural network was 1 and an F-measure of 0.875.

Dedication

I dedicate this research to the almighty God for the grace, the strength and the good health. To all the people who believed in me: my parents Beatrice and Samuel Maina, my sister Maryann Wanjiku, Aunt Alice, Aunt Hilda, cousins Rita and Tim and all my friends who did not cease to encourage me.

Acknowledgement

I would like to express my gratitude to Dr. Joseph Orero whose feedback and guidance went a long way in ensuring the completion of this project on time. I would also like to thank Prof. Ateya for the support and guidance as we learnt the loops of undertaking research. I would like to acknowledge my parents Beatrice and Samuel, my sister Maryann, Aunt Alice, Aunt Hilda who were supportive and who encouraged me when the going was tough. I also acknowledge the help of my colleagues Royford, Roselinda and Caroline Magiri whom I could lean on with my work. To Carolyne Wanja, a friend who gave me reason to keep on with the journey. To the Faculty of IT team for all the words of encouragement. Finally, I thank George, an extension worker who provided me with feedback that informed my research.

Abbreviations/ Acronyms

CMD	**-**	Cassava Mosaic Disease
FAO	**-**	Food and Agriculture Organization of the United Nations
HSI	-	Hue Saturated Images
KARI	**-**	Kenya Agricultural Research Institute
KJAR	**-**	Kenya Joint Assessment Report
GLS	**-**	Gray Leaf Spot Disease
MLND	**-**	Maize Lethal Necrosis Disease
MSV	**-**	Maize Streak Virus
PNN	-	Probabilistic Neural Networks
SVM	-	Support Vector Machine

Table of Contents

List of Figures

List of Tables

List of Equations

Chapter 1: Introduction

1.1 Background of Study

In Kenya, maize is highly relied upon as a source of labor, income and food (Short, Mulinge & Witwer, 2012). Maize is affected by various biotic and abiotic stress factors. Biotic stress includes fungi, bacteria, viruses, weeds and pests. Abiotic stress are caused by the surrounding environment, for example, water, temperature. There are various leaf diseases that affect maize which include: Maize Gray Leaf Spot disease, Maize stem borer, Maize Lethal Necrosis Disease, Ear Rot, Stem Borers, and Maize Streak Virus. The diseases affect both the quality and quantity of the maize yields. The leaf diseases that affect maize reduce the photosynthetic region leading to high loss in yields (KJAR, 2012).

Pests and diseases cause heavy losses through deaths, reduced productivity and loss of markets for products. Crop pests and diseases reduce yields substantially, sometimes by over 50 per cent or even total crop failure. Measures to prevent, control and eradicate diseases and pests in livestock and crops play a major role in improving productivity (Government of Kenya, 2010).

Human visual examination is the most commonly used method in identifying crop diseases. This method gives room for a lot of errors as diagnosis of the diseases is based on farmers or extension workers experience (Ghaiwat & Arora, 2014). It takes a great deal of effort and time to identify crop diseases based on their visually observable characteristics (Harnsomburana et al., 2011). In developing countries, farmers may have to go long distances to contact experts which makes consulting too expensive and is time consuming.

Computer vision technique has been studied and applied in agriculture due to its nature of accuracy and speed (Owomugisha, Quinn & Mwebaze, 2014). The technique has been applied in agriculture for quality inspection and sorting (Hasankhani & Navid, 2012). However a little attempt has been made using the techniques to detect diseases that affect maize. The delay and misdiagnosis in detection of disease in crops leads to losses in yields. Anything that results in lower yields in maize affects the food security in the country.

Computer vision has been applied in various aspects of foods and agricultural sector. It has been applied for distribution of pesticides based mainly on segmentation (Tellaeche, BurgosArtizzu, Pajares, Ribeiro, & Fernández-Quintanilla, 2008). It has been applied in the detection of defects in potatoes (Hassankhani R, 2012). Computer vision has also been used in: classification of grains based on their morphological characteristics, sorting and grading of fruits based on their color as well as measuring the quality of meat based on color and muscle (Gomes & Leta, 2012).

Vision-based detection of plant diseases is beneficial in monitoring large fields of crop and symptoms that appear on the plant leaves (Kanjalkar & Lokhande, 2013). Since symptoms are not obvious, it is difficult to distinguish the characteristics so the delay leads to late control of crop diseases. Images of leaf diseases are processed by using computer image processing technology and extraction of the spots according to colour, texture, type, size and number of lesions/spots. A disease that is diagnosed in time leads to the disease being prevented or controlled based on the circumstances of the crop.

The research aims at combining the benefits of computer vision and artificial neural networks in identifying the diseases that affect the leaves of the maize crop.

1.2 Problem Statement

Farmers and extension workers in rural areas employ the visual examination method to identify diseases affecting their crops. These farmers have limited knowledge of the diseases that stress their maize crops. However, changes observed on the leaf surface due to presence of lesions or streaks, changes on the stem, and stunted growth of the crops are some of the ways that inform farmers of the presence of diseases. The farmers use visual examination to identify infections on their crops. The visual examination method gives room for a lot of errors as the diagnosis is based on the experience of the farmer or the extension worker (Ghaiwat & Arora, 2014). The method also takes a great deal of effort and time to identify crop diseases based on the visually observable characteristics (Harnsomburana et al., 2011).

Vision-based model will aid in a faster method of identifying the diseases affecting maize leaf. The accuracy of the model will lower the error of misdiagnosis. As a result of this, there will be easier detection and management of the diseases. Artificial neural networks operate at a fast processing speed thus increasing the rate of obtaining the output. The algorithm is able to work with incomplete and inconsistent data thus improving the efficiency of the proposed model. The artificial neural network model has been proven to provide high levels of accuracy in identification (Koné Tadiou, 2013).

1.3 Research Objectives

i. To establish characteristics associated with identification of crop diseases

ii. To investigate the problems associated with the current methods applied in the recognition of maize diseases

iii. To review the existing models, mobile applications, techniques and architectures for disease identification in crops

iv. To develop a vision-based model for identifying maize leaf diseases

v. To validate the vision-based model

1.4 Research Questions

i. What characteristics are associated with the identification of crop diseases?

ii. What problems are associated with the current methods applied in recognition of maize diseases?

iii. What are the existing models, mobile applications and architectures used for crop disease identification?

iv. How will the vision-based model be designed?

v. How will the vision-based model be validated?

1.5 Justification

Visual examination involves a lot of effort, diagnosis of diseases is dependent on the experience of the farmer or extension worker and this often leads to inconsistent results. Loss in yield are experienced due to late detection of disease as farmers are sometimes forced to for extension workers to get to their fields to identify the problem affecting their crops (Ghaiwat & Arora, 2014).

Computer vision is applied in the field of agriculture because it provides accurate, non-destructive and consistent results. This research focused on mobile camera machine vision since it's cheap, has a rapid inspection rate, is simple and can be applied in broad areas. Artificial neural networks operates at a fast processing speed thus increasing the rate of obtaining the output. The back propagation algorithm is able to work with incomplete and inconsistent data thus improving the efficiency of the proposed model. The artificial neural network model has been proven to provide high levels of accuracy in identification.

1.6 Scope

The research lay its focus on two leaf diseases that greatly affect maize yield which in turn result to serious food insecurity. The diseases in focus were Maize Lethal Necrosis Disease (MLND) and Maize Streak Virus (MSV). Back Propagation Neural Network algorithm was used to identify the diseases affecting the maize leaves based on the pixel features obtained from the images of the leaf.

1.7 Limitations

Use of Unmanned Aerial Vehicles would have been more appropriate for image capture over large-scale maize farms. However, due to the complexity of their implementation, it was not discussed as part of the research. Another limitation of the research was that available maize disease datasets were not inclusive of the diseases under research. This thus led to the data obtained from the captured images being used for training and testing the neural network.

Chapter 2: Literature Review

2.1 Introduction

This chapter will lay it focus on discussing the contribution of agriculture to the livelihood of Kenyans, a brief on Nyeri County, the challenges that farmers face in the identifying the diseases that affect their crops. The research will provide more details on the diseases affecting maize and the symptoms of the diseases. The methods, models and mobile applications that are associated with the identification of the diseases will be discussed further. The chapter will also focus on computer vision and the different machine learning techniques that aid in the identification of diseases.

2.2 Agriculture in Kenya

Agriculture contributes 24 percent of national GDP directly and 27 percent indirectly making it the mainstay of Kenya economy. Agriculture is a means of livelihood for most Kenyans contributing income to more than 80 percent of the population (UNEP, 2015). The agricultural sector comprises of crops, livestock, fisheries, land, water, cooperatives, environment, regional development and forestry subsectors. The sector also includes the development of arid and semi-arid lands (Government of Kenya (GoK), 2010).

The agriculture sector is faced with many challenges including pests and diseases that affect the yield of the agricultural produce. Kenya's agriculture is mainly rain-fed making the sector vulnerable to fluctuations (UNEP, 2015). Pests and diseases cause heavy losses through deaths, reduced productivity and loss of markets for products. Crop pests and diseases reduce yields substantially, sometimes by over 50 per cent or even total crop failure. Measures to prevent, control and eradicate diseases and pests in livestock and crops play a major role in improving productivity (Government of Kenya, 2010)

Crop pests and diseases continue to lower the potential crop yields both pre- and post-harvest. Invasive pests like locusts, army worms and quelea birds are controlled by the Government. Other pests and diseases are controlled on the farms by farmers. However, pest and disease identification and management still poses a major challenge to most farmers, especially small- and medium-scale operators, due to high cost of pesticides and control equipment (Government of Kenya (GoK), 2010).

2.2.1 Nyeri and Agro-ecological Zones

Nyeri is located in agro-ecological zone II (AEZ II. This zone is characterized by steep terrains and receive rainfall all year around, save for one to two dry months. Consequently, they support crop cultivation. Nyeri County among other counties in the ecological zone produce large quantities of tea and coffee, among other crops such as maize and potatoes ("Nyeri County," n.d.).

Kenya's agro-ecological zones are as illustrated in Table 2.1

Table 2.1 Kenya Agro-Ecological Zones (Adimo, 2016)

Zone	**Appr. Area (km^2)**	**% Total**
I. Agro-Alphine	800	0.1
II. High Potential	53,000	9.2
III. Medium Potential	53,000	9.2
IV. Semi-Arid	48,200	8.5
V. Arid	300,000	52.9
VI. Very arid	112,000	19.8
Rest (waters etc)	15,600	2.6

2.3 Challenges Faced by Farmers in Identifying Crop Diseases

Farmers encounter a number of challenges while trying to understand the stress that affects their crops leading to low yields. (Ghaiwat & Arora, 2014b) assert that farmers tend to misdiagnose the diseases affecting their crops due to lack of knowledge. The diagnosis of the disease affecting the crop is based on the farmers experience with dealing with the various diseases. (Owuor, Wambui, Argwings-Kodhek, & Poulton, 2009) affirmed that farmers are forced to rely on the assistance of the in expertise of the extension workers in their area to identify the diseases. The number of extension workers Nyeri is limited and this leads to a delay in farmers receiving assistance when they need it. The extension workers also have limited technology to assist them in their day to day operations. In Nyeri, the division agricultural officers are equipped with Tablets which they use to browse the disease that is affecting the crop. Other extension workers lack these equipment and this rely on visual examination method as well as photo sheets that are provided to them. The method of visual examination by the extension workers is prone to errors and this leads to the farmer treating the wrong disease.

Plant clinics have been put up in Kenya where farmers can take their sample of infected crop for the plant doctor to diagnose the problem. The clinics may be quite far from the farmers and this discourages the farmers from accessing these results. It also takes a lot of time for the disease to be detected thus late management and control of the diseases affecting the crops.

2.4 Phenotypic Characteristics in Crops

Phenotypes are the observable characteristics of an organism("Phenotype | Define Phenotype at Dictionary.com," 2016). The characteristics include color, shape, size, frequency and distribution as well as spatial relationships, leaf angle, leaf rolling, leaf elongation. (Harnsomburana et al., 2011) describe some of the phenotypes as: Shape denotes the structures of leaves, fruits and substructures, *Size is described as s*mall, medium or large, *Frequency and Distribution* is described as the frequency of the lesions on the plant are described by words such as few, a few, moderate or many. Spatial relationships refer to the relativity of the phenotypes to other objects and the relativity to a particular plant. C*olor is also used to* characterize many plant traits including the leaf color, stem color, fruit color, and analysis of disease and description of mimic mutants on crops. Different mutants in crops produce different colored lesions. Necrotic lesions are brown in colour while chlorotic lesions are yellow in color. There are several color models that are used including RGB color model, HSI color model, YCbCr color model, HSV color model. The research made use of the RGB and HSI color models.

2.4.1 RGB Color Model

This color model specifies the coordinates system in which each color represents a single point. The scale of the red, green and blue component is between 0 and 255.

2.4.2 HSI Color Model

This model corresponds to the closest way that humans interpret and describe color. Hue is used to describe a pure color that also represents the dominant color perceived by observer. The intensity refers to the amount of white light added to hue. Saturation refers to the amount of white light added to hue.

2.5 Maize Farming in Kenya

According to **(**Short, Mulinge & Witwer, 2012), maize is an important food crop in Kenya that plays a key role in food security. Kenya produces at least 3 million metric tonnes of maize every year that comes from middle and large scale farmers. (ACDI/VOCA, 2015) acknowledges that inefficient production and marketing in the maize subsector contribute to economic stagnation and poverty in Kenya.

Maize is widely grown in Kenya and ranks highly in food security, meeting dietary preferences of many communities in Kenya. In 2011, the area under maize production was 2,131,887 Ha; realized yields were 37.5 million bags of dry maize and 4.6 million bags of green maize. Total maize crop was worth 87.8 billion. Hence any factor that threatens maize production becomes a national food security issue (KJAR, 2012).

Maize production is affected by both biotic and abiotic stress. Environmental conditions such as drought and soil fertility have had major effects on the yield. The infestation by weeds, pests and diseases lead to a decrease in the production of maize. These factors have often contributed to significant loss in the yield of the crop which has been estimated at 30% annually. Adoption of new technology has been low as well as use of certified seeds which has increased the rate of infections (KJAR, 2012). The research will focus on identification of three leaf disease that commonly affect maize in Kenya.

2.6 Diseases Affecting Maize

Maize is affected by various by different diseases which include Maize Grey Leaf Spot disease, Maize stem borer, Maize Lethal Necrosis Disease, Ear Rot, Stem Borers, Maize Streak Virus. The research focuses on three diseases that affect the leaves of the maize.

2.6.1 Maize Lethal Necrosis Disease

This disease is caused by Sugarcane Mosaic Virus (SCMV) and Maize Chlorotic Mottle Virus (Gekone et al., 2013). The disease was first observed in September 2011 in the low altitude area of Bomet. It later spread to neighboring high altitude areas (KJAR, 2012). The disease shows a number of symptoms which include:

- Yellowing and whitening at edges of leaves on young plant

- Dying leaf edges after ear formation
- Severe necrosis lesion on affected maize (Kessy, Bukalasa, Akonaay & Lema, 2013)
- Lesions on the leaf margins progress to the mid-rib resulting in the drying up of the whole leaf

The Maize Chlorotic Mottle Virus (MCMV) is transmitted mechanically by several insect vectors such as maize thrips, maize rootworms and leaf beetles. The Sugarcane Mosaic Virus is spread by maize aphids (KJAR, 2012). Figure 2.1 represents maize that is infected by Maize Lethal Necrosis Disease (MNLD).

Figure 2.1: MNLD Infected Maize (Kenya Joint Assessment Report, 2012)

. The disease was first detected in Kenya in I995.

2.6.2 Gray Leaf Spot Disease

The gray leaf spot disease is a fungal disease that greatly lowers the yield of maize (Kinyua, Smith, Kibata, Simon & Langat, 2010). The disease is caused by *C. Zeina* but there may be a few instances of *Sorghi Var Maydes*. The disease is characterized by

- Small tan spots often with a yellow halo that appear on the leaves
- The leaves develop pale brown or grey to tan long narrow streaks which result from elongation of the spots
- Dark greyish-brown rectangular lesions
- Under favorable conditions, lesions can coalesce to form large, irregular areas of dead tissue on the leaves.

The leaves of maize affected by gray leaf spot disease are as illustrated in Figure 2.2

Figure 2.2 Plantation affected by Gray Leaf Spot disease (Kinyua et al., 2010)

Gray leaf spot disease has been described as a disease of global importance. The disease was first detected in Kenya in I995.

2.6.3 Maize Streak Virus

Maize streak virus (MSV) disease is mainly transmitted by *Cicadulina mbila* (maize leaf hopper) but other leafhopper species such as *C. storeyi, C. arachidis and C. dabrowski* have also been found to transmit the virus. Leaf hopper have sucking mouth parts that enable them penetrate the plant cells by use of salivary and gut enzymes or mechanical force. Yield loss is caused by plant stunting and the termination of ear formation, development and grain filling in infected plants. Plants can die prematurely due to severe infection (Muriithi et. al, 2014).

Early disease symptoms begin within a week after infection and consist of very small, round, scattered spots in the youngest leaves. The number of spots increases as the plant grows and they enlarge parallel to the leaf veins. Development of a chlorosis with broken yellow streaks along the veins, contrasting with the dark green color of normal foliage is observed on fully elongated leaves. Maize that is severely affected experiences stunted growth as illustrated in Figure 2.3.

		
Healthy plants	Stunting of severely infected maize plant	Pale spots form on the leaf that become longer streaks that eventually coalesce.

Figure 2.3 Maize Streak Virus (Muriithi et al., 2014)

Once identified, the following immediate course of action can be undertaken. Plant maize in an open area to avoid shade as leafhoppers prefer shade. Minimize weeds that can harbor MSV vectors, particularly grasses. Chemical insecticides can be effective in control of the vector.

This research will focus on classifying three leaf diseases that affect maize. The research aimed at enabling the farmer to identify the disease that has affected the maize crops a lot more easily and provide a recommended course of action.

2.7 Current Techniques for Identification Crop Diseases

There are various techniques that have been applied in identification of crop diseases. Diseases can be detected directly through use of sensors, through visual examination and through use of applications.

2.7.1 Sensors

According to (Mahlein, 2016), the sensors discusses aid with identification and detection of crop diseases.

2.6.1.1 Thermal Sensors

This assesses the plant temperature and is related to co- related with plant water status. The sensor can detect thermal infrared range from 8 to 12 UM. The thermal sensors are useful in detecting changes in temperature due to pathogen infections. To ensure effective use of the sensors, the heterogeneity between and within leaves must be detected. The main temperature difference within single leaves, plant and crops are an indicator for appearance of plant diseases.

2.6.1.2 Fluorescence Imaging

This method is used to estimate the difference in the photosynthetic activity of plants. Chlorophyll is measured on the leaves and analysis is done based in the change of inflorescence parameter. Florescence imaging consists of an active sensor with an LED. The challenge with this method is that plants must follow strict protocol making it difficult to implement in normal green houses or field setting.

2.6.1.3 Multi and Hyperspectral Reflectance Sensors

These were the first sensors to be invented that are based on the special scale and type detector. They analyze the spectral information of the objects and across a wide band range. The sensor provide users with spectral and spatial information for the imaged object. They look at the optical properties of the leaves. The reflectance is complex as it depends on the leaf structure, internal scattering process, absorption of water by leaf and shortwave infrared. The methodology is used in detection of plant disease as well as monitoring fruit health and quality.

2.7.2 Human Visual Examination

The human visual examination is the most commonly used method of disease identification. This method involves deductively eliminating the possible causes of the diseases. Insect damage can be identified early since they are easily spotted on the plant or around the plant. If a thorough search is done, the next stage is to eliminate any abiotic disorder (environmental factors). The abiotic factors include temperature, moisture, soil pH, air quality, light regime, and nutrition (Monica Elliott, 2014). This method is implemented either by the farmer or the extension worker.

Human visual examination is considered a very tedious and ineffective way of leaf diseases detection. Currently, the identification of crop diseases is mostly dependent on manual recognition, but it can be erringly diagnosed by farmers because they usually judge the symptom by their experiences. The misclassifying leads to some erroneous control measurements such as desultory and untimely use of pesticides (Rothe & Kshirsagar, 2014).

2.7.3 Mobile Applications Used in Agriculture

Several mobile applications have been developed to assist farmers in their agricultural activities (Omolayo, 2015).

i. iCow

This application has a series of agri-products that farmers are able to access by dialing *285#. Farmers are receive messages from the application depending on the product that they choose. iCow also provides farmers with knowledge from experts. The applications runs on all mobile devices (Omolayo, 2015).

ii. vetAfrica

This mobile application is popular for diagnosis of livestock diseases and suggestion of suiTable medication for the farmer to apply. vetAfrica enables farmer to monitor and record animal data (Omolayo, 2015).

iii. KilimoSalama

KilimoSalama was created by Syngenta foundation and Safaricom Kenya. The application farmers with up-to-date and full climate data through text messages. The farmers receive updates on how on increase productivity, ensure food security and protect their crops from bad weather (Omolayo, 2015).

iv. M-Shamba

M-shamba has its accessibility on both smart and low end phones. SMS is used to provide information to a farmer based on production, harvesting, marketing, credit and weather. The application is customized based on the location of the user. Forums are available for farmers to share information with each other. The application is mainly used by farmers in Kenya to help them adapt to new technologies (Omolayo, 2015).

2.7.4 Particle Swarm Optimization (PSO) Model for Disease Pattern Detection on Leaf

This is an automatic unsupervised algorithm used for better segmentation of the image of the leaf. The method is simple to use and occupies bigger optimizing capability. PSO model faces challenges in segmenting scattering and non-coordinate problems (Muthukannan & Latha, 2015)

2.7.5 Active Contour Model

This model is used for segmentation of images. It has the ability to delineate an object outline from a noisy image. It provides a unified solution to several image processing problems. The model is however prone to oscillation. It incorporates edge information and thus ignoring the image characteristics. The model must also be initialized close to the region of interest to avoid being distracted by the noise and clutter (Rothe & Kshirsagar, 2014).

2.7.6 Computable Visually Observed Phenotype Ontological Framework

Data collection has had its own scope, terminology and descriptions that vary across domains by research group and individuals. Various ontologies have been developed to standardize the data. The characteristics used for identification include size, color, shape, frequency distribution and spatial relationships. The framework was developed based on various ontologies and is efficiently usable for organization, quantification and annotation of plant phenotypes. The framework is however limited since computer algorithms must exist that can measure the semantics of interest, the framework requires sufficient training examples, semantic labelling is a laborious task, optimal use of the framework would require extensive time, effort and resources from the plant community to decide upon the species and phenotype images (Harnsomburana et al., 2011).

2.8 Computer Vision Systems

Computer vision has been applied in agriculture for various things: quality inspection example, qualitative sorting of potatoes by color analysis machine (Hasankhani & Navid, 2012). They have also been used for evaluation purposes as they provide suitably rapid, economic and consistent objective assessment (Brosnan & Sun, 2004). Owomugisha et al., (2014), describes machine vision as having superior speed and accuracy which has significantly lead to its application in crop disease detection. According to Owomugisha et al., (2014), computer vision is an area of interest for most researchers. The rising opportunity is the need to bring mobility and flexibility to the already developed model. Computer vision is not limited by physical, personal and environmental factors making it more effective (Kanjalkar & Lokhande, 2013).

2.8.1 Computer Vision Implementation

Computer vision algorithms are implemented through various stages. The images that are obtained are processed to establish the quantity of the area affected by the disease, the colour of the affected area, the boundary of the affected area and the texture of the area that is affected (Gavhale & Gawande, 2014)

The step by step description of the plants disease detection is as discussed by (Shire, Jawarkar & Manmode, 2015) in Figure 2.4.

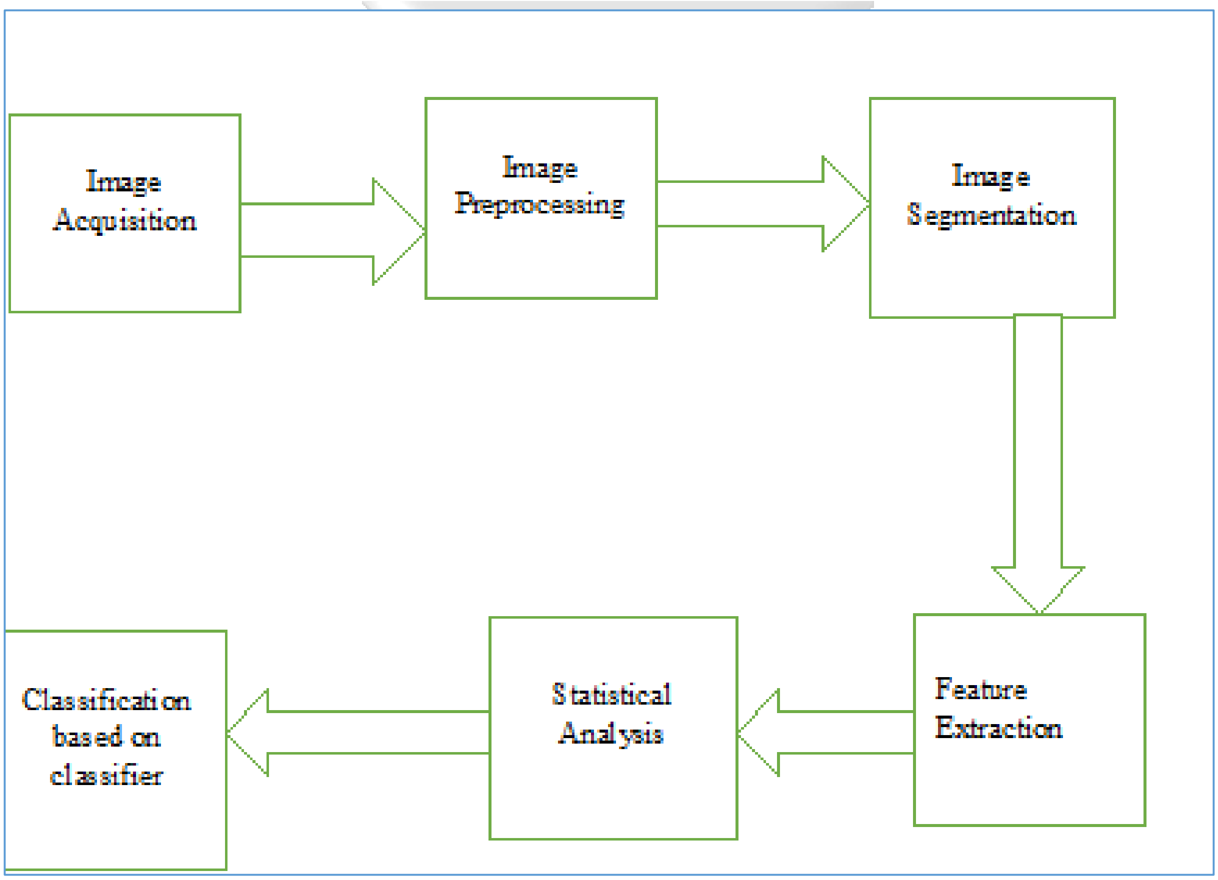


Figure 2.4 Disease detection algorithm (Shire, Jawarkar & Manmode, 2015)

Image Acquisition- The digital images are acquired from the environment. Can be done through the use of a cameras or sensors.

Color Transformation- RGB images are converted to a standard color space which is Hue Saturated Images (HSI). HSI is popular as it is based on human perception. The color features that are involved include: the skewness of the image, the mean, the standard deviation, and kurtosis. M represents are the dimension of the image. P_{ij} are the values of the color on the i^{th} and j^{th} columns (Kadir, Nugroho, Susanto, & Santosa, 2013).

Image Segmentation- Used to distinguish objects from their backgrounds or to partition their images to related portions. Image segmentation simplifies representation into something that can be understood. Segmentation involve removing the image from the background. The adaptive threshold has been proven to work. An intensity histogram was built consisting of 20 major bins. Two peaks that represented the leaf and background were built. The least values lying between the background and the leaf were obtained and used as the threshold to separate the leaf from its background(Kadir et al., 2013).

According to Gavhale & Gawande (2014) the techniques of image segmentation include:

a) Region based – This technique involves grouping of related pixels. The boundaries of the area are then identified for segmentation and at least one pixel related to the particular region is considered. The edge flow is the converted into a sector and other edges detected for segmentation.
b) Edge based – In this technique, the boundary to segment is identified. The edges help with identification of discontinuities in the image. Support vector machine is used for the classification.
c) Threshold based – The segmentation is done based on the value that are obtained from the histogram. Of the edges from the original image. Accurate edge detections result in accurate threshold. The advantages of this method is that it involves fewer computations compared to other methods. However it is not suiTable for complex images.
d) Feature based clustering – This method involves conversion of images into histogram upon which clustering is done. The pixels of the image are clustered using the Fuzzy c technique images.
e) Model based – referred Markov Random Field. It works in combination with edge detection to accurately identify the edge.

Feature Extraction- This is the process of simplifying the amount of resources required to describe a large data set accurately. It can also be viewed as transforming images into features. In cases of leaf disease identification, the features would be selected according to the diseases thus more discriminative. Feature extraction involve color, shape and texture. Gavhale & Gawande, (2014) describe texture as the main point of focus for most researchers in leaves disease detection and detail the process as

Texture Analysis Methods- Textures are patterns of non-uniform spatial-distribution of differing image intensities mainly focusing on the individual pixels making an image. The features of texture include uniformity, regularity, density, linearity, directionality, roughness, coarseness and frequency. The features are characterized through:

a) Statistical- includes gray-level histogram, grey level, concurrence matrix and run length matrix
b) Structural- presumes texture as a combination following the two-dimensional transform and Gabor transform.
c) Fractals- These model the quality of roughness and self-similarity
d) Signal- processing- suiTable for analysis of wavelets

Texture Feature Extraction Methods- Extraction involves color co-occurrence method which encompasses relationship between pixel pairs in the image. It computes contrast, correlation, energy, entropy and homogeneity. It is applied on different color space. Gabor filter consists of parameters such as radial center frequency, standard deviation and orientation. Since signal processing methods result in large feature size, they require to be downsized to prevent dimensionality issues. The principal component analysis is used to downsize the feature space. Another method in used is the wavelet transform which makes use of the frequency domain. They this method is a bit slow, the wavelets produce features with a higher frequency.

Classification- This is the process of identifying and labelling the pathology that is affecting the plant (Barbedo, 2013). Different machine learning classifiers can be used for classification of the diseases. These include K- Means, Artificial Neural Networks, Support Vector Machines, Probabilistic Neural Network, and Fuzzy Logic

2.9 Machine Learning Identification Techniques for Classifying Crop Diseases

Machines make use of algorithms to quantify patterns (Witten and Frank 2005). According to Mitchell (as cited in Behmann, Mahlein, Rumpf, Ro¨mer & Plu¨mer, 2014), a machine is able to improve in performance in a specific task T given its experience E. Buxton and Breiman (as cited in Behmann et al., 2014) in their study concluded that machine learning allows important data and patterns to be extracted from raw data even if the model is unknown.

Unsupervised learning involves the ability to learn and organize information without providing an error signal to evaluate the potential solution (Sathya & Abraham, 2013). Examples of unsupervised learning method is the K-Means clustering.

Supervised learning involves the process of training a data sample from data source with correct identification already assigned (Sathya & Abraham, 2013). Supervised learning is used to solve identification problems Examples of supervised learning algorithm include K-Nearest Neighbor, Bayesian Learning, decision trees, Rule based learning, support vector machines, Neural Networks and Model ensembles. The research will use supervised learning to classify the maize leaf diseases. The algorithm applied in the research was Artificial Neural Network.

2.9.1 Artificial Neural Networks

(Koné Tadiou, 2013) described a Neural Network (NN) as a system that is comprised of several *artificial* neurons and *weighted* links binding them. The artificial neurons processing the information are organized into interconnected layers along chosen *patterns*. Every neuron in its layer, receives some type of stimuli as input, processes it and sends through its related links an output to neighboring neurons. The networks heavily rely on learning in order to adapt to their environments. The neural network is composed of four main sections:

- Input which is a node that activates upon receiving a trigger from incoming signals
- Interconnections between nodes
- An activation function (rule) which transforms inside a node, input into output
- An optional learning function for managing weights of input-output pairs

Figure 2.5 illustrates the architecture of an artificial neural network.

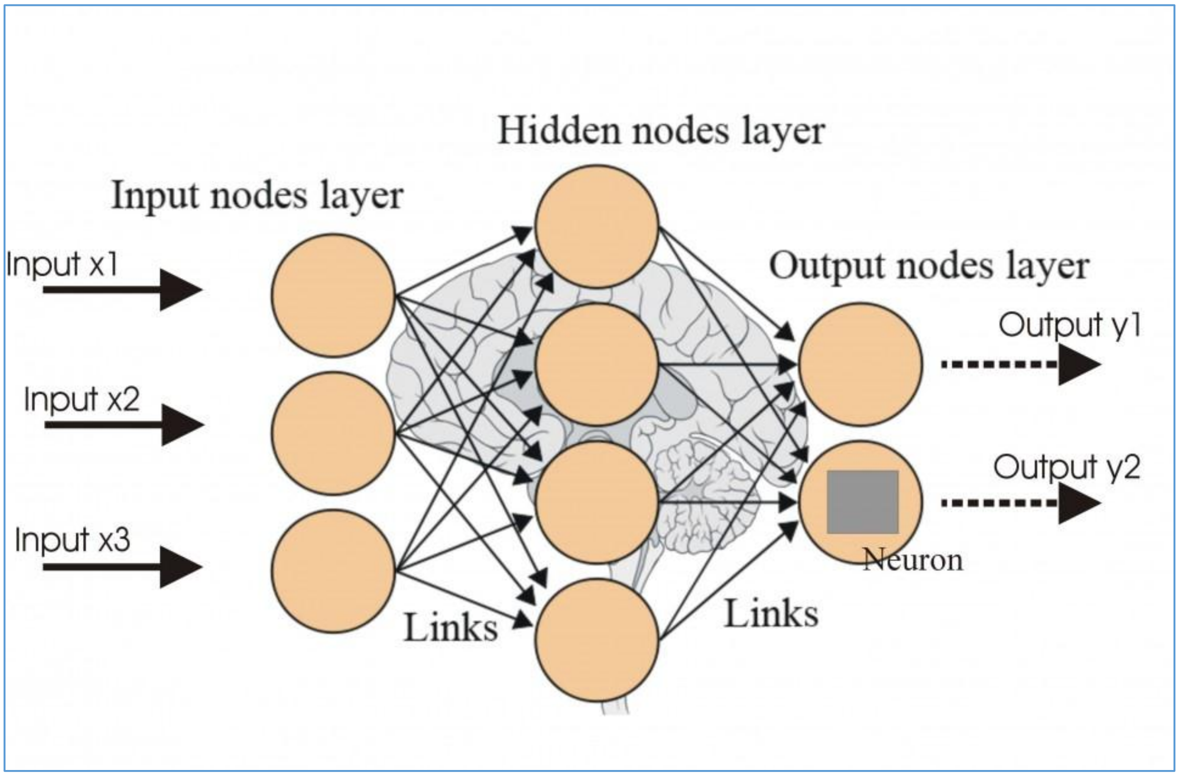


Figure 2.5: Basic structure of an Artificial Neural Network (ANN) (Koné Tadiou, 2013)

Neural networks are designed to learn through inductive learning. Once the network is initialized, it can be modified to improve its performance through learning.

2.9.2 Architectures and Methods of Artificial Neural Networks

Artificial Neural Networks has a number of structures that are designed to deal with abstract and poorly defined problems**.** Artificial Neural Network has several classes including the *feed-forward*, the *feedback* and the *back propagation* neural networks, *probabilistic* neural networks. The research will employ the back propagation neural network for classification of the diseases affecting the maize leaf.

2.9.3 Back Propagation Neural Network

Back propagation neural network is a supervised learning method. The process of backpropagation is carried out in two phases (Devi, Reddy, Kumar, Reddy, & Nayak, 2012):

i. *Propagation*

Forward propagation- The training patterns are propagated forward in order to generate propagation output activation.

Backward propagation of the output activation through the neural network using the training patterns target in order to generate the delta of all the outputs and hidden neurons.

ii. *Weight update*

This involves multiplying data and input activation to get the gradient of the weight. The weights are brought in the opposite direction of the gradient by subtracting a ratio from the weight. This ratio influences the speed and quality of learning which is referred to as the learning rate. The sign of the gradient of a weight indicates whether the error is increasing as a consequence of being updated in the opposite direction.

iii. *How Back Propagation Neural Network works*

It is initialized by setting all its weight to small random numbers mainly between -1 and +1. The forward pass involve the input pattern being applied and the output calculated. The computation gives an output that is different from the target. The error of each neuron is calculated by deducting the actual output from the target. The computer error is used to change the weights. In the reverse pass, the output of each neuron gets closer to the target (Devi, Reddy, Kumar, Reddy, & Nayak, 2012).

2.9.4 Applications of Artificial Neural Networks

Artificial neural networks are preferred because:

- Useful in pattern recognition, identification, generalization, abstraction and interpretation of incomplete and noisy inputs. Example of areas of application include handwriting recognition, image recognition, and voice and speech recognition, weather forecasting.
- Useful in financial applications as it provides some human characteristics to problem solving that are difficult to simulate using the logical, analytical techniques of expert systems and standard software technologies.

2.9.5 Limitations of Artificial Neural Networks

Artificial Neural Networks do not produce an explicit model making it difficult to justify result (Ghaiwat & Arora, 2014).

2.9.6 Applications in Research

The artificial neural network algorithm was selected for use in the research due to the following reasons

- Artificial Neural Networks have the ability to solve new kinds of problems. ANNs are particularly effective at solving problems whose solutions are difficult, if not impossible, to define.
- Robustness- They have the ability to cope with incomplete or fuzzy data. ANNs can be very tolerant of faults when properly implemented.
- Fast processing speed- Since they consist of a large number of massively interconnected processing units, all operating in parallel on the same problem, ANNs can potentially operate at considerable speed when implemented on parallel processors.
- Flexibility and ease of maintenance. ANNs are very flexible in adapting their behavior to new and changing environments. They are also easier to maintain, with some having the ability to learn from experience to improve their own performance.

2.10 Conceptual Model

Figure 2.6 demonstrates how the proposed model works. The farmers/extension worker makes use of their mobile phones to take the images of the maize leaf. Pixel features of the images will be used as the input for the neural network model. The neural network model will be trained and validated to work. The maize leaf disease identification denotes the output of the model.

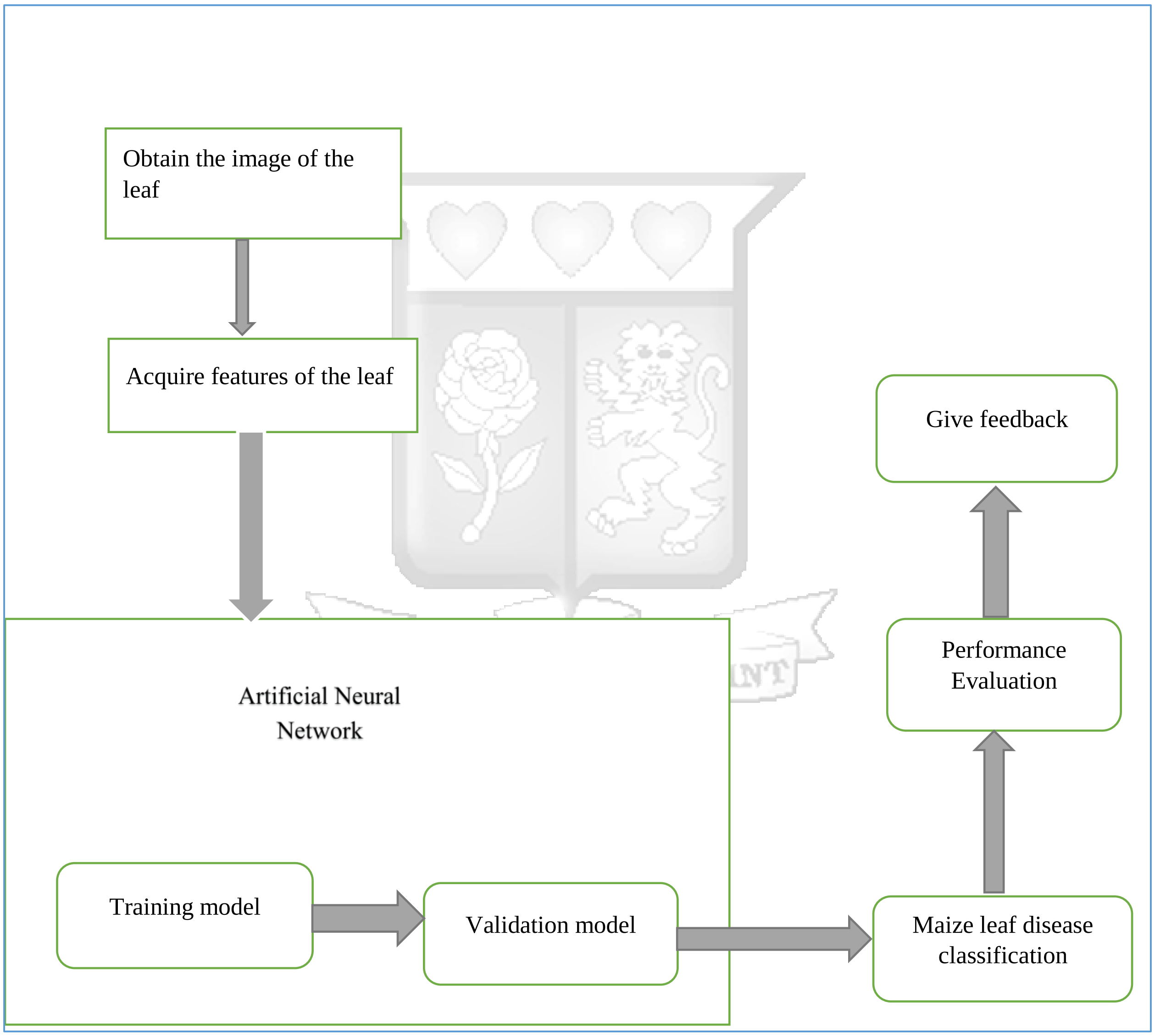


Figure 2.6 Conceptual Design of the Proposed Model

Chapter 3: Research Methodology

3.1 Introduction

Research Methodology is defined as the process of systematically solving problems. It can be considered as the science of doing research (Bhatnagar &Singh, 2013). The research will be guided by the objectives that the author is proposing to meet at the end of the research. It will be greatly informed by the nature of the problem being studied, research designs that have been used in the related work reviewed in chapter 2. Of importance in the research design will be the population that will be studied, the sample that will be selected based on a certain criteria and the method that will be used to obtain the data. A classifier for categorizing the data will also be identified and the validity of the classifier determined.

The research employed an applied approach with the use of IT test bed implementing Machine Learning Neural Networks. Primary and secondary data was used to facilitate the research. Secondary data was obtained from KALRO fact sheets as well as from the Plant wise website. The primary data, image pixel features, was obtained from the images of the maize leaf that were captured.

3.2 Research Site

Maize being a major food crop in Kenya is grown by small scale farmers, middle scale and large scale farmers. Taking into consideration the time constraints and reach of the regions that grow maize, the research will lay its focus on Nyeri County. Based on ASDP (2014) productivity of crops, maize returns the highest yield in the region. The county's proximity to the user also informs the decision of selecting the region.

Majority of the population lives in the rural area and engages in farming of the major food crops that are favored by the climate on Nyeri County. Extension services as well as education are provided to the farmers through churches and other social gatherings. Most farmers do not go for the said meetings and end up with losses due to lack of information on old and new diseases affecting their crops.

3.3 Research Design

The research was aimed at developing a neural network model that would aid in identifying the leaf diseases that affect maize. In a related research on automated diagnosis of Banana Bacterial Wilt Disease and Black Sigatoka Disease, Owomugisha et al. (2014), effectively use acquisition of images, feature extraction, disease classification and calculation of the classification performance.

The prototype was developed using rapid application development (RAD) methodology. This methodology is appropriate due to its shorter development cycle compared to other methodologies and was thus suitable for the limited duration of time.

3.3.1 Data Acquisition Methods

Various data collection tools were made use of in the research. Non-participant observation was used for obtaining the images of the leaves of the maize. Mobile phone camera was used to capture the image of the maize leaves. In recognition of the leaf diseases, images were captured and features of the images extracted. Pixel values of the image were obtained for use in training and testing of the model. Classification of the diseases affecting the maize was done using a back propagation algorithm. The advantage of this method is that there is no bias in the output that is obtained.

Interview as a tool of collecting data was used to obtain information from the farmer and extension worker on the awareness of the disease, the measures they take and the challenges they face. The method was appropriate as it enabled the farmer and the extension worker to highlight things that may have been left out when non-participant observation is used.

Internet sources were used to gather data on related information to the researcher's area of study. This was useful in identifying the gaps that would be filled in by the research.

3.3.2 Sample Split

This process involved identifying the data that would be used in training model, the test data to test model training and the validation data used to measure the output error. The sample taken was representative of the identified population.

3.3.3 Model Training

Model training involved providing inputs to the model for processing in order to train the model on the type of input data and the expected output of the training session. The training data was fed into the neural network model through the identified model neurons. A number of iterations were undertaken during the training process with each iteration being aimed at reducing the error rate and adjusting the input weights.

3.3.4 Model Testing and Validation

The process involved the use of test data to check whether the system was properly trained by observing the actual model output versus the expected output. In using the validation data set, the disparities in the output captured by error performance measures were used to adjust the weights of the neuron for the purpose of fine tuning the model.

3.3.5 Presentation of Output

Tables and graphical representations were used to illustrate the model outputs. The tables were used to display the accuracy, precision as well the recall ratio that was obtained during the classification. Graphical representation were used to display the model that was obtained from the implementation.

3. 4 Population

At least two out of every three farms are involved in maize production in Kenya (Brooks et al, 2009). Farmers are the first people to come face to face with different diseases, plants and insects that affect their production. The extension workers also play a major role in identifying biotic as well as abiotic stress facing the crops. The population used in the research constituted of the maize leaves images that were captured as well as images obtained from the KALRO. The population consisted of the images of healthy maize, MLND infected leaves and MSV infected leaves.

3.5 Sampling

The research will make use of stratified sampling in which the image population will be divided into strata (Bordens & Abbot, 2011). One strata will consist of the images that will represent healthy leaves, the second sample will be comprised of MLND infected leaves images. The third strata will consist of MSV infected leaves images. Different split ratios will then be applied to the images feature sets to determine the training set as well as the test set. The split ration that produces the highest accuracy will be used in the implementation

$$n_h = (N_h / N) * n$$

Eq. (3.1)

n_h represents the total sample of the strata h, N_h is the population size of the stratum, N represents the total population size and n is the total sample size. Based on a similar research that has been done in 2 banana leaves disease in Uganda, the proposed sample will consist of 100 images (Owomugisha et al., 2014).

A farmer will be interviewed to determine the factors they consider to identify the diseases affecting their maize. An extension worker will also be interviewed to determine the factors they put in place as well as to test the usability of the system.

3.6 Research Quality

In their research of disease plant leaves using artificial neural networks, Muthukannan, Latha, Selvi & Nisha (2015) applied the following measures to carry out their performance evaluation. Accuracy, Precision, Recall ratio and F- Measure will be used to carry out the evaluation.

3.6.1 Accuracy

The accuracy (AC) is defined as the proportion of the total number of predictions that were correct. *tp* represents the true positive, *tn* represents the true negative, f*p* represents the false positive and *fn* represents the false negative in the equations that were used to measure performance. Eq. 3.2 below was used to determine accuracy

$$\text{Accuracy (AC)} = \frac{tp+tn}{tp+tn+fp+fn}$$

Eq, (3. 2)

3.6.2 Recall ratio

The recall or true positive rate (TP) is the proportion of positive cases that were correctly identified as shown in the Eq 3.3

$$\text{Recall ratio} = \frac{tp}{tp+fp}$$

Eq, (3. 3)

3.6.3 Precision

Precision (P) is the proportion of the predicted positive cases that were correct, as computed using the Eq, 3.4

$$\text{Precision (P)} = \frac{fp}{fn+fp}$$

Eq, (3.4)

3.6.4 F_Measure

The F-measure computes some average of the information retrieval precision and recall metrics.

3.7 Ethical Considerations

The researcher ensured that consent was obtained from the participants before she began her survey. This ensured that both parties were aware of what was happening at any one particular time during the process. The researcher maintained confidentiality of the data obtained and the personal details of the respondents. Data obtained was used for the sole purpose of the research. The researcher cited work obtained from other authors giving them the credit that was due to them.

.

Chapter 4: System Design and Architecture

4.1 Introduction

This section of the paper details the design structure of the proposed solution by incorporating the various requirements collected in the previous chapter through the various interactions with the potential users and experts. To achieve this, design diagrams under the Unified Modelling Language were drawn and detailed information for each design diagram put down. Design diagrams and structures put down for the purpose of the model included a use case diagram with detailed follow-up use case descriptions, System Sequence Diagrams, and an activity diagram.

4.2 Requirements Analysis

Based on the objectives as well as the user requirements, this section outlines the various requirements to be met in the research.

4.2.1 Functional Requirements

i. The application should accept the uploaded image in the form of image only. The image should be in the form of .JPG, .TIFF, .PNG or .GIF. Files of different formats should be rejected.
ii. The application should extract the pixel features of the images presented to it.
iii. The application should classify the diseases affecting the leaves of the maize using back propagation neural network
iv. The application should return a correct classification of the disease based on the pixel features of the images presented to it.
v. The application should provide a recommendation to the farmer on the course of action they should take.
vi. The recommendations should be based on factsheets from Kenya Agricultural Livestock Research Organization (KALRO).

4.2.2 Usability Requirements

Since the research site of this research is based on Nyeri County which is considered a rural area, the application should be simple and easy to use.

4.2.3 Reliability Requirements

i. A backup of the entire application should be provided regularly.

ii. In the event of failure, the administrator should be able to restore the system

iii. The application should have the ability to convert the images provided to it into pixel values

4.2.4 Supportability Requirements

The application should run on a standard smart phone without a need to change any settings of the phone.

4.3 System Architecture

Figure 4.1 gives an illustration of the system architecture. The process begins by capture of the leaf images of the maize by the farmer. The image is then converted to its pixel values by the application. The pixel values are considered as the independent variables. The value of the RGB and alpha are normalized in order to handle the variation of the values obtained by the cameras of varying resolutions. The pixel values are converted to Hue Saturation Intensity (HSI) values. The HSI values are then used by the application to train and test the neural network specifically the back propagation neural network. The back propagation neural network classifies the disease and gives a prescription to the farmer to enable him avert the disease affecting the crop in the least amount of time.

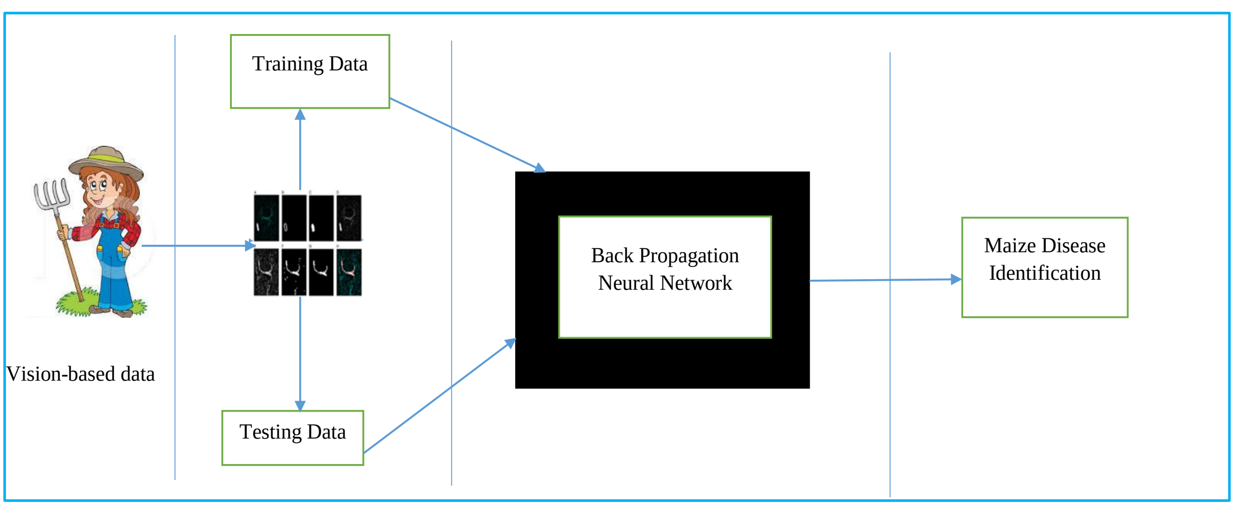


Figure 4.1 System Architecture

4.4 Diagrammatic Representation of Model

4.4.1 Use Case Diagram

The use case diagram described in Figure 4.2 describes the user's interaction with the system. Use case diagram comprises of the actors (something with a behaviour or a role) who in the research were the System users, Boundary which represents the limits to which the system operates and the use case which are a collection of success or failure scenarios. The scope which is represented by the boundary is identification of the leaf diseases that affect maize.

Use Case Description

Use Case: Obtain Image of the Maize Leaf

Primary Actor

Farmer

Precondition

Mobile phone has a camera

Post condition

The image obtained belongs to the leaf of a maize crop

Main Success Scenario

Actor Intention	System Responsibility
1. Farmer prepares the camera settings	
	2. Capture the image of the leaf
	3. Save Image of the leaf
4. View image obtained	
5. Exit camera mode	

Extensions

At any time that the system fails to capture the image

Restart the camera

Confirm the camera setting are correct

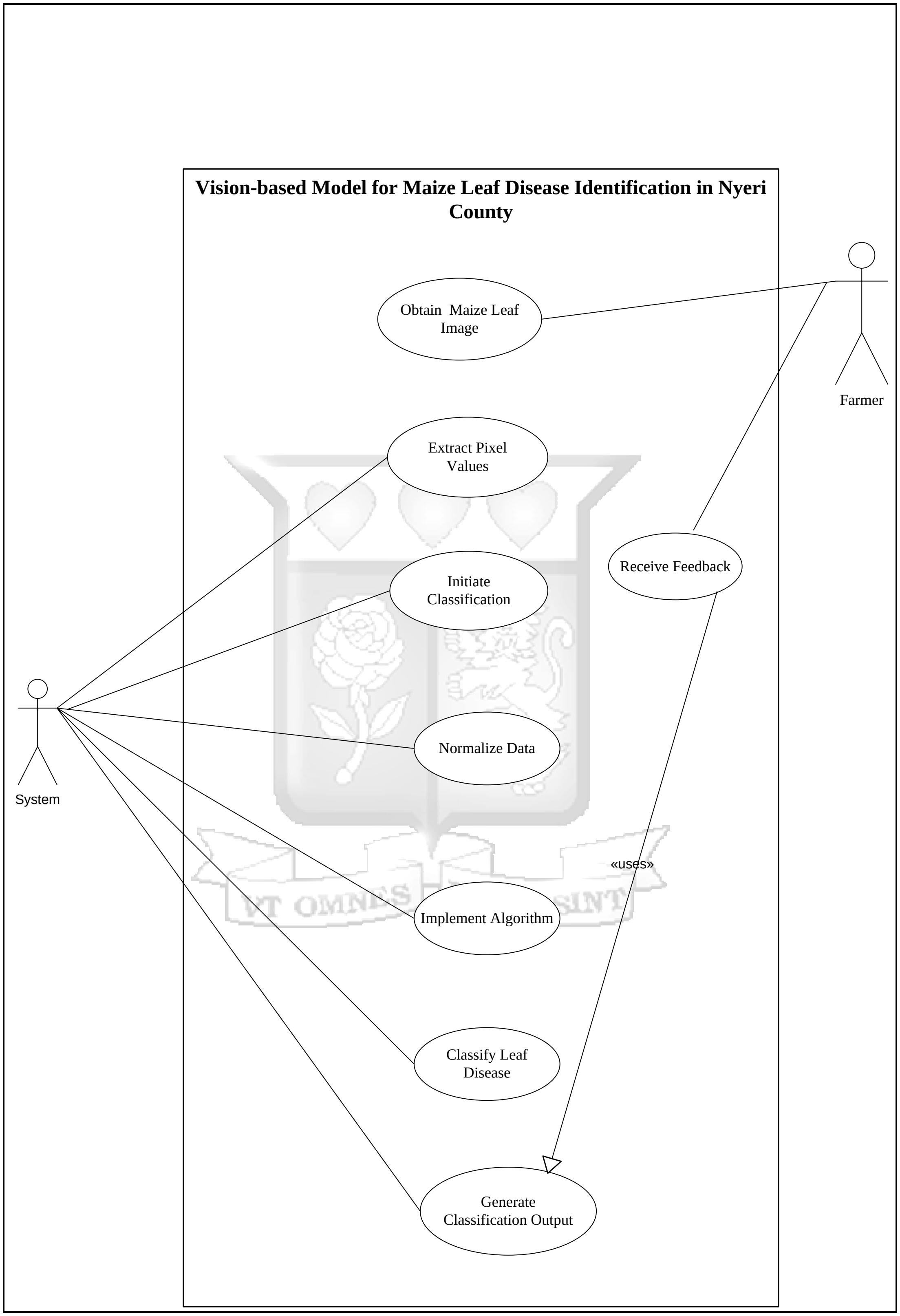


Figure 4.2 Use Case Diagram

Use Case Description

Use Case: Obtain Image of the Maize Leaf

Primary Actor
Farmer

Precondition
Mobile phone has a camera

Post condition
The image obtained belongs to the leaf of a maize crop

Main Success Scenario

Actor Intention	System Responsibility
1. Farmer prepares the camera settings	
	2. Capture the image of the leaf
	3. Save Image of the leaf
4. View image obtained	
5. Exit camera mode	

Extensions
At any time that the system fails to capture the image
Restart the camera
Confirm the camera setting are correct

Use Case Description: Extract Pixel Values, Initiate Classification, Normalize Data, Implement Algorithm, Classify Leaf Disease, Generate Classification Output

Primary Actor

System

Preconditions

Images of the maize leaf were successfully captured

Post Condition

Disease affecting the leaf was accurately classified

Main Success Scenario

The system is associated with several use cases involved in the process of identification of the leaf diseases that affect maize.

Extracting pixel values- the model obtains the binary values of the leaf images that were captured using the mobile phone camera. The values that were extracted included the red, green and red components of the two dimension maize leaf image that was captured.

Initialize classification- This use case demonstrates the process of identifying the attributes that would be used as the inputs for the neural network as well as the number of hidden layers that would be used. The expected outputs were also identified during the initiation function. The activation function used was also identified.

Normalization of the data

The data used in the process varied in the range of values obtained from the leaf. The noise in the data was removed to allow for more accurate predictions.

Implement the algorithm- The system implemented the back propagation neural network algorithm. This algorithm was used due to its proven ability to learn with any kind of data presented to it, its adaptability and highly accurate classification output at a fast speed.

Classify leaf disease- This process was effected by providing to the system the set of training data from which it was expected to learn the environment and expectations. A test set was then presented to the system to validate that the system was classifying correctly.

Generating classification output- The system presented the classification of the disease based on the new images provided to it.

Use Case: Receive Feedback

Primary Actor

Farmer

Precondition

Successful classification of the disease
by the system

Post condition

Recommended course of action

Main Success Scenario

Actor Intention	System Responsibility
1. Request for disease identification	
	2. Return the output of classification
3. View feedback and course of action	
	4. Display time taken and the level of accuracy
5. Exit the system	

Extensions

At any time feedback is requested and not obtained

Retry the request for feedback

Send email request to admin

4.5 Domain Model

The domain model in Figure 4.3 is useful in visually illustrating meaningful conceptual classed or real-world objects in a domain of interest. Domain model makes use of class diagrams without methods. The domain model consists of the contents of the domain model, the association between the model and the attributes of the conceptual classes. In the domain model association are shown from the capture of the image by the farmer through the use of his mobile phone. The image is converted into pixels values. Several pixel values were used as input into the neural networks model. The neural network model is presented with training data after which the model is tested to work well. The classification of the disease is done based on the examples provided to the neural network model. The farmer then receives the feedback on the disease affecting his maize crops and the recommended course of action. The expert provides expert knowledge to the knowledge base upon which the output can be compared for accuracy.

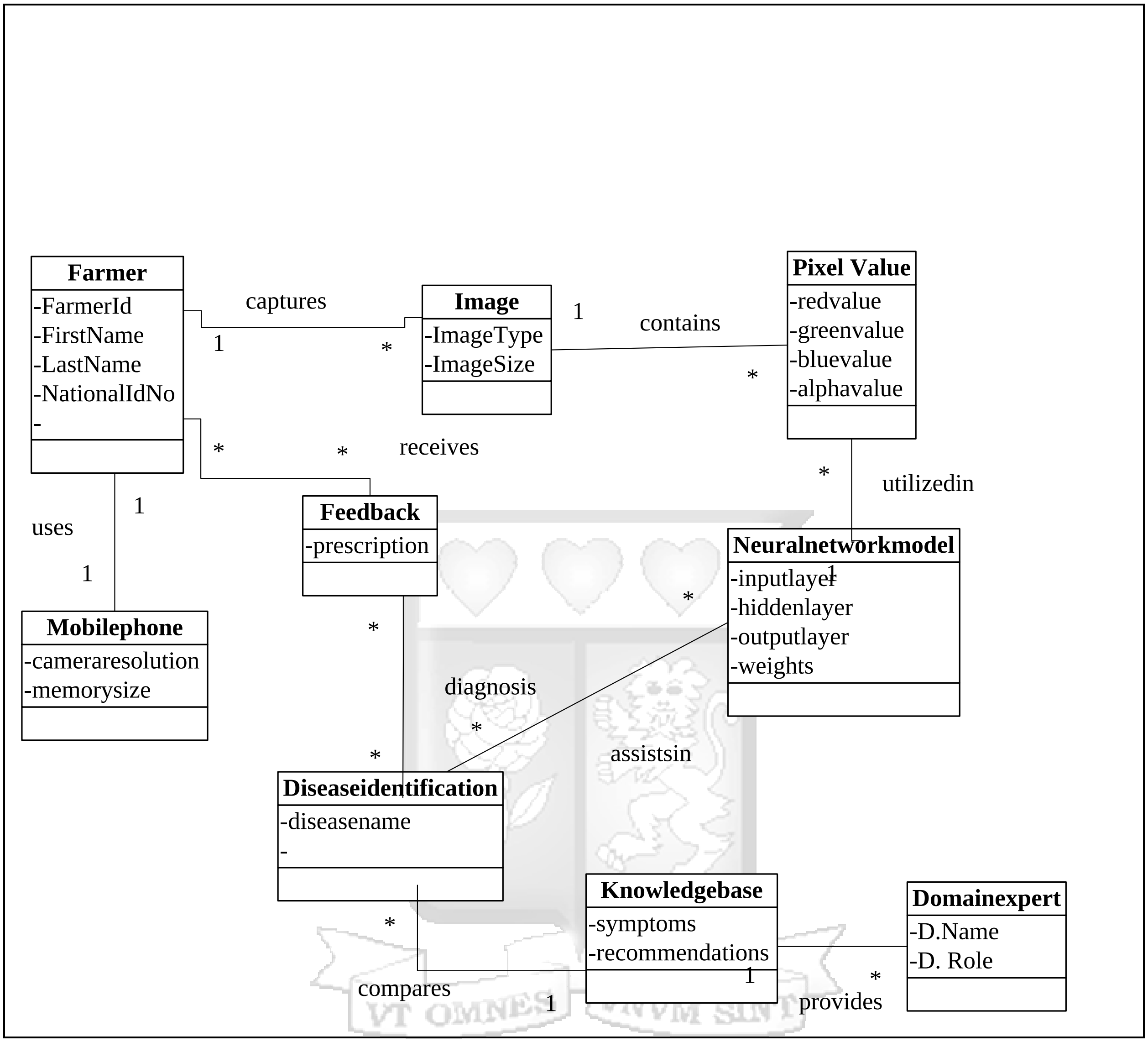


Figure 4.3 Domain Model

4.6 Activity Diagram of the Neural Network

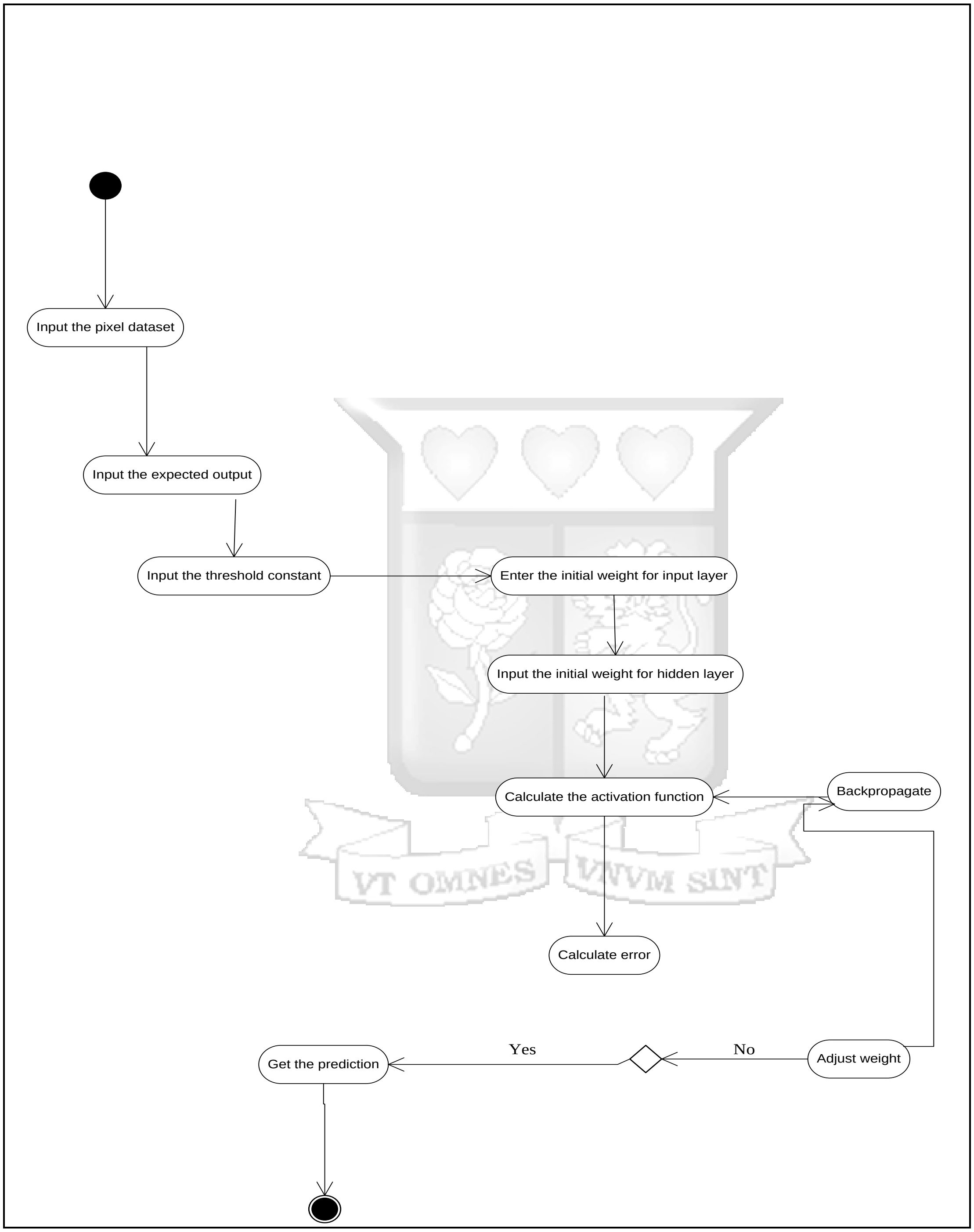


Figure 4.4: Activity Diagram

The activity diagram in Figure 4.4 represents the conditional logic for the sequence of the system activities used to meet the needs of a business process. The activity diagram represented is used to represent the parallel and alternative behaviour and was further used to represent a use case process.

In the vision-based model, the activities that were involved in meeting the objective of classifying the maize leaf diseases based on their visually observable characteristics involved the following activities

a) Input the pixel values-This data represents the visually observable symptoms on the leaf area of the maize crop. The presence of the lesions on the leaf affected the photosynthetic area on the leaf. The lesion attributes that were used included the lesion color, the lesion shape, the lesion coverage/distribution and the lesion size.

b) The expected output for the model was the identification of the diseases affecting the leaves inclusive of the severity of the disease. For example, the leaf could be classified as affected by mild GLS, MNLD or severe MSV at the initial stages of development.

c) The activation function that was used was the sigmoid activation function. The weight was readjusted in order to allow the machine to learn through its experience which led to an increase in the accuracy of the algorithm used. Once the training of the algorithm was complete, a test set was introduced to check if the algorithm was giving the expected output.

4.7 Sequence Diagram

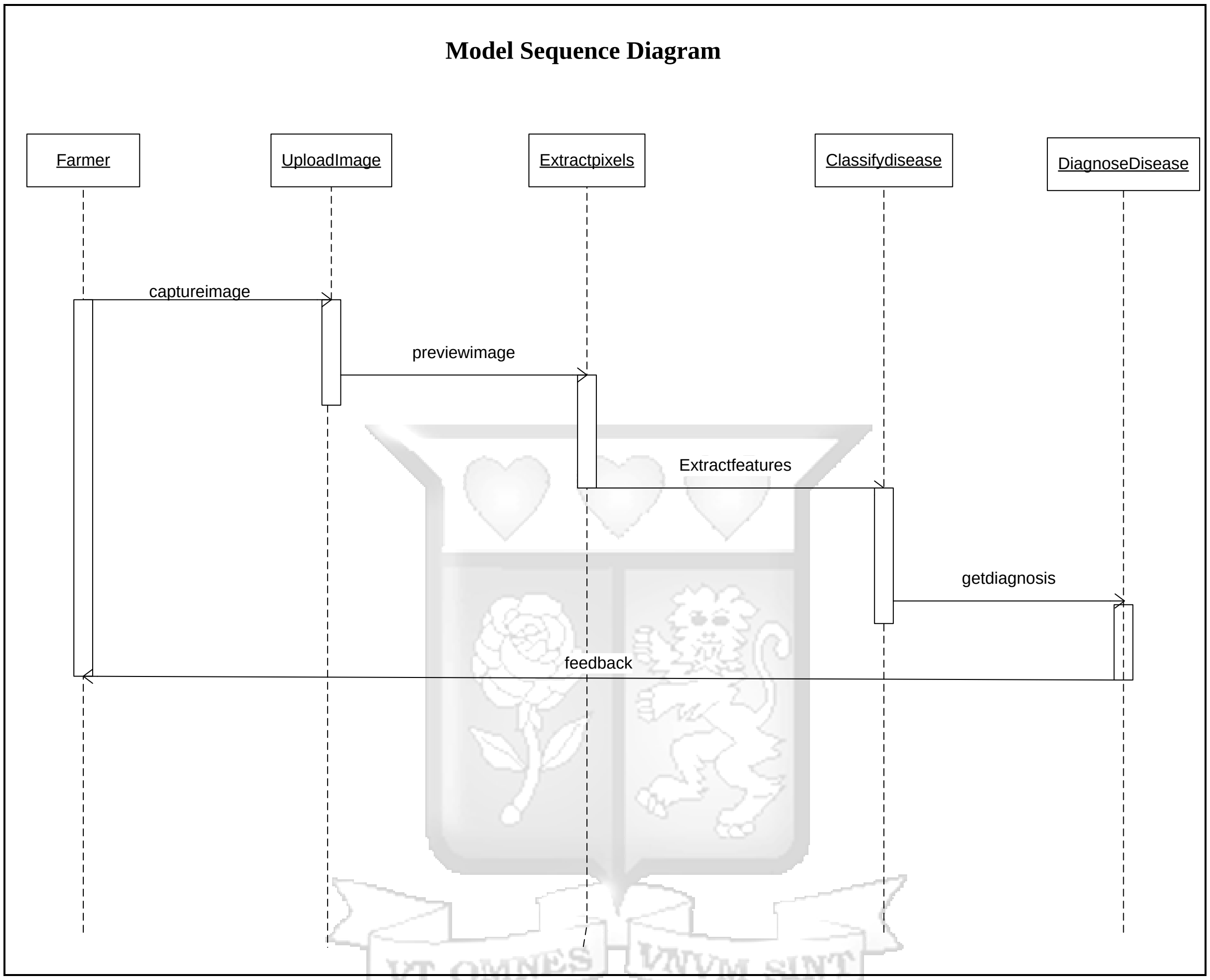


Figure 4.5: Sequence Diagram

Figure 4.5 illustrates the sequence of activities that are followed. The farmer captures the image and uploads it. Once the image is uploaded, the pixels values are extracted and the normalized. The normalized data is then used as input and test set for the artificial neural network algorithm. The network is trained and tested to work well. The classification of the disease affecting the maize based on the input values is provided and a message is sent back to the farmer showing the identified disease and the course of action.

4.8 Model Design

4.8.1 Context diagram

The context diagram in Figure 4.6 illustrates the flow of the data to and from the entities (users) to the application. The main users of the proposed system are the domain expert and the farmer. The farmer provides an input to the system and receives the output from the system. The domain expert provides the expert knowledge on the diseases affecting maize and the recommended course of action.

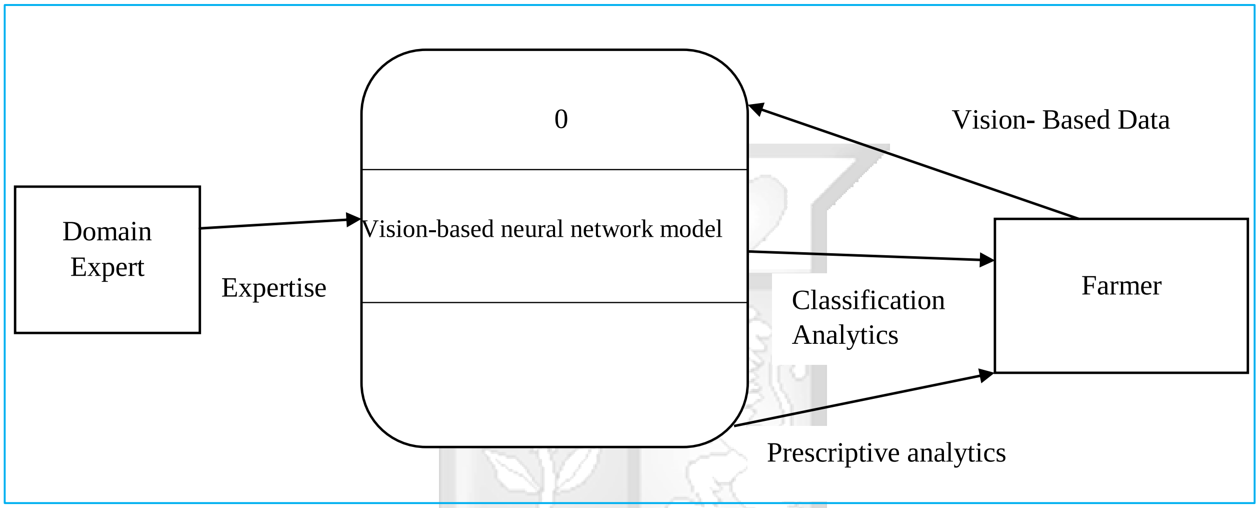


Figure 4.6 Context Diagram

4.8.2 Level 0 DFD

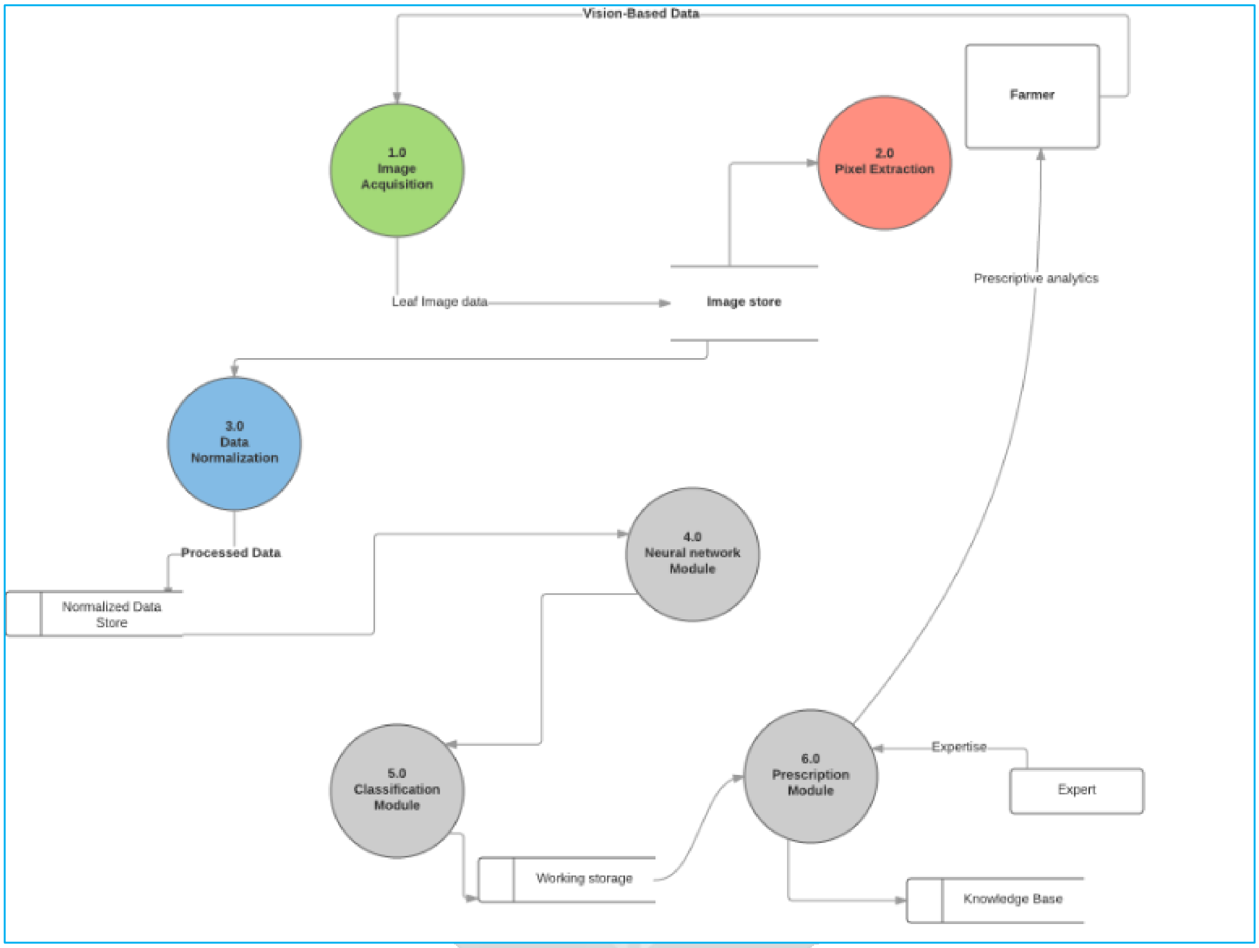


Figure 4.7: Level 0 DFD diagram

The level 0 DFD in Figure 4.7 illustrates the various processes in the application, the data stores for storing various types of data and the entities that interact with the processes. The arrows connecting the various processes and entities show the origin and destinations of the various messages. The level 0 DFD exhibits the process that take place in achieving the vision-based neural network model for classifying the diseases. The process starts with the image capture by the farmer. The image is sent to an image store after which the image for preprocessing. The image is down sampled to enable faster feature extraction. Pixel values are extracted and presented to the neural network for classification. The disease are identified by the classification of the neural network. Feedback is sent to the farmer with the diagnosis and a course of action provided.

Chapter 5: Implementation and Testing

5.1 Introduction

The model was implemented through three main activities. It began with the capture of the images by the farmer using a mobile phone camera. Once the images were captured, they were pre-processed to remove the noise elements. The region of interest was obtained and features were extracted specific for the specific region. The pixel values were extracted from the images captured and stored in a database. Back propagation neural network algorithm was used to implement the classification of the diseases. To initiate the learning of the multilayer perceptron neural network, a training set was presented to the network. Once the classification was completed, the outputs were provided and the farmer received a message of the disease and the recommendation of course of action.

5.2 Model components

5.2 1 Image Processing Components

Camera- The model requires the use of a digital camera to obtain images of the leaves. The camera is a physical devise that is sensitive to the energy of the leaf we image.

Software for image processing- The purpose of this software was to allow the image to be manipulated to different formats from which the information that was obtained would be used for various computer vision functionalities.

Storage of the image- This involves short-term storage which enables fast access and allows a user to preview an image. It may involve online storage that is essential for relatively fast access. Archival storage involves storing the images for infrequent access. Image are stored in bytes during the archival storage for example bytes, kilobytes. Frame buffers enhance storage of one or more images for frequent access.

5.2.2 Neural Network Components

The back propagation neural network algorithm consisted of a number of components as detailed below

5.2.2.1 Input Layer

This is the first layer of the network. The layer consists of one neuron for each specific attribute used by the network to classify. The number of neurons determine how the input layer is structured. The input layer interacts with the external environment and consists of an independent variable interacting with the environment.

5.2.2.2 Output Layer

This is the last layer on the neural network. The output is provided after all the inputs have been processed and presents a pattern to the external environment. In the classification process, the output layer consists of the specific groups for which the inputs are assigned into.

5.2.2.3 Hidden Layer

The purpose of the hidden layer is to enable the neural network to produce better results of the expected output for the given input. The hidden layer provides an intermediate layer for which the activation function has been implemented. The hidden layer consists of hidden neurons. The number of hidden neurons are critical as this prevents the problem of overfitting or under fitting.

Critical error- This is the desired mean square error for stopping network training

The neural network was implemented by attaching a weight to the input variables in the network. A transfer function was selected for the model. The model used sigmoid transfer function which lies between 0 and 1. Once the network was trained it classified the target class as the output. A learning rate of 0.2 and momentum of 0.5 was used to train the network. A test was carried out to validate that the network perform as was expected.

5.3 Model Implementation

5.3.1 Image Capture of the Leaf

The image was captured using the mobile phone camera. The image was then set as a preview enabling the user to see the image they had captured. The camera that was used during the research was 13 megapixels. The application scaled the image down to so as to ensure it did not consume a lot of memory space. The user took the image at a close range since some diseases had breaks between the streaks and this could only be captured if the distance between the camera and the object was too large. A distance of in the range of 1meter to 1.5 meters was preferred so as to ensure that the features of the image were clearly captured. The camera allowed for the object to be zoomed in or out thus ensuring that the farmer obtained a clear view. If the captured image was unclear, the application allowed the user to capture another image. The image was then saved to a folder in the gallery.

Images of maize leaf affected by maize streak virus that were capture for use in the study are as shown in Figure 5.1

Figure 5.1 Maize Leaves Affected by Maize Streak Virus

Maize that were affected by Maize Lethal Necrosis Disease captured are illustrated in Figure 5.2

Figure 5.2 Leaves Affected by MLND

Figure 5.3 provides a sample illustration of the images of maize leaves that were infected by Gray Leaf Spot disease.

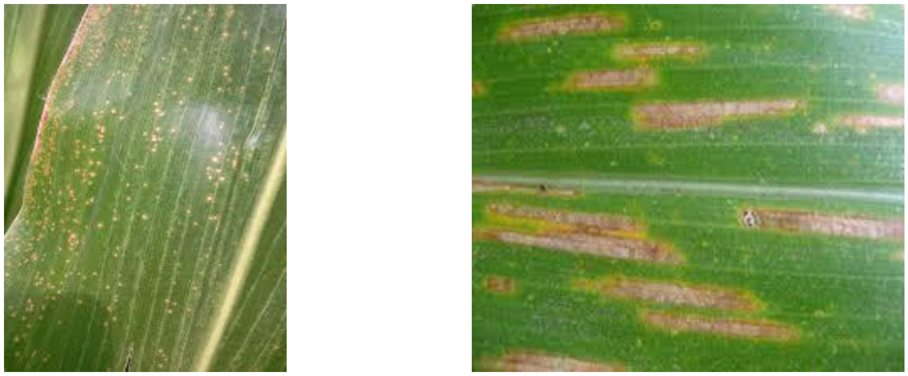


Figure 5.3 Images of Leaf infected by Gray Leaf Spot Disease

5.3.2 Image Processing

Next, the images were converted to their binary values in the RGB (Red, Green and Blue format) and uploaded to the MySQL database.

Sample RGB Values obtained from the images are as illustrated in Table 5.1

Table 5.1: Table of RGB values

R	G	B
40	107	40
159	226	70
48	116	70
146	213	98
57	126	10

5.3.3 Conversion of RGB to HSI

The image was then converted to the HSI (Hue, Saturation, and Intensity) values which is friendly to the human eye. The HSI values were stored and used as the input values for the neural network. Sample RGB values that were converted to HSI using a color converter as shown in Table 5.2.

Table 5.2 Sample HSI Values

H	S	I

120	0.358	0.244
85	0.538	0.595
138	0.385	0.306
95	0.357	0.597
96	0.845	0.252
77	0.916	0.515
95	0.562	0.403
77	0.912	0.535
94	0.624	0.396
94	0.372	0.587
92	0.683	0.383
94	0.473	0.461

5.3.4 Data for Use in Model

The data was obtained from the images that were captured. The pixel values were extracted from images that were presented. The images were converted into RGB values that were later converted to the Hue, Saturation and Intensity (HSI) values. The labels of the various disease were based on the information obtained from KALRO factsheets.

5.3.5 Data Normalization

Neural networks are designed to accept floating point numbers as their input. The range of the values 0 to +1 was used since the model made use of the sigmoid function. The process of normalizing causes all the attributes being superior to another. The numeric range for all attributes must be known. To obtain the minimum and maximum value for each pixel attribute. The minimum maximum normalization method was used where:

$$f(x) = \frac{(x-dl)(nH-nL)}{(dH-dL)} + nL \qquad \text{(Eq. 5.1)}$$

Equation 5.1 was used to normalize a value x. D represented the high and low values of the data while n represented the high and low normalization range desired. Sample normalized data based on the minimum maximum normalization is as illustrate in Table 5.3

Table 5.3: Normalized data set

Normalized RGB value		
-0.79	-0.32	-0.79
0.05	0.52	-0.58
-0.73	-0.25	-0.58
-0.04	0.43	-0.38
-0.67	-0.18	-1

5.3.6 Implementing the Neural Network Algorithm

The backpropagation algorithm was implemented in the research to aid in the identification and resulting classification of the diseases affecting the maize leaf. The algorithm consisted of several elements as detailed.

5.3.4.1 Input Layer

This is the first layer of the network. The layer consisted of one neuron for each specific attribute used by the network to classify the diseases. The number of neurons determine how the input layer is structured. The layer consisted of the Hue, Saturation and Intensity values used as input values. The values that were used in the classification were based on the normalized values that were obtained.

5.3.4.2 Output Layer

This was the last layer on the neural network. The output is provided after all the inputs have been processed. The output layer consisted of the target diseases that were affecting the maize leaves. The target classes were the either GLS, MLND, MSV or Healthy. The healthy class was introduced to act as a control target value in the implementation of the neural network.

5.3.4.3 Hidden Layer

The purpose of the hidden layer is to enable the neural network to produce better results of the expected output for the given input. The hidden layer was defined similarly to the input and output layers. One hidden layer was used in the implementation of the model.

Further to the major components of the neural network algorithm, other factors that were considered in the implementation were:

i. Initial weight range- The weights were initialized in a range between -1 and + 1.
ii. Number of hidden Layers- One hidden layer was used in the classification of the maize leaf diseases.

iii. Number of nodes in the hidden layer- The model was constituted of three hidden nodes in its hidden layer.

iv. Number of Epochs- An epoch refers to one clean sweep through all records in the training set. Increasing the number of epochs increases the accuracy of the model. 1500 iterations were used in the training thus increasing the accuracy of the model.

v. Step size or learning rate for gradient descent- This is the multiplying factor for error correction during back propagation. Low step size produces slow but steady learning. High value produces rapid but erratic value. The value of the step size ranges from 0.1 to 0.9. A learning rate of 0.2 was used in the model. This ensured steady learning by the network.

vi. Hidden layer sigmoid- The outputs of the hidden node passed through the sigmoid function. The range of the sigmoid function was between 0 and 1.

5.4 Training and Testing the Model

A training data set was constructed from the pixel values that were obtained from the images of the leaves. Training was done through adjustment of the weights to produce more accurate results. The training continued until the error rate (epoch) reached an acceptable level. The epoch was set to 1500 iterations. With a higher epoch rate, more accurate results were obtained compared to using a lower epoch rate. Once the normalized data set was obtained, a split ratio of 70: 30 was used to separate the training set from the test set. The training set was used to train the neural networks on how to classify. The test set was used to validate that the neural network provided the desired output. The neural network was trained using the back propagation algorithm. The process involved feed forward approach after which a back pass was carried out to give a more accurate result. A sigmoid activation function was used.

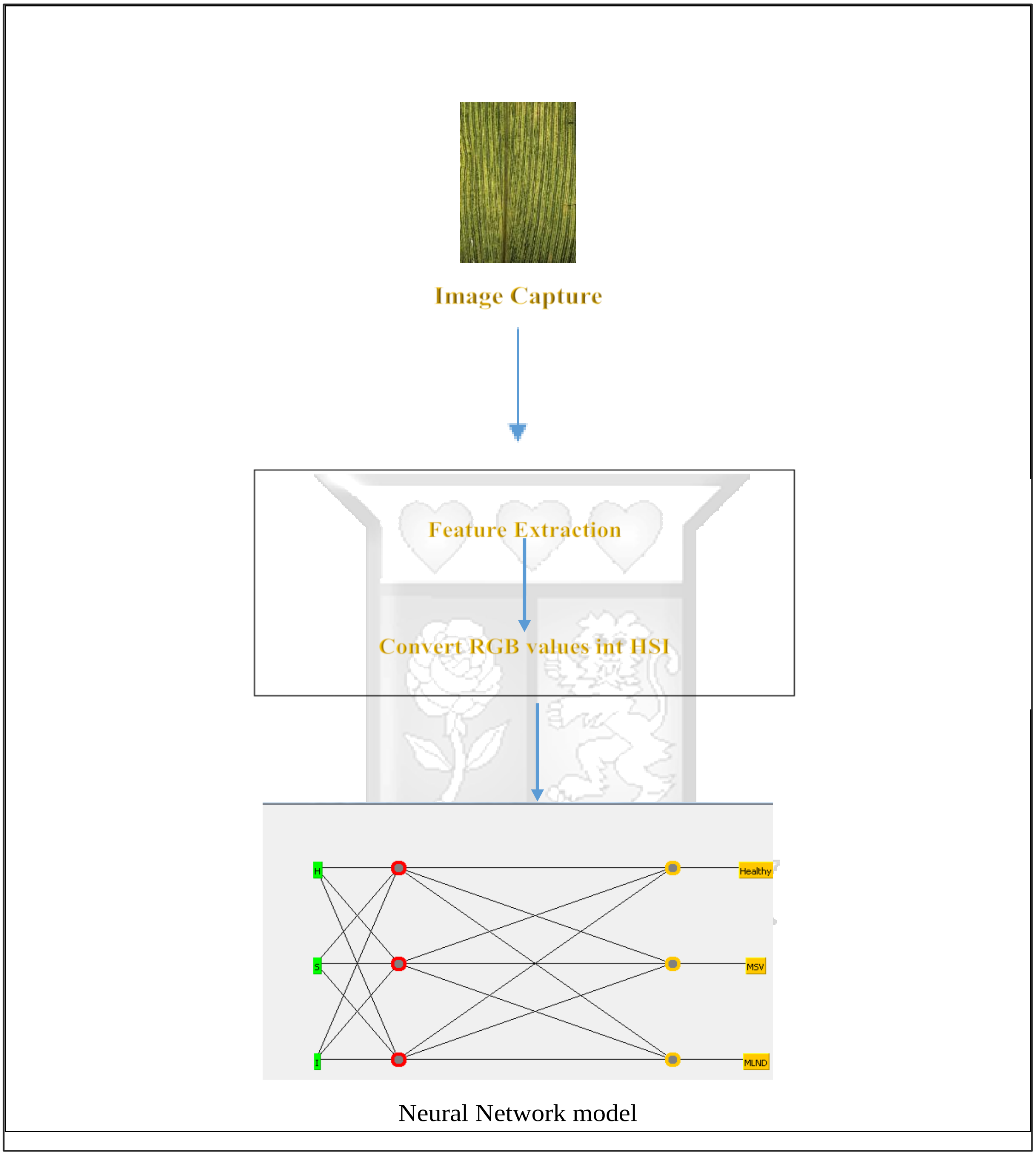
5.5 Software Flow

An android application was built within which the farmers were able to capture images of their maize leaves with the camera module. When the application was launched, a button was provided that enabled the user to capture the image of the leaf. Once the farmer captured the image, an option to save or discard the image was provided. Upon saving the image, a preview of the captured image was displayed. A check disease button was displayed on the screen. When the button was pressed, the image pre-processing and feature extraction was activated. Once the classification is complete, the results were displayed on the screen alongside the prescriptive course of action. A popup message appeared requesting the farmer would like to capture another image, a yes activated the image capture option. If the farmer selected the no option, he was prompted to exit the application. The development environment for the application is:

- Android
- Java
- Mobile Phone Camera
- Windows 7

5.6 Model Architecture

Figure 5.4 below illustrates the architecture of the model proposed in this research. Images were captured and features of the images, pixel features were extracted and converted to binary values. The binary values were fed in to the neural network as training and testing algorithm. The output values obtained were the target classification of the diseases affecting the leaf of the maize.



;

5.4 Model Architecture

5.7 Model Testing

This process involved testing the functionality and reliability of the proposed model. Several parameters were used to test the model developed as illustrated in Table 5.4.

Table 5.4 Model Testing

Test class	Inspection Check	Priority
Functional	Does the application validate user input to ensure only images are uploaded	High
Functional	Are the different farmers who uploaded different images able to receive separate analysis of their respective images	High
Reliability	Is there a different in the presence of noise between the original image uploaded and the images used for analysis	Medium
Functional	Does the identification model abide by the KARLO symptoms of various diseases affecting maize and their action taken	High

5.7.1 Model Testing Results

Application successfully validated the user input to ensure that only acceptable image formats were uploaded. A unique identifier was uploaded to prevent an image that is not uploaded by the farmer to be analyzed.

Table 5.5: Model Results

F. ID	Test Results	Comment
Functional	Pass	Pixel based classification as used to identify diseases
Reliability	Pass	Prescriptions are based on recommendations from Kenya

		Agricultural livestock Research Organization

5.8 System Testing

The system was tested to check how the developed model performed in comparison to having the limited number of extension workers visiting the various farms as illustrated in Table 5.6.

Table 5.6 System Testing

Test Class	Inspection Check	Priority Level
Performance	Does the entire process involved from to the point of which the farmer recover the predictive and prescriptive analytics take a short duration of time	High

5.8 Acceptance Testing

This testing illustrated in Table 5.7 involved checking whether the key aspects of the applications were achieved.

Table 5.7 User Testing

Test Class	Inspection Check	Priority
Usability	Have user requirements been met?	High
Usability	Are users satisfied with the application's output	Medium

Chapter 6: Discussions

6.1 Introduction

The vision-based model was implemented by using features of the leaf images that were captured in order to classify the diseases affecting the maize crop. The model was tested for correct classification on the basis of accuracy, precision, and the error that was obtained. The model was considered quite suitable in comparison to the other methods discussed in the research. Farmers in Nyeri County largely depended on the visual examination for identification of the disease affecting the leaves. This method was prone to errors and inaccuracies when farmers were trying to identify the diseases that were ailing their crops. The vision-based model discussed provided more accurate output compared to the human visual examination method. Farmers in Nyeri County also depended on extension workers to help the m identify the diseases affecting their maize. Since the extension workers in the county are few, it would take a long time for the extension workers to reach the farmers. The delay would lead to late diagnosis thus control of the diseases was delayed. The computable visually observable ontological framework on the other hand involved experts manually annotating the images that were acquired through hyperspectral imaging. This approach demanded a lot of time and effort from the expert and would consequently lead to high chances of errors. Based on the results obtained from the structured interview with the extension worker in Nyeri County, the researcher found out that a mobile application would come in handy in identifying the diseases affecting the maize. The maize disease common in the region were gray leaf spot disease, maize streak virus, rust, maize lethal necrosis disease as indicated in Appendix C. The main methods used in identifying diseases in the region were visual examination and referencing to photo sheets by the extension workers. The two methods rely on the visually observable characteristics to identify the diseases. Plant clinics have also been introduced in the region which give farmers an alternative avenue to have their crops assessed for diseases affecting them.

The model developed in this research gives more accurate classification based on the fact that it was implemented based on the neural network algorithm. The use of the machine learning algorithm enabled the model to work much faster thus the feedback to the farmer was provided almost immediately. Combining the strengths of computer vision and machine learning enabled the model developed in the research to provide accurate results while providing a prescription to the farmer.

6.2 Validation of the Model

The model was validated for accuracy, precision, recall ratio using the confusion matrix. A cross validation of 10 folds was used to test the model. 15 out of 19 instances presented to the network were correctly classified. This resulted to an accuracy 78.94%.

Table 6.1 Classification Output

Correctly Classified Instances	15	78.9474 %
Incorrectly Classified Instances	4	21.0526 %
Mean absolute error	0.1688	
Total Number of Instances	19	

The absolute error refers to difference between the classified value and the expected value. The mean absolute error in Table 6.1 is an average of all the absolute errors obtained from the classification.

The performance evaluation for the classification of the diseases based on the precision, recall ratio and F-Measure per class as stated in Chapter 3 obtained are as shown

6.2.1 Detailed Accuracy by Class

Table 6.2 Detailed Accuracy by Class

TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC Area Class**	**Target**
1	0.167	0.778	1	0.875	1	Healthy
0.8	0.143	0.667	0.8	0.727	0.943	MSV
0	0	0	0	0	0.771	GLS
1	0	1	1	1	1	MLND
Weighted Avg.	0.789	0.099	0.673	0.789	0.724	0.949

TP represents the true positive, FP represents the false positives. The Receiver Operator Characteristics (ROC) refers to the classification the false positives and false negatives. T

6.2.2 Confusion Matrix

The confusion matrix that was obtained from the classification was as illustrated in Table 6.3. The confusion matrix contains information on the actual and predicted classifications. There were a total of 19 instances that were used to train and test the network. 7 instances of the healthy target were presented to the network. All instances were correctly classified as healthy. 5 instances of leaves infected with MSV were contained in the set. 1 instance of MSV was incorrectly classified as infected by MSV while the remaining 4 were correctly classified as infected by MSV. 3 instances of leaf features affected by the GLS were presented to the network. One instance was incorrectly classified as infected GLS while the remaining were correctly classified as infected with GLS. The final classification done was for leaves infected by MLND. All the instances of MLND presented to the neural network were correctly classified.

Table 6.3 Confusion Matrix

Input	**Healthy**	**MSV**	**GLS**	**MLND**
Healthy	7	0	0	0
MSV	1	4	0	0
GLS	1	0	2	0
MLND	0	0	0	4

6. 3 Contributions of the Model to Research

Considering the challenges that the farmers faced in identifying maize diseases, the model offered an improved solution compared to the commonly used visual examination method. The vision-based neural network model provided the farmer with reliable results of the disease affecting their crops. The models also helped the farmers reduce the reliance on the extension workers and thus control diseases affecting their maize in time to ensure a higher yield.

6.4 Shortfalls of the Research

The model that was developed had the following limitations

i. The model did not consider all the features of the image including texture and edges that would enable the neural network provide more accurate results
ii. The model was limited to maize crops
iii. The location of the farmers/farms was not put into consideration
iv. The model did not consider the abiotic stress factors that also lower the yield of the maize crops

Chapter 7: Conclusions and Recommendations

7.1 Conclusions

As highlighted in the interview with an extension worker who works closely with farmers in Nyeri County, farmers face a number of challenges in identifying and consequently managing the diseases affecting their crops. One of the major problems is misdiagnosis which is based on the experience of the farmer. Misdiagnosis results in the farmer taking the wrong action and thus obtains a low yield. (Ghaiwat & Arora, 2014a) emphasized that misdiagnosis of crop diseases is as a result of overreliance of experience based on the visually observable characteristics of the diseases affecting the crops. It also takes some time for the limited number of extension workers to visit various farms in the rural areas.

The research lays its emphasis of taking advantage of computer vision techniques as well as machine learning algorithms for the classification of the various diseases. The farmers are able to act as soon as they receive the prescription from the application.

The research relies on pixel features of the images that are captured by the farmers. Once the pixel values are obtained, the data is normalized so as to minimize the range of values returned. Back propagation neural network is used to classify the neural network and the diagnosis and prescription are provided to the farmer.

Kenya Plant Health Inspectorate recognizes the need to apply technology to aid in early detection of crop diseases. Early detection of crop diseases leads to control and management. This results in higher yields and thus a more food secure nation.

With an estimate of every two out of three households in Kenya growing maize, there is a high overreliance on maize for consumption. It is therefore pertinent to ensure that the high yields are obtained from the crop. Technology can therefore be applied to detect biotic stress that affects the crops early enough thus enabling the farmers to take a course of action as recommended.

7.2 Recommendations

The recommendations below are made from the results obtained from this research:

i. The process of feature extraction can take into account more features in addition to pixel values so as to improve the output of the model.
ii. The use of the application can be expanded to focus on more disease affecting the maize crops as well as the vectors (pests) that aid in the spread of the diseases.
iii. Use of Unmanned Aerial Vehicles would be an added advantage especially in large fields
iv. Location information should also be added to the application since different zones are prone to different kinds of disease based on the ecological zones.

7.3 Suggestions for Future Research.

i. The researcher recommends that abiotic stress such as pH of the soil, weather should be considered as inputs to the system so as to give a more accurate classification.
ii. Since most small scale farmers carry out mixed cropping on their farms, the application can be extended to consider the other crops grown alongside the maize and a link if any is identified between the crops and the diseases affecting the maize.
iii. The application can be built to support local languages since most farmers in the rural areas are elderly. The applications can also be built to support speech to text.
iv. The application can be extended to have forums in which farmers are able to share information based on their location. An expert should also be a part of the forum so as to give expert information on various issues raised by the farmer.

References

Adimo, O. (n.d.). Kenya - Global Yield Gap Atlas. Retrieved March 19, 2016, from http://www.yieldgap.org/kenya

Aduwo, R. J., Mwebaze, E. & Quinn, A. J. (2010). Automated Vision-Based Diagnosis of Cassava Mosaic Disease. Workshop on Data Mining in Agriculture (DMA 2010), Berlin

Agricultural Cooperative Development International and Volunteers in Overseas Cooperative Assistance (ACDI/VOCA). Boosting Household Incomes, Raising Productivity. Retrieved on 24th August 2015 from http://acdivoca.org/our-programs/project-profiles/kenya-kenya-maize-development-program-kmdp

Agricultural Sector Development Support Programming (ASDSP) (2013). Nyeri County. Retrieved from http://www.asdsp.co.ke/index.php/nyeri-county.

Bashish, A. D., Braik, M., & Bani-Ahmad, S. (2010). A Framework for Detection and Classification of Plant Leaf and Stem Diseases. International Conference on Signal and Image Processing

Bhatnagar M. & Singh K. P. (2013). Research Methodology as SDLC Process in Image Processing. International Journal of Computer Applications. Vol. 77 No 2

Bordens, K. S., & Abbott, B. B. (2011). Research design and methods: a process approach (8th ed). New York: McGraw-Hill.

Brooks, S., Thompson, J., Odame, H., Kibaara, B., Nderitu, S., Karin, F. and Millstone, E. (2009) Environmental Change and Maize Innovation in Kenya: Exploring Pathways In and Out of Maize, STEPS Working Paper 36, Brighton: STEPS Centre

Brosnan, T., & Sun, D.W. (2004). Improving quality inspection of food products by computer vision-a review. Journal of Food Engineering, 61(1), 3-16. http://dx.doi.org/10.1016/S0260-8774 (03)00183-3

Devi, C. J. D., Reddy, B. S. P., Kumar, K. V., Reddy, B. M., & Nayak, N. R. (2012). ANN Approach for Weather Prediction using Back Propagation, 3(1). Retrieved from http://www.internationaljournalssrg.org

Fang, Y., & Ramasamy, R. P. (2015). Current and Prospective Methods for Plant Disease Detection. Biosensors, 5(3), 537–561. http://doi.org/10.3390/bios5030537

Ganatra, A., Kosta, Y. P., Panchal, G., & Gajjar, C. (2011). Initial Classification Through Back Propagation In a Neural Network Following Optimization Through GA to Evaluate the Fitness of an Algorithm. International Journal of Computer Science and Information Technology, 3(1), 98–116. http://doi.org/10.5121/ijcsit.2011.3108

Gang, W. S., Bao, S. F., Xu, Y. E., Yu – Xuan, W. & Yi – Fan C. (2007). A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network. IEEE 7th International Symposium on Signal Processing and Information Technology

Gavhale, R. K. & Gawande, U. (2014). An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques. IOSR Journal of Computer Engineering, Vol 16. No. 1

Gekone, M., Otipa, M. & Kamau, R. (2013). Maize Lethal Necrosis Disease on Maize. Retrieved from http://www.plantwise.org/FullTextPDF/2013/20137804328.pdf

Ghaiwat, N. S., Arora, P. (2014). Detection and Classification of Plant Leaf Diseases Using Image processing Techniques: A Review. International Journal of Recent Advances in Engineering & Technology. 2347 - 2812, Vol. 2, No. 3

Gomes, J. F. S., & Leta, F. R. (2012). Applications of computer vision techniques in the agriculture and food industry: a review. European Food Research and Technology, 235(6), 989–1000. http://doi.org/10.1007/s00217-012-1844-2

Government of Kenya (GoK). (2010). Agricultural Sector Development Strategy 2010- 2020. Retrieved February 10, 2016, from http://faolex.fao.org/docs/pdf/ken140935.pdf

Gurjar, A. A., Viraj, A. G. (2012). Disease Detection on Cotton Leaves by Eigenfeature Regularization and Extraction Technique. International Journal of Electronics, Communication & Soft Computing Science and Engineering (IJECSCSE) Vol. 1, No. 1

Hansankhani, R., & Navid, H. (2012). Qualitative Sorting of Potatoes by Color Analysis in Machine Vision System. Journal of Agricultural Science, Vol. 4, No. 8. doi:10.5539/jas.v4n4p129

Harnsomburana, J., Green, J. M., Barb, A. S., Schaeffer, M., Vincent, L., & Shyu, C.-R. (2011). CompuTable visually observed phenotype ontological framework for plants. BMC Bioinformatics, 12, 260. http://doi.org/10.1186/1471-2105-12-260

Hassankhani R. (2012). Potato surface defect detection in machine vision system. African Journal of Agricultural Research, 7(5). http://doi.org/10.5897/AJAR11.2049

Kadir, A., Nugroho, L. E., Susanto, A., & Santosa, P. I. (2013). Leaf classification using shape, color, and texture features. arXiv Preprint arXiv:1401.4447. Retrieved from http://arxiv.org/abs/1401.4447

Kanjalkar, P. H. & Lokhande, S. S. (2013). Detection and Classification of Plant Leaf Diseases using ANN. International Journal of Scientific & Engineering Research, Vol. 4, No. 8

Kenya Food Security Steering Group (KFSSG) (2008). The Impact of Rising Food Prices on Disparate Livelihoods Groups in Kenya

Kenya Joint Assessment Disease (KJAR) (2012). Report on Status of Maize Lethal Necrosis Disease in Kenya and General Maize Performance. Retrieved from http://www.fao.org/fileadmin/user_upload/drought/docs/Maize/Lethal/Necrotic/Disease/in/Kenya_Joint/Assessment/Report

Kenya National Bureau of Statistics (KNBS) (2015). Economic Survey. Retrieved from www.knbs.or.ke/index.php?...economic

Kessy, J. G., Bukalasa, J., Akonaay, B. H. & Lema, R (2013). Maize Lethal Necrosis Disease. Plantwise

Kinyua, Z.M., Otipa, M.J., Muriithi, C.W., Amata R.L & Mbaka J.N. (2014). Maize Grey Leaf Spot factsheet.

Koné Tadiou. (2013, May 26). Artificial Neural Networks. Retrieved from http://futurehumanevolution.com/artificial-intelligence-future-human-evolution/artificial-neural-networks

Kumar, M., Kamble, M., Pawar, S., Patil, P., & Bonde, N. (2011). Survey on Techniques for Plant Leaf Classification. International Journal of Modern Engineering Research, 1(2), 538–544.

McCann, C. J. (2009). Maize and Grace. Africa's Encounter with a New World Crop 1500-2000. Harvard University Press

Monica Elliott, K. P. (2014, June 19). Guidelines to Identification and Management of Plant Disease Problems: Part I. Eliminating Insect Damage and Abiotic Disorders. Retrieved February 9, 2016, from http://edis.ifas.ufl.edu/mg441

Muthukannan, K., & Latha, P. (2015). A PSO MODEL FOR DISEASE PATTERN DETECTION ON LEAF SURFACES. Image Analysis & Stereology. http://doi.org/10.5566/ias.1227

Nyeri County. (n.d.). Retrieved March 19, 2016, from http://www.kccap.info/index.php?option=com_phocadownload&view=category&id=25&Itemid=66

Owomugisha, G., Quinn, A. J. & Mwebaze, E. (2014). Automated Vision-Based Diagnosis of Banana Bacterial Wilt Disease and Black Sigatoka Disease. Proceedings of the 1st International Conference on the use of Mobile ICT in Africa

Patil, .B. S. & Bodhe, K. S. (2011). Leaf Disease Severity Measurement Using Image Process. International Journal of Engineering and Technology. Vol.3 (5), 2011

Phadikar, S. & Sil, J. (2008). Rice Disease Identification Using Pattern Recognition Techniques. Proceedings of 11th International Conference on Computer and Information Technology, 25-27

Phenotype | Define Phenotype at Dictionary.com. (n.d.). Retrieved February 18, 2016, from http://dictionary.reference.com/browse/phenotype

Rothe, P. R., & Kshirsagar, R. V. (2014). A Study and Implementation of Active Contour Model For Feature Extraction: With Diseased Cotton Leaf as Example, 4(2). Retrieved from http ://inpressco.com/category/ijcet

Rothe, R. P. & Kshirsagar, V. R. (2014). SVM-based Classifier System for Recognition of Cotton Leaf Diseases. International Journal of Emerging Technologies in Computational and

Applied Sciences. Retrieved from http://iasir.net/IJETCASpapers/IJETCAS14-224.pdf

Sathya, R., & Abraham, A. (2013). Comparison of supervised and unsupervised learning algorithms for pattern classification. Int J Adv Res Artificial Intell, 2(2), 34–38.

Shire, A., Jawarkar, U. & Manmode, M. (2015). A Review Paper On: Agriculture Plant Leaf Disease Detection Using Image Processing. International Journal of Innovative Science, Engineering & Technology, Vol. 2 No. 1

Short, C., Mulinge, W. & Witwer, M. (2012). Analysis of Incentives and Disincentives for Maize in Kenya. Technical notes series, MAFAP, FAO, and Rome

Sing, K., Malik, D., & Sharma, N. (2011). Evolution of K-Means Algorithm in Data Mining and their Removal. International Journal of Computational Engineering & Management, Vol. 12, No. 4

Tellaeche, A., BurgosArtizzu, X. P., Pajares, G., Ribeiro, A., & Fernández-Quintanilla, C. (2008). A new vision-based approach to differential spraying in precision agriculture. Computers and Electronics in Agriculture, 60(2), 144–155. http://doi.org/10.1016/j.compag.2007.07.008

UNEP. (2015). Green Economy Sector Study on Agriculture in Kenya. Retrieved February 10, 2016, from http://www.unep.org/greeneconomy/Portals/88/documents/Kenya%20agriculture_Final.pdf

World Health Organization (WHO) (2015). Trade, Foreign Policy, Diplomacy and Health: Food Security. Retrieved from http://www.who.int/trade/glossary/story028/en/

Ying, G., Miao L., Yuan Y. & Zelin, H. (2008). A Study on the Method of Image Pre-Processing for Recognition of Crop Diseases. International Conference on Advanced Computer Control, IEEE.

Appendix A: Originality Report

Turnitin Originality Report

Vision-Based Model for Maize Leaf Disease Identification. A Case Study in Nyeri County by Carolyn Wanja Mwangi

From MSc IT Thesis Proposal (MSc. IT Thesis Proposal 2015)

- Processed on 01-Apr-2016 2:34 PM EAT
- ID: 652121032
- Word Count: 13688

Similarity Index
24%
Similarity by Source

Internet Sources:
18%
Publications:
11%
Student Papers:
14%

sources:

1 1% match (Internet from 05-Jun-2014)
http://www.kilimo.go.ke/kilimo_docs/pdf/ASDS_Final.pdf

2 1% match (student papers from 27-May-2015)
Submitted to Strathmore University on 2015-05-27

Appendix B: Interview Guide

Vision-based Neural Network Model for Identification of Maize Leaf Disease in Nyeri County

i. What are the common diseases that affect Maize leaves in Nyeri?
ii. What are the methods that are used to identify the diseases that affect maize?
iii. Are plant clinics efficient in early diseases detections?
iv. Is there a phenotype database that exists in Kenya? If yes which organization is charged with the management of the phenotype?
v. What challenges are faced by the extension workers in identifying the diseases affecting various crops?
vi. Would use of a mobile application make the work of identifying diseases easier for the extension worker?

Your assistance will be highly appreciated

Appendix C: Interview Feedback

Interview Guide

Vision-Based Model for Maize Leaf Disease for Identification: A Case Study in Nyeri County

What are the common diseases that affect Maize leaves in Nyeri?

- Gray Leaf Spot Disease
- Leaf Blight
- Maize Streak Virus
- Maize Smut
- Maize Lethal Necrosis Disease

What are the methods that are used to identify the diseases that affect maize?

- Observation
- Photo Sheet (Used to compare but not always used)

Are plant clinics efficient in early diseases detections?

- Yes they are. When a farmer encounters a new disease, they go to a farm doctor who provides them with advice on the next course of action
- The doctors are equipped with a tablet that helps them to access the internet and obtain the information from the internet.

Is there a phenotype database that exists in Kenya? If yes which organization is charged with the management of the phenotype?

- Not sure
- In Kenya -KEPHIS
- In the District Agriculture Officer photo sheets are used
- Consulting Plant wise and bioinformatics website

What challenges are faced by the extension workers in identifying the diseases affecting various crops?

- Lack of technology that can aid them in fast and accurate detection of the diseases.
- The extension workers are few thus not able to attend to all the needs of the farmers in the region

Would use of a mobile application make the work of identifying diseases easier for the extension worker?

- Yes it would

Your assistance will be highly appreciated