

**MALTING BARLEY SEED IDENTIFICATION USING MACHINE LEARNING**

**A Thesis Presented**

**By**

NIYONIZEYE CYRIAQUE

**To**

**The Faculty of Informatics**

**Of**

**St. Mary’s University**

**In Partial Fulfillment of the Requirements**

**For the Degree of Master of Science**

**In**

**Computer Science**

**June, 2023.**

ACCEPTANCE

Ethiopian Malt Barley Seed Identification Using Machine Learning

By

NIYONIZEYE CYRIAQUE

Accepted by the Faculty of Informatics, St. Mary’s University, in partial fulfillment of the requirements for the degree of Master of Science in Computer Science

Thesis Examination Committee:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Internal Examiner

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

External Examiner

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dean, Faculty of Informatics

June, 2023.

DECLARATION

I, the undersigned, declare that this thesis work is my original work, has not been presented for a degree in this or any other university, and all sources of materials used for the thesis work have been duly acknowledged.

Student

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature

Addis Ababa

Ethiopia

This thesis has been submitted for examination with my approval as an advisor.

Alembante Mulu (PhD)

Advisor

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature

Addis Ababa, Ethiopia

June, 2023.

# AKNOWLEDGEMENT

I want to start by expressing my great gratitude to my supervisor, Dr. Alembante Mulu, for his unwavering support of my thesis and for his immense expertise, patience, and encouragement. His advice was extremely helpful to me during the entire thesis. In addition to my adviser, I want to thank everyone who has helped me over the past years in any way they can. Finally, without the encouragement and support of my dear family, nothing is impossible.

Thank You All.

Table of Contents

[AKNOWLEDGEMENT iii](#_Toc138972680)

[LIST OF FIGURES vii](#_Toc138972681)

[LIST OF TABLES viii](#_Toc138972682)

[LIST OF ACRONYMS ix](#_Toc138972683)

[ABSTRACT x](#_Toc138972684)

[CHAPTER ONE 1](#_Toc138972685)

[INTRODUCTION 1](#_Toc138972686)

[1.1. Background of the Study 1](#_Toc138972687)

[1.2. Statement of the Problem 2](#_Toc138972688)

[1.3. Motivation 3](#_Toc138972689)

[1.4. Research Question 4](#_Toc138972690)

[1.5 Objective of the Study 4](#_Toc138972691)

[1.5.1. General objective 4](#_Toc138972692)

[1.5.2. Specific Objectives 4](#_Toc138972693)

[1.6. Scope and Limitations of the Study 4](#_Toc138972694)

[1.7. Significance of the Study 5](#_Toc138972695)

[1.8. Organization of the Thesis 5](#_Toc138972696)

[LITERATURE REVIEW 7](#_Toc138972697)

[2.1. Overview 7](#_Toc138972698)

[2.2. Malt-barley Selection and Grading Technique 7](#_Toc138972699)

[2.3. Image processing 8](#_Toc138972700)

[2.3.1. Image Acquisition 9](#_Toc138972701)

[2.3.2. Image Preprocessing 9](#_Toc138972702)

[2.3.3. Segmentation 10](#_Toc138972703)

[2.3.5. Image Identification 12](#_Toc138972704)

[2.4. Review of Related Works 13](#_Toc138972705)

[2.2.1. Fruit Varieties Identification 13](#_Toc138972706)

[2.2.2. Rice Varieties Identification 15](#_Toc138972707)

[2.2.3. Wheat Varieties Identification 15](#_Toc138972708)

[2.2.4. Barley Varieties Identification 16](#_Toc138972709)

[CHAPTER THREE 20](#_Toc138972710)

[3.1. Dataset Preparation 20](#_Toc138972711)

[3.2. Preprocessing 20](#_Toc138972712)

[3.2.1. Grayscale Conversion 21](#_Toc138972713)

[3.2.2. Noise Removal 21](#_Toc138972714)

[3.2.3. Binarization 22](#_Toc138972715)

[3.2.4. Edge detection 24](#_Toc138972716)

[3.3. Feature Extraction 24](#_Toc138972717)

[3.4. A General Approach for Building Identification Model 26](#_Toc138972718)

[3.4.1. Data set Partitioning 26](#_Toc138972719)

[3.4.2. Building Model 26](#_Toc138972720)

[3.5. Design of Malt-barley Identification Model 26](#_Toc138972721)

[3.5.1. Image Acquisition 28](#_Toc138972722)

[3.5.2. Image Preprocessing 30](#_Toc138972723)

[3.5.3. Feature Extraction 32](#_Toc138972724)

[3.6. Constructing Identification Model 36](#_Toc138972725)

[3.6.1. Artificial Neural Network Classifier 36](#_Toc138972726)

[3.6.2. K-Nearest Neighbor Classifier 38](#_Toc138972727)

[3.6.3. Hybrids of Artificial Neural Network and K-Nearest Neighbor 40](#_Toc138972728)

[3.7. Malt-barley seed Identification Model 41](#_Toc138972729)

[3.8. Evaluation Techniques 42](#_Toc138972730)

[CHAPTER FOUR 44](#_Toc138972731)

[RESULT AND DISCISSION 44](#_Toc138972732)

[4.1. Experimental Result 44](#_Toc138972733)

[4.1.1. Artificial Neural Network Classifier 44](#_Toc138972734)

[4.1.2. K-Nearest Neighbor Classifier 47](#_Toc138972735)

[4.1.3. The Hybrid of ANN and KNN Classifiers 50](#_Toc138972736)

[4.2. The Overall Performance of Identification Model 52](#_Toc138972737)

[4.3. Discussion of Results 54](#_Toc138972738)

[CHAPTER FIVE 56](#_Toc138972739)

[CONCLUSION AND RECOMMENDATION 56](#_Toc138972740)

[5.1 Conclusion 56](#_Toc138972741)

[5.2 Recommendation 57](#_Toc138972742)

[References 58](#_Toc138972743)

# LIST OF FIGURES

[Figure 5 Sample scanned Sabini malt-barley seed; (a) RGB image, (b) Gray level image 21](#_Toc138972744)

[Figure 6 **Sample Sabini malt-barley seed image; (a)** gray level before MF applied, **(b)** gray level image after MF applied 22](#_Toc138972745)

[Figure 7 Sample binary Sabini malt-barley seed image 23](#_Toc138972746)

[Figure 8 **Sample Sabini malt-barley seed; (a)** segmented image, **(b)** edges of each segmented image using Sobel edge detection 24](#_Toc138972747)

[Figure 9 **Sample features computed from malt-barley seeds** 25](#_Toc138972748)

[Figure 1 Architecture of the proposed model for Ethiopian malt-barley identification 28](#_Toc138972749)

[Figure 2 Sample malt-barley seed images 29](#_Toc138972750)

[Figure 3 Major and minor axis lengths of barley seed 33](#_Toc138972751)

[Figure 4 A typical Multilayer perceptron architecture 37](#_Toc138972752)

[Figure 10 **Prototype of malt-barley seed identification model** 42](#_Toc138972753)

[Figure 11 The performance of each feature set using ANN model 45](#_Toc138972754)

[Figure 12 **The correlation between the four varieties of malt-barley seeds** 47](#_Toc138972755)

[Figure 13 The performance of each feature set using KNN model 48](#_Toc138972756)

[Figure 14 **The correlation between the four varieties of malt-barley seeds using KNN** 49](#_Toc138972757)

[Figure 15 **The performance of each feature set using hybrid model** 50](#_Toc138972758)

[Figure 16 **The correlation between the four varieties of malt-barley seeds using hybrid model** 52](#_Toc138972759)

[Figure 17 Model performance comparisons; (a) Model performance in each verity of barley seeds, 53](#_Toc138972760)

[Figure 18 Sample challenging seed image 54](#_Toc138972761)

# LIST OF TABLES

[Table 1 Number of sample malt-barley images taken from each variety 24](#_Toc137567947)

[Table 2 Templates of Sobel mask 26](#_Toc137567948)

[Table 3 Parameters used for ANN classifiers 46](#_Toc137567949)

[Table 4 The performance of ANN model for the selected feature sets 47](#_Toc137567950)

[Table 5 Summary results of malt-barley seeds using ANN with combination of morphology, color and texture features 48](#_Toc137567951)

[Table 6 The performance of KNN model for the selected feature sets 49](#_Toc137567952)

[Table 7 summary results of malt-barley seeds using KNN with a combination of morphology, color and texture features 50](#_Toc137567953)

[Table 8 The performance of Hybrid model for the selected feature sets 52](#_Toc137567954)

[Table 9 **Hybrid of ANN and KNN summary results of malt-barley seeds using a combination of morphology,** 53](#_Toc137567955)

[Table 10 Performance comparison of related works 56](#_Toc137567956)

# LIST OF ACRONYMS

|  |  |  |
| --- | --- | --- |
| EBC |  | European Brewery Corporation |
| GLCM |  | Gray Level Co-occurrence Matrix |
| GMF |  | Gondar Malt Factory |
| JPEG |  | Joint Photograph Expert Group |
| KNN |  | K-Nearest neighbors |
| MATLAB |  | Matrix Laboratory |
| MBIM |  | Malt Barley Identification Model |
| MF |  | Median Filtering |
| MSE |  | Mean Square Error |
| RGB  MLP  ANN  ICT  TP  TN  FP  FN  WTO  SVM  SFFS  FL |  | Red Green Blue  Multi-Layer perceptron  Artificial Neural Network  Information and Communication Technology  True Positive  True Negative  False Positive  False Negative  World Trade Organization  Support Vector Machine  Sequential forward feature Selection  Fuzzy Logic |

# ABSTRACT

*The key step in making beer is choosing the malt-barley. Every malt house requires that the types of grain be checked before being purchased. The creation of premium malt depends on varietal consistency. It might be challenging to distinguish between different malt-barley kinds during inspections because it takes knowledge and practise. To identify several kinds of Ethiopian malt-barley, a computerised image processing technique based on combined morphological, texture, and colour aspects has been investigated. The one local location for taking the varieties is the Gondar Malt Factory, where sample malt-barleys were taken. Each of the four variations yielded an average of 52 photos (Holker, Propino, Sabini and Misikal). A total of 208 pictures were captured, and each one included 5408 malt-barley seeds. Each image of a scanned barley seed was used to extract nine morphological, five texture, and six colour features for identification. K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and the combination of the two techniques are examined for the construction of identification models for the prediction of maltbarley types. According to experimental findings, an ensemble model combining ANN and KNN performs better on combined characteristics of morphology, texture, and colour when utilising the Sequential Forward Feature Selection (SFFS) technique than a single model built using either ANN or KNN. For the Holker, Propino, Sabini, and Misikal varieties, a quantitative accuracy of 86% is attained utilising the ensemble of ANN and KNN with the combined feature sets of morphology, colour, and texture. This indicates a positive outcome for creating a useful malt-barley identification model. Malt-barley photos with non-uniform size and overlap significantly impair the effectiveness of the identifier; as a result, this area of future research requires an examination of general segmentation and noise removal techniques.*

*Keywords****:*** *K-Nearest Neighbor, seed identification, machine learning****,*** *Malt barley, ANN*

# CHAPTER ONE

# INTRODUCTION

## 1.1. Background of the Study

Barley is thought to have been planted in Ethiopia. It makes up 10% of Ethiopia's cropland and is a significant cereal crop [1] [2]. According to how much they are consumed, Ethiopian barley types are divided into food and malt kinds; food barley is grown in the majority of the country. Highlands are the key growing regions for food barley. Several varieties of barley can be used to make a variety of regional specialties and beverages [1]. More than 60% of Ethiopia's high-fat diet, which is a significant source of calories, comes from barley [3]. In the 1950s, studies on barley farming in Ethiopia got under way. To increase grain production potential and stabilize good grain quality, introduce unique germs and regional aggregates [7]. Due to the weak link between research and extension, technologies are still too late to make the desired impact [8].

Machine learning is a technique where computers use specialized algorithms to analyze data, learn from it, and apply what they have discovered to carry out particular tasks [2]. Several fields, including bioinformatics, seed identification, financial market analysis, recommendations, fraud detection, malware identification, and others, find use for machine learning techniques. Building models for training data, which comprises of inputs and anticipated outputs, is the focus of supervised learning. The tagged data is used by these algorithms. Examples include k-Nearest Neighbors, Logistic Regression, Support Vector Machines (SVM), Decision Trees, and Random Forests. Supervised Learning algorithms can be used to solve various problems that can be mainly categorized as identification and regression [3]. Identification deals with differentiating entities based on certain patterns or features of the entities. Some examples include image identification, text identification, handwriting analysis, face detection, spam detection etc. This these is can be reduced to a problem of image identification as we focus on using the text derived from the seed content of malt barley to categorize the seed species. For a given identification problem, the performances of the classifier models are different for different training datasets [1]. Although models such as KNN are known to be used for identification, the accuracy depends on the input data. In previous years there were a number of problems which was affected machine learning due to its‟ computationally intensiveness. Hence, pre- processing is used for improving the visual appearance of images to a human viewer and preparing images for extraction of the features and representations for future analysis [2] [3].

Agricultural goods can be identified, classified, and graded using machine learning, but these processes are typically inconsistent, error-prone, and inefficient, which is why research is focused on replacing human operators with automated systems for faster and more accurate analysis [1]. Breweries and malt factories are currently being produced by Ethiopian and foreign enterprises. Local barley is utilized as the primary raw material in the brewing of malt, which is then used to make beverages.

Due to the improved processing capability of malt barley and the growth of the nation's breweries and breweries, malt factories are now in higher demand [8]. Therefore, malt barley is regarded as one of the cash crops. A model for identifying malt barley is now being developed by researchers to assist local experts. Malnutrition, on the other hand, alters the nature of barley by causing barley seeds to depart from the norm, as do climatic and soil conditions. Preparing a data set is therefore a requirement for the activities to be carried out when a pop-up gene identification model is constructed due to the distortion.

## 1.2. Statement of the Problem

Presently, Ethiopia is home to two malt plants, six major brewers, 12 breweries, and at least 24 different beer brands. While some of these breweries use malt produced domestically, some use malt from elsewhere. The primary ingredient used in the manufacturing of beer is malt. Malt, which is utilized in Ethiopia, accounts for 90% of the cost of all raw materials used to make beer [9]. The cost of importing malt products from outside created economic problems while using local malt products allowed breweries to be produced at minimal cost.

Experts in the field aim to identify malt barley seeds using specific physical and chemical features. The existing ASSALA and Gondar malt plants employ manual procedures for barley seed identification. Although professionals make an effort to distinguish between each species of barley, the process and success of doing so heavily depend on the knowledge and expertise of the local experts. This frequently leads to inefficiency, inequality, and a propensity for mistakes. Yet, barley varietal homogeneity is essential for the production of high-quality malt and beer in the brewing industry and malt plant [12]. Thus, the manufacturers make an effort to find a solution by choosing malt barley seeds, educating experts through providing them with training, and setting up a well-equipped laboratory by investing additional funds.

Outside of Ethiopia, most study on identification and ratings is done. Even the accuracy of the existing algorithm needs to be checked and improved, along with new metaphors for environmental and developing species. To develop such a model, no research has been done based on the researcher's understanding of the present and emerging malt barley species.

Consequently, the goal of this work is to investigate if malt barley seed identification might be automatically developed using machine learning algorithms in the image domain.

To this end, this study seeks to explore and answer the following research questions

* To what extent does the model work in malt-barley seed image identification?
* Which method of preprocessing from noise removal and edge detection technologies, is more effective in producing malt-barley seed images?
* What are the ideal characteristics for building a malt-barley seed identification model?
* Which behavioral methods are most effective for malt barley seed representation?

## 1.3. Motivation

the motivation for malt barley classification using machine learning is to improve the quality and consistency of the malt barley used in beer-making, while also reducing costs, optimizing yields, and promoting sustainability in the agriculture industry.

Malt barley is an important ingredient in the beer-making process, and the quality of the barley can have a significant impact on the flavor and aroma of the final product. Machine learning algorithms can be trained to classify barley based on various quality parameters such as protein content, moisture content, kernel size, and germination rate. Classifying malt barley manually can be time-consuming and labor-intensive. By using machine learning algorithms to automate the classification process, breweries and maltsters can save time and money. Machine learning can help optimize the yield of malt barley by identifying the most suitable varieties for a particular region or climate. By analyzing data on soil type, weather patterns, and other environmental factors, machine learning algorithms can predict which barley varieties are most likely to thrive in a particular location. Malt barley classification using machine learning can also help promote sustainability in the agriculture industry. By identifying the most suitable barley varieties for a particular region or climate, farmers can reduce the use of pesticides and fertilizers, and minimize the environmental impact of their operations.

## 1.4. Research Question

Towards assessing the aforementioned problem, this study attempts to explore and answer the following research questions: -

• What are the features that are more important to describe and represent images of malt barley?

• Which machine learning algorithm is more suitable for malt barley classification?

• To what extent the model is effective in malt barley detection and classification?

## 1.5 Objective of the Study

### 1.5.1. General objective

The general objective of this research work is to construct a predictive model for Ethiopian malt barley seed identification using machine learning techniques so as to facilitate and improve malt- barley seed selection.

### 1.5.2. Specific Objectives

To address the research issues, the following specific objectives are formulated.

* To review different global and local researches to understand the basic concept and technologies of machine learning.
* To collect, organize and prepare different sample images as training and test dataset.
* To select suitable machine learning techniques, feature extraction schemes and identification algorithms.
* To develop a user-friendly GUI prototype for malt-barley seed selection and identification of the model.
* To evaluate the performance of the proposed model using test data set.

## 1.6. Scope and Limitations of the Study

The major goal of this project is to investigate KNN and ANN methods and strategies to build a model for malt-barley identification. The study's major focus was on feature extraction methods, which are essential for malt-barley identification models built with hybrid artificial neural networks and K-Nearest Neighbor (KNN). The extraction of characteristics includes morphological, textural, and color features. Several types of malt-barley gathered from the Gondar Malt Factory have been used to assess the system's performance. Additionally, since malt-barley seed variants are available in malt factories and are identified by domain specialists using various physical and biological properties, non-barley components were not intermingled with the data. Since the accuracy was obtained outside of a lab setting, it was subject to various limitations, including the quality of the scanner, sample procedures, and the environment in which the seeds were grown. Also, the time from the end of August to the beginning of November when Ethiopian malt-barley seeds are produced, and the time from the end of November to the end of February when these new items are offered. Our samples were collected in May, when there is a shortage of some varieties of barley. As a result, there is currently no opportunity to choose samples based on our interests and the likelihood that one may be blended with the other. Experimentation revealed this characteristic. The production of barley seed is further influenced by climatic factors, soil types, fertilizers, and other factors. As a result, this also alters the barley seeds' natural state. Since these variables could lead to the misidentification of malt-barley seeds, preparing the atmosphere for dataset creation was a crucial and difficult task during our experiment. We chose a small number of datasets for the experiment because there isn't a well-organized barley seed picture collection for both training and testing purposes.

## 1.7. Significance of the Study

We anticipate that the target populations in the selected domain region will gain varied advantages from our study activity. The two main players, the malt factories and brewery sectors, have benefited from the findings of this research activity. The Gondar Malt Factory will use the model of malt-barley seed identification as a platform to aid the specialists in malt-barely seed identification. In addition, any brewery industry that employs a Gondar Malt Factory product as an ingredient in the production of alcoholic beverages would receive pure malt derived from just malt-barley kinds. Also, with the aid of the identification model, it enables the expertise to save both time and money. Furthermore, this study can serve as a foundation for future research in the field and will be cited by them. The researcher will also profit by acquiring the required research background and finishing the academic program requirements.

## 1.8. Organization of the Thesis

This research work is organized into the following five chapters:

**Chapter one:** includes the background of the study, a statement of the problem, objective of the study, the scope and limitation, the research methodologies and design, the significance and used to conduct the research. **Chapter two:** covers literature review which gives a detail overview of the study area, various techniques, review of related works and challenges of the domain. **Chapter three:** discusses the newly integrated techniques algorithmically and the design decisions made in the research work. **Chapter Four:** deals with the experimentation activity undertaken to implement the methods and techniques described in chapter three. **Chapter Five:** the dataset used experimental setup and the results of the experimentations and interpretation are discussed. **Chapter Six:** provides conclusion and recommendation research direction for scholars interested in the area.

**CHAPTER TWO**

# LITERATURE REVIEW

## 2.1. Overview

The main crop utilized in the creation of malt is barley. This might have happened historically because it was more readily available than other cereals [10]. The brewing business places a high priority on the identification of barley seed types since variation affects factors such as germination energy, enzyme production, speed of course relaxing, malt performance, protein content, and resistance to microbial contamination, among others. Nowadays, expert is responsible to identify varieties manually using identification keys and tables as the base of the variety diagnosis through which identification is largely dependent on the skills and experience of the evaluator [11] [12]. Therefore, there should be a means to identify barley seed varieties. Designing a predictive model for seed identification is one of the solutions for the problem [11] [12] [13]. A predictive model facilitates the seed purchasing process on the field and allows for a fast and simple procedure for malt-barley assessment and selection [15]. By creating seed image identification models that can convert the provided image into a row by column matrix and compare this result with each other, several attempts have been made to close the gap in the selection and identification of seed verities. For several seed pictures, including those of wheat, maize, corn, barley, and coffee, a seed recognition model has been created [3][19][31]. Barley was domesticated about 10000 years ago from Near East in Iran, and primarily it was used for making alcoholic beverages such as barley wine in Babylonia, 2800 BC [12] [15]. Then after it was cultivated in many dry areas of North Africa, West Asia, Afghanistan, Pakistan, Eritrea, Yemen, Ethiopia, and other countries. Barley is a major staple food crop in the highlands of northern Ethiopia. The crop is used for preparing various types of traditional food such as Kita, Kolo, Beso, Enjera, and for Malt [15]. Based on different biological and physical criteria such as quality, grain yield and the nature of barley can be used either for food or malting purpose [16].

## 2.2. Malt-barley Selection and Grading Technique

In developing countries Malt is the second largest use of barley [6] [17]. In Ethiopia barley is an important cereal crop which covers 10% of the land under crop cultivation [1]. At the current time malting barley is considered as one of the cash crops and its demand by malt factory has increased due to its increased capacity of malt-barley processing and the expansion of breweries and beer consumption level in the country [8] [19]. As long as the barley cultivated and used, it has lost some of its important agronomic and brewing quality (change in shape, color, texture), due to the effect of environmental factors such as soil type, fertilizer usage, rainfall amount and others, which has greater impact for the final added value in crop production [2] [19].

According to World Trade Organization (WTO) crop grading and selection-based standards which are primary depend on origin of the crop, defects, and physical characteristics [3]. In the process of barley selection for malt producers or the brewery industry, the level of impurities must be minimized, varieties, and the evaluation of the kernel size of barley is regarded as the most important step [16] [20].

Barley grain hardness determines the quality of barley so that it is an important characteristic of barley. Malting barley varieties are usually soft, whereas non malting varieties are usually hard. The relationships between hardness of barley grain as assessed using the particle size index and hot water extract of malt as well as the malt quality index of malt-barley. Generally, sound barley grain has a bright light yellow or off-white color. The grain color of barley can vary from light yellow to purple, violet, blue, and black [22] [22].

According to Szczypiński [14], immunological analysis, DNA analysis, high-performance liquid chromatography, protein electrophoresis and enzyme analysis are some of the methods used for cereal grain testing. But these methods are labor intensive and need a specialized laboratory.

## 2.3. Image processing

Image processing is the technique through which images are manipulated in various ways using computer technology; to get clear and enhance image. Image processing improves the visual appearance of images and simplifies the extraction of useful information for further image analysis [1] [24] [24].

The basic steps of image processing can be broadly classified into five stages [3] [36] [39]; Image acquisition, Image pre-processing, Segmentation, Feature extraction and Image identification.

Image processing starts by image acquisition using camera or scanner. Sometimes we may receive noisy images due to various effects. So, in preprocessing stage, the raw image noise and distortion is reduced and the image is refined and simplify for the next stages processing load. After preprocess we segment objects of interest within the image. And then extraction of a set of representative features which are segmented. Finally, based on the extracted features decision is made and each segmented object are identified.

### 2.3.1. Image Acquisition

Image acquisition is the process of obtaining a digital image of object in real world using camera or scanner. During the acquisition process of each image, we may introduce random changes into the values of pixels in the image which is called noise [25]. Atmosphere degradation, motion of image or camera, quality of scanner, distance between camera lens and image are some of the causes for pixel values changes in image acquisition [1] [25].

### 2.3.2. Image Preprocessing

Image preprocessing technique is the preliminary steps for enhancing the quality of raw image [23] [26] [27]. Preprocessing stage includes collection of operations that are applied for successive transformations of image. These are color image to gray-scale conversion, noise removal, binarization, boundary enhancement and edge detection [28] [29] [30]. The details of each preprocessing stage operations are discussed in the following section:

Converting Color Image to Grayscale: color image is an RGB image which is stored as an m-byn-by-3 data array that defines red, green and blue color components for each individual pixel. Gray scale image pixel values range from 0 to 245 [23]. The conversion process of color image to gray scale is done using a Matlab function which is called rgb2gray() and it works by forming a weighted sum of the R, G, and B components; as shown in equation 2.1 below [68].

0.2989 \* R + 0.5870 \* G + 0.1140 \* B……………………………….………………….… (2.1)

Noise Removal/Filtering: Noise occurs due to the random changes into the values of pixels in the image [25]. In digital images we may face a variety of noises types. The most common types of noises are Impulsive noise (thermal noise), Additive noise (Gaussian Noise) and Multiplicative noise (salt and pepper noise) [3] [25] [31]. The cause for thermalnoise are photo-electronic sensors while salt and pepper noise usually the result of an error in transmission or an atmospheric disturbance to occur a black and white color on the image and Gaussian noise contains intensity variations that are drawn from a Gaussian distribution and is caused by camera electronics errors [25]. Gaussian filter and median filter are some of the common methods to reduce such noises [25] [32]. In the median filtering operation, the pixel values in the neighbors are ranked according to the intensity level, and then the middle value which is the median becomes the value for the nearby pixel [45].

Image Binarization: is process of converting either a color image or a grayscale image, to a bi-level image through which each pixel are categorized either as a foreground or a background pixel. And it is an important preprocessing method in image analysis and boundary detection [25]. Image binarization or thresholding is used to distinguish the content from the background.[33]. During the process of binarization, image pixels are classified into two values one and zero, which represents foreground and background classes respectively. Binarization is done based on the image pixel value; for a threshold value T, if the value of image pixel is greater than or equal to the threshold then it becomes one otherwise it is zero [25] [34].

According to Stathis et al [35], bad binarization would reduce the performance of further processing steps and causes a failure. Binarization algorithm can be classified as global and local. The global algorithms such as Otsu, Histogram peaks, K-means and others calculate one threshold for the entire image while the local algorithms like: Kohonen SOM, Niblack, Abutaleb set different threshold values depending on the local regions of the image [32] [35].

Edge Enhancement/Detection: Edges are the boundaries between segments, and edge detection describes the general problem of determining the location, magnitude and segment of boundaries in an image [25]. Edge detection is the process of locating the edge pixels while edge enhancement is the process of increasing the contrast between the edges and the background [33]. In this case, Sobel edge detection strategy is employed. Sobel edge detection is a 3 x 3 neighborhood based gradient operator which uses a vertical and a horizontal mask. Then by applying these two masks separately on the input image to yield two gradient components Gx and Gy in the horizontal, and in the vertical orientations respectively [32].

### 2.3.3. Segmentation

Segmentation is the process of partitioning an image into a set of homogeneous and meaningful regions with set of identical properties or attributes. It can also be defined as a subdivision process of an image, in image analysis, into different parts or objects. Segmentation can be done at both gray level and binary images [25] [34].

During segmentation the image is usually subdivided until the objects of interest are isolated from their background [39]. Segmentation algorithms use two methods that work on gray level values. The first one is based on the discontinuity of gray level values which partitions an image based on abrupt changes in gray level while the second is based on the similarity of the gray-level values which uses thresholding and region growing. The threshold value, which is supposed to be constant in all images with the same lighting conditions, is generated based on the results of the histogram analysis. Therefore, a fixed threshold value determined from the histogram of the gray plane could separate the content of the object from its background [34] [39].

2.3.4. Feature Extraction

Feature extraction is the process of retrieving relevant information that can characterizes an image. It can also be defined as the process of distinguishing the primitive characteristics or attributes of an image. One of the key concepts in image analysis is the extraction of sufficient information which leads to describe the type and nature of image [3] [36].

In the process of images feature extraction, the retrieved information is in the form feature vectors. These feature vectors have an observable impact on efficiency during identification since identifiers use these vectors to identify the input image with target output unit [36].

According to Kumar [36], different features are associated with different feature extraction methods such as Template matching, Deformable templates, Graph description, Projection Histograms, Contour profiles, Fourier descriptors, Gradient feature and Gabor features.

In order to simplify the process of image analysis and reduce time consumption of feature extraction methods, some quantitative information can be extracted from the object’s region of interests so that the computational cost of object analysis is greatly reduced [39] [40].

For object identification or detection, the features should have sufficient information of the object in the image and any irrelevant information should be reduced to avoid redundant knowledge from the extraction and to the identifier. Identification with large information is computationally intensive so that information should be easy to compute [38] [40].

There are several types of image features that have been proposed for image identification; morphology, color and texture are some of the basic image features [1] [3] [14] [23].

Morphology feature: is the geometric property and physical dimension (shape and size) measure of an image [3]. Morphological features are used to characterize the appearance of an image; it can be expressed in terms of area, perimeter, aspect ratio, roundness, and Equivalent diameter. Each morphological feature is computed from a binary image which has a pixel value 1 and 0 to represents the foreground (contents) and the background respectively [3] [47]. Accordingly, area can be computed through counting the number of pixels in an object which has a pixel value of 1 while perimeter is computed from the identified object boundary whose pixels cover an area by counting the number of pixels around the boundary of the object starting at an arbitrary initial boundary pixel and to the initial pixel back [47].

Color feature: is expressed in terms of the level of red, green and blue light. Most of the available algorithms use the geometrical properties such as kernel size for identification which results misclassifications because kernel size depends on different factors such as maturity and growing conditions and also growing region of the crop [14] [40].

The RGB system is the dominating representation of color due to its importance for television and camera systems. Other representations that can be constructed from the RGB values like mean, mode, standard deviation, variance etc. of RGB are used for image identification [39] [41].

Texture feature: is suitable for visualizing pattern and surface properties. It provides information about intensity variation of an image surface [29]. Energy, contrast, entropy, homogeneity, correlation is some of the texture features used in image identification [38] [42] [43].

Morphology, color and texture are the recommended features for crop identification using image analysis [40].

### 2.3.5. Image Identification

Identification is the process of finding a model that describes and distinguishes data classes to predict the class of objects whose class label is unknown [3] [44]. According to Han et al [45], the predicted model is based on the analysis of a set of training data whose class label is known. And then a given new input, whose class level is unknown, is predicted.

In the process of identification, training and testing are the two basic phases. At training phase, the classifier learns all extracted feature and the association between each sample from labeled sample class while testing phase applied data, that it has never seen, to the trained system to check the performance of the classification. A classifier with minimal test error is the desired of most researchers [3] [23] [44].

There are different types of pattern classification techniques which have been used in machine learning for identification of patterns. These classification methods are mainly categorized as supervised learning and unsupervised learning type [23].

During supervised classification, the classifier is trained with a set of training pattern samples and the new set can be classified as a member of known class while in unsupervised classification, the system partitions or clusters the entire data set based on some similarity criteria, each cluster of patterns belongs to a specific class [23] [61].

As diagrammatically shown in Figure 2.1 below, image identification is the process of learning a target function (identification/predictive model) which can maps a set of attributes to one of the predefined class labels. The rapidly growing and available of computing power has facilitated the use of sophisticated and diversified methods for data analysis and classification. At the same time, demanding on automatic identification systems is raising enormously due to the availability of large databases that require high performance system [20][58]. In many of the emerging applications, it is clear that no single approach for classification is optimal. Hence, combining several classifiers is now a commonly used practice in identification [82]. The most commonly used classification techniques are as follow: Neural networks, Bayesian, K-nearest neighbor and Ensemble approach.

#### 2.3.5.1. K-Nearest Neighbor Classifier

K-Nearest Neighbor (KNN) is a machine learning technique where the classification is achieved by identifying the nearest neighbors, for determination of the class of the given sample, based on the calculation of the minimum distance between the given point and other points where the distances are computed using; Euclidean, Manhattan, Minkowski, Supremum and Cosine similarity [60]. During classification, test sample is classified based on the highest number of votes from the K neighbors, with the sample being assigned to the class most common among its K-nearest neighbors. K is a positive integer, which is determined through trial-and-error method among which gives smallest error rate. Then the sample is simply assigned to the class of its nearest neighbor [41].

## 2.4. Review of Related Works

The application of Information and Communication Technology (ICT) has been increasingly implemented as a solution to improve and support different sectors such as health, agriculture, government and education. The idea of integrating ICT with medicine, agriculture and food industry sectors plays a major role for the development of an automated system for identification, identification, recognition, selection and grading of agriculture products [11] [18][24] [27].

Identification of seed varieties (wheat, corn, barley, maize, oats, rice and others), identification and grading of fruits (apple, banana, mango, orange and others), detection of defects and diseases of fruits and seeds are some of the applications of image processing in agriculture [25][26][30].

### 2.2.1. Fruit Varieties Identification

In Malaysia, image processing has been applied for identification and grading of four varieties of fruits (apple, banana, mango and orange) and a vegetable (carrot) with 54 sample images of each fruit type in production year of 2009 [27]. The model was developed using Support Vector Machines (SVMs) for identification and Fuzzy Logic (FL) approach for grading. The developed model was used fruit shape and size as the features and 96.25%, 81.25%, 0%, 98.75% and 6.52% identification accuracies of Apple, Banana, Carrot, Mango and Orange were obtained respectively from experimental results. The result shows that identification accuracies of apples, bananas and mangoes are relatively good while oranges and carrots were misclassified as apples and bananas respectively due to the similarity in size and shape of these fruits. The scholar suggests that to achieve better result, features such as color and texture should be considered [27].

In India, identification and identification model for bulk fruits using color and texture feature were developed [33]. 1000 sample images of five types of fruits (Apple, Chickoo, Mango, Orange and Sweet lemon) in production year of 2011 were taken and Back Propagation Neural Network (BPNN) is used to classify sample images. Best identification accuracy is achieved by combining color and texture features. On the average, 93%, 94%, 92%, 92% and 93% respective identification accuracies were obtained for Apple, Chickoo, Mango, Orange and Sweet lemon fruits, respectively. The performance of the system shows that the accuracy obtained from a combination of color and texture features are out performed the individual color and texture features in identification and identification of different bulk fruit image samples. And also, the scholar suggests that, using combined feature helps to achieve relatively better result than using features separately [28].

In Poland, identification model based on artificial Kohonen-type neural network to classify tomatoes were developed [29]. 9 distinctive input parameters such as width, height, color A, color R, color G, color B, converted ARGB colors, area, Feret shape ratio RF were extracted. To design and generate Kohonen-type neural networks, we had constructed two training sets containing 729 samples each, Tomato1 (tomatoes with stemmed) and Tomato2 (tomatoes without stems). And they conducted experiments using Tomato1 and Tomato2 datasets, with the possible combination of the 9-feature set. The best identification accuracy was identified in the Kohonen network model made with the use of the Tomato1 training set using the 3 inputs of converted ARGB, area and Feret shape ratio RF with training set error and validation set error of 0.1148 and 0.1148 respectively. And also, the scholar suggests that, using non-parametric identification technique performed by the Kohonen method turned is well-suited for the quality-based identification of tomatoes with the use of the graphic information encoded in digital photographs [29].

### 2.2.2. Rice Varieties Identification

Mousavi et al [23] attempts to classify five different varieties of Iranian rice, using a back propagation neural network-based classifier with color and texture features. 100 sample images of each rice kernel varieties were acquired by a flatbed scanner used 8-bit grayscale and at resolution of 600 dpi and stored in JPG image format. From each rice kernels image, they were extracted sixty colors and texture features which was used for identification stage. To optimize the number of features that contributed significantly to their identification model, they were implemented four feature selection algorithms namely: branch and bound, standard sequential forward (SFS), standard backward sequential (SBS), and plus-l-takeaway-r algorithm. And finally, they used the selected 22 features to build the classifier and come up with an overall identification accuracy of 96.67%. Finally, for further research the use of different feature extraction and selection algorithms are suggested for further experimentation.

In China, digital image analysis technique has been applied to discriminate rice varieties [25]. In this work, a two-layer tan-sigmoid/log-sigmoid network was developed to identify six varieties (ey7954, syz3, xs11, xy5968, xy9308, z903) of rice seeds. For experimentation two hundred and forty kernels used as the training and sixty kernels as the test data set were used. To build the proposed model 7 color and 14 morphological features were extracted. Based on this experimental setup they achieved 90.00%, 88.00%, 95.00%, 82.00%, 74.00%, 80.00% identification accuracies for ey7954, syz3, xs11, xy5968, xy9308, z903 rice varieties respectively. Experimentations using large quantity of rice seeds were recommended as future research direction.

### 2.2.3. Wheat Varieties Identification

Zapotoczny [31] attempts discrimination of wheat grain varieties using image analysis and neural networks on images acquired from a flatbed scanner interfaced to a personal computer. The sample images were taken from a grain of common spring and winter wheat of four quality classes (elite wheat, prime quality wheat, bread wheat, forage wheat) and three cultivation years (2004, 2005, 2006), with 11 varieties. A total of 552 sample kernels image and texture feature and wheat images arranged in a specially designed matrix were used for experimentation and they achieve 100%, accuracy. According to their experimental result the production year did not affect the result while as the number of input variables increase identification accuracy were improved.

Other investigations have been conducted to separate wheat from barley in this research work, machine vision system was presented to classify objects between two possible classes. For experimentation 10 images from both varieties which contain 543 total kernel images were taken. Discriminant analysis (parametric classifier based on normal distribution function) and K-nearest neighbor (the nonparametric classifier) were used with three different features (morphologic, color, and texture) and 99% accuracy was achieved from a combination of the three features. From the result that the researchers achieved, it was observed that using a higher number of features increases the computational cost and may also reduce the identification accuracy [30].

In Iran, digital image analysis technique has been applied to discriminate wheat varieties [22]. In this work, various topologies of artificial neural networks (ANN) with different number of neurons in the hidden layers were developed to classify the four varieties of wheat seeds. The experiment was based on multilayer perceptron and morphological features to classify the wheat varieties. To find the optimum number of neurons in the hidden layers, trial and error method was used. The best model was selected with minimum value of MSE. Based on the selected model they were achieved 85.7% overall identification accuracy of wheat varieties discrimination. During experimentation, acceptable identification accuracies were obtained for ``Goha``,``Dehdasht`` and ``Seimareh``cultivars while lowest identification accuracy was obtained for``Koohdash`` and ``Seimare`` cultivars from each other. So that to achieve better identification accuracy in these two varieties, color and texture features were recommended as future research direction.

### 2.2.4. Barley Varieties Identification

In Portugal, image analysis was applied to predict European Brewery Corporation (EBC) barley kernel weight distribution [15]. The researcher collected 830 sample image of Scarlett, Nevada, Esterel and Prestige barley varieties. The barley varieties were digitized using a Hewlett Packard digital camera with 2.048 megapixels. The research was developed using different seed image identification techniques including image acquisition, preprocessing, segmentation, feature extraction and identification. For the determination of the kernel weight, and weight distribution of barley varieties into four different size (<2.2, 2.2–2.5 and >2.5mm classes) morphological feature and multivariate Partial Least Squares (PLS) analysis was performed and regression coefficient of 0.991of weight distribution accuracy for Scarlett and Prestige barley varieties were achieved. But for Esterel and Nevada barley varieties accuracy were not satisfactory since there was a great heterogeneity in most important parameters of each barley variety. Hence, the scholar suggested further experimentation to achieve better prediction accuracy of these two barley varieties [15].

According to Nowakowski et al [11], the technology of image analysis was employed to support and facilitate the process of malt-barley selection. The researchers tried to evaluate the feasibility of image analysis in malt-barley selection and identification using four barley varieties of 1200 images. Artificial neural network classifier with Radial Basic Function Network (RBF) and features like morphology and color were used and achieved on the about 94 % identification accuracy at different level of preprocessing. And the scholars suggested that, further experimentation should do, by using different feature extraction techniques that can extract a best representative feature used in kernels image analysis.

In Poland, neural image analysis was employed to elaborate complete methodology for the identification of varieties, the level of contamination and other visual features of malting barley [34]. Wes tried to distinguish between three barley varieties (Beatrix, Sebastian and Xanadu) which are quite common in their specification. For experimentation, 700 image samples from each barley variety in the cultivation year of 2011 were taken using Nikon D90 camera, and then geometrical parameters, color values, and texture features were extracted. According to their experimental result, 96.7% accuracy was achieved using color feature. For further research the use of more varieties of barley and additional feature extraction techniques and classifier are suggested for experimentation [34].

Another investigation has been conducted to identify eleven barley varieties which were taken from selected farms in the Regions of Poland. The researchers tried to evaluate the feasibility of image driven identification of barley varieties and to identify subsets of attributes with the highest discriminatory ability and to establish whether information related to a kernel's orientation, such as its dorsoventral and germ-brush direction, improves identification performance. In their experimental setup 3 sample images were used from each 11 varieties with 13000 individual seeds in each image. These images were acquired using 400 dpi, Epson 4990 flatbed scanner. Artificial neural network classifier with back propagation algorithm and features like morphology, color and texture were used and achieved on the average 76.5% identification accuracy at different level of preprocessing. The study demonstrated that level of image preprocessing, orientation and maturity of kernels significantly influences identification results. Therefore, the scholars suggested further experimentation at preprocessing and image acquisition stage [12].

In summary, the above studies noted that malt-barley grain size is an important factor regarding the uniformity of malting process in brewery industries [15] [33]. As a result, morphological feature has got more identification performance than color. It was also shown that the texture feature has less discriminating power than the two [3] [11] [32]. And also, the usage of a combination of different categories of features achieves better identification results for variety identification than individual features do. Sometimes as the number of features increase the identification accuracy may decrease so that researchers should care on feature extraction and selection when they use both individual and combined feature for identification of seed varieties [11][12][17][34].

In addition to this, as shown by many researchers’ current works, neural network is widely used due to its high performance in the identification accuracy of agricultural products than other identification techniques [3][12]. In the previous research works, identification accuracies were high when features are distinctly different among tested varieties while features are high similar among groups to be discriminated, the identification accuracies were less than that of distinct varieties. In those researchers work, such problem can be improved by using the hybrid of neural network model and k-nearest neighbor rather than other statistical classifier.

Even though there are many attempts made to facilitate the selection, identification and grading of crops, in Ethiopia, only few attempts are made to develop prediction model for coffee beans identification and flower disease detection. These researchers developed a prediction model to classify and detect only limited botanical region coffee varieties and flowers using Artificial neural network and Bayesian classifier and they need further studies [3]. Most of the identification and grading researches are done outside Ethiopia; for local and emerging varieties, new paradigms should be researched and even accuracy of existing algorithm should be verified and optimized. To the best of the researcher’s knowledge, still there is no any attempt in the area of Ethiopian existing and emerging malt-barley varieties. This initiates the current study to construct a predictive model for malt-barley seed identification using image processing techniques so as to facilitate and improve malt-barley seed selection. It is therefore the objective of the current study to explore different image preprocessing and identification algorithms suitable for malt-barley varieties used in Ethiopia.

# CHAPTER THREE

**METHODOLOGY**

Experts have identified and chosen malt-barley seeds utilizing identification keys and tables as the foundation for the variety diagnostic. Germination energy, enzymes, speed of course relaxing, malt performance, extractives, protein content, and resistance to microbial contamination are all influenced by variation. These malt-barley cultivars are sold separately to the market for the brewing industry. In order to preserve the unique essence of each type, a lot of emphasis was placed on maintaining shipments from several sorts. Hence, choosing malt-barley seeds based on variety is an essential step in the manufacture of beer [11]. As shown in the design process of malt-barley varieties identification using machine learning approaches, the process of malt-barley identification includes several stages, starting with the acquisition of barley seed images and ending with the identification of seeds.

In this study, a model for the process of identifying malt-barley seeds is attempted. Here are detailed descriptions of the implementation process, dataset preparation, and experimental findings.

## 3.1. Dataset Preparation

Since there is no readymade dataset for this type of research, we have prepared our own dataset for training and performance evaluation. To do this, malt-barley seeds have been collected from Gondar Malt Factory.

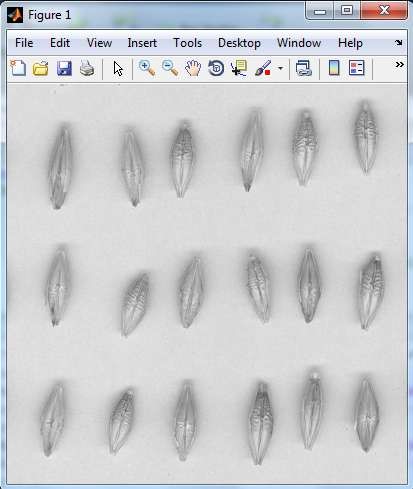
After gathering malt-barley seed, a 300-dpi flatbed scanner, the HP Scanjet 4041, was used to digitize the data. Malt-barley seed scanning was required so that digitization could be carried out and a digital image of the scanned barley seed was created. As flat-bed scanners are inexpensive and simple to use, and since they are less affected by environmental conditions than cameras are, they have been employed for digitizing processes [1]. The image has been scanned with a resolution of 300 dpi and saved in JPG format to ensure clarity.

## 3.2. Preprocessing

The accuracy of identification models highly depends on the effectiveness of the preprocessing stages. The main goal of preprocessing, from the point of view of this research, is to reduce the noise from the image for further analysis of the next phases. The preprocessing stage is a collection of operations that apply successive transformations on an image to modify and prepare the pixel values of the digitized image for subsequent operations. This includes gray-scale conversion, binarization, edge detection, noise removal and size normalization of malt-barley seed image and the details are discussed in the following sub sections.

### 3.2.1. Grayscale Conversion

At this stage, we convert the color image of digitized malt-barley image into grayscale (image which has only one bit value). The sample image which is scanned in RGB format and its equivalent gray level image which are prepared for subsequent processes are shown below in Figure 5 (a) and (b) respectively.

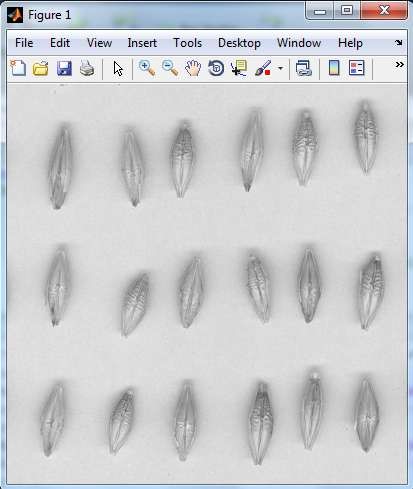
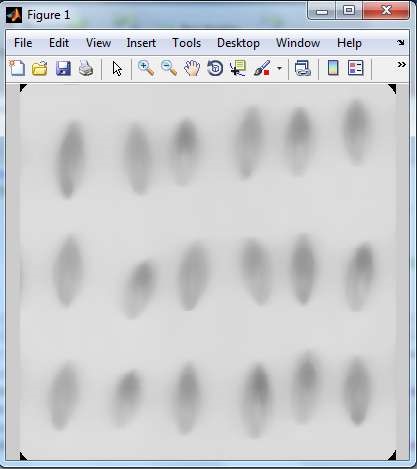
 

**(a) (b)**

Figure 5 Sample scanned Sabini malt-barley seed; (a) RGB image, (b) Gray level image

### 3.2.2. Noise Removal

There are several techniques which are employed for noise removal in image processing [21]. These are Median Filter, Gaussian filter, Prewitt and Adaptive Median filtering. In order to clean noise from malt-barley seed images, we consider the median filtering techniques since it offers advantages such as; no reduction in contrast across steps, it does not shift boundaries, and it is less sensitive than the mean to extreme values or outliers, it preserves edges while removing noise and for certain types of random noise, they provide excellent noise-reduction capabilities, with considerably less blurring over the others [20] [21]. In addition to this reason, it has good capability of removing impulsive (salt-and-pepper) noise [43]. Median filter replaces the value of a pixel by the median of the gray level values in the neighborhood of that pixel. This technique is implemented using a MATLAB function **medfilt2 (I, [m n])** performs median filtering of a gray image **I**, with each output pixel containing the median value in the m-by-n neighborhood. **Medfilt2()** pads the image with 0s on the edges, so the median values for the points within [m n]/2 of the edges might appear distorted. The sample malt-barley seed image, after the implementation of median filtering, in gray level format and its equivalent filtered image which are prepared for subsequent processes are shown below in Figure 6(a) and (b) respectively.

**(a) (b)**

Figure 6 **Sample Sabini malt-barley seed image; (a)** gray level before MF applied, **(b)** gray level image after MF applied

### 3.2.3. Binarization

After we applied filtering algorithm to remove noise and enhance the quality of the image, we convert the filtered grayscale malt-barley seed image into binary (0s and 1s) format which is binarization [1]. Since filtering improves the performance binarization it was applied first. After binarization, we also tested median filtering. But the result is the same as the binarized image. Filtering techniques should be applied to gray scale or color images since this type of images contains different levels of intensity.

A grayscale image is converted into a binary image according to whether its gray value is greater than or less than threshold value T. Since, no fixed threshold value for all images in different application domains [1], threshold value should determine for each application using thresholding techniques. Hence, in this study we implement Otsu thresholding techniques, which choose the threshold to minimize the intra-class variance of the black and white pixels [45]. According to equation 3.2, for an input filtered grayscale image with a threshold value T, the pixels become 1 (white) if gray level greater than threshold T otherwise 0 (black).

Otsu binarization technique is implemented using the built-in MATLAB function **graythresh().** This function computes a threshold level which can be used to convert an intensity image to a binary image with function **im2bw().** A threshold level is normalized and its value lies between zero and one [1]. The function **bw = im2bw(I, level)** converts the filtered grayscale image I to a binary image.

The sample binarized malt-barley seed image, after the implementation of Otsu thresholding technique, on filtered gray level format is shown in Figure 7 below.

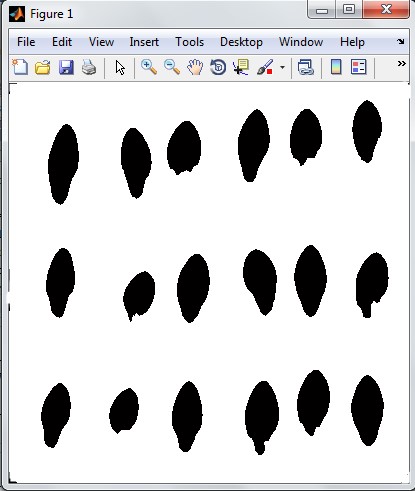
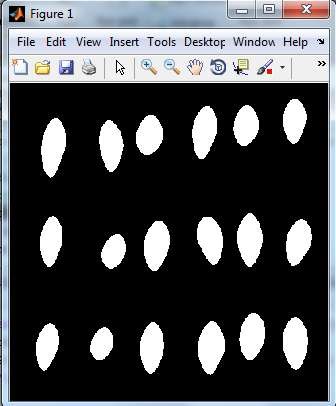
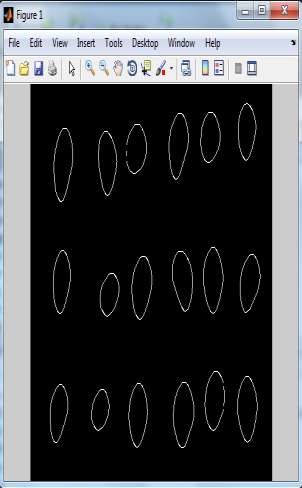


Figure 7 Sample binary Sabini malt-barley seed image

### 3.2.4. Edge detection

Edge detection is the process of localizing pixel intensity transitions or it is a process of finding sharp discontinuities in malt-barley seed image. The edge detection has been used by object recognition, target tracking, segmentation, and object identification. Hence, the edge detection is one of the most important parts of image preprocessing [35] [42].

Different operators were used for edge detection such as Prewitt, Sobel, Laplacian of Gaussian, and Canny. In this work, we use Sobel edge detection since it gives more sharp and clear edges as compared to other operators and it is simple to use [42]. Sobel edge detection strategy is a 3 x 3 neighborhood based gradient operator. The Sobel edge detector uses two masks, one vertical and one horizontal. The Sobel method uses the derivative approximation to find edges. Therefore, it returns edges at those points where the gradient of the image is considered maximum. The sample segmented Sabini malt barley seed image and their equivalent edges, using Sobel edge detection, are depicted below in Figure 8 (a) and (b) respectively.

**(a) (b)**

Figure 8 **Sample Sabini malt-barley seed; (a)** segmented image, **(b)** edges of each segmented image using Sobel edge detection

## 3.3. Feature Extraction

The features of images and the extraction method used during experimentation are decided by the output of the image-preprocessing steps. In the traditional system, human vision attempts to identify malt-barley varieties using structural forms like shape and size, their visual color differences and intensity variation of their surface. Hence, the current identification model is proposed based on morphology, texture and color analysis, which considers an assessment of human visual inspection and physical characteristics of each malt-barley seeds as starting point. Based on this, we extract 20 features from each malt-barley seed image. These are morphological (9 features), textural (5 features) and color (6 features) totally 20 features.

The function **regionprops (concomp, properties)**, measures a set of morphological properties for each object in connected component **concomp** like Area, Perimeter, Major and Minor Axis Length.

The function **graycoprops (GLCM, properties)**, measures a set of textural properties from the graylevel co-occurrence matrix **GLCM** like Contrast, Correlation, Energy, Entropy, and Homogeneity.

The function **rgb(:, :, c)** computes Red, Green and Blue components from the RGB images where the value of **c** is 1,2 and 3 respectively. And the value of Hue, Saturation and Intensity is computed by using equations described in section 3.4.2. Then the color features used for experimentation are computed from the mean value of each color. The main reason to compute the mean of each color is the existence of pixel value variation of individual color across the image [3].

The sample features, extracted are depicted below in Figure 9.

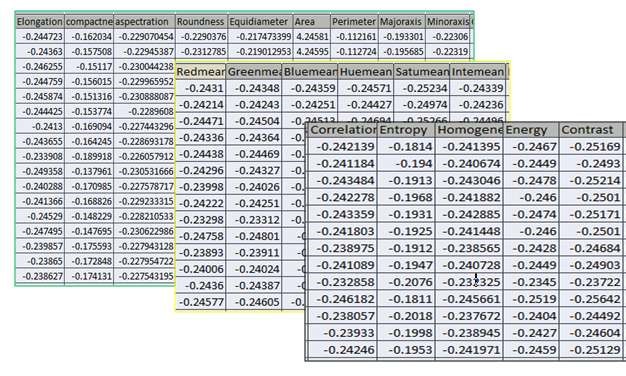


Figure 9 **Sample features computed from malt-barley seeds**

All the feature values are measured in pixel and normalized to the mean and standard deviation of each row for simplicity of processing. After we have extracted those 20 features, we have used each feature for identification separately as an input for the identification model and also all the possible combination of each of the three already extracted features of malt-barley seeds were used during experimentation. Sequential Forward Feature Selection (SFFS) technique is used to select the most dominantly discriminated features during combination.

## 3.4. A General Approach for Building Identification Model

A general model for identification of malt-barley seeds image is depicted above section. This model can be explained in four phases which are data set partitioning, model building, model testing, and finally we measure identification model performance.

### 3.4.1. Data set Partitioning

In order to build malt-barley seed image identification model the dataset is divided in to training and testing set, using percentage split technique, 80% of the total images (100 images) are used for training, and the remaining 20% (17 images) for testing. Such percentage splitting technique is done using cross validation technique which is recommended in order to avoid problems, like overfitting during training [39]. In k-fold cross-validation, the training set is divided into k subsets of equal size. In this study the dataset is partitioned into 5 subsets that are roughly of the same size.

### 3.4.2. Building Model

our identification model is constructed by feeding the training data set which was partitioned in data set partitioning phase. In order to build the model, we have used k-Nearest Neighbors (KNN).

## 3.5. Design of Malt-barley Identification Model

Designing a malt-barley seed image recognition model for Ethiopian malt-barley seed variants is the focus of this research project. The photos of the digitalized barley seeds are first preprocessed to convert to gray scale, get rid of noise, and find edges. The next step after preprocessing a picture is segmenting it into a collection of meaningful, homogeneous sections that share a set of characteristics. The segmented images are then given to the stage of feature extraction. At the feature extraction stage, features of the segmented barley seeds are extracted. And then the best features are selected using forward feature selection techniques. After this step, the best features which are selected by the forward feature selection technique are delivered to the identification stage in order to classify the barley seeds into the correct class using K-Nearest Neighbor (KNN). Finally, the identified malt-barley seeds are aligned to a varietal class that it belongs to. To identify malt barley varieties a series of steps are required. The details of these steps, main focuses of this study, are depicted in Figure 1 below.

**Scanner**

Malt barley seeds

predictive model

**Digitization**

**Identified seed varieties**

**Testing segment**

**Barley image**

**Processed image**

**Segmented image**

**Selected features**

**Training segment**

**Barley image**

**Processed image**

**Segmented image**

**Selected feature**

**Identification**

KNN, ANN, HYBRID

**Malt**

**-**

**barley**

**Varieties**

**Identification**

**Segmentation**

**Feature extraction**

**Preprocessing**

Grayscale

Noise removal

Edge detection

Binarization

Labeled Malt

-

barley

seed image

Scanned Malt

-

barley seed image

Figure 1 Architecture of the proposed model for Ethiopian malt-barley identification

### 3.5.1. Image Acquisition

In seed identification, digitization enables to produce a digital image of the scanned seeds in the form of JPEG (Joint Photographers Expert Groups) file format. Acquiring images from malt-barley seeds can be done using scanner or digital camera [1] [44]. In this study, the device used for image acquisition is a flat-bed scanner instead of digital camera. This is because flat-bed scanner is a cheap alternative and factors such as lighting, movements, and winds have lower effect than in camera [1]. So that it helps to record clear image with minimal noise level.

For this study, images of four malt-barley seed verities (Holker, Propino, Misikal, and Sabini) used in Gondar Malt Factory, Ethiopia, are considered. During acquisition the varieties of the sampled barley seeds were certified by the domain experts in the Factory`s laboratory. All sampled variety seeds were the products of 2014/2015 production year and the samples were taken in the first week of May 2015. The total number of images taken for this study, from each variety; Holker, Propino, Sabini and Misikal, are shown in Table 1 below.

Table 1 Number of sample malt-barley images taken from each variety

|  |  |  |
| --- | --- | --- |
| **Varieties** | **Total number of images** | **Total number of barley seeds** |
| Holker | 67 | 1742 |
| Propino | 45 | 1170 |
| Misikal | 46 | 1196 |
| Sabini | 50 | 1300 |
| **Total** | **208** | **5408** |

All malt-barley seed images, for training and testing the identifier, are collected from Gondar Malt Factory. Among these, Figure 2 depicts sample Holker, Propino, Misikal and Sabini malt-barley seed images taken at 300 DPI.



Figure 2 Sample malt-barley seed images [47]

### 3.5.2. Image Preprocessing

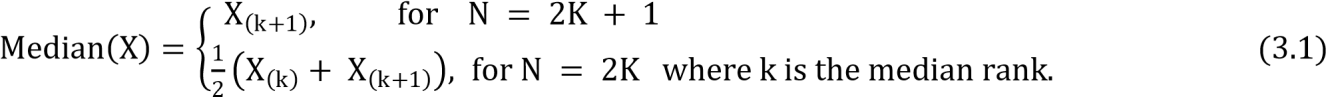
Image processing is a technique that manipulates images in various ways to enhance image quality [3]. Image preprocessing stage consists of several sub-stages; the detail of each sub-stages has been described in section 2.3.2.

Image preprocessing stage takes raw image as input and produce enhanced image as an output. In this study, there are different preprocessing algorithms used for noise removal, binarization and edge detection. The preprocessing algorithms which will be used during the implementation of this study are described in the following sub sections.

#### 3.5.2.1. Noise removal

Noise reduction or smoothing is one of the most important processes in image processing. Images are often corrupted due to positive and negative impulses stemming from decoding errors or noisy channels. Images are often degraded by noises. Noise can occur during image acquisition or transmission [26]. Noise removal is an important task in image processing. In general, the results of the noise removal have a strong influence on the quality of the image processing technique. Several techniques for noise removal are well established in image processing [34]. The nature of the noise removal problem depends on the type of the noise corrupting the image. In this research, we consider the median filtering techniques since it offers advantages such as, no reduction in contrast across steps, it does not shift boundaries, and it is less sensitive than the mean to extreme values or outliers [33][34].

According to Kirchner and Fridrich [46], for a given set of random variables, X = (X1, X2, - - -, XN) with the order X1‟, X2‟, ---, XN‟ are random variables, which sorted in an increasing order. The median value can be calculated as shown below in equation 3.1.



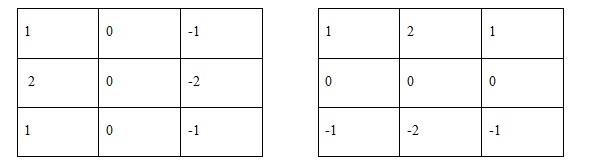
The median operator is usually implemented using a template. The algorithm of median filtering for 3 × 3 template is depicted in Algorithm 3.1 [33].

#### 3.5.2.2. Edge Detection

Edge is the line separating two places having relative different gray level characteristics. Edge detection is the process of locating abrupt changes in pixel intensity that define the boundaries of objects in a scene, or it is the process of locating sharp discontinuities (i.e., pixel intensity transitions) in an image. Target tracking, segmentation, object identification, and object recognition have all employed edge detection. Hence, one of the most crucial aspects of image processing is edge detection [45] [40] [44].

The Sobel edge detection approach is a 3 x 3 neighborhood-based gradient operator, according to Nixon and Alberto [33]. One vertical mask and one horizontal mask are used by the Sobel edge detector. Separate applications of the two masks on the input image result in two gradient components, Gx in the horizontal direction and Gy in the vertical orientation. These two mask templates are displayed in Table 2, respectively, (a) and (b).

Table 2 Templates of Sobel mask

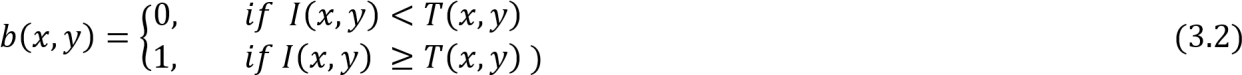


(a) Templates for horizontal orientation (GX) (b) Templates for vertical orientation (GY)

Different operators were used for edge detection such as Prewitt, Sobel, Laplacian of Gaussian, Zero cross, Canny and others [39][40]. In this study, we use Sobel edge detection since it performs better in edge structuring and orientation in noise environment so that it gives more sharp and clear edges as compared to other operators and it is simple to use [44][46].

#### 3.5.2.3. Binarization

Image binarization is a process of separation of pixel values of an input image into two-pixel values; black as foreground and white as background. Hence, it is an important step of preprocessing method in image identification and edge detection. It is difficult to select the fixed threshold value for all images in different application domains hence each thresholding techniques are associated with a certain intensity value called threshold [1]. To separate the pixel values of the input grayscale image into background and foreground, each and every pixel should be compared with the threshold value and transformed to its respective class [3]. Thresholding is a process of finding an appropriate threshold value for binarization. For an input grayscale image with a threshold value T(x, y), the output binarized image b(x, y) is as follows [1] [33].



Where I(x, y)  [0, 1] and is an image pixel intensity at location(x, y).

Based on equation 3.2, b(x, y) = 1 represents the object (foreground) pixels and b(x, y) = 0, represents the background pixels. 

### 3.5.3. Feature Extraction

Feature extraction is the process of retrieving meaningful information from an image that is used for identification /identification of images to different categories [1]. The feature of images and the extraction method decide the nature and the output of the image-preprocessing and segmentation steps. As far as the format of the extracted features match the requirements of the classifier, the image features and extraction methods work on color images, gray level or binary images [46]. For identification of malt-barley seed image, three identification parameters are identified. Hence, the proposed predictive model in this study is based on, domain experts, an assessment of human visual inspection and physical characteristics of malt-barley seeds as starting point. The identified features are morphological, texture and color features. The details of these features are discussed in the following section.

#### 3.5.3.1. Morphological Feature

The geometric properties of images, such as area, height, width, perimeter, maximum diameters, minimum diameters, and others, are known as morphology features [15]. In this instance, it alludes to the malt barley seeds' main axis, minor axis, area, aspect ratio, elongation, compactness, equivalent diameter, and roundness.

Morphological features extracted from the barley seeds which are calculated from the basic feature are discussed as follows [3] [23] [24] [45] [43]:

1. Major Axis Length (MaxL): It is the distance between the end points of the longest line that could be drawn through the barley seed. The major axis end points are found by computing the pixel distance between every combination of border pixels in the barley seed boundary and finding the pair with the maximum length.
2. Minor Axis Length (MinL): It is the distance between the end points of the longest line that could be drawn through the barley seed while maintaining perpendicularity with the major axis. The visualization about MaxL and MinL distances are shown in Figure 3 below.

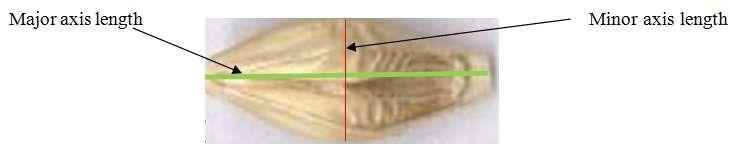


Figure 3 Major and minor axis lengths of barley seed [47]

1. Area (A): The number of pixels inside the region covered by a barley seed, including the boundary region.
2. Perimeter (P): The length of the outside boundary of the region covered by a barley seed.
3. Aspect Ratio (AR): it is the ratio of the length of the major axis to the length of the minor axis.
4. Elongation (E): - is the inverse of Aspect Ratio which is calculated as a ratio of minor axis length to major axis length.
5. Compactness(C): is the ratio of perimeter square to the area of the object.
6. Equivalent diameter (Deq): is the diameter of the circle with the same area as the object and is expressed as:



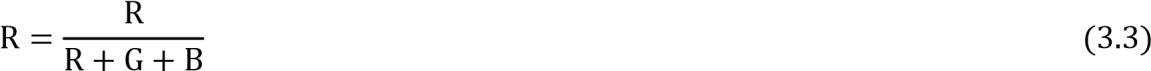
1. Roundness(R): Measures the degree of roundness or circularity. It is calculated as:

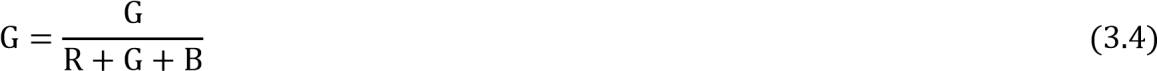


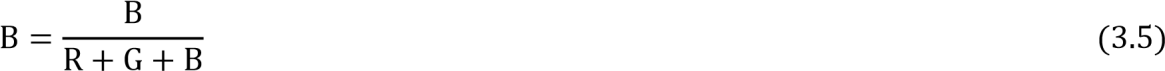
#### 3.5.3.2. Color Feature

According to Littman and Ritter [43], in television and camera systems, the Red (R), Green (G) and Blue (B) system is the dominating representation of color. And this representation is similarly realized in the human retina that consists of three color-sensitive photoreceptor types with a maximum spectral sensitivity corresponding to red, green, and blue.

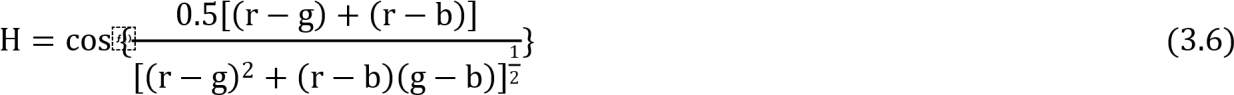
The default images taken from the camera are in RGB color model and these RGB components are separated from the original image. The extracted colors are calculated from the normalized components of R, G and B where the normalized colors are Red(R), Green (G) and Blue (B) as shown in equation 3.3-3.5 below [43][45].

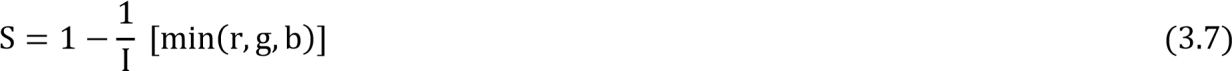


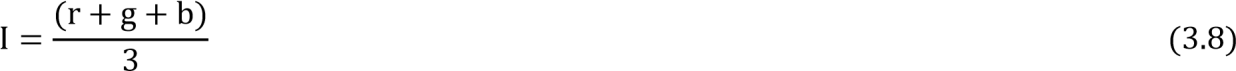




HSI colors are the common perceptual descriptors of a light sensation; Intensity is a measure of the brightness while saturation is the amount of whiteness of a light source in a given image. The hue is also an attribute of light that distinguishes one color from the other; for example, a red color from green or yellow color [1]. The HSI components extracted from RGB components are calculated using equation 3.5-3.8.





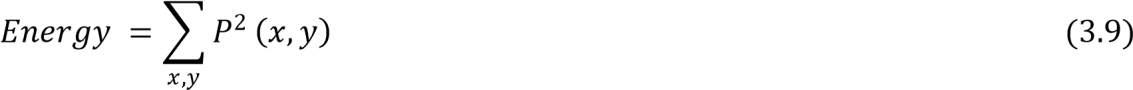


In our case we consider a total of 6 color features which are the mean of both RGB and HSI color components.

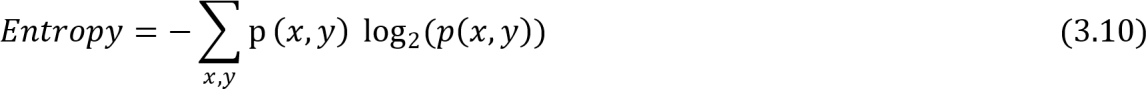
#### 3.5.3.3. Texture Feature

Texture feature is a basic characteristic of image that is related to its measures of properties such as smoothness, coarseness, and regularity of pixel structure [41]. An image segmentation, identification and interpretation are used to get useful information such as energy, entropy, contrast and homogeneity. Textural features can be defined and calculated using equations 3.9-3.13 below [41] [42] [45] [43].

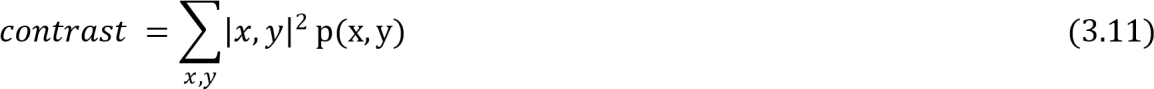
1. Energy: it measures the concentration of intensity pair in co-occurrence matrix and is calculated as:



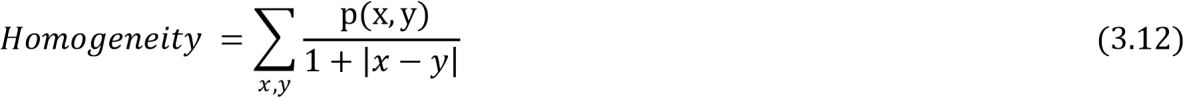
1. Entropy: it is used to calculate the degree of randomness of intensity distribution and is calculated as:



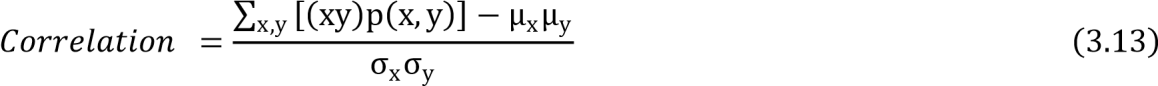
1. Contrast: it measures the strength difference between intensity in an image and is calculated as follows:



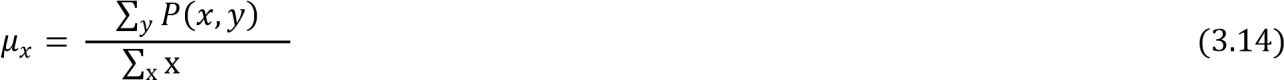
1. Homogeneity: it measures the homogeny feature of the intensity variation within the image and calculated as:

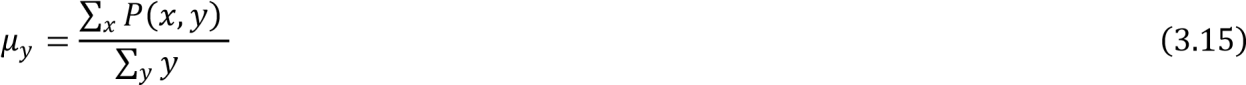


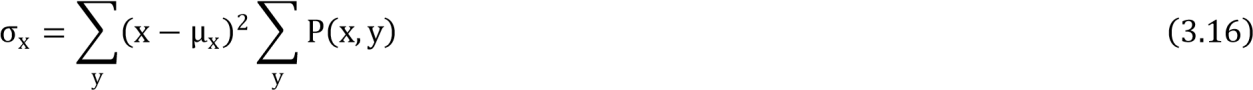
1. Correlation: the connection or link between intensity variation within the image and is calculated as:

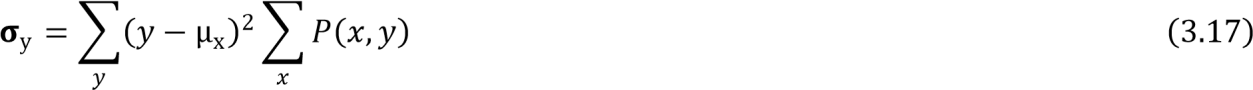


Where **p** denotes the number of occurrences of gray levels within a given image, which shows the value of the element within co-occurrence matrix, while **x** and **y** show the intensity couple from the neighboring intensity. This neighboring couples in co-occurrence matrix act as row and column matrix and **μx**, **μy** are means and **σx**, **σy** are the corresponding standard deviations. And the means and standard deviation can be computed as follows:









#### 3.5.3.4. Sequential Forward Feature Selection

Sequential forward feature selection algorithm is often used to reduce feature space dimensionality in pattern identification so as to enhance identification performance due to the removal of noisy or unreliable features [46]. Generally, feature selection algorithms are used to reduce computational costs of model training and identification since the dimensionality of extracted features is reduced through the discovery of relevant and irrelevant features [44] [45].

The procedure of sequential forward feature selection starts with an empty set of attributes as the reduced set. Then the best of the original attributes is determined and added to the reduced set. In each of the iterations the best of the remaining original attributes is added to the set sequentially until the performance starts to decrease.

## 3.6. Constructing Identification Model

There are various types of identification algorithms, some of the common identification algorithms used in image processing and data mining are: neural networks, rule-based classifier, support vector machines, hybrid techniques, nearest neighbor classifier and naïve Bayes classifier [41][42][43]. The details of the algorithms that are proposed in this research are presented in the following sections.

### 3.6.1. Artificial Neural Network Classifier

Artificial neural network is a biologically inspired computational model formed from several of single units, artificial neurons, connected with weights which constitute the neural structure. These neurons are organized in layers so that every neuron in a layer is exclusively connected to the neurons of the preceding layer and the subsequent layer [14] [39]. ANN is similar to human brain; hence, it emulates some of the amazing working powers of brain and they termed as neural networks. ANN acquires knowledge with the network through a learning process and store the knowledge in the form of weights which is known as the strengths of interneuron connection [3] [14] [24] [38].

Among most identification algorithms neural network classifier with back propagation algorithms is the most commonly used because it is robust, user friendliness, can handle noisy data and well suited to analyze complex problem [41][42].

There are many types of neural networks designed by researchers frequently but all can be described by the transfer functions of their neurons, by the training or learning algorithm, and by the connection formula [3] [4]. The details are presented as follows:

#### 3.6.1.1. Architecture of Artificial Neural Network (ANN)

The architecture of ANN is the arrangement or layout of the network which describes the pattern of connections between neurons [28]. Even though there are no easy answer what arrangement or architecture to be used for a particular problem, neurons can mostly be arranged in multilayer fashion where the choice of the size of layer depends on the type of problem [38] [43]. Figure 3.4 depicts a multilayer neural network with input, hidden and output layers [38].

In a Neural net with Multilayer Perceptron (MLP) the Input layer is used to communicate with the external environment that presents an input pattern to the neural network and the output layer presents a pattern to the external environment. The third layer in MLP is hidden layer which act as an intermediate layer between the input layer and the output layers [38] [43].

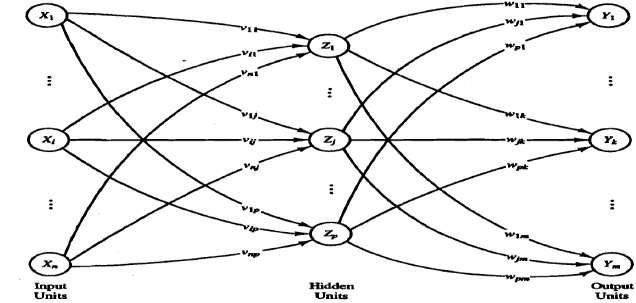


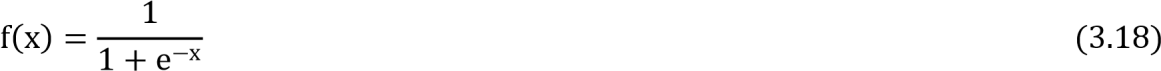
Figure 4 A typical Multilayer perceptron architecture

#### 3.6.1.2. Learning paradigms and algorithm for ANN

Learning paradigm is just a method of setting the values of the weights or simply labeling of training. It is important to model the environment in which the neural network operates to learn or train the system. Based on this, networks can be supervised or unsupervised. In supervised learning, the desired output is available for all of the samples needs to be trained. In unsupervised learning, the system must determine class structure, mainly the optimal numbers of classes and their properties [41] [38] [43].

In general, learning in ANN indicates the adaptation of weights between the connections of two neurons [3]. In this research work back propagation is used as a learning algorithm since it is popular, it avoids cross talk (occurs when the input vectors are not orthogonal, instead of recalling the associative target, the response will include a portion of each of their target values.), minimize the total squared error, a backpropagation algorithm is a supervised type of learning algorithm which operates by propagating an identification errors from the output layers back towards the input layers and modify the weight to minimize the occurrence total errors [41] [38][43]. The algorithm for backpropagation is depicted in algorithm 3.5 [44]. This algorithm involves three stages: the feed forward of the input training pattern, the calculation and backpropagation of the associated error and the adjustment of the weights [41] [38].

As described in equation 3.18 below, sigmoid function is used as an activation function since activation function used for a backpropagation net should be continuous, differentiable, and monotonically nondecreasing [24] [38].



Where **x** is an input to the function and e is epsilon

In summary, the input features are the morphological, color and texture features and the outputs units are the available malt-barley verities (Holker, Propino, Sabini and Misikal). Depending on the selected features and the situation in the experiment that will be conducted, the value of input units will be between one and the maximum number of selected features. To determine the number of hidden layers, several methods are used till now but none of them has been provided the exact formula for calculating the number of hidden layer as well as number of neurons in each hidden layer [38]. In addition to the number of hidden layers, ANN mainly depends on activation function, initial weight and bias, and learning rate [38] [38]. Hence, after attempts of conducting a serious of experimentations the numbers of neurons, in the hidden layers and the value of each parameter are decided as a best fit.

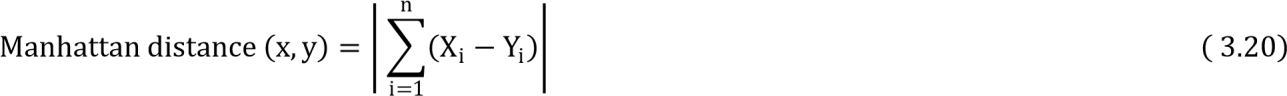
### 3.6.2. K-Nearest Neighbor Classifier

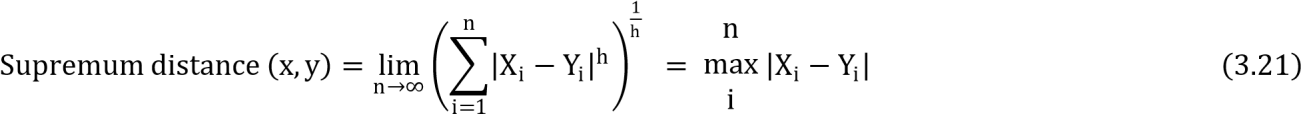
K-Nearest Neighbor (KNN) is a statistical classifier that focuses on similarity of samples measured by a distance metric. It is a machine learning technique where the identification is achieved by identifying the nearest neighbors to query examples and then make use of those neighbors for determination of the class of the query based on the calculation of the minimum distance between the given point and other points and it assigns data to the most represented category within its closest k neighbors [43] [44].

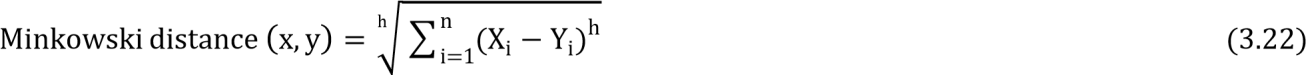
KNN method does not include any training process. Hence, it is easy to implement and also it is possible to score good results if the features are chosen carefully. Therefore, it works well on basic identification problems [43] [44].

Nearest-neighbor classifiers search the pattern space for the k training tuples that are closest to the unknown tuples. The value of k is determined experimentally, starting with k =1. This process can be repeated each time by incrementing k to allow for one more neighbor, using a test set to estimate the error rate of the classifier. Then k value that gives the minimum error rate is selected. In general, for the larger number of training tuples, the value of k will be the larger [43].

These k training tuples are the k nearest neighbors of the unknown tuple and its closeness is defined in terms of a distance metric, such as Euclidean, Manhattan, Supremum, and Minkowski distance [43]. Each distance that can be calculated between two points, X = {x11, x12... x1n} and Y = {y11, y12... y1n} is shown below.







Where i run from 1 to n, and h is a real number greater or equal to one, X and Y are tuples in which their distance to be calculated. As shown in equation 3.19 to 3.22 above, KNN uses these formulas to calculate the minimum distance between two points for classifying objects in to its closest neighbors. In this research work we consider only Euclidean distance which is more popular [43].

### 3.6.3. Hybrids of Artificial Neural Network and K-Nearest Neighbor

Hybrid methods are learning algorithms which is constructed from a set of classifiers. A single identification model, in most of the emerging applications, doesn’t behave efficiently due to this it is difficult to build a machine learning device, one which would give the best possible answer every time [35]. An identification system which aims at utilizing a best individual classifier have some drawbacks and some classifiers work better in some kinds of application in the presence of specific sorts of noise. Hence, multiple identification algorithms have to be combined together giving result to hybrid models which have relatively better identification accuracy [34] [44].

It is very difficult to identify a best classifier unless doing a deep experimentation on different data set or having prior knowhow about it. Using different feature sets and different identification techniques, identification systems have different performance [46]. To enhance the identification performance of a system, one can use a set of individual classifiers and combiner to make the final decision. Multiple classifiers can be used in several ways to enhance the system performance. Each classifier can be trained in a different region of feature space and can provide probability estimation and decision based on the analysis of individual results [44]. There exist numerous methods for model combination; the linear combiner, the product combiner, and the voting combiner are the most commonly used methods. Though a combiner could be specifically chosen to optimize performance in a particular application [34] [35].

The identification performance of multiple classifiers not only depends on the combination strategy, but also relies on the competency, complementariness, and the number of the base classifiers. By analogy, when designing a hybrid is the same as when establishing a committee of people; number of committee members may differ according to the duty that they composed for and each member of the committee should be as competent as possible, but the members should be complementary to one another. If they always agree, then the committee is unnecessary, any single member is sufficient. If the members are complementary, then when one or a few members make an error, the probability is high that the remaining members can correct this error [9] [46]. Identification of pixels in image and identifying their relevant class is mainly depends on the feature extraction and classifier selection process [39].

The hybrid of K-Nearest Neighbor ((KNN) and Multilayer Perceptron (MLP) classifier models is built to identify the type of malt- barley seed images. During hybrid learning, instead of generating multiple models, the hybrid passes the training set to each of its multiple base models, obtains their predictions, and then combines them in some appropriate manner [44] [46].

## 3.7. Malt-barley seed Identification Model

Based on the features extracted from the malt-barley seed image, one supervised machine learning algorithms (KNN) and there are used for training and constructing the identification model for two barley seed images. The total data set is splitted into training and test set using percentage split where 80% is taken as training dataset and the remaining 20% as the test set. In this research work, morphological, color and texture features, and their combination are used for training purpose. For feature combination, we use a sequential forward feature selection method which can select the best feature set through which features are sequentially added to an empty candidate set until the addition of further features does not decrease the performance.

The prototype of Ethiopian malt-barley seed identification model, depicted in Figure 4.6 is designed and developed using MATLAB R2014a with the following components. The detail of components of the prototype and their corresponding functionalities are described as follows:

**Load Image Button:** the model starts by loading a malt-barley seed image from the path of the image file where it is stored, into the Axes which used to display the image.

**Image Preprocess Button:** after the image loaded to the Axes then all the preprocessing activities are performed using this button. When the Image Preprocess button has been pressed different preprocessing functions such as imResize(), imGrayscale(), imNoiseFilter(), binarization() and imEdgedetction() are invoked. In this stage, these functions ready the images so as to feed into feature extraction phase.

**Feature Extraction Button:** After the images are preprocessed, Feature extraction is taken place on the processed images. When the Feature Extraction button is pressed the function of the regionprops(), graycoprops() and rgb(:, :, c) are instantiated.

**FSS Button:** used to select the best feature so as to combine and prepare best predictor. Hence, these features will be employed for identification using feature selection push button.

**Feature Selection Push Button:** the identification features are selected from the push button and used as a identification criterion of the model.

**Identification Algorithm Push Button:** the identification algorithm is selected from this push button and then pressed to initiate the function of the selected algorithm.

**Test Button:** This button is used to activate the model built by the selected algorithm which is used for testing and then confirm the actual predictive power of the selected algorithm.

**Radio Buttons:** these buttons are viewed in view state as a group and helps to view the state of the image at different preprocessing stages; for example, **RGB** to view the RGB image, **filtered** to view the binary image after nose removal and **GrayScale** radio button to view the image at gray level.

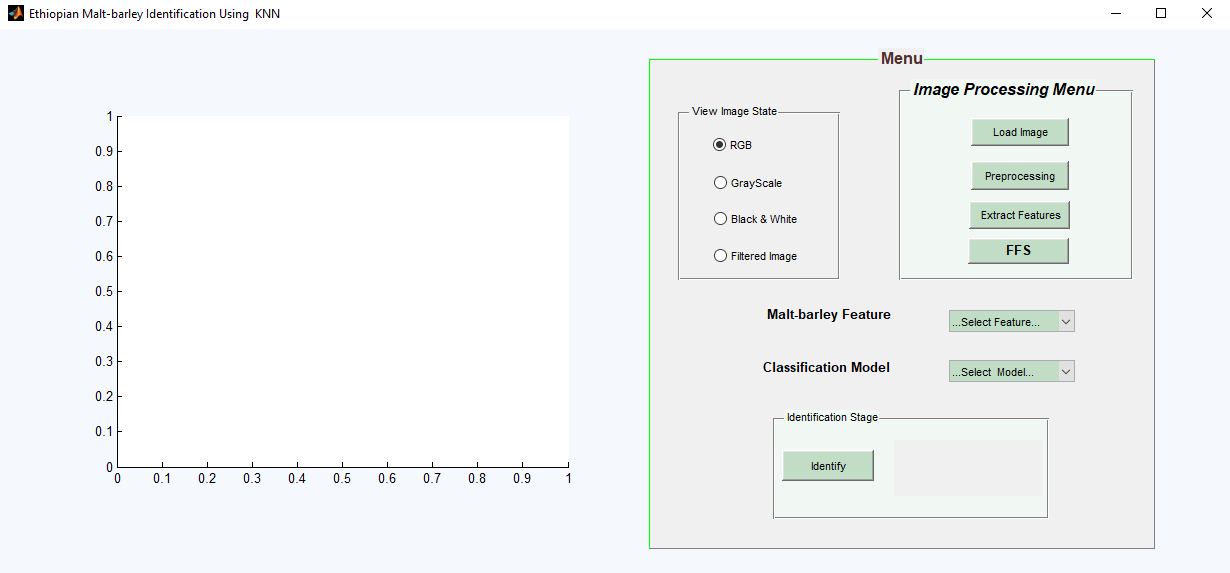
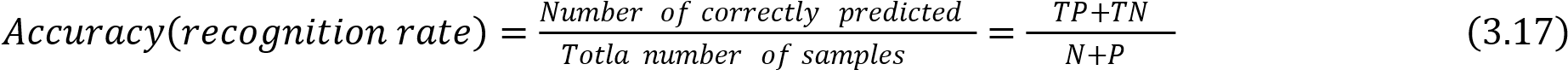
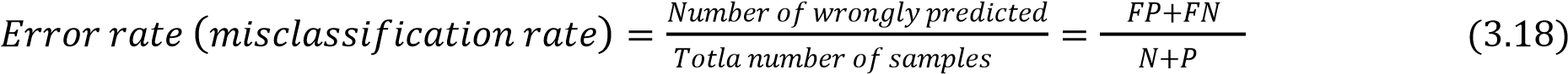


Figure 10 **Prototype of malt-barley seed identification model**

## 3.8. Evaluation Techniques

The performance of identification model constructed in this study is evaluated based on the counts of test records which are correctly and incorrectly predicted by the model. Based on this, the performance of each individual classifier and their hybrid can be evaluated by using statistical measures such as accuracy and error rate. Correctly and incorrectly classified counts of the test model are tabulated in a table known as a confusion matrix. The statistical measures which are used to evaluate the performance of the model are formulated as follows [45].





Where **P**: positives which refer to the total number of positive tuples.

**N**: negatives which refer the total number of negative tuples.

**TP**: True positives which refer positive tuples that were correctly labeled by classifier.

**TN**: True negatives which refer negative tuples that were correctly labeled by classifier.

**FP**: False positives which refer the negative tuples that were mislabeled as positive.

**FN**: False negatives which refer the positive tuples that were mislabeled as negative.

# CHAPTER FOUR

# RESULT AND DISCISSION

## 4.1. Experimental Result

An experiment is conducted once the prediction model has been developed to determine how effectively it can distinguish between the various varieties of malt-barley grown in Ethiopia. The morphology, color, and texture features, as well as their combinations, are tested for each classifier in order to accomplish this. The combinations are made using the Forward Feature Selection (FFS) approach, which builds a combined feature set by systematically adding candidate sets beginning with an empty set until the performance is unaffected by the addition of more features. In each trial, the best performance for recognizing Ethiopian malt-barley seeds is selected using the MLP, KNN, and their hybrid identification algorithms. In this work, the segmented malt-barley seeds were first used to extract nine morphological, five textural, and six color features. The feature set selection during combination was then carried out using the FFS approach. By combining each of the three features, we were able to calculate every potential feature set. The experiment is conducted in the chosen dimensional space of the feature vectors of the data sets after the combined features have been computed.

### 4.1.1. Artificial Neural Network Classifier

The first identification algorithm for constructing malt-barley seed variety identification model is artificial neural network. The parameters used for the classifier with their value are depicted in table 3 below.

Table 3 Parameters used for ANN classifiers

|  |  |
| --- | --- |
| **Parameters** | **values** |
| Number of hidden layers | 2 |
| Network type | Feed-forward with backpropagation |
| Number of neurons in hidden layer | 10 |
| Activation function | Sigmoid |
| Algorithm | Multilayer perceptron |
| Initial weight | 1 |
| Initial bias | 1 |

The artificial neural network classifier uses the characteristics collected from images of malt-barley seeds using the feature extraction methods we've just covered to determine how accurate each method is. Then, the accuracy of each method is evaluated by dividing the number of malt-barley seed varieties that were successfully recognized by the total number of barley seed variants., depicted in table 4 below.

Table 4 The performance of ANN model for the selected feature sets

|  |  |  |  |
| --- | --- | --- | --- |
| **Features extracted** | **Correctly identified** | **Incorrectly identified** | **Accuracy (%)** |
| Morphology | 33 | 8 | 80.48 |
| Texture | 23 | 18 | 56.1 |
| Color | 24 | 17 | 58.5 |
| Morphology + Color | 37 | 4 | 89.2 |
| Morphology + Texture | 35 | 6 | 84.5 |
| Morphology + Color + Texture | 39 | 2 | 95.12 |

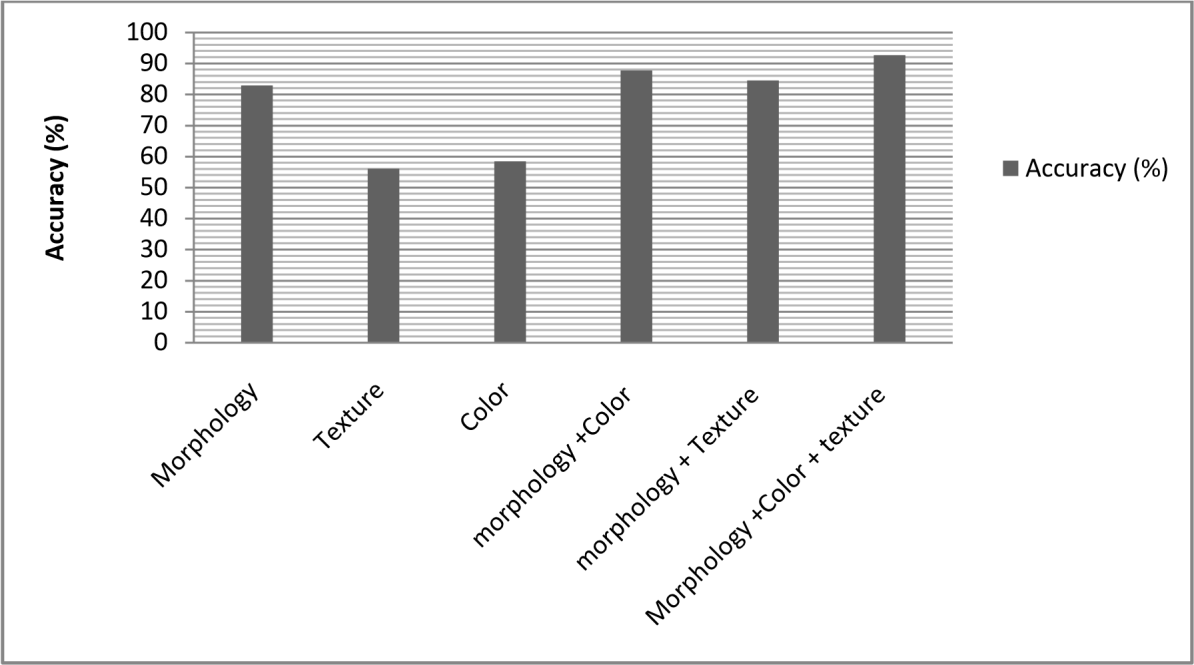


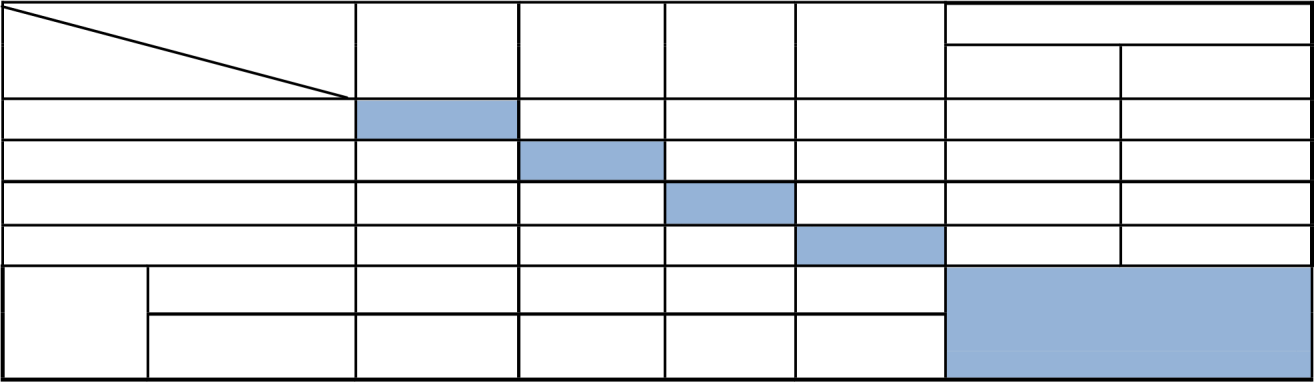
Figure 11 The performance of each feature set using ANN model

Table 4 shows that, with an accuracy of 95.12%, the combined feature set of morphology, texture, and color registered better results than the other feature set. Our experimental finding led us to the conclusion that the neural network model performed relatively better at discrimination when given combined features as input than when given separate features. The experimental results achieved using these feature sets are presented using a confusion matrix as follows in Table 5.

Combining morphology, texture, and color features using an ANN confusion matrix: Our neural network model in this experiment included 13 (combinations of morphological, texturing, and color) features. The neural network model, where the input units are equal to the number of features and the output units are equal to the number of classes, contains 13 inputs, 10 hidden, and 4 output units (13:10:4). The results of the experiment are compiled and displayed in Table 5 below.

Table 5 Summary results of malt-barley seeds using ANN with combination of morphology, color and texture features

**Actual class Total (%)**



**Holker**

**Propino Sabini Misikal**

**Correctly**

**Incorrectly**

**Predicted class**

**Holker**

9

1

0

0

90.0

10.0

**Propino**

0

9

0

0

100.0

0.0

**Sabini**

0

0

12

2

85.7

14.3

**Misikal**

0

0

0

8

100.0

0.0

**Total**

**Correctly**

100.0

90

100

80.0

**(**

**%)**

**Incorrectly**

**92.7**

**7.3**

0.0

10

0.0

20.0

According to the experimental findings, 92.7% of the 41 images were correctly identified, whereas 7.3% of the images in the sample data set were wrongly identified. A review of the findings revealed that while none of the Holker or Sabini kinds were incorrectly classified, 1 image of Propino was mistaken for a Holker, and 2 photographs of Misikal were mistaken for Sabinis. According to the findings of this experiment, Holker and Sabini varieties can be distinguished completely, while one image of Propino is misidentified as Holker and two images of Misikal are mistakenly identified as Sabini. The morphological similarity between those misidentified types is the primary source of confusion.

The correlation between each of the four kinds of malt-barley seeds is shown using a bar chart in Figure 12 below, in accordance with the experimental results presented in Table 5.



0



2



4



6



8



10



12



14



holker



propino



sabini



misikal



holker



Propino



sabini



misikal

Figure 12 **The correlation between the four varieties of malt-barley seeds**

Figure 12's correlation chart shows that Misikal and Propino varieties are more linked with Sabini and Holker, respectively. As we can see from the stand-alone bar of Sabini and Holker barley varieties, there is no link between them, nor is there any association between Misikal and Holker barley varieties.

### 4.1.2. K-Nearest Neighbor Classifier

Both the input and their target variables are utilized to develop the model in the case of K-nearest neighbor identification, the training data set that has been employed by neural network so far. Then, by comparing the distance of the unknown tuple to K nearest neighbors of that reference model, it is possible to establish the class of the test data, which only includes the input variables. We take into account Euclidean distance as the distance metric parameters whereas 3 is the value of k, as described in section 3.5.2. Based on this, table 6 shows the correctness of each feature set.

Table 6 The performance of KNN model for the selected feature sets

|  |  |  |  |
| --- | --- | --- | --- |
| **Features extracted** | **Correctly identified** | **Incorrectly identified** | **Accuracy (%)** |
| Morphology | 30 | 11 | 73.1 |
| Texture | 29 | 10 | 70.73 |
| Color | 30 | 11 | 73.2 |
| Morphology + Color | 34 | 7 | 82.9 |
| Morphology + Texture | 31 | 10 | 75.6 |
| Morphology + Color + Texture | 36 | 5 | 87.8 |

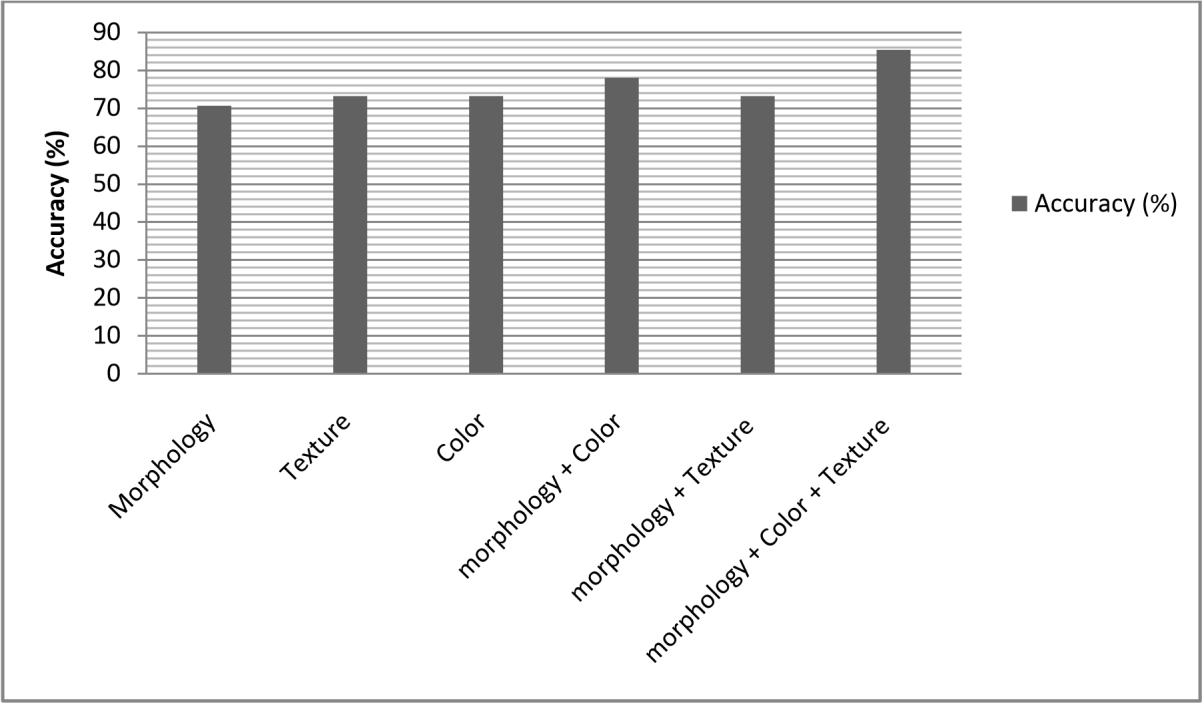
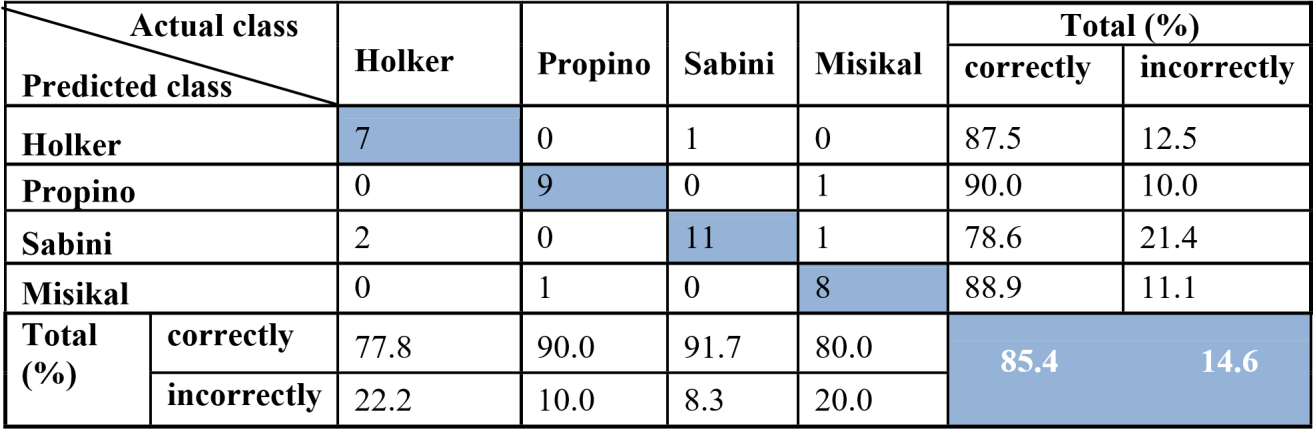


Figure 13 The performance of each feature set using KNN model

The combined feature sets of morphology, texture, and color features registered better results than the other feature set, according to the experimental findings given in Table 5.4, with an accuracy of 87.8%. According to the experimental findings, utilizing the combined features with a KNN classifier to distinguish between Ethiopian malt-barley seeds produced somewhat better results than using the individual features alone. The top outcome from the experiment is shown using a confusion matrix as follows in Table 6. Combining morphology, texture, and color features using a confusion matrix: We combined all three of the features used in our neural network experiments during this trial. Table 6 following provides an overview of the experimentation's recorded results.

Table 7 summary results of malt-barley seeds using KNN with a combination of morphology, color and texture features



According to the experimental findings, 14.6% of the images in the sample data set were mistakenly identified while 85.4% of the 41 images were correctly identified. An analysis of the data revealed that 2 photographs of Holker variety were mistakenly labelled as Propino, while 1 image from each of the Propino and Sabini was mistakenly identified as Misikal and Holker, respectively. Moreover, Propino and Sabini were mistaken for two representations of Misikal. The results of this experiment show that no variety can be entirely differentiated, but Sabini kinds are the best at it, and Holker is more connected with Sabini.

The correlation between each of the four varieties of malt-barley seeds are depicted in Figure 14 below.



0



2



4



6



8



10



12



holker



propino



sabini



misikal



holker



Propino



sabini



misikal

Figure 14 **The correlation between the four varieties of malt-barley seeds using KNN**

According to Figure 5.5's association chart, Propino and Sabini barley varieties are both connected with Misikal varieties, whereas Sabini and Holker barley types are correlated with one another.

### 4.1.3. The Hybrid of ANN and KNN Classifiers

During training the hybrid classifier parameters used in both k-nearest neighbor and artificial neural network are used. In this hybrid model, we use the same dataset which is used to construct the ANN and KNN model. The summery of the result, obtained using each and combined feature sets, is depicted in table 8 below.

Table 8 The performance of Hybrid model for the selected feature sets

|  |  |  |  |
| --- | --- | --- | --- |
| **Features extracted** | **Correctly identified** | **Incorrectly identified** | **Accuracy (%)** |
| Morphology | 36 | 5 | 87.8 |
| Texture | 25 | 14 | 60.9 |
| Color | 28 | 13 | 68.3 |
| Morphology + Color | 36 | 5 | 87.2 |
| Morphology + Texture | 37 | 4 | 90.2 |
| Morphology + Color + texture | 38 | 3 | 92.1 |



Figure 15 **The performance of each feature set using hybrid model**

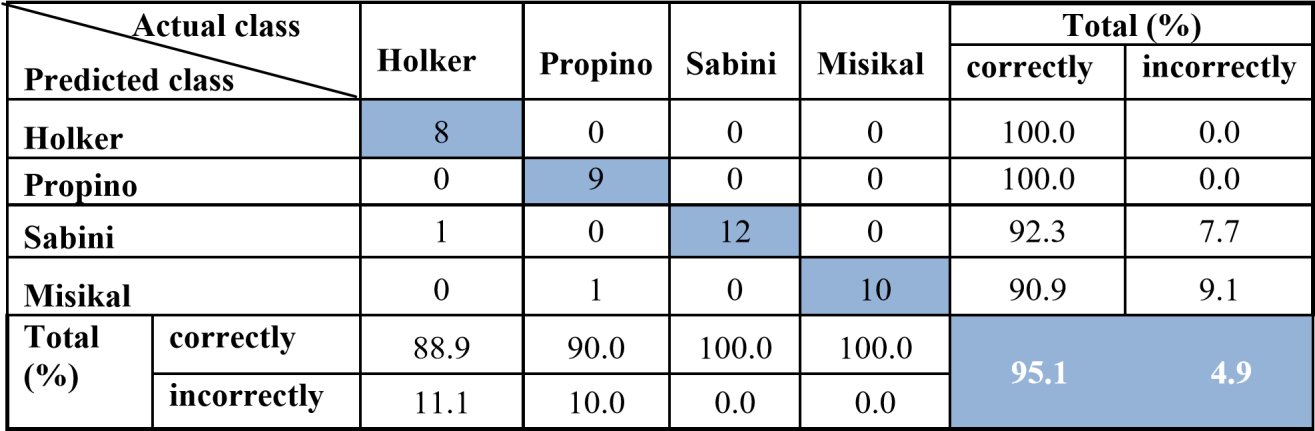
Table 8 shows that, with an accuracy of 92.1%, the combined feature sets of morphology, texture, and color reported a superior result than the other feature set. According to the experimental findings, utilizing a combination of an ANN and KNN classifier with combined characteristics produced superior discriminating results than using individual features to identify Ethiopian malt-barley seeds. As shown in Table 8, the best experimentation outcome is represented using a confusion matrix.

**A confusion matrix for the hybrid ANN-KNN classifier's morphology, texture, and color features:**

We used a combination of all three characteristics that were tested in both the ANN and KNN classifiers throughout this experiment. Table 9 following provides an overview of the experimentation's recorded results.

Table 9 **Hybrid of ANN and KNN summary results of malt-barley seeds using a combination of morphology,**

**color and texture features**



Based on this experimental result, out of 41 images, 95.1% were correctly identified and 4.9% of the sample data set images were incorrectly identified. An investigation of the results indicated that 1 image in each of Holker and Propino were incorrectly identified as Sabini and Misikal respectively. And all the images of both Sabini and Misikal varieties are perfectly identified. The correlation between each of the four varieties of malt-barley seeds registered from experimental results of the hybrid of KNN and ANN classifier using the combination of morphology, texture and color features are depicted in Figure 16 below.



0



2



4



6



8



10



12



14



holker



propino



sabini



misikal



holker



Propino



sabini



misikal

Figure 16 **The correlation between the four varieties of malt-barley seeds using hybrid model**

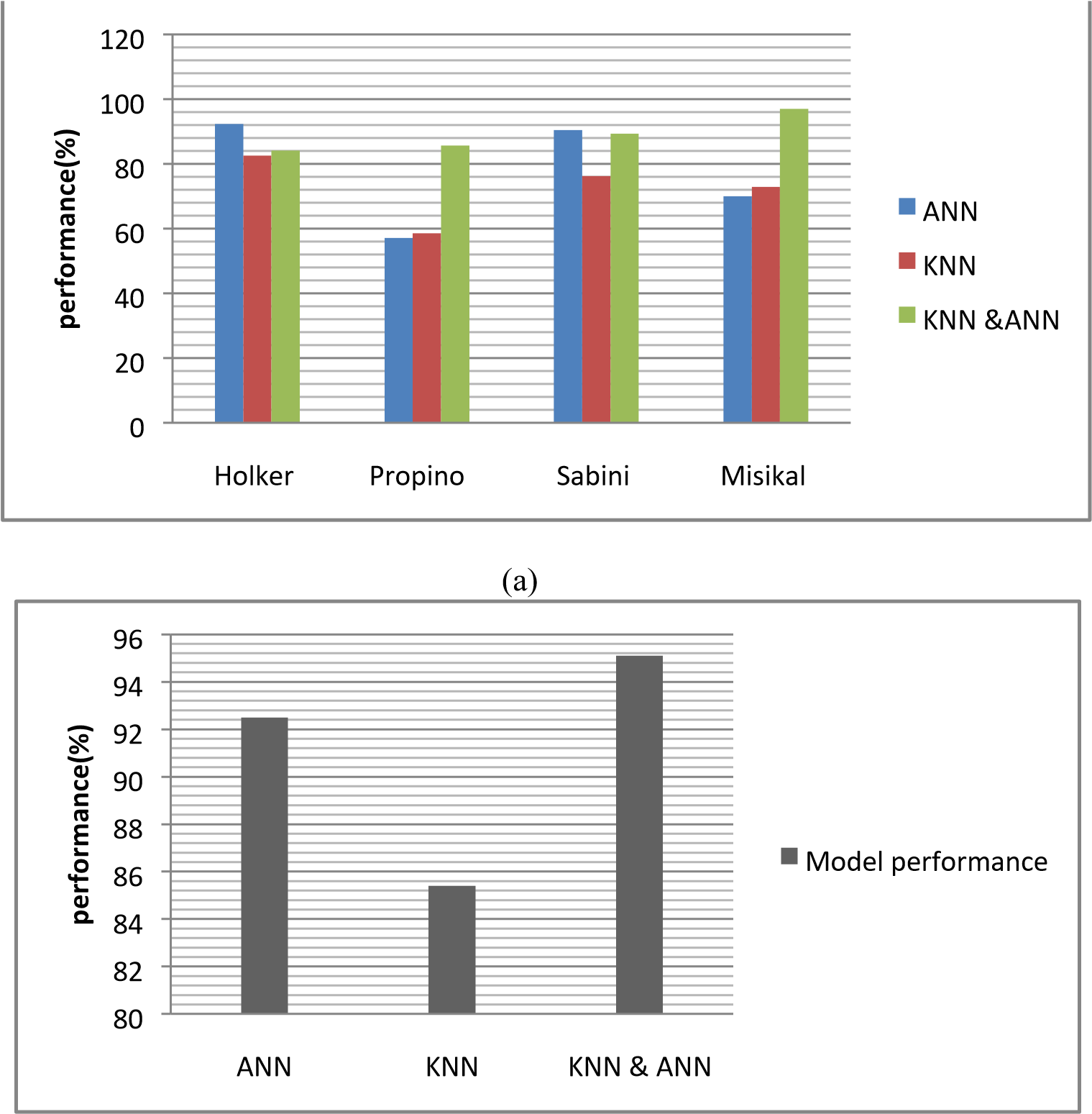
Using a mixture of morphology, texture, and color variables as the input for the hybrid of ANN and KNN model, Figure 16's bar chart illustrates the association between wes and each variety of malt-barley. The bars for Sabini and Misikal are independent in this graph since they don't correlate with other varieties, although Holker and Propino malt-barley types correlate with Misikal and Sabini varieties, respectively.

## 4.2. The Overall Performance of Identification Model

As we have described in detail in the preceding sections, we conducted experiments in six different scenarios using feature sets of texture, morphology, and color separately, as well as combining color and morphology and texture with morphology, before using all three feature sets together using the Forward Feature Selection method. Likewise, experimental findings indicate that texture is superior to nearest neighbor while morphological feature has the largest discriminating potential using neural network. Although many types of malt barley have more or less identical colors, color was not as good as morphology and texture in terms of relative quality.

The K-nearest neighbor identification model achieved 85.4% accuracy, and the hybrid of KNN and ANN identification model registered 95.1% accuracy using a combination of color, morphology, and texture. The best accuracy of the neural network model is 92.7% using combined features of morphology, texture, and color. Hence, combining characteristics and algorithms as we did to identify Ethiopian malt-barley seeds yields better results than doing it alone.

The identification performance of artificial neural networks was superior to that of k-nearest neighbors, as shown in figure 5.6 below, and using the hybrid of the two techniques performed relatively better than using the two techniques alone. Figure 17 (a) displays each model's performance for a specific variety of malt-barley seeds, whereas Figure 17 (b) displays each model's overall performance.



(b)

Figure 17 Model performance comparisons; (a) Model performance in each verity of barley seeds,

(b) overall performance of models

The regularity of seed size and seed arrangement present the biggest obstacles to the identification model's performance. In the photos, seeds may not all be of the same size and may be highly interconnected with one another. The justification is that barley seeds are occasionally combined into a single linked component during preprocessing in an effort to decrease noise. Consider the examples provided in Figure 18 below as an example.

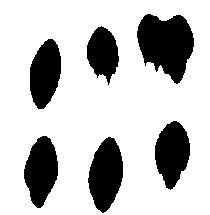


Figure 18 Sample challenging seed image

## 4.3. Discussion of Results

This study represents the first effort in the field of Ethiopian malt-barley identification, to the best of the researchers' knowledge. In order to compare the findings of this research with other research conducted globally, we did so.

In this study, during experimentation combined features in the hybrid of ANN and KNN model have been found to work very well in identification of Ethiopian malt-barley varieties. The results of the current identification model identification techniques, and the previous researches [14] [15] [44] [44] are depicted in Table 10.

Table 10 Performance comparison of related works

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Researchers** | **Dataset size** | **Features used** | **Classifier** | **Performance** |
| Nowakowski et al [14] (2012) | 1200 image | Color and texture | ANN | 98.2% |
| Barbara et al [44] (2012) | 2100 seeds | Color, texture morphology, and shape factors | ANN | 94.9% |
| Aliresapazoki et al  [60] (2013) | 270 images | Color and morphology | ANN | 82.2% |
| Szczypinski [15] (2015) | 33 images  (13000 seeds) | Color, texture and shape factors | ANN | 86% |
| Current study | 208 images  (5408 seeds) | Combined color, morphology,  texture features | The hybrid of ANN and KNN | **95.1%** |

As can be seen in Table 5.8, the performance of malt-barley seed identification model of the current study shows promising result. The features combined and selected using FFS improved the accuracy of the classifier, as shown in section 5.1 above, for Ethiopian malt-barley seeds since it selects the best features of a seed through iterative testing.

The test results of the proposed identification model may have had ambiguities or errors that were either segmentation errors or identification errors. While identification errors are caused by morphological, color, and textural similarities (similarities that occurred as a result of this condition, soil type, and fertilizer usage) between the input barley seed image and its identification result, segmentation errors are errors that were caused by the segmentation algorithm.

Also, the algorithm that was employed in our experiment to eliminate noise is unable to do so for severely blurred and shady author seed images. Before using segmentation, feature extraction, and identification procedures, this calls for sophisticated noise reduction algorithms that are efficient in identifying blurring and shadow in seed images.

# CHAPTER FIVE

# CONCLUSION AND RECOMMENDATION

## 5.1 Conclusion

Malt-barley is a cash crop that is mostly utilized to make local alcoholic beverages and beer. Many biological and physical traits of malt grains have a significant influence in the creation of beer. Well-equipped laboratories and highly experienced expertise pick the barleys with the qualities that are utilized in the manufacture of beer. But occasionally the laboratory may not be well-equipped, and even though the expertise is highly skilled, they may feel fatigued and exposed for bias. The construction of a malt-barley identification model can aid in the selection and marketing of malt-barley, which is a crucial step in the early stages of brewing beer. As far as the researchers are aware, there has been no attempt to encourage the selection of Ethiopian malt-barley seeds by researchers. An attempt has been made in this work to create the best model possible for identifying the different types of Ethiopian malt barley seeds.

In order to achieve this, the study used experimental research, which entails the development of data sets for the training and evaluation of a model for identifying malt-barley seeds. Images of the Sabini and Holker varieties of Ethiopian malt barley were taken at the Gondar Malt Factory. The preprocessing and segmentation of the images are followed by feature extraction. The identification model is created using KNN after we extracted the morphological, color, and texture properties of each variation. To choose the ideal combination of morphology, texture, and color features, the forward feature selection technique is utilized. The test findings demonstrate that a KNN hybrid model with an overall performance of 86% accuracy can be built utilizing the combined color and texture data to identify seed kinds. Overall, this study's results for the identification of malt-barley types are encouraging. The difficulties encountered by the segmentation and noise removal procedures from non-uniform size malt-barley seed pictures, which results in having inadequate features during feature extraction, are blamed for the bulk of misidentification errors. Malt-barley seeds in a photograph are more frequently misidentified because of their irregular size.

## 5.2 Recommendation

Based on the investigation and findings of the study, the following recommendations are forwarded for future and further research works:

* In removing noises and detecting edge from barley seed image median filtering and Sobel operator used respectively and it works well to clean noise. However, these noise removal and edge detection techniques cut if seeds are overlapped and have different size. Therefore, to improve the performance of the model, future works need to integrate advanced noise removal techniques that clearly separate noises from seeds in noise, overlapped and non-uniform size barleys.
* This study considers regionprops, rgb components and graycoprops Matlab functions as feature extraction techniques. However, their performance greatly affected depending on the variety of barley seed images used. It is therefore necessary to conduct further research to identify feature extraction techniques that are effective to extract better representative features of malt-barley seeds.
* In this study properly arranged and scanned seed images have been used. The arrangement of seeds affects the process of segmentation. As a result, future work should investigate randomly distributed and overlapped barley seed.
* A malt-barley seeds have both the dorsal side and ventral side. In this study we only have been considered seeds in ventral side. As a result, future work needs to consider dorsal side and/or both dorsal and ventral sides of the seeds too.
* In this study, we consider the malt-barley seeds for identification; as a result, future work should consider the malt from which barley verities it is produced.

# 

# References

|  |  |
| --- | --- |
| [1] | Zemede Asfaw, "The Barley of Ethiopia," *in Proceedings of the 1st Barley Research Review Workshop, addis ababa,* p. 182–192, 2000. |
| [2] | Tinku Acharya and Ajoy K. Ray, "Image Processing Principles and Applications," *New Jersey, India: WILEY & SONS,* 2005. |
| [3] | Granitto P. M., Navone H. D., Verdes P. F. and Ceccatto H. A, "Automatic identification of weed seeds by color image processing," *in VI Argentine Congress on Computer Science, Ushuaia, Argentina,* pp. 91-103, 2000. |
| [4] | Habtamu, M, "Image Analysis for Ethiopian Coffee Classification," *Habtamu, M. “Image Analysis for Ethiopian Coffee Classification.” M.Sc. Thesis, Addis Ababa University, Addis Ababa, Ethiopia,* 2008. |
| [5] | Bayeh Mulatu and Stefania Grando, "Report on the Preliminary Results of Area, Production and Yield of Temporary Crops (Meher Season, Private Peasant Holdings)," *CACC (Central Agricultural Census Commission), Addis Ababa preliminary Report,* 2002. |
| [6] | Ceccarelli S., Grando S., Shevstov V., Vivar H., Yayaoui A., El-Bhoussini M. and Baum M., , "ICARDA Strategy for global barley improvement,” in Rachis: Barley and Wheat Newsletter," pp. 3-12, 1999. |
| [7] | Hailu Gebre and J.A.G. van Leur , "Food barley breeding," *in Proceedings of the 1st barley research review workshop Oct 16-19, Addis Ababa,* pp. 9-23, 1996. |
| [8] | BedadaBegna, Mesay Yami and Eshetu Lemma , "Characterization of a Barley-based Farming System in the highlands, Ethiopia.," *Journal of Agricultural Extension and Development,* pp. 309-320, 2014. |
| [9] | Getachew Legese, Tolosa Alemu and Tesfaye Zegeye, "Unexploited opportunity in agro processing: the case of malt-barley in Ethiopia," *Proceedings of the 2nd National Barley Research and Development Review Workshop, Assela, Ethiopia,* 2011. |
| [10] | Bayeh Mulatu and Berhane Lakew, "Barley Research and Development in Ethiopia an overview," *Proceedings of the 2nd National Barley Research and Development Review Workshop, Holetta, Ethiopia,* 2006. |
| [11] | Nowakowski K. Boniecki P., Tomczak R.J, Kujawa S. and Raba B., "Identification of malting barley varieties using computer image analysis and artificial neural networks," *Fourth International Conference on Digital Image Processing,* pp. 833425-833430, 2012. |
| [12] | Szczypiński, P. M., Klepaczko, A., and Zapotoczny, P, "Identifying barley varieties by computer vision," *Computers and Electronics in Agriculture,* pp. 1-8, 2015. |
| [13] | Nowakowski, K., Boniecki, P., Tomczak, R. J. and Raba, B, "Identification process of corn and barley kernel damages using neural image analysis," *Third International Conference on Digital Image Processing,* pp. 80090C-4, 2011. |
| [14] | Antonio L. Amaral, Orlando Rocha, Cristina Goncalves, Antonio A. Ferreira, Eugenio C. and Ferreir,, "Application of image analysis to the prediction of EBC barley kernel weight distribution," *Industrial Crops and Products, Elsevier,* p. 366–371, 2009. |
| [15] | Araya, Solomon Habtu, Mitiku Haile, Sisay, F and Tadesse Dejenie, "Determination of Local Barley (HordeumVulgare) Crop Coefficient and Comparative Assessment of Water," *CNCS, Mekelle University,* vol. 3, pp. 65-79, 2011. |
| [16] | Kemelew Muhe, "Opportunities and challenges in malt barley research and production in North," *Proceedings of the 2nd National Barley Research and Development Review Workshop, Holetta, Ethiopia,* 2006. |
| [17] | Amaral, A.L., Rocha, O., Goncalves, C., Ferreira, A.A., Ferreira and E.C, "Development of image analysis methods to evaluate barley/malt grain size," *In Proceedings of the CHEMPOR 2005-9th International Chemical Engineering Conference, Coimbra, Portugal,* 2005. |
| [18] | Vinita Dixit and Jyotika Pruthi, "Review of Image Processing Techniques for Automatic," *in International Journal of Computer Science and Mobile Computing,* vol. 3(3), pp. 371-378, 2014. |
| [19] | Torsten Seemann, "Digital Image Processing using Local Segmentation," *PhD Thesis, Monash University, Monash, Australia,* 2002. |
| [20] | Kiratiratanapruk K. and Sinthupinyo W., "Color and texture for corn seed classification by machine vision," *in Proceedings of the 19th IEEE International Symposium on Intelligent Signal ,* 2011. |
| [21] | Mark Nixon and Alberto S. Aguado, "Feature Extraction and Image processing," *2nd ed. London, UK: Elsevier,* 2008. |
| [22] | Parker J.R. , "Algorithms for Image Processing and Computer Processing on Vision," *Second Edition ed., Kostas Terzidis, Ed. Indianapolis, India,Wiley Publishing,* 2011. |
| [23] | Alireza Khoshroo, Arman Arefi, Asad Masoumiasl and Gholam-hossein Jowkar, , "Classification of Wheat Cultivars Using Image Processing and Artificial Neural Networks.," *Agricultural communications,* vol. 2(1), pp. 17-22, 2014. |
| [24] | Mousavi R. S. J., Akhlaghian T. F. and Kaveh M., , "Classification of rice varieties using optimal color and texture features and BP Neural networks," *IEEE,* 2014. |
| [25] | Mayur P. Raj, Priya R. Swaminarayan. , "“Applications of Image Processing for Grading," *International Journal on Recent and Innovation Trends in Computing and Communication,* vol. 3, p. 1194 – 1201, 2015. |
| [26] | Liu, Zhao-yan, Cheng Fang, Ying Yi-bin and Rao Xiu-qin, "Identification of Rice Varieties Using Neural Networks.," *Journal of Zhejiang University Science,* pp. 1095-1100, 2005. |
| [27] | Yang, C.C., S.O. Prasher, J.A. Landry, H.S Ramaswamy and A. Ditommaso, "Application of artificial neural networks in image recognition and classification of crop and weeds," *Canadian Agricultural Engineering,* vol. 42(3), pp. 147-152, 2000. |
| [28] | Nur Badariah Ahmad Mustafa, Syed Khaleel Ahmed, Zaipatimah Ali, Wong Bing Yit, Aidil , "Agricultural produce Sorting and Grading using Support Vector Machines and Fuzzy Logic," *IEEE International Conference on Signal and Image Processing Applications (ICSIPA),* pp. 391-396, 2009. |
| [29] | Dayanand Savakar, "Identification and Classification of Bulk Fruits Images using Artificial," *International Journal of Engineering and Innovative Technology (IJEIT),* vol. 1(3), pp. 36-40, 2012. |
| [30] | Boniecki P., Nowakowski K., Tomczak R., Kujawa S., Piekarska-Boniecka H., "The application of the Kohonen neural network in the non-parametric quality-based classification of tomatoes," *Fourth International Conference on Digital Image Processing ,* vol. 8334, pp. 833427-5, 2012. |
| [31] | Guevara-Hernandez F. and Gomez-Gil J., , "A machine vision system for classification of wheat and barley grain kernels," *Spanish Journal of Agricultural Research,* pp. 672-680, 2011. |
| [32] | Piotr Zapotoczny, "Discrimination of wheat grain varieties using image analysis and neural networks. Part I. Single kernel texture," *Journal of Cereal Science, Elsevier,* pp. 60-68, 2011. |
| [33] | AlirezaPazoki, ZohrehPazoki and BehzadSorkhilalehloo, "Rain Fed Barley Seed Cultivars," *World Applied Sciences Journal ,* pp. 755-762, 2013. |
| [34] | Alena Sykorova1, Evžen Šarka, Zdeněk Bubn ik, Matyaš Schejba and Pavel Dostalek, "Distribution of Barley Kernels," *Czech J. Food Sci., ,* p. 249–258, 2009. |
| [35] | Barbara Raba1, Krzysztof Nowakowski, PiotrBoniecki, "Visual quality evaluation of malting barley with use of neural image analysis," *international Conference on Agricultural Engineering, Valencia,* 2012. |
| [36] | Benjamaporn Lurstwut and Chomtip Pornpanomchai, "Plant Seed Image Recognition System," *IACSIT International Journal of Engineering and Technology,* vol. 3(6), 2011. |
| [37] | André Aichert, "Feature extraction techniques," *Camp Medical Seminar,* 2008. |
| [38] | Biniyam Girma, Getachew Abebe and Girma Goro, "Characterization and Classification of Hararghe Coffee (Coffea arabica L.) Beans Using Morphological Attributes Based on Classical Images," *East African Journal of Sciences,* vol. 7(1), pp. 11-16, 2013. |
| [39] | Birhanu Turi, Getachew Abebe and Girma Goro,, "Classification of Ethiopian Coffee Beans Using Imaging Techniques," *East African Journal of Sciences ,* vol. 7(1), pp. 1-10, 2013. |
| [40] | Savita N. Ghaiwat, Parul Arora, "Detection and Classification of Plant Leaf Diseases Using Image Processing Techniques," *International Journal of Recent Advances in Engineering & Technology (IJRAET),* vol. 2(3), 2014. |
| [41] | Jiawei Han, Micheline Kamber and Jian Pei, "Data Mining: Concepts and Techniques," *Third Edition. 225Wyman Street, Waltham, Elsevier Inc,* 2012. |
| [42] | Elif AYBAR, "Sobel Edge Detection Method for Matlab," *Anadolu University, Porsuk Vocational School,* 26410 Eskisehir. |
| [43] | Daniel Kim, "Sobel Operator and Canny Edge Detector," *ECE 480,* 2013. |
| [44] | Matthias Kirchner and Jessica Fridrich, "On Detection of Median Filtering in Digital Images," *SPIE Conference on Media Forensics and Security,* 2010. |
| [45] | K. S. Fu, "Syntactic pattern recognition and application," *Prentice Hall,* 1982. |
| [46] | Nobuyuki Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 9, pp. 62-66, 1979. |
| [47] | https://www.researchgate.net/publication/304625358\_Applying\_Image\_Processing\_for\_Malt-barley\_Seed\_Identification. |