T.R.
ONDOKUZ MAYIS UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF AGRICULTURAL MACHINERY AND TECHNOLOGIES ENGINEERING

MACHINE-VISION-BASED CLASSIFICATION OF CASHEW NUTS USING COLOUR FEATURES

Master's Degree Thesis

Geofrey Prudence BAITU

Supervisor
Prof. Dr. Y. Benal ÖZTEKİN

This study was supported by Ondokuz Mayis University with PYO.ZRT.1904.22.010 project number.

SAMSUN
2022

ACCEPTANCE AND APPROVAL OF THE THESIS

The thesis entitled **"MACHINE-VISION-BASED CLASSIFICATION OF CASHEW NUTS USING COLOUR FEATURES",** prepared by **Geofrey Prudence BAITU** under the supervision of **Prof. Dr Y. Benal ÖZTEKİN** was found successful and unanimously accepted by committee members as a Master's thesis of the Department of Agricultural Machinery and Technologies Engineering, following the examination held on 27/05/2022.

	Title Name Surname **University** **Department**	**Signature**	**Final decision**
Chairperson	Assoc. Prof. Dr. H. Kürşat ÇELİK Akdeniz University Department of Agricultural Machinery and Technologies Engineering		☒Accept ☐ Reject
Member	Prof. Dr. Y. Benal ÖZTEKİN Ondokuz Mayıs University Department of Agricultural Machinery and Technologies Engineering		☒Accept ☐ Reject
Member	Assoc. Prof. Dr. K. Çağatay SELVİ Ondokuz Mayıs University Department of Agricultural Machinery and Technologies Engineering		☒Accept ☐ Reject

This thesis has been approved by the committee members already stated above and determined by the Institute Executive Board.

APPOVAL

… / … / 2022

Prof. Dr. Ali BOLAT

Head of Institute of Graduate Studies

DECLARATION OF COMPLIANCE WITH SCIENTIFIC ETHIC

I have prepared this research work in accordance with the rules of the Institute of Graduate Studies, Ondokuz Mayıs University for Master's Degree Thesis. It contains correct and complete information, which was gathered at different stages of the thesis. I also adhered to scientific ethics and all information sources from which I benefited are indicated in the reference section.

Is an Ethics Committee Necessary?

Yes ☐

No ☒

…./…./2022

Geofrey Prudence BAITU

DECLARATION OF THE THESIS STUDY ORIGINALITY REPORT

Thesis title: MACHINE-VISION-BASED CLASSIFICATION OF CASHEW NUTS USING COLOUR FEATURES.

As a result of the originality report taken by myself from the plagiarism detection program on the 21.04.2022 for the thesis that is mentioned above, it was found as follows;

Similarity rate: 18%

Single resource rate: 1%

21/04/2022

Prof. Dr. Y. Benal ÖZTEKİN

ÖZET

KAJU FISTIĞININ RENK ÖZELLİKLERİ KULLANILARAK MAKİNE GÖRME İLE SINIFLANDIRILMASI

Geofrey Prudence BAITU
Ondokuz Mayıs Üniversitesi
Lisansüstü Eğitim Enstitüsü
Tarım Makinaları ve Teknolojileri Mühendisliği Anabilim Dalı
Yüksek Lisans, Haziran /2022
Danışman: Prof. Dr. Y. Benal ÖZTEKİN

Kaju, Tanzanya'nın ülke ekonomisine dış gelir olarak katkı sağlayan başlıca ticari ürünlerden biridir. Kaju çekirdeklerinin işlenmesi, halen büyük ölçüde el emeği kullanılarak yerel olanaklarla yapılmaktadır. İdeal koşullarda iyi işlenirse kajuların beyaz renkte olması beklenir. Ancak, buhar odalarında uzun süre kavurma veya aşırı kurutma gibi çeşitli faktörler nedeniyle, bazı kaju çekirdekleri hafif kahverengi bir renge dönüşebilmektedir. Renk değiştirmiş bu kajulara kavrulmuş kaju denir. Besin kalitesi de dahil olmak üzere beyaz kaju çekirdekleri ile aynı özelliklere sahip olmasına rağmen, renk ve görünüm tüketicilerin kalite algısını etkilediği için bu kaju çekirdeklerinin ayrılması gerekmektedir. Tanzanya başta olmak üzere dünyanın pek çok yerinde kaju çekirdeklerinin ayırma ve sınıflandırma işlemi elle yapılmaktadır. Uluslararası ticarette, kaju sınıflandırması çok önemli olup ürün kalitesini artırmak için üretimin bu aşamasında daha etkili ve tutarlı yöntemlerin uygulanması gerektiği anlamına gelir. Bu çalışmanın amacı, kaju çekirdeklerinin beyaz veya kavrulmuş olarak sınıflandırılmasında renk özellikleri kullanılarak geleneksel Makine Öğrenmesi tekniklerinin kullanımının değerlendirilmesidir. Bu çalışmada, görüntülerden farklı renk özellikleri çıkarılmıştır. Çıkarılan özellikler, RGB ve HSV renk uzaylarında kanalların ortalamaları (μ), standart sapmaları (σ) ve çarpıklığını (γ) içerir. Python'da Boruta Kütüphanesi kullanılarak sarmal (*wrapper*) yöntemi uygulanarak bu sınıflandırma problemi için ilgili özellikler seçilmiş ve ilgili olmayanlar çıkarılmıştır. Bu çalışmada 5 model çalışılmış ve verimlilikleri analiz edilmiştir. Değerlendirme teknikleri Lojistik Regresyon, Karar Ağacı, Rastgele Orman, Destek Vektör Makinesi ve K-En Yakın Komşu (KNN) yöntemleridir. Karar Ağacı modeli, %98,4 ile en düşük doğruluğu vermiştir. 100 ağaçlı Rastgele Orman modelinde maksimum %99,8 doğruluk elde edilmiştir. Uygulamadaki basitliği ve yüksek doğruluğu nedeniyle Rastgele Orman bu çalışma için en iyi model olarak önerilmektedir. Bu çalışma, üretilen kaju fıstığının kalitesini artıracak kaju fıstığı işleme sistemlerinin geliştirilmesine katkı sağlayacaktır.

Anahtar Kelimeler: Lojistik Regresyon, Karar Ağacı, Rastgele Orman, Destek Vektör Makinesi ve K-En Yakın Komşu (KNN), Kaju fıstığı

ABSTRACT

MACHINE-VISION-BASED CLASSIFICATION OF CASHEW NUTS USING COLOUR FEATURES

Geofrey Prudence BAITU
Ondokuz Mayis University
Institute of Graduate Studies
Department of Agricultural Machinery and Technologies Engineering
Master, June /2022
Supervisor: Prof. Dr. Y. Benal ÖZTEKİN

Cashew is one of the major commercial commodities contributing to the national economy of Tanzania as foreign revenue. And yet still the processing of cashew kernels is run locally using manual labour for a big part. If processed well under ideal conditions, cashews are expected to be white in colour. But due to various factors like prolonged roasting in the steam chambers or over-drying, some cashew kernels tend to have a slight brown colour, and these are referred to as scorched cashews. Despite sharing the same characteristics with white cashew kernels, including nutritional quality, these cashew kernels are supposed to be graded differently. In many places around the world, particularly in Tanzania, the sorting and grading process of cashew kernels is performed by hand. In international trade, cashew grading is very important and this means more effective and consistent methods need to be applied in this stage of production in order to increase the quality of the products. The objective of this study was to evaluate the use of traditional Machine Learning techniques in the classification of cashew kernels as white or scorched by using colour features. In this experiment, various colour features were extracted from the images. The extracted features include the means (μ), standard deviations (σ), and skewness (γ) of the channels in RGB and HSV colour spaces. The relevant features for this classification problem were selected by applying the wrapper approach using the Boruta Library in Python, and the irrelevant ones were removed. 5 models are studied and their efficiencies analysed. The studied models are Logistic Regression, Decision Tree, Random Forest, Support Vector Machine and K-Nearest Neighbour. The Decision Tree model recorded the least accuracy of 98.4%. The maximum accuracy of 99.8% was obtained in the Random Forest model with 100 trees. Due to simplicity in application and high accuracy the Random Forest is recommended as the best model from this study. This study will contribute to the improvement of cashew nuts processing systems, which will improve the quality of the produced cashew nuts.

Keywords: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbour, Cashews.

ACKNOWLEGMENT

Throughout the preparation of this thesis, I have received tremendous assistance and support from different people who played a significant role in my school life.

First and foremost, I am enormously grateful to my supervisor, Prof. Dr. Yeşim Benal ÖZTEKİN, for her invaluable support, continuous advice and patience during my Master's study. Her immense knowledge and abundant experience have helped to motivate and encourage me for the whole duration of my academic research and daily life. Her insightful feedback compelled me to sharpen my thinking and further my work to a more elevated level.

I express my gratitude to Ondokuz Mayis University for financially supporting this research study under the project number PYO.ZRT.1904.22.010. Also, I sincerely thank Prof. Dr. Kubilay Kazım VURSAVUŞ from Çukurova University for his support.

I would also like to thank Assist. Prof. Dr Hüseyin DURAN for his support in my study and daily life. Also, I would like to thank all the staff members in the Department of Agricultural Machinery and Technologies Engineering at the University of Ondokuz Mayis. It is their kindly support and guidance that have made my study and life in the Republic of Turkey a wonderful experience.

I would like to acknowledge all of my colleagues at the Department of Agricultural Machinery and Technologies Engineering for their remarkable collaboration. I would specifically like to mention Ms Omsalma Alsadig Adam GADALLA, Mr Mahtem Teweldemedhin MENGSTU and Ms Züleyha KARASU, for they were with me on my journey since the first day to the end. In addition, I would like to thank Mr Mashenene Charles MALIMA, a research officer at the Tanzania Agricultural Research Institute and a PhD candidate at Ondokuz Mayis University, for his support.

Finally, I would like to express my heartfelt gratitude to my parents and family for their sympathetic ear and wise counsel. Without their tremendous support and encouragement over the past few years, it would be impossible for me to complete my studies.

Geofrey Prudence BAITU

CONTENTS

ABBREVIATIONS AND ACRONYMS

AI	: Artificial Intelligence
CCD	: Charge-Coupled Device
CMOS	: Complementary Metal-Oxide Semiconductor
CNN	: Convolution Neural Networks
DL	: Deep Learning
EM	: Electromagnetic
HSV	: Hue - Saturation - Value
KNN	: K-Nearest Neighbours
ML	: Machine Learning
NPK	: Nitrogen-Phosphorus-Potassium
RGB	: Red-Green-Blue
RNN	: Recurrent Neural Network
SVM	: Support Vector Machine

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

1. INTRODUCTION

1.1. Introduction

Cashews are low in sugar, rich in fibre, heart-healthy fats, and plant protein. Cashews are considered as one of the topmost edible crops in the world, competing in the same market as other edible nuts including, hazels, pistachios, macadamias, peanuts and almonds (Oluwaseun Kilanko et al., 2020). It is estimated that only about 40% of cashew nut kernels are included in confectionery, while about 60% are consumed in the form of snacks. This means that the post-harvesting processes should ensure that the cashew kernels are extracted with minimum damage and at the highest possible quality. Traditionally, semi-skilled workers are being used in performing various processing operations manually. Thanks to the mechanization revolution, various equipment has been developed to perform different tasks mechanically. Mechanization in the cashew processing chain has been applied in different operations like roasting, shell liquid extraction and shelling in many places in the world. But for the most part, the cleaning of raw cashew nuts and sorting and grading of the kernels is still manually operated by human labour (Azam-Ali & Judge, 2001).

Tanzania is one of the biggest cashew nut producers in Africa and in the world in general. Up to this time, many farmers are forced to sell their cashews in raw form due to a lack of improved technology to process the nuts after harvesting. This study will open doors for more studies to help solve this issue by developing more sophisticated systems to help process the cashew nuts after harvesting so as to improve their quality and price in the market. As for Turkiye, despite not being a cashew nut producing country, this study can have some advantages to the country. There are some whole sellers who buy raw cashew nuts in bulk from other countries and process them here, then distribute them inside the country or export them abroad. So, this study will be useful in improving the systems they are using to process the cashew nuts by substituting human labour for computer vision in the sorting and classification phase. Also, this study can be an eye-opening to the improvement of other systems used to process other nuts grown in Turkiye, like pistachio and hazelnuts.

As mechanization is for human physical labour, so is Artificial Intelligence (AI) for human mind labour. The development of AI technology has helped simplify works that require human thinking by enabling the machine to think and make its own decisions. AI comes with the advantage of fast, consistent and improved decision-making accuracy than humans. In this technologically fast-developing world, it is imperative that we understand and apply these technologies in solving different problems (Maurice, 2021). For the computer to have the ability to think and perform like humans, we need to train it first. Now, this is the part where Machine Learning (ML) comes into play. It is through ML algorithms that we can train the computer to think and make decisions like a human brain. This can be achieved by feeding data to the computer, having it learn and understand the pattern and relationship between them then, it can perform different operations on new data like classification or prediction (Mich, 2020). Different traditional Machine Learning models are available and are being applied in solving various problems. Savoy (2020) mentioned K-Nearest Neighbour (KNN) model, which is based on the similarity (distance) measure between points to classify them, the Support Vector Machine (SVM) model, which uses a linear border to split and divide the data into different categories, the Naïve Bayes model, which uses the probability theory, the Decision Tree model, the Random Forest model and the Regression models as the commonly used ML algorithms. In some situations, where we want to learn a complicated set of data, the traditional ML techniques fail to perform well. And due to these setbacks, a new and more sophisticated approach called Deep Learning (DL) was born (Kubat, 2021). DL models have numerous advantages over traditional ML models. These advantages include better self-learning ability, the ability to generate features automatically, the ability to work well with unstructured data and many others (Janiesch, Zschech, & Heinrich, 2021).

Post-harvest operations and food quality control are essential activities in ensuring the safety of the consumers and the quality of the produce. AI has proven to be very helpful in both of these operations. In this thesis, we are going to focus on sorting and grading cashew kernels into white and scorched categories using colour features. Numerous studies have been performed to apply different AI techniques in sorting and grading cashew kernels in different ways. Given the sophistication and high efficiency of DP models in solving problems, many studies performed recently have focused much on the application of these models (Drugman, Huybrechts, Klimkov, & Moinet, 2018). Even though DL is a trending topic, their requirement for large amounts of data makes them unsuitable as general-purpose algorithms. In fact, they can be outperformed by simple traditional ML algorithms. Also, not all problems can be solved by DP models. Furthermore, even though recent DP approaches have achieved considerably better accuracy, this comes at the cost of increased innumerable math operations, increased processing power requirement, and associated costs (Walsh et al., 2019). Instead of just focusing on one approach, it is necessary to try out the simple techniques and see how they perform.

1.2. Objectives

As we have seen in the introduction that many studies done focused on the application of DP algorithms. Even though there are some studies which used the traditional ML algorithms, there is no any study which used the colour features to classify the cashew kernels into white and scorched categories. The main object of this research was to study the feasibility of using the traditional ML algorithms in sorting and classifying cashew nuts as white or scorched using colour features.

The specific goals were:

- To extract different colour features from the cashew kernel images and study their usefulness in classifying the cashew kernels into white and scorched categories. Then only the most effective ones will be selected and used.
- Use the selected colour features on different ML models and measure their efficiency in classifying the cashew kernels.
- Compare all the ML models and their performance in order to identify the most appropriate model in sorting cashew nuts.
- To develop an algorithm to sort cashew nuts by using the most effective ML model and the appropriate set of colour features.

1.3. An Overview of the Thesis Chapters

The work presented in this thesis is divided into five different chapters. The opening chapter gives us a brief view of our problem. It constitutes the introduction and the objectives of the research.

Chapter Two of this thesis explains in depth the history, cultural and processing activities for cashew nuts. It continues by introducing the concept of machine vision and its application in the food industry in general. A complete background of traditional ML algorithms and their applications in the food industry is explained clearly. More emphasis is put on cashew kernels classification and various examples are explained showing how this technology has been applied in the classification of cashew kernels.

Chapter Three gives a detailed description of the methods and materials used in this research. This includes everything from the first step of collecting samples, a thorough description of the image acquisition system, feature extraction and selection, and finally, the used classification algorithms.

The results and discussion of the methods used in this research are presented in Chapter Four. Here, the flow follows the experiment's flow from beginning to end. The results of the feature selection process are explained, and the selected features are outlined. Also, the classification efficiencies of the used traditional ML algorithms are presented and discussed.

Chapter Five, summarizes the research methods, findings and presents the conclusions made from the results obtained from the conducted experiment. It also proposes suggestions for future research.

CHAPTER TWO

2. BACKGROUND AND REVIEW OF LITERATURE

2.1. Nuts

Nuts can be defined in different ways, but from a botanical point of view, a nut is a fruit or seed with a hard or tough shell protecting a kernel inside of it, which is usually edible. These include Pistachios, Hazelnuts, Cashews, Walnuts, Marcona Almonds, Macadamia Nuts, Peanuts, Almonds, Brazil Nuts, Pine Nuts and Pecans (Wilkinson, 2005). Even though peanuts fall in this category, they are not true nuts. They are actually legumes rather than nuts. Humans have been consuming nuts for ages since the early times when man lived in hunter-gatherer societies and has remained a gourmet food up to this day. Different studies have revealed that several varieties of nuts have been around for quite a very long time and were used as food in the past. Recently researchers found evidence proving that nuts formed a major part of man’s diet 780,000 years ago in an archaeological dig that was done in Israel (Anonymous, 2021a). Nuts have plenteous health benefits to humans due to their fat, fibre and protein content and because of this, there has been a significant increase in the consumption of nut and nut products lately, which is also highly recommended by medical authorities. There is a persistent great deal of evidence from clinical and epidemiologic studies of the beneficial effects of nut consumption on the risk of Congenital Heart Defects (CHD), which includes major and emerging cardiovascular risk factors, diabetes in women, sudden cardiac death as well as other chronic diseases (Ros, 2010). A study was done by Bao et al. (2013) to find the association between nuts consumption and mortality, and it was found out that after adjusting other known and suspected risk factors, nut consumption had an inversely proportional relationship with the total mortality for both men and women.

Given all the advantages that nuts have, it is clear that people would like consume more nuts. Over the last decade, we have witnessed a major shift in people's priorities in their feeding habits. This change in feeding habits is caused by the demand for a healthier and disease-free life. Nowadays, people care much about their health and well-being and are very careful with what they should and should not consume. Plant-based proteins, which are abundant in nuts, have proven to be very helpful to our bodies, and the more people become aware of the benefits associated with nuts consumption, the more they tend to consume more and more. The increase in nut consumption can be proved with the increase in nut production all over the world. Literally, there has been an increase in the production of all types of nuts every year. An International Tree Nut Council – Conseil International des Fruits Secs body (INC-CIFS), which was established in 1982 in an international world tree congress held in Sorrento, Italy, has been correcting data about the production of nuts worldwide. In the statistical yearbook 2020/2021, the amount of nuts produced worldwide over the past ten years has been presented as shown in (Figure 2.1). From this data it can be clearly seen that the production of nuts has been rising at an increasing rate.

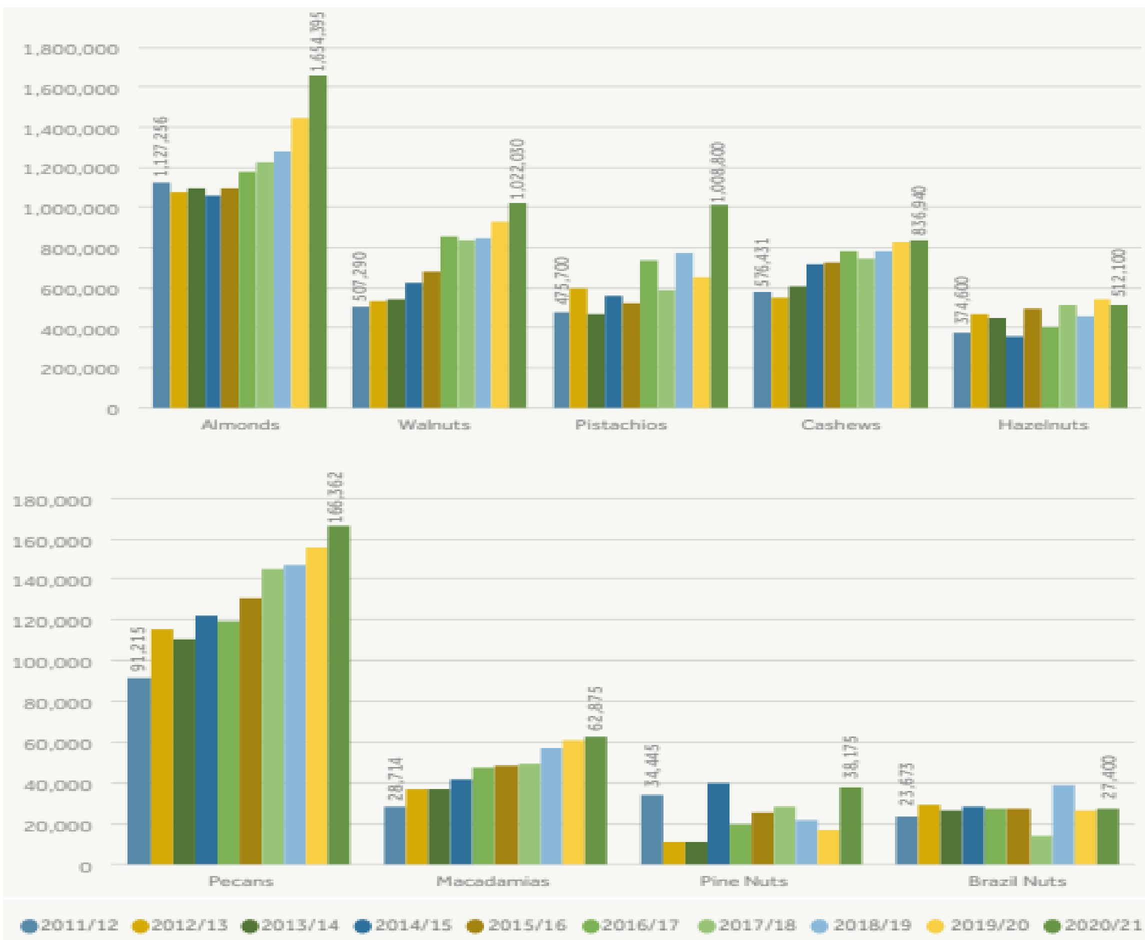

Figure 2.1. Kernel basis world tree nut production 2011/2021 in metric tons (Anonymous, 2021b)

2.2. Cashew Tree

2.2.1. Introduction of Cashew Tree

The cashew tree (*Anacardium occidentale*) of the *Anacardiaceae* family is an evergreen tropical tree that produces cashew apples and cashew nuts. Depending on different factors like soil characteristics and climate, the cashew tree's height can vary between 8-20 m, but for the dwarf cultivars, they can be as short as low as 6 m (Catarino, Menezes, & Sardinha, 2015). The cashew nut trees normally start flowering when they are about three years old, while the maximum production can be obtained by the eighth year, and this period of full production may last for 20-30 years (Jøker, 2003). The cashew originated in the north-eastern part of Brazil and was domesticated by the locals in those areas. When the Portuguese arrived in Brazil in the 16th century, they recognized the value of the cashew tree and took the crop to their other colonies around the world (Smith, Williams, Talbot, & Plucknett, 1992). During this time, the Portuguese missionaries managed to spread the cashew tree through East Africa and India along the seacoast in places of low altitudes. Soil erosion had been a long-time problem along these coastal regions, thanks to the introduction of the cashew tree in these areas, which helped to control soil erosion (Woodroof, 1979).

Figure 2.2. Cashew tree plantation

2.2.2. The Fruit (Cashew Nut)

The cashew fruit is composed of a cashew apple and a nut that is attached to the bottom of the apple, see (Figure 2.3) below (Britannica, 2020). When ripe they fall together and are separated thereafter. In spite of the nuts' dietary popularity, still, the pseudo fruits cashew apple) do not get much attention. The cashew apple looks like an oval-shaped mini boxing glove. From a botanic point of view, cashew apples are accessory fruits since they grow after the full development of the cashew seeds (Bhuwad & Pawaskar, 2019). While different researches are still ongoing to understand the health advantages of the cashew apple, current evidence gives a promising hope that consuming the cashew apple may help our physical fitness due to its good composition of vitamin C.

The cashew seed (nut) hanging at the bottom has a conical c-shape. Inside the outside shell of the cashew is an edible kernel, which is called "the nut". The cashew kernel is generally soft, meaty and white in its raw form but changes its taste and colour after being roasted. Containing vitamin C five times more than orange, vitamin B1, iron and calcium more than other fruits like avocados, citrus and bananas makes the cashew nut an excellent source of different nutrients that would benefit our bodies (Kluczkovski & Martins, 2016).

Figure 2.3. Cashew fruits hanging from a cashew tree

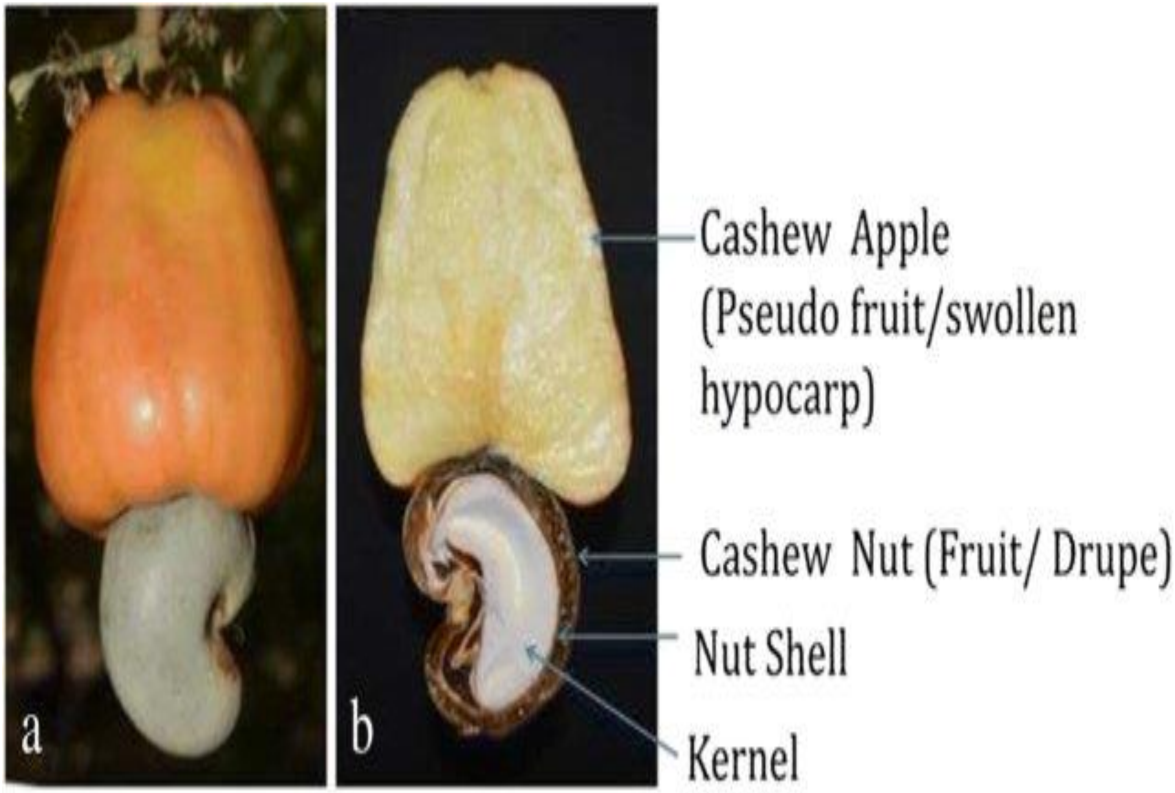

Figure 2.4. Cross section view of the cashew fruit (Savadi, Bm, & Preethi, 2020)

Another constitute of the cashew fruit is a dark viscous brown fluid called cashew nut shell oil/liquid. This important by-product of cashew nut is located within the soft honeycomb shell of the fruit. (Figure 2.5) shows the major components of the nut shell liquid. The cashew nut's shell is roughly 1/8 inch thick. This pericarp fluid of the cashew nut can be used as raw material in several applications. In the petrochemical industry, cashew nut shell oil is being used as a raw material. In recent years, petroleum resources have been overstretched due to the rapid growth in population around the world. Together with a number of other factors, this has contributed to the fast drainage of the global petroleum reserves. Therefore, in order to maintain the living standard and the continuation of industrial operations, which are crucial for human survival in this era, an alternative source of fuel has to be brought forward, and cashew nut shell oil has proven to be the best back-up plan (Taiwo, 2015).

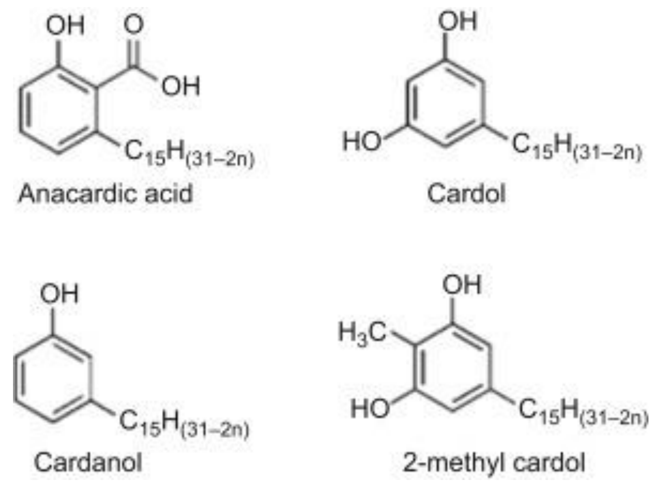

Figure 2.5. Major components of cashew nut shell liquid (Rwahwire, Tomkova, Periyasamy, & Kale, 2019)

2.2.3. Cashew Cultivation Practice

A cashew tree is famous for producing cashew nuts, which is sometimes referred to as "white gold" due to its economic importance (Mahajan, 2011). Apart from the food production purpose, cashew trees can be helpful in different ways, for example in reforestation projects and controlling soil erosion. Many crops are very sensitive to environmental conditions, but luckily for the cashew tree, which can survive even if planted in places where many crops cannot survive. As surprising as it may seem, this crop can even be planted in hilly and marginal areas. Moreover, it is drought-resistant and can stand wet season given that proper drainage systems are well installed on the farm. Despite being tolerant to different conditions, still, it is recommended to consider the following conditions and cultivation practices in order to get the ultimate yield. Below are the recommended conditions and cultivation practices for cashew production.

2.2.3.1. Climate requirement for cashew production

As we already know, the cashew is a tropical tree that can even survive in high temperate areas. Frost can affect young cashew plants as they tend to be sensitive to this kind of weather. Cashew cultivation works well in places of altitude 700 m above the sea level and with temperatures above 20 °C for the most time of the year. Areas with annual precipitation ranging from 1000-2000 mm and temperature ranges between 20-30 °C are ideal. Even though the cashew tree may grow in places with an even distribution of heavy rainfall throughout the year, these areas are not favourable. A dry season for at least four months is ideal for ultimate yields. Excess and heavy rains during the flowering period may affect the plant's progress by causing fungal diseases and flower/fruit drops.

2.2.3.2. Suitable soil for cashew production

The good thing about cashew is its adaptability to different soil conditions without affecting its productivity to a significant level. But we should keep in mind that, even though it can be planted in poor soils, good soils would give better results. Cashew can thrive in pure sandy soils, but deep sandy loams and well-drained is the best choice to overcome the problem of mineral deficiency. Soils with poor drainage are not suitable, for example, heavy clay soils with pH above 8.0. Soils with high values of salinity and alkalinity are not suitable too. Costal sands, lateritic soils, and red sandy loam soil are the best for ultimate yield in cashew production.

2.2.3.3. Land preparation and planting procedures

When preparing the land for cashew plantation, it should be ploughed and levelled thoroughly. If it is a forest, the bushes should be cleared and burned in advance. Once the bushes are cleared out, soil trenches should be dug across the contours to guarantee more satisfactory moisture conversation. After preparing the soil, then the seeds can be planted. First, the seeds are put into a planting bag with a sterilized and loose soil mixture. Usually, one hole may take up to 3/4 seeds, and only the strongest one will be left to develop. The taproot of the cashew plant tends to grow very fast, which means a depth of 350 to 400 mm is ideal for the plant bags. About 8 x 5 m plant spacing is recommended during the first three years and alternate trees should be removed until a permanent spacing of 10-12 m is attained. Since cashew grows quickly in the first three years, it can thrive well even with little attention dedicated to the orchards.

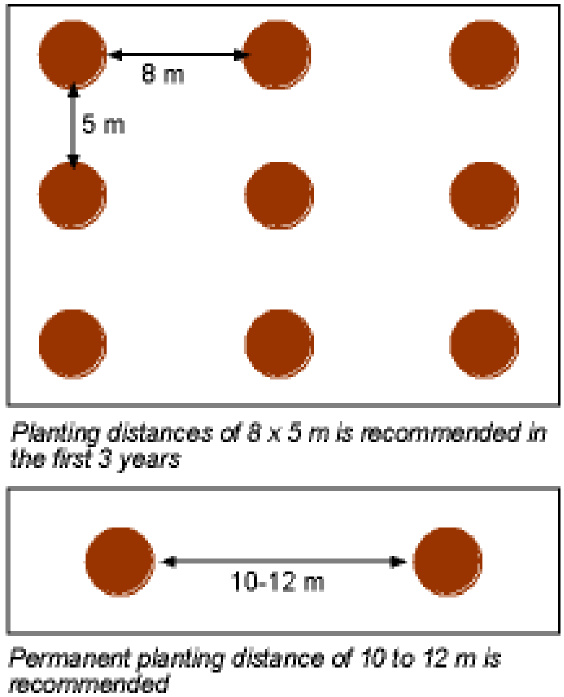

Figure 2.6. Recommended plant spacing for cashew plantation

2.2.3.4. Irrigation requirement

We have seen before that cashew does not grow well in places with heavy rains throughout the year. A cashew tree can go months without water, but this does not mean that it does not need water at all. Timely irrigation specifically for the young trees is essential to assure adequate returns. On average, 1800 litres of water are required for a mature tree in two weeks.

2.2.3.5. Mulching requirement

Mulching is essential in cashew cultivation when there is no adequate moisture content in the soil. Mulching also serves in preventing soil erosion. Materials such as dry or green grass, which can be easily obtained on the farm, can be used as mulching materials. Using these organic materials also helps to improve soil fertility. However, many experts recommend using black polythene materials due to their advantages in increasing the growth and yield of cashew. It also helps to protect the soil from soil-borne insects and pests attacks.

2.2.3.6. Manure and fertilizers requirement

Even though the use of manure and fertilizers is not mandatory, the use of manure and fertilizer can help boost the soil's productivity by increasing more nutrients in the soil. Proper utilization of manure and fertilizer in cashew cultivation has a great impact on crop yield. An adequate ratio of NPK should be maintained in the soil to get the best results. As a general rule, 10-15 kg of farm manure is recommended per plant to provide sufficient organic matter in the soil. For fertilizers, different ratios of NPK may be used depending on the soil's fertility. Fertilizer may be applied twice or once a year; the most important thing is that it should be applied when there is enough moisture content in the soil.

2.2.3.7. Training and pruning in cashew cultivation

To ensure a healthy and strong plant, training and pruning are very important processes. During the early stages of the plant, a single stem should be maintained until it grows to about 0.75-1.0 m above the ground. This is essential to help reduce the competition for nutrients. Usually, many sprouts come out from the rootstock and tend to compete for nutrients, so these sprouts have to be frequently pruned out. Also, training and pruning the cashew tree helps to give it a proper shape, this involves chopping off criss-cross and weak branches. In addition, the flowers that appear in the first two years should be removed so that the plant can be allowed to mature well. After the tree has matured enough, usually after three years, there is no minimum to no pruning necessity. At this point, only regular removal of dead/dried branches and criss-cross branches is performed to maintain the plant in a healthy state.

2.2.3.8. Weed control in cashew cultivation

Weeds are known for being a problem in agriculture which means weed control is among the necessary tasks that must be carried out on the farm. The same applies to cashew cultivation. It is generally recommended to conduct the first weeding before the first dose of manure or fertilizers are applied on the farm later a light digging weeding operation can be carried out before the rainy season ends as a second phase. While the slashing technique may be used, cutting the weeds and hoeing are more effective. Chemical weeding is also applied in different places, but according to various feedbacks from farmers, it has not proven any significant importance yet. However, in cases where there is a labour shortage or high wages, chemical weeding might be helpful.

2.2.3.9. Pest and disease control in cashew cultivation

Another important thing that affects the yield of cashew is pest and disease control. For higher yields and high-quality cashews, proper pest and disease control measures should be done in the plantation. There are almost 30 different species of insects that attack cashews causing about 30% losses of cashew yield. The most-reported pests include fruit and nut borer, root and stem borer, flower thrips and tea mosquitos. In terms of disease, lucky for the cashew tree, it does not have many serious diseases affecting it except one disease caused by fungus. This disease is called powdery mildew and affects the young cashew trees. It generally occurs during cloudy weather. By applying 2% of sulphur W.P., this disease can be controlled. Other diseases which can be observed in cashew trees are Anthracnose and Dieback.

2.2.3.10. Harvesting of cashew nuts

Cashew fruit is made of a cashew apple and a cashew nut attached at the bottom. As a general rule, the perfect time for harvesting is two months after the formation of the first fruit when the nut turns grey and the apple changes to pink or red. The nuts can be harvested directly from the tree, or you can wait to correct them from the ground after they fall. The main disadvantage to plucking nuts from the tree is that you may end up plucking nuts which aren’t fully matured and this will affect the quality of the nuts. So, it is usually recommended to collect the nuts after they fall down. After harvesting, the nut can be separated from the apple by twisting it off. The nuts should be dried and prepared for the following processes. One thing to keep in mind while detaching the nut from the apple is that the nut should be removed clearly and clean without any remains of the apple because this may affect the quality of the nut. A sample of the neatly separated nut and an unacceptable separated nut is shown in (Figure 2.7).

Figure 2.7. A properly separated nut on the left and an unproperly separated nut on the right

2.2.4. Post-harvest Handling Operations

Living organisms are divided into two major groups, one of which is plants. Plants can be found in different sizes and forms, from trees, flowering plants, bushes, herbs, grasses to the tiniest forms like mosses and algae. Plants are a fundamental piece to the functioning of the biosphere (Fernando, 2012). Since they are also living organisms, this means they have some necessities for their survival. These include nutrients, water and air. Plants are capable of making their foods through photosynthesis (Usman, Abubakar, Alaku, & Nnadi, 2014). The fun fact is that these agricultural products stay alive even after being detached from their parent plant, which literally means their cells will continue with their daily activities. And for a living cell to perform its activities, it will need energy, the energy that can only be provided by the parent plant, which is no longer available. And it's at this point when agricultural products start to deteriorate. So, in order to prevent this from happening, post-harvest handling operations need to be done to elongate the shelf life of the product (Opara & Mditshwa, 2013). From an economic point of view, post-harvest handling operations are essential as they help maintain the quality of the food product from the time it is harvested until it reaches the final consumer (Silva, Phillips, & M, 2009). Various activities can be performed in the post-harvest handling phase depending on several factors, including the type of the agricultural product, the expected shelf life and others. The most standard ones are; cooling, cleaning, sorting, grading and packaging (Banks, 2009). Let's look at some post-harvest handling operations done in cashew production.

2.2.4.1. Raw Cashew Nuts Drying

Various physical properties of agricultural products like density, weight, conductivity and many others are influenced by moisture content. The Moisture Content of an agricultural product is the measure of the quantity of water possessed in the product (Gone & Vp, 2017). Moreover, Moisture Content has a significant effect on the shelf life or durability of agricultural products. We can alter (increase or reduce) the amout of moisture retained in the agricultural products by merely adding water or reducing water (drying) from the product. After the collected nuts have been detached from the cashew apple, they must be dried to reduce the moisture content. In the nut industry, moisture content less than 5% is preferred (Ajith, Stephen, Kumari, & Potty, 2014). Different methods can be used to dry raw cashew nuts. A traditional method used by local farmers is sun-drying, where the nuts are spread on a drying mat or concrete floor in an open area for 2-4 days. When using this method, the nuts should be turned around now and then to ensure that all parts dry well. The use of solar panels has proved to have improved efficiency in dry raw cashew nuts. Using this method helps to save time, reduce the cost of operation and improve the quality of nuts compared to the traditional technique (Huddar & Kamoji 2019).

2.2.4.2. Roasting/Steaming of Raw Cashew Nuts

Before removing the shell of cashew nuts, the nuts need to be roasted because this facilitates the process of shelling. A technique that is widely adopted in medium-scale cashew processing plants for roasting nuts is the use of boilers to steam them. Two diverse approaches are practical for this process. The cashew nuts can be directly or indirectly steamed. The direct approach involves boiling the nuts in the same place with water, while the indirect approach uses steam from the water boiling in a different chamber to heat the nuts (Mohod, Khandetod, Sengar, & Shrirame, 2012). Typically, it takes roughly between 20-40 minutes for nuts in the drum to get satisfactorily roasted. After the roasting process is finalised, the roasted nuts are cooled in the open air for almost 12 hours, after which the shelling process can be initiated (Kosoko, Sanni, Adebowale, Sokabi, & Oyelakin, 2009).

2.2.4.3. Shelling

The edible cashew kernel is enclosed inside a unique kidney-shaped shell which is difficult to crack open. Moreover, inside the shell, is a liquid called cashew nut shell liquid that can be significantly hazardous if not handled with great care. So, the shelling process is particularly crucial to ensure that both workers and consumers are free from the effect of this liquid (Balasubramanian, Sherfudeen, Kaliannan, & Murugesan, 2016). Traditionally, cashew nut shelling has been conducted manually. With the aid of a wooden mallet and a sharp object, experienced workers can open the shell to extract the kernel inside it. This method has a lot of drawbacks, including poor quality of the extracted nuts since the workers are likely to break the cashew kernels, labour intensity and high operation cost. Also, it is not ergonomic as there are a lot of reported cases of workers suffering from eye irritation and burning rushes. Due to these setbacks of manual shelling, mechanical shelling methods have to be developed (O. Kilanko et al., 2019). A compression test to investigate the mechanical and physical properties of roasted cashew nuts was conducted by (Ojolo, Damisa, Orisaleye, & Ogbonnaya, 2010), where the cashew nuts were subjected to altering impact loads at different orientations until a critical force, which would crack open the shell without impairing the kernel inside it, was found. To obtain this critical force, parameters like average nut mass and kernel/shell moisture content were taken into account. All components were designed and assembled. The machine was tested and found to have 95 per cent efficiency. Another significant study was done by Tran et al. (2011) to design a robotic system to automatically shell cashew nuts with very little to no human intervention at all.

2.2.4.4. Cashew Nut Kernel Drying

After the cashew kernels have been extracted from their shells, there is still a thin skin to peel out. Putting the nuts in a heating oven maintained at a particular temperature simplifies the peeling process and makes the skin brittle. Eang & Tippayawong (2018) investigated three variables that affect the efficiency of the heating process. The variables examined were temperature, the velocity of rotating the drum and the drying duration. The results in this study suggested that a constant temperature of 115 °C, a velocity of 4 m/s and 30 min of drying duration were ideal for obtaining the ultimate yield in kernel quality. If no sophisticated machines are available, the kernels can be sun-dried naturally.

2.2.4.5. Peeling the Tesla

A thin layer that is left covering the cashew kernel after it has been removed from the outer shell is called the cashew nut testa. It is red-brownish in colour (Lin Hu et al., 2017). Even though the testa does not have many uses for human nutrition, its high fibre content makes it useful as a feed ingredient for monogastric animals' food manufacture (Donkoh, Attoh-Kotoku, Kwame, & Gascar, 2012). This inner shell has to be removed to remain with a white nut. After the kernels have been left to cool for some time, they can be easily peeled. Each nut is peeled separately then sent to the grading section.

2.2.4.6. Grading

At this point, the cashew kernels are ready for grading into different grades. The grading of cashew kernels is done based on various factors such as colour, size and shape. Based on colour, cashew kernels are divided into two groups, i.e., white and scorched. Based on size and shape, we have, i.e., whole nuts, split nuts, butts and baby bits. For the purpose of business, there are internationally acceptable grades that are being used. These grades incorporate colour, size and shape to categorize a nut to a particular grade. To make it simple to understand, the cashew grades have been grouped into 4 main categories, which are;

a) White grades

As we all know, cashew nuts are white. This group of white nuts has two main classes, i.e., white wholes and pieces. The tables below give more details of the grades in each of the classes.

Table 2.1.Cashew kernel grades for white wholes

Grade	Pieces per pound (453.59 g)	Colour and characteristics
WW-180	140-180 Nuts	This is the highest standard of cashew, internationally known as the King of Cashew. Large in size and very expensive. White/pale ivory in colour.
WW-210	180-210 Nuts	The 2nd standard of cashew nuts internationally known as Jumbo Cashew Nuts. Large in size and very expensive. White/pale ivory in colour.
WW-240	220-240 Nuts	The 3rd standard of cashew nuts internationally known as Premium Large Nuts. Large in size and very expensive. White/pale ivory in colour.
WW-320	300-320 Nuts	The 4th standard of cashew nut internationally known as Standard Large Nuts. White/pale ivory in colour.
WW-450	400-450 Nuts	The 5th standard of cashew nut internationally known as Standard Nuts. White/pale ivory in colour.
WW-500	460-500 Nuts	The 6th standard of cashew nut. White/pale ivory in colour.

Table 2.2. Cashew kernel grades for white pieces

Grade	Pieces per pound (453.59 g)	Colour and characteristics
WS (white split)	N/A	White/pale ivory or light ash. Kernels Split naturally lengthwise
WB (white butts)	N/A	White/pale ivory or light ash. Kernels broken cross-wise (evenly or unevenly) naturally attached
LWP (Large White Pieces)	Kernels broken into more than two pieces and not passing through 4 mesh 16 SWG sieve/4.75 mm. I.S.Sieve	White / Pale ivory / Light ash. Broken Pieces
SWP (Small White Pieces)	Broken Kernels smaller than those described on LWP but not passing through 6 mesh 20 SWG Sieve/2.80 mm I.S. Sieve	White / Pale ivory / Light ash. Small broken Pieces
BB (Baby bits)	Plemules and broken Kernels smaller than those described as SWP but not Passing through a 10 mesh 24 SWG Sieve/1.70mm I..S. Sieve	Light brown / Light, Deep ivory / Light ash-grey. Broken Pieces, Baby bits, Granules.

b) Scorched grades

Generally, all nuts have to be white, but due to some factors like over-heating, some nuts burn during the processing, which leads to scorched nuts. Scorched nuts have no nutritional difference from white nuts. The only difference is appearance and a little bit difference in taste due to over burning. The tables below show the grades for scorched nuts. Just like the white nuts, here too we have two main classes, i.e., scorched wholes and pieces.

Table 2.3. Cashew kernel grades for scorched wholes

Grade	Pieces per pound (453.59 g)	Colour and characteristics
SW-180	140-180 Nuts	Have all the other characteristics of white kernels and have the same nutritional qualities. known Internationally known as SW180 cashew. Slight brown color due to longer roasting.
SW-210	180-210 Nuts	Have all the other characteristics of white kernels and have the same nutritional qualities. known Internationally known as SW210 cashew. Slight brown color due to longer roasting.
SW-240	220-240 Nuts	Have all the other characteristics of white kernels and have the same nutritional qualities. known Internationally known as SW240 cashew. Slight brown color due to longer roasting.
SW-320	300-320 Nuts	have all the other characteristics of white kernels and have the same nutritional qualities. Internationally known as SW320 cashew. slight brown color due to longer roasting.
SW-450	400-450 Nuts	have all the other characteristics of white kernels and have the same nutritional qualities. Internationally known as SW450 cashew. A slight brown color due to longer roasting.

SW-500	460-500 Nuts	have all the other characteristics of white kernels and have the same nutritional qualities. Internationally known as SW500 cashew. A slight brown color due to longer roasting.

Table 2.4. Cashew kernel grades for scorched pieces

Grade	**Pieces per pound (453.59 g)**	**Colour and characteristics**
SS (scorched splits)	N/A	Kernels split naturally lengthwise. Kernels may be scorched/slightly darkened due to over-heating while Roasting or drying in drier/borma
SB (scorched butts)	N/A	Kernels broken crosswise (evenly or unevenly) and naturally attached. Kernels may be scorched/slightly darkened due to over-heating while roasting or drying in the drier/borma
SP (scorched pieces)	Pieces passing through a mesh 16 SWG sieve/4.75 mm I.S. Sieve	Kernels may be scorched/slightly darkened due to over-heating while roasting or drying in drier/borma
SSP (scorched small pieces)	Pieces smaller than SP but not passing through a 6mesh 20 SWG Sieve/2.80 mm I.S. Sieve	Kernels may be scorched/slightly darkened due to over-heating while roasting or drying in drier/borma

c) Dessert grades

Table 2.5. Cashew kernel grades for dessert wholes

Grade	Pieces per pound (453.59 g)	Colour and characteristics
SSW (scorched wholes seconds)	N/A	Kernels may be over-scorched, immature, shriveled (Pirival), speckled (Karaniram) discoloured and light blue.
DW (dessert wholes)	N/A	Kernels may be deep scorched, deep brown, Deep blue, speckled, dis-coloured and black spotted

Table 2.6. Cashew kernel grades for dessert pieces

Grade	Pieces per pound (453.59 g)	Colour and characteristics
SPS (scorched pieces seconds)	Kernels broken into pieces but not passing through a 4 mesh 16 SWG sieve/4.75 mm I.S. Sieve	Kernels may be over-scorched immature, shrivelled (Pirival) speckled(Karaniram) discoloured and light blue
DP (dessert pieces)	Kernels broken into pieces but not passing through a 4 mesh 16 SWG sieve/4.75 mm and I.S.Sieve	Kernels may be deep scorched, deep Brown, blue, speckled, discoloured and black spotted

2.3. Cashew Production in Tanzania and Its National Importance

2.3.1. Introduction about Tanzania

Tanzania is a country found within the Great Lakes Region of Africa. Its official name is the United Republic of Tanzania (Swahili: Jamhuri ya Muungano wa Tanzania). It is located in the eastern part of Africa bordered by Kenya in the northeast, Uganda in the north, Malawi and Mozambique in the south, Zambia in the southwest, Rwanda, Burundi and the Democratic Republic of the Congo in the west, and finally, the Indian ocean in the east (Mugnier, 2008). Having been located near the equatorial region (Latitude: -6° 22' 22.17" S Longitude: 34° 53' 32.94" E) means the country has a tropical climate for the most part (Mascarenhas, Bryceson, Ingham, & Chiteji, 2021). However, it is worth pointing out that the climate in Tanzania differs considerably from place to place. During the hottest season, which lasts from November to February, the temperature in most parts of the country lies between 20-30 °C. However, in the highland regions, the temperature rarely exceeds 20 °C, it usually ranges between 10-20 °C. During the coldest season, which is between May and August, the temperature lies between 15–20 °C. Different parts of the country experience diverse forms of rainfall seasons. The northern part experiences a bi-modal rainfall season, where it rains twice a year; October–December and March-May. Whereas the western, central and southern parts of the country experience a uni-modal rainfall season with only a single rainfall period a year between October-April (Zorita & Tilya, 2002).

Just like many other developing countries, agriculture is the backbone of the economy of Tanzania. By benefitting the nation with a diverse production of products including staple food crops, cash crops and a variety of livestock and with almost three-quarters of the nation's workforce involved in it, the agriculture sector represents approximately 30 per cent of the nation’s GDP. It is indeed the most important and the largest sector of the Tanzanian economy, and different strategies are being implemented to strengthen and improve it (Mihayo & Swai, 2019). Depending on various factors including, a) economic factors, such as production cost, seed prices and marketability of the products; b) physical factors such as the availability of water and the quality of soil; c) crop characteristics such as crop yield and resistance to pests and drought; d) availability of resources like fertilizers and agricultural implements e) and finally personal preference; different crops are preferred across the country. The most common cash crops produced in Tanzania include coffee, cashew nuts, cotton, sisal, tea and tobacco. For the food crops, bananas, beans, sorghum, sugar cane, maize, potatoes, rice, and wheat are predominantly grown in many places (Makundi, Massawe, & Laswai, 2006). The graph below shows the most produced commodities in Tanzania between 1994-2020 according to Food and Agricultural Organisation (FAO) reports.

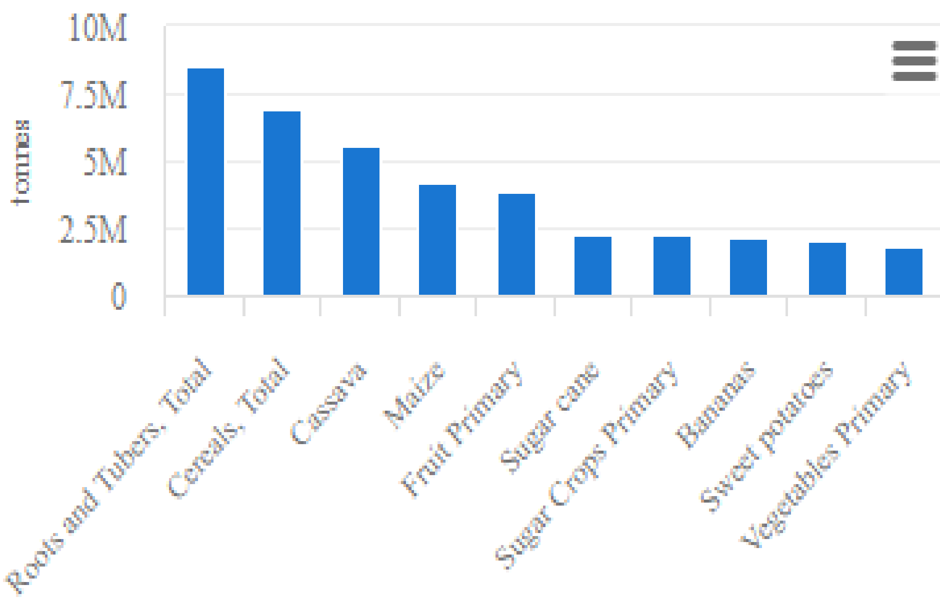

Figure 2.8. Most produced commodities in Tanzania, 1994-2020 (FAO, 2022)

2.3.2. Cashew Production

Cashew is one of the major crops produced in Tanzania that contribute much to the country's economy through foreign exchange earnings. Cashew was introduced in Tanzania in the early 1930s by the European Missionaries (Ngatunga, Cools, Dondeyne, & Deckers, 2001). According to FAO statistics, Africa produces about 50% of the world's raw cashew nuts, followed by Asia, producing about 40%. According to the data published in 2020 by the FAO, Tanzania contributes about 20% of the cashew produced in Africa, following behind Burundi, Cote D'Ivoire and Nigeria.

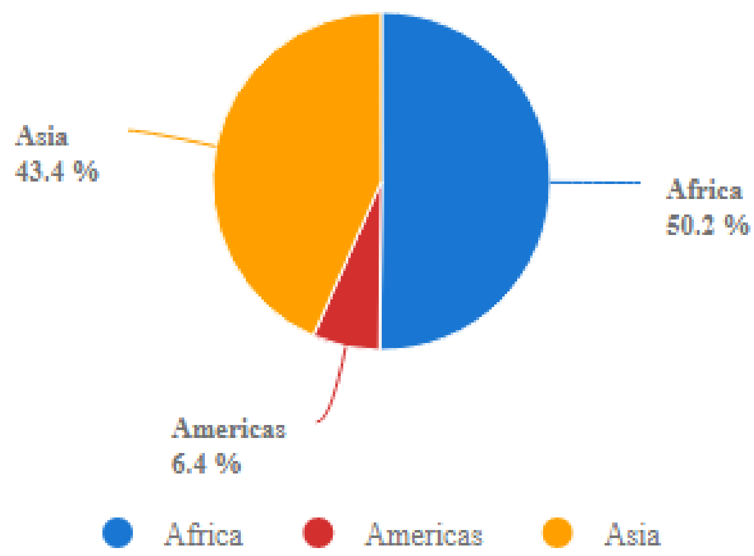

Figure 2.9. Production share of cashew nuts, with shells by region 1994-2020 (FAO, 2022)

The production of cashew in Tanzania is still conducted locally by small farm holders using local methods, the main reason for this problem being lack of government strategies and poor regulations. To demonstrate this point, let's look back in time to the 1980s when the industry was on the verge of crumpling with a tremendous decline in the annual production to as low as 20,000 tones. After the independence of Tanzania in 1961, the country adopted a socialist ideology called (Ujamaa in Swahili). In simple words, the Ujamaa ideology meant people were to live in a community as a unified division with cooperative economics working jointly in all economic activities and making profits from them as a unit to help supply the necessities of living in their community (Gordon, 1974). During the early times of post-colonization, the country needed a strategic plan to follow, and as a national development project, Ujamaa was taken as the way forward. The Ujamaa ideology was transformed into a political-economic management model in various ways. One of which was the villagization and collectivization of production (Pratt, 1980). The implementation of this strategy witnessed a tremendous relocation of people as well as the collectivization of villages. It was due to these government's interventions that many people were forced to abandon their farms, which eventually affected the cashew yield.

Moreover, the government interfered with the harvesting and marketing activities and nationalized several cashew processing industries denying the farmers and cashew producers the flexibility to perform their activities freely. After the abolishment of the Ujamaa ideology, various economic reforms were implemented to revive the cashew industry in the country. These reforms include abolishing the state monopoly and liberalizing the trade. Since these actions were taken, there has been a fantastic recovery in the industry. According to FAO statistics, over 300,000 tonnes of cashew nuts were produced in Tanzania in 2018. This marks the peak in cashew yield ever since cashew started being grown in Tanzania.

As we have seen before, the majority of cashew growers in Tanzania are small farm holders, these account for more than 85% of all cashew growers. Most of these farmers cannot afford to use developed technologies. Hence, most of the farm activities rely on manual labour. Cashew is mostly grown in the coastal region of Tanzania. About 70% of the total yields come from the Mtwara Region. Behind Mtwara is Lindi Region contributing about 18%, Pwani Region producing 8%, Ruvuma Region giving 4%, and the rest comes from Tanga, Dodoma and Iringa regions (Masawe & Z, 2014). Despite the tremendous increase in the cashew yield over the past years that we have just seen, the cashew industry in Tanzania is still facing one major problem. This problem is the lack of enough processing units. Currently, the country has only four running facilities to process cashew nuts, which have a maximum capacity of 98,000 tonnes only, which makes about 30% of the total yield. These facilities are distributed in Dar es Salaam, Tunduru and the other two in Mtwara. However, due to some issues, these facilities cannot operate up to their maximum capacity, hence, of all the produced raw cashew nuts in Tanzania, only about 10% are processed inside the country. The remaining amount, about 90% of the total yield is exported abroad as raw cashew nuts. India has been the main destination for raw cashew nuts from Tanzania for a very long time. But with the recent shift in the economy, the Vietnam market has been growing at a fast rate. By exporting such a massive quantity of raw cashew nuts, the nation misses out on a very big foreign income that it would have earned if it had processed the cashew nuts domestically and sold them as processed products (Krepl, Kment, Rajdlová, & Kapila, 2016).

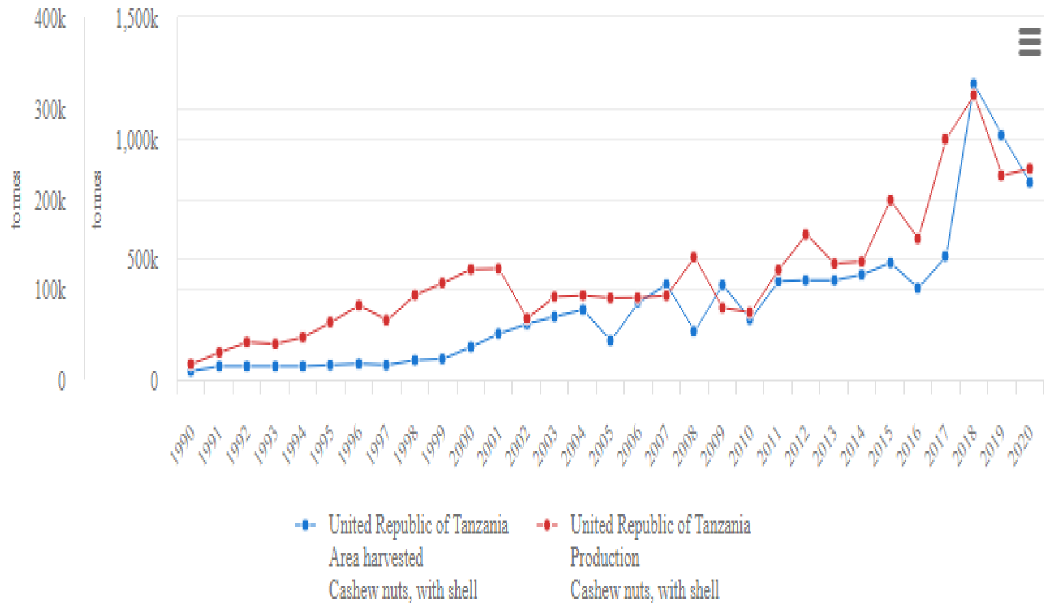

Figure 2.10. Production/Yield quantities of cashew nuts, with shell in Tanzania 1990-2020 (FAO, 2022)

2.4. Inspection by Machine Vision

Vision is one of the most vital senses of humans that we humans have happened to rely on more than other senses in performing our day-to-day tasks. Vision provides us with a peculiar sense of sight, enabling us to perceive and interpret things, which helps in decision making and problem-solving. Human vision is a very complicated system that is yet to be thoroughly understood (Kim, Nam, & Suk, 2004). As important vision is for humans, so it is for computers, too. For computers to be able to make intelligent decisions, they need to be able to see. So, this ability of a computer to see is what is referred to as "Machine Vision" in the process of simulating human vision, a computer avails itself of cameras, signal converters (analogue to digital) and signal processing units (Labudzki, Legutko, & Raos, 2014). Figure 2.11. shows the hardware components of a typical machine vision system. In any vision system, two critical specifications that are worth mentioning are the resolution and the sensitivity. The degree to which the computer can distinguish between objects is its resolution, whereas sensitivity is the capability to notice weak impulses in dim light or invisible wavelengths. In general, it is hard to have both resolution and sensitivity at their highest quality because they are inversely related to each other, meaning the increase in one of them reduces the other one. However, a study performed by John demonstrated that the overall quality of an image can be improved by improving the resolution of the image at the cost of sensitivity (Fahey et al., 1992).

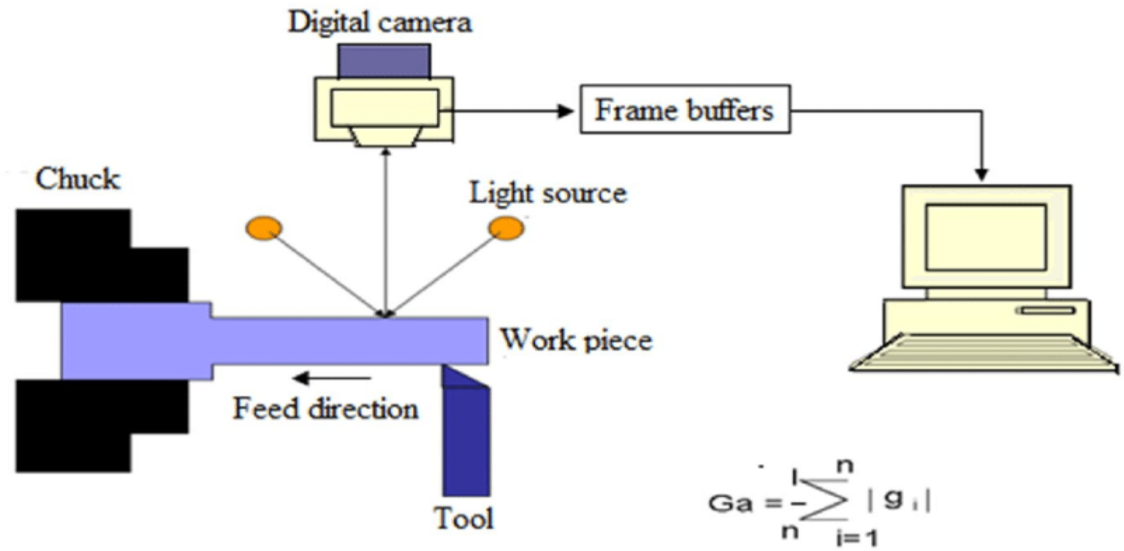

Figure 2.11. Hardware components of a typical machine-vision system (Subbiah, 2013)

Food safety is the number one priority in the food industry as it strengthens customers' trust, thus maintaining the company's revenue. Thoroughly inspection is the essence of food safety since the consumers and government regulations have zero tolerance for error. Generally, traditional inspection has been performed by skilled labour who check and access the quality and safety of the products. This approach has proven to be ineffective because it is prone to human errors, expensive and time-consuming. While the human eye's sensitivity is limited to electromagnetic wavelengths ranging between 390-770 nm, a video camera can push the limits of sensitivity of X-ray, ultraviolet or infrared wavelengths. So, in order to overcome the challenges associated with traditional inspection methods, machine vision systems, which guarantee increased levels of safety and consistency of the products at a relatively low cost of operation, are necessary. Over the past few decades, machine vision has acquired enormous attention in product inspection in the food industry. Numerous studies are being carried out to try and sophisticate this technology more. Machine vision systems can be useful in several ways in the food industry, including; measuring, counting, location, orientation detection, defect detection, sorting and grading. Moreover, machine vision systems also present considerable advantages over traditional inspection processes. They can operate for quite a long time without getting tired like how human labour would. They have improved accuracy, and by incorporating them within an automated line, they can work at an extraordinary speed. Another advantage of machine vision systems is the fact that there is very little chance of products getting damaged since the inspection is carried out without contact and by non-destructive procedures (Sarkar, 2017).

Machine vision systems are more adaptable and more manoeuvrable than other inspection approaches. In this approach, multiple-feature, which might be signals sent by various sensors or acquired through an algorithm or a combination of both, can be processed at the same time. The classification parameter (feature) can easily be modified to adapt to new conditions enhancing the performance of the system or using it for the inspection of a different product without making many modifications to its hardware. The algorithms enforced in machine vision systems perform a multi-category classification of the products.

2.5. Image Acquisition

Before any automated inspection task can commence a good image of an object under investigation must be captured by a camera and undergo some conversions into a manageable entity. This process is called image acquisition which involves the collection of images and transmitting them to an associated processing unit. The highest quality of the image is desired to ensure the effectiveness of the inspection system. For a camera to capture an image, it needs some sort of measurable energy. In order to achieve this, we need to have a proper source of energy which in this case will be light or just simply electromagnetic (EM) waves. An EM wave can simply be defined as a sinusoidal variation of the magnetic and electric fields of the massless entity called a photon. This source of light will cast light over the object of interest which will reflect this light to the detector of the optical system. Improper illumination may cause the light reaching the detector from the object to have insufficient intensity which may affect the key features of the object. So, it is crucial to always seek for the best illumination system. Figure 2.12. shows an overview of how a typical image acquisition system works.

After illuminating the object of interest, light will be reflected from the object and captured by the camera. Inside a camera, there is a detector which is the basic element for capturing images. For analogue cameras they use a film whereas for digital cameras sensors are used. All digital camera sensors operate based on the photoelectric effect to convert light into electricity by collecting the light charges and turn them into current or voltage as output. There are two main types of sensors used in the digital cameras industry today, which are CCD and CMOS image sensors. These two types of sensors work on different principles to capture images, and even though have no essential quality difference CMOS have been more preferred over the last few decades due to its promising performance compared to the CCDs (Mhd Zafer al Hosny, 2017). In the food inspection industry, most of the time the inspection is done in operation lines where the object of interest is in motion, there we need to use line-scan cameras. Using these cameras makes it possible to capture the image of an object as it passes through without stopping (Steger & Ulrich, 2021).

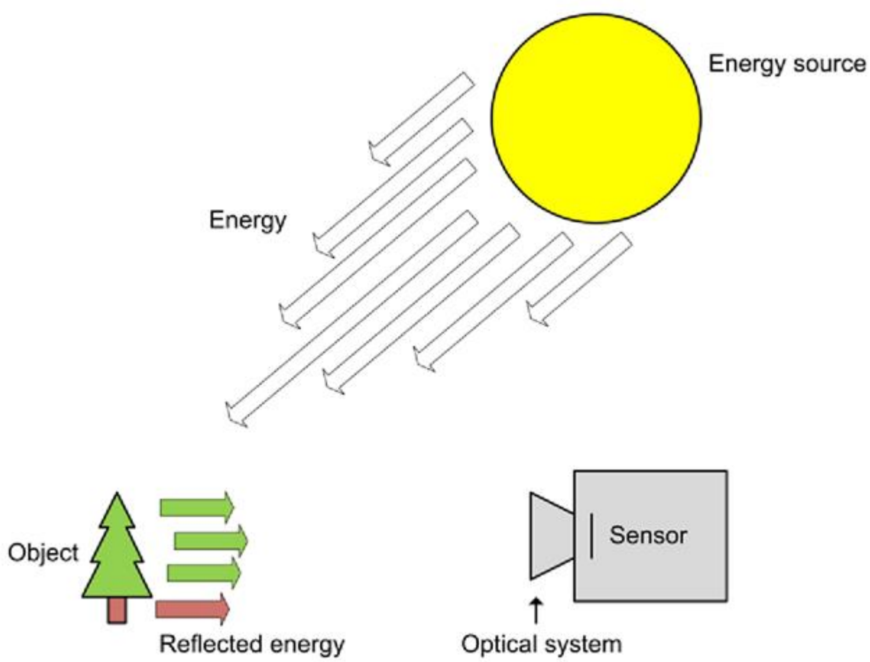

Figure 2.12. Overview of the typical image acquisition process, (Paulsen & Moeslund, 2020)

2.6. Image Pre-processing

As a data analyst, pre-processing or cleansing of data is a crucial step after collecting data. So, most of the data analysts spend a considerable amount of time in pre-processing the data before building the model. These data pre-processing activities may include outlier detection of outliers, removal of unwanted or noisy data and the treatments of the missing data. Just as important as data pre-processing is in data analysis so is, Image pre-processing in image processing. These operations are not performed with the aim of increasing the information content in image, rather they aim to improve the information in the image by suppressing undesired distortions and enhancing the features which are relevant for further analysis and processing tasks (Demant, Garnica, & Streicher-Abel, 2013).

There are 4 types of image pre-processing techniques as follows (Masters, 2018);

- Pixel brightness transformations/ Brightness corrections
- Geometric transformations
- Image filtering and segmentation
- Fourier transform and image restauration

2.7. Segmentation

Let’s consider a certain scenario in order to have a better understanding of image segmentation. Consider an image with of a nut. Since there is only object in this image i.e., a nut, then we can easily build a straightforward classifier model to predict that there is a nut in the given image. But what is we have two nuts of different kinds or a nut and another material like a piece of wood? So, in these situations where we have multiple objects, we rely on the idea of object detection (OD). With the object detection technique, we can predict the location as well as the class of all the objects in the images. This means that we would first need to know the location of the nut/object of interest in the image. And this brings us to the concept of localization. After locating our object of interest in the image we need to study its characteristics and know what it is made of, and this is what segmentation is all about

Generally, an image is a set of pixels arranged like a matrix. These pixels have different values depending on the information they contain. Some parts of the image may contain the information that is irrelevant to us or useless or may not contain any information at all (Semmlow & Griffel, 2021). So, when dealing with images it is not recommended to process the entire image at the same time because there are parts of the image which are not useful to use. By grouping pixels with similar attributes together we can divide the image into parts called segments and then we can process those segments which contain the information that is relevant to us (Jindal, Joshi, Jangwal, Rathi, & Jain, 2021).

Different types techniques can be used to segment images. These are as listed below;

- Thresholding Segmentation
- Edge-Based Segmentation
- Region-Based Segmentation
- Watershed Segmentation
- Clustering-Based Segmentation Algorithms
- Neural Networks for Segmentation

2.7.1. Thresholding Segmentation

By far the threshold method is simplest method for image segmentation in image processing. It is mostly used when the object of interest has a higher interest than the unwanted part in the image (background). The way this method works is by comparing the pixels in the image to a certain value called the threshold and partition the image accordingly. The threshold value can be kept constant or dynamic depending on the requirements. Keeping the threshold value constant works better when the image has very little unwanted information (noise). This method divides the image into two segments (object of interest and the not required sections) hence converting a grey-scale image into a binary image (Qingge, Zheng, Zhao, Wei, & Yang, 2020). Depending on the threshold value we can categorize the threshold method into the following categories;

- Simple Thresholding

As its name goes, this is a very simple thresholding approach and usually suitable for beginners. In this approach the image's pixels are replaced by either black or white. A certain threshold value will be set and all the pixels will be compared to it, those pixels whose intensity is less than this threshold value would be replaced by black, and those with a higher value would be replaced with white (Reddy & Swathi, 2017).

- Otsu's Binarization

As we saw above in the simple thresholding, a constant threshold value is picked and used for image segmentation. However, there is a challenge to this method, and this is to determine the right threshold value. A most commonly used approach is to test different values and pick the best one, but we can do better than this. Let's assume we have an image with two peaks in its histogram, in the foreground and the background. With this Otsu binarization approach, the approximate value in the middle of these two peaks can be taken as a threshold value. Regardless of having many limitations including not applicable for images that are not bimodal (images with multiple peaks in their histograms). This approach is quite useful in recognizing patterns, removal of unnecessary colours from a file and the scanning documents (Nyein & Khaung Tin, 2021).

- Adaptive Thresholding

Different factors like lighting and temperature may affect the quality of the image and hence make an image have background conditions. So, when an image has different lighting conditions it would not be a nice idea to use a constant threshold value to perform segmentation on over the entire image. This method performs better in these conditions because it allows us to apply different threshold values on the same image depending on the characteristics on specific areas of the image. The way it works is that it divides the image into smaller parts and calculate the threshold value depending on the conditions in this particular area (Woitha & Janich, 2004).

2.7.2. Edge-Based Segmentation

In image processing one of the common implemented techniques in segmentation is the one works on the concept of detecting edges of the objects in the image, and this is called Edge-based segmentation (Zhou, Du, & Wang, 2021). Edges contain crucial information which helps in finding the features of different objects in the image. Apart from that, this method helps in removing unnecessary and unwanted information from the image and also reducing then image size which makes it widely popular. Differences in contrast, saturation, texture, grey level and other properties help the algorithms used for edge-based segmentation to detect edges in an image. Edges in an image can be connected into edge chains matching the image borders and this can help increase the accuracy of segmentation process (Riad, Al-Taie, Basma, Saleh, & Abdalhasan, 2021).

2.7.3. Region-Based Segmentation

In this approach the segmentation algorithms partition the image into regions with similar features. The algorithm first locates the seed point in the image and from there it forms these regions. The seed point may be a small section or a large portion of the image. After locating the seed points, the algorithm either adds more pixels on to the seed points or shrinks them by merging them with other seed points to form regions with similar features (Abdul, 2019).

2.8. Classification features

An image is a bunch of pixels collected together to give a meaningful view/object. Hence, the smallest unit of an image is a pixel. A computer stores an image as an array of numbers defining the pixel values on every point. In inspection operations, we need to study the attributes of all pixels making up the image so that we can decide what action can be taken depending on different situations. But this sequential procedure is time-consuming and not appropriate for high-speed in-line operation. Hence, we need to have a set of features defining the image, which will be used as inputs in the classification algorithm to decide the category of the object. Features in this context are defined as quantitative pieces of information extracted from the image by applying various image analyses or image processing algorithms on that image. A pattern, which is a numerical definition of the image, is formed when various features of an image are merged.

There are many types of image features that can be used in image classification depending on the given task and conditions. The most common features include; edges, corners, blobs, colour features, morphological features, etc. As matter of a fact, there are too many of them to list all of them here. Pavlidis (1978) conducted a survey and concluded that image features can be categorized as external and internal features. External features are the ones that encode the boundary details of the image. Image segmentation is usually required to get the pixels on the outward shape of the image then mathematical operations follow to extract the useful features. Grey-level knowledge of the internal pixels is not required here. External features include; morphological features, boundary sequence, boundary chain code and Fourier descriptors. After dealing with the pixels on the boundary now, let's take a look at the pixels within the boundaries of the image. These features, which are acquired through analysing the pixels within the boundaries of the image, are what we refer to as Internal Features. Here both the grey level of the pixels and their location has a significant role to play. The most commonly applied Internal features include; texture features, colour moments, and grey-level histograms.

2.8.1. Feature Extraction

Feature extraction can be defined as the procedure of converting raw data into numerical values that can be processed while maintaining the details in the original data set. It produces more pleasing outcomes than using machine learning directly to the raw data. We can perform feature extraction manually or automatically. In the manual approach to extracting features, we need to identify and describe the relevant features for a particular problem and implement a method to extract them. It is always valuable to have adequate knowledge of the matter in concern as it helps to select the appropriate features for the particular problem. A more sophisticated approach for feature extraction is one that uses deep networks or technical algorithms to automatically extract the features from the images or signals without depending on human intervention. Since this approach does not need human intervention, this means that it is less prone to human errors hence more effective. With the rise of deep learning in the machine learning industry, the first layers of deep networks have almost completely taken over the feature extraction activity. However, for some applications, an expert is still needed to manipulate the feature extraction process in order to develop an effective model (Jain & Kumar, 1998).

2.8.2. Feature Selection

As we know, machine learning models need to be trained first before they can predict new data. Thus, we need a couple of features to train our models. Generally, we will have already extracted an enormous amount of features to train our model. But most of the time, the original dataset may contain some irrelevant and noisy data, which may not be of much use in training our model. Training a model with a huge amount of irrelevant features in our dataset will have a negative effect on the efficiency of the model. The model is likely to falsely learn the pattern of the data hence making wrong predictions in the future. Moreover, a massive amount of data would require more memory capacity for storage and slow down the training process. Thus, it is very essential to remove such features from our dataset. And this is what we refer to as feature selection. Using only the selected relevant features have many advantages, as listed below;

- It aids in dodging the curse of dimensionality.
- It simplifies the model, making it easily understood by researchers.
- It reduces the training time.
- It reduces overfitting.

There are two principal types of feature selection techniques: unsupervised approach, which can deal with unlabelled data, and supervised approach, which may further be categorized into; filter method, embedded method and wrapper method.

2.8.2.1. Filter Method

Filter methods are normally applied in the pre-processing phase, where the machine learning algorithm that will be used to build the model is not taken into account when selecting the features. Rather, the features are selected based on their statistical measures in relation to the dependent variable. The filter method uses various metrics to rank the features according to their significance in predicting the dependent variable then filtering out the irrelevant/nonessential and redundant feature columns. One of the advantages of using the filter method is that they need less computation time since they don't have to train the model. Also, this method is less likely to over fit the data. However, these methods may not be able to extract the best subset of features in some situations. Some of the common filter methods include; Chi-square Test, Information Gain, Fisher's Score, Missing Value Ratio (Sun et al., 2013).

2.8.2.2. Wrapper Method

In wrapper methods, random features are selected at random to form a certain combination of features which is then tested on the training model to evaluate its performance. Based on the deductions drawn from the previous model, some features are added or removed, creating another combination that is also tested on the training set to evaluate its performance. This process is repeated over and over until the combination with the highest performance is obtained, which will be accepted as the final combination of features. Some of the common examples of wrapper methods include; recursive feature elimination, backward feature elimination and forward feature selection, etc. The computation of these methods is usually very expensive.

2.8.2.3. Embedded Method

Embedded methods come as a hybrid of both the filter and wrapper methods. Even though these methods also iterate the features during their operations, they can be computed at a low cost, which means they carry the advantages of both wrapper and filter methods. In simple words, these methods are more like the filter methods in terms of processing speed, only with improved accuracy (Liu, Zhou, & Liu, 2019). Algorithms with feature selection built-in methods are used to implement this type of method. Some of the common applied embedded methods include; RIDGE and LASSO regression.

2.9. Machine Learning

2.9.1. Introduction to Machine Learning

In order to have a good understanding of ML, let's first take a quick look at AI. AI is the imitation of the human brain's operation process, such as problem-solving and learning by machines, particularly computer systems (Karthikeyan, Hie, Jin, &

Adgaonkar, 2022). AI systems are utilised to accomplish complex tasks in a very similar manner to how humans unravel problems. Some of the most common applications of AI include; autonomous vehicles, speech recognition, natural language processing, expert systems, and machine vision. If you ever wonder how YouTube can get your favourite playlist right or how your Instagram feed is always filled with your favourite videos, then that's what AI does. Having been founded just in the1950s as an academic principle, AI has had numerous developments and improvements in this short period of time that make it one of the best innovations of humankind of all time. So, for the computer to be able to mimic the human brain, it needs to be taught. And this is where ML comes into play.

Arthur Samuel defined ML as “the subfield of computer science that gives computers the ability to learn without being explicitly programmed.” ML is a data analytics strategy that uses special programming algorithms to teach computers to do what humans do intrinsically in solving different problems and making decisions. In ML, a built model learns information directly from a set of sample data called training data without depending on predetermined equations like other ordinary models. After the model has been trained by the training data set, it can now make certain decisions or predictions on new data not by being programmed to do so but based on the pattern it learned from the training data set (Singh, Gehlot, Prajapat, & Singh, 2021). The model's performance can improve through experience and the amount of data available for training. Generally, the more the training data set is increased, the more the model's performance is improved. ML algorithms come in handy in solving tasks that cannot or are difficult to be solved by conventional algorithms such as speech recognition, e-mail filtering, medicine and computer vision. In order to make agriculture more effective and precise, ML models can be utilised in all stages of production from preparing the soil, planting, manure and fertilizers addition, irrigation, pre-harvesting, harvesting to post-harvesting operations (Meshram, Patil, Meshram, Hanchate, & Ramkteke, 2021).

2.9.2. Types of Machine Learning

As we have seen before, ML is about training a computer so that it can perform certain activities. Now, depending on the training approach used, we get various types of ML. Each type has its advantages and disadvantages. Before we dive deep into these types of ML, let's first talk about the different kinds of data used as input in ML models. There are two kinds of data used which are labelled and unlabelled data. The input data usually comes as a combination of many features describing the object. For instance, we have two kinds of nuts, white and scorched nuts. Here the features would be something like length, width, colour features, texture features, etc. Now the label is a column that defines the category of each nut. In our case, it will be either white or scorched. That being said, Labelled data contain a column with labels while unlabelled data do not contain a label column. Thus, using either labelled or unlabelled data gives us three types of ML. These are; supervised learning, unsupervised learning and reinforcement learning.

2.9.2.1. Supervised Learning

Supervised learning is among the commonly used ML approach which uses labelled data to train the model. It is called supervised because the labelled data supervise the model in the learning process. If applied in the appropriate situations and with accurately labelled data, this approach gives extremely satisfying outcomes. During the training stage, a portion of the dataset is fed into the model. Then the model learns the relationship between the features and the labels. After understanding the pattern, the model can make predictions or classifications on new data. When working with new data, the model can still learn new patterns if it detects a different pattern in the new data. Hence, continuing to improve its efficiency as it receives new data. Supervised learning can be used in classification and regression applications.

- **Classification problems** use the model to assign test data into particular categories. Classification is very important in food inspection operations. For example, to separate white cashew nuts from scorched cashew nuts.

- **Regression** is the second type of supervised learning approach that tries to study and understand the relationship between the independent and dependent variables. These can be used to predict numerical values such as projected yield for a given crop based on different factors like rain and others. Some common regressions include; linear regression, logistic regression and polynomial regression.

2.9.2.2. Unsupervised Learning

Unsupervised ML comes in handy when we are dealing with unlabelled data. Also, this approach can handle much larger datasets. Because of labelled data, supervised learning models are able to easily understand the pattern between the dependent and independent variables. However, the situation is not the same for unsupervised learning models since they do not use labelled data. Instead, they have to create hidden structures to study the relationship between variables, and because of this, they are more adaptable. This implies that instead of sticking to a defined problem statement, unsupervised learning models are able to adapt to the data by dynamically adjusting the hidden structures. Hence, having more post-deployment improvement than supervised learning models. Unsupervised learning models are used for: clustering, dimensionality reduction and association:

- **Clustering** is a data mining method that groups unlabelled data according to their similarities or differences. For example, K-means clustering algorithms put data points with similar characteristics into the same groups, where K stands for the size of the grouping and granularity.
- **Dimensionality reduction** is a method used when there are too many features in our data set. This algorithm is capable of reducing the number of features to a manageable size without affecting the data integrity.
- **Association** uses a different approach to find the relationship between the variables.

2.9.2.3. Reinforcement Learning

Reinforcement learning uses the trial-and-error approach to learn from new situations, hence imitating directly the way human beings learn from data in their lives. Favourable outcomes are promoted or 'reinforced', while non-favourable outcomes are demoralised or 'punished'. According to the psychological concept of conditioning, reinforcement learning operates through an interpreter and a reward system. In every iteration, the interpreter checks whether the outcome is favourable or not. If the output is found to be correct, the interpreter reinforces it by giving a reward to the system. In case the outcome is not favourable, the system will reiterate until it gets the correct outcome. Most of the time, the reward system is directly linked to the effectiveness of the outcomes.

2.9.3. Traditional and Representational Machine Learning

As we discussed before, ML algorithms learn from some data and later make their own predictions or choices. Based on the form in which the data is presented to the model, ML models can be categorised into two categories. These are Traditional ML and Representational ML.

Traditional ML comprises all fundamental algorithmic structures for solving particular problems. In these kinds of models, the choice of the algorithm to be used and the input data (features) that should be fed into the model are subject to an expert. The input data should be in a structured format like numbers for the traditional ML models. Let's say you have wanted to classify images. Firstly, you have to identify the important features for your classification problem. Then you have to extract these features from images in the form of numbers and feed them into your model. These models can be applied for regression, classification, clustering and dimensionality reduction problems. Examples of the traditional ML models include; Logistic Regression, Linear Regression, K-means, Naive Bayes, etc.

On the other hand, Representational ML models can identify the significant features to pay attention to from the data given without relying on the expert's intervention. Also, Representational ML models can deal with unstructured data as inputs, like videos or photos. Literally, this means that if we have to classify images, we will feed them into the model directly without needing to extract the features first. These models can be applied for regression, classification, clustering and dimensionality reduction problems. A good example of Representational ML models is Deep Learning (DP).

Table 2.7. Comparison between Traditional and Representational Machine Learning Models

Traditional Machine Learning	**Representational Machine Learning (Deep Learning)**
Features should be chosen by an expert and fed into the model	The model can identify important features by itself
A lot of classifiers available	Limited number of classifiers
Model can be trained with small amount of data	Model requires larger amount of data for training
Training process is fast	Training process can take a long time

2.10. Application of Machine Learning in Cashew Processing Industry

ML models have had a wide range of applications in the cashew processing industry. Most of the applications are in the sorting and grading stage, which is a very critical point because there are various grades of cashew. In order to comply with the international standards, the kernels should be precisely categorised into their respective grades. Also, another application would be sorting out foreign materials. Let's have a look at some of the related studies done in this area.

A study was conducted by Arora & Devi (2019) aiming at creating a supervised learning model to precisely identify and classify cashew kernels into various grades. Cashew images were captured, and various pre-processing operations were run to improve the quality of the images. Then the cashew kernel was segmented from the background by using K-means clustering. Morphological features, texture features and colour features were extracted from the image. Two supervised learning models (i.e., Support Vector Machine (SVM) and Random Forest) were trained and compared. At first, the SVM operated with a One-vs-All algorithm. An efficiency of 85% was recorded in this approach due to some difficulty in detecting the WW-180 grade. The model mistook them for WW-450 or WW-320. When a One-vs-One algorithm was used, the efficiency increased to 90.6%. Maximum efficiency of 94.28% in this study was recorded in the Random Forest.

The study recommended the Random Forest Model for this classification problem due to its high accuracy and simplicity in training.

In the study performed by Aran, G Nath, & Shyna (2016), five classifiers were compared in classifying cashew kernels into various grades. The studied classifiers were Random Forest, Multi-Class Classifier, Classification Via Regression, Multi-Layer Perception, and Back Propagation Neural Network (BPNN). External features of the cashew kernel like; texture, colour, shape and size were used in this study. Moreover, this work also studied the effect of different pre-processing operations in grading the cashew kernels. And the results revealed that operations such as Lucy filter and contrast enhancement had a significant improvement in the classification accuracy. Among the studied classifiers in this study, the highest recorded accuracy, which was 96.8%, was obtained in the BPNN model.

The classification of whole and split cashew kernels can be a bit tricky at times. A novel machine vision approach was used for the classification of whole and split cashews in a study done by (Sunoj, Igathinathane, & Jenicka, 2018). In this study, the shadows cast by the cashew kernels were used to identify the split-up and whole kernels, while the idea of grayscale-intensity-profile was used for split-up kernels. Associated image processing operations and the classification algorithms were run using the ImageJ plugin. A 100% accuracy was acquired in this study.

Nagpure & Joshi (2016) used the K-Nearest Neighbour algorithm to present a novel defect of cashew nuts based on colour and texture features. In this study, the Support Vector Machine algorithm (SVM) was used for background removal and colour classification in MATLAB. After the image segmentation process, the healthy and infected portion of the cashew nuts were calculated and the classified as good or defect accordingly.

A study was carried out by Ahmadabadi et al., (2017) to sort out peeled pistachio kernels from shells. The intelligent system developed in this study based on combined machine vision (MV) and Support Vector Machine (SVM). Initially, images were converted from RGB colour space to HSV colour space for segmentation purposes. The H-component in the HSV colour space and Otsu thresholding method were applied. A feature vector containing 30 colour features was extracted from the captured images and based on the sensitivity analysis carried out, only the superior features were selected and fed into the SVM classifiers for training. Various SVM models having a different kernel function were developed and tested. The SVM model having cubic polynomial kernel function and 38 support vectors achieved the best accuracy (99.17%) and then was selected to use in online decision-making unit of the system.

Another study was done by Babu et al. (2012) to classify cashew kernels. In this study, featured were extracted from images of cashew of known and unknown grades and passed into a neural network, which learned the pattern of the data and created an intelligent model to classify new samples. The accuracy of the developed model was tested and found to be 86%. Image processing and data mining were the basis for this model. This model was faster and more cost-effective. Hence it was recommended for fully automated high-speed cashew processing plants. Moreover, such a model would not only be applicable in cashew grading but can also be used in other quality control operations in cashew-based processing facilities.

Convolution Neural Networks (CNN) with deep layers have had incredible achievement in various image classification applications. One of the additional advantages of deep CNN is its ability to identify and extract important features from the image by itself without relying on an expert. Sivaranjani, Senthilrani, Ashokumar, & Murugan (2018) performed a research study to develop a computer vision-based cashew grading system using deep CNN. Various parameters were considered in this study to optimize the CNN model. The study suggested that, in order to improve the classification accuracy, the CNN can be combined with the Recurrent Neural Network (RNN) and trained together using reinforcement learning. By doing so, computer vision systems could be implemented in other products as well by using the developed framework in this study.

Vidyarthi, Singh, Tiwari, & Rai (2020) also did a study to classify cashew kernels using deep CNN and incorporating image processing. Cashews were classified into five categories including; whole cashew, butts and pieces. Four deep CNN models, namely ResNet50, VGG-16, Inception-V3 and a custom model, were studied. The performance of the models was analysed and compared by utilising various model evaluators, like specificity, sensitivity, accuracy, precision and F1-score. Generally, all models gave promising results with the minimum performance accuracy was 95.1% and was obtained from the Custom-built model. The highest accuracy of 98.4% was obtained from the ResNet50 and Inception-V3 neural networks. According to the results obtained in this study, it showed that the developed deep CNN models were capable of achieving accurate, fast and automatic cashew classification.

Ganganagowdar & Siddaramappa (2016) performed a study to utilize Artificial Neural Networks (ANN) in developing a novel intelligent automated model that recognizes and classifies white whole grade cashew kernels. The developed model worked in two phases. The first phase was intended to extract important features. 24 colour features and 16 morphological features were extracted from the input images. In the second phase, a Multilayer Perceptron ANN was used to recognize and classify the cashew kernels into their respective white-wholes grades. A back-propagation learning approach was used in this model. Classification accuracy of 88.93% was achieved with this model.

Another study was done by Ganganagowdar & Siddaramappa (2011), which aimed to classify cashew kernels using colour features. Colour features were extracted from images in the RGB (red-green-blue) colour space and used to train a Feed-Forward Neural Network. The developed model in this study was able to classify cashew kernels into six different grades with an accuracy of 80%. The model was cost-effective, hence can be useful in places with low-income farmers.

2.11. Summary

Cashew is of significant economic importance because of its numerous benefits. In this chapter, a brief review of the history of cashew and its cultural and post-harvest practices was given. During the post-harvest handling of cashew nuts, drying and roasting cashew kernels are among the main operations. There are various grades of cashew kernel available, as we have seen in this chapter. These various grades are determined by both colour (i.e., white, scorched and desserts) and shape and size of the kernel (i.e., whole, splits, butts and pieces). The drying and roasting processes may affect the quality of the nuts in terms of colour. Normally, cashew nuts are expected to be white in colour. But over roasting or over-drying the nuts turns them to slight brown or reddish colour.

Currently, the sorting out of scorched nuts is still performed manually in many places, particularly in Tanzania. Manual classification is prone to human errors hence making it less effective. Moreover, this process is time-consuming and lacks consistency. The concepts of ML and machine vision were explained in chapter two. In general, machine vision has acquired tremendous awareness within the past few decades because of its high potential for in-line processing. In machine vision, the number one key to achieving a successful classification is the ability to extract appropriate features from the image. This technology has been applied in various areas, and still, more and more studies are being performed to improve it even better.

Machine vision technology has been utilised for many classification problems in various areas because of its sophistication and improved accuracy. Agriculture is one of the areas where this technology has been widely applied. Some of the applications of machine vision technology in the classification of cashew kernel were also reviewed in this chapter. The review of literature on the classification of cashew kernel using machine vision systems revealed that in order to have the best performing classification system, much attention should be put on the lighting system, the position of the object under the camera, estimating and extracting the required features for classification. The following chapter presents the materials, experimental procedures, equipment, and analysis methods followed in this research to classify cashew kernels using a machine vision system.

CHAPTER THREE

3. MATERIALS AND METHODOLOGY

3.1. Experiment Samples

One thousand samples of raw cashew kernels were obtained from the Tanzania Agricultural Research Institution centre in Naliendele (TARI Naliendele), located at 10° 22′ 20“S, 40° 10′ 34“E along the coastal belt of the Indian Ocean. The obtained samples included the scorched and white whole grades of cashew kernel. The cashew kernels were inspected and stored at room temperature.

3.2. Image Acquisition System

The image acquisition system used in this study was comprised of a camera connected to a computer equipped with an image acquisition board, see (Figure 3.1). A proper lighting setup was tested and used in this study. The system is clarified in more detail in the sections that follow.

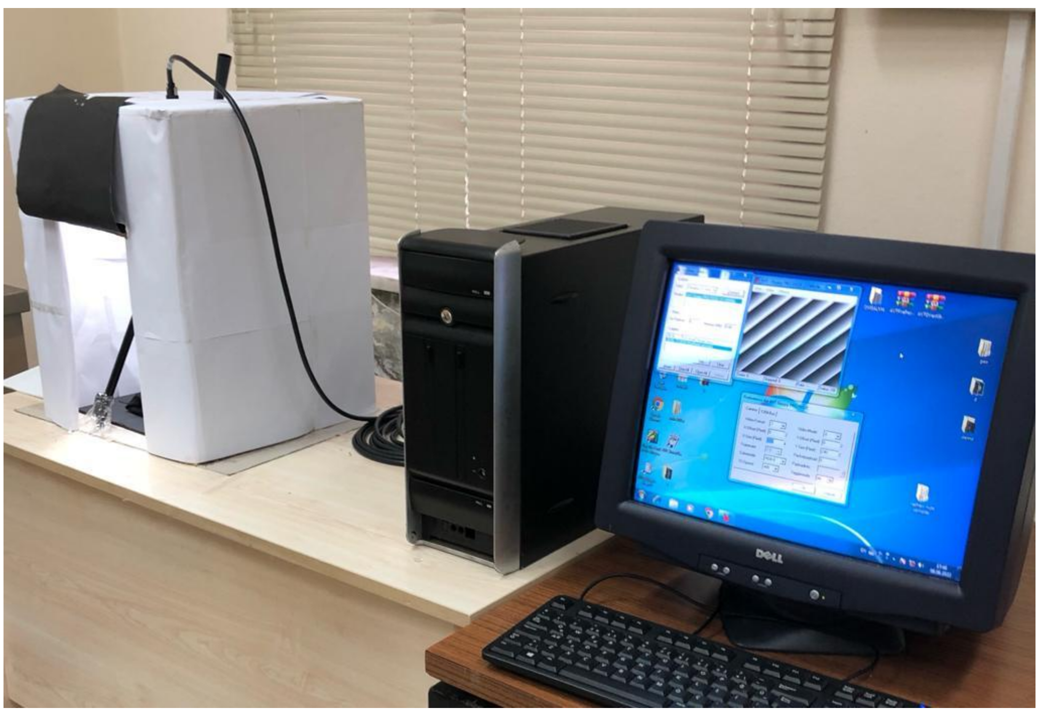

Figure 3.1. Image acquisition system

3.2.1. Image Capturing Camera

The image acquisition system was equipped with a high-resolution Guppy camera model Pro F-032, which had a CCD progressive Sony ICX424 sensor, 4.9 mm x 3.7 mm, consisting of 656 horizontal by 492 vertical cells. The camera had an Analogue to Digital Converter capacity of 12 bits. Electric charges were produced by the cells proportional to the amount of incident light striking on them. And then, the generated analogue electric signals were transmitted to the image acquisition system installed on the computer. The camera was held stationary at a fixed distance of 30 cm above the base plate by a tripod stand. The focal length of the camera was adjusted to improve the quality of the image by playing with the zoom feature of the camera.

Figure 3.2. Image capturing camera

3.2.2. Lighting

Proper lighting is one of the crucial elements in capturing good quality images. A proper lighting system basically should fulfil the following requirements; i.) It should provide sufficient light to make the object clearly detected by the camera. ii.) It should be installed in such a way that; it gives out uniform light in order to avoid or reduce the effect of shades. In this experiment, two 8.5-watts fluorescent tube lights were placed on opposite sides and were able to cast light on the object with a very minimum to totally no shadow problem. This lighting system was able to produce heavily contrasted images with contour well-segmented from the light background.

3.2.3. Computer System and Image Acquisition Board

The computer had an 80 MB hard disk drive, 8 MB RAM, and a 40 MHz dock speed. The AVT SmartView programme and the necessary image processing software were installed on the computer. AVT SmartView is a Fire Package camera viewer optimised with AVT IEEE1394 cameras to gain access and manipulate various features of the Allied Vision Technology cameras. Through this SmartView, various settings were calibrated. The target grey level was set to 125. The blue colour balance (UB) was set to 394, and the red colour balance (VR) was set to 450.

3.2.4. Image Capturing

In order to make the segmentation process easy, a perfect background should be used that can be easily separated from the object of interest. In this case, a black material seemed suitable to work as a background, and various black materials were tested. Out of all the tested materials, the best results were observed from the black cloth. So, this was used as the background for our experiment. The images of the cashew kernels were captured by using the "controlled position approach". In this approach, the camera was adjusted to focus on a specific point on the base. Then, the cashew kernel was placed on this spot, and its image was captured. The aim of using this approach was to only focus on the object of interest. Hence, the captured images were 120 x 120 in size. The images were then stored in the computer as Portable Network Graphics file format (PNG). Out of the 1000 cashew kernel samples that were collected from the Tanzania Agricultural Research Institution centre in Naliendele (TARI Naliendele), 812 kernels were selected, and their images were captured for this experiment. 402 of the captured images were scorched, and 410 were white. 67% of these samples were used for training the model, the rest 33% were spared for testing the model.

3.3. Image Segmentation

After the images were captured, segmentation was done to extract the object of interest from the background. In this experiment, the simple thresholding approach was applied to segment the images. First of all, various thresholding values were tested to find out the one that effectively separated the foreground from the background. Let's assume this value is "T". After selecting the most appropriate threshold T value, it was fed into the segmentation algorithm. Because a black background was used in this experiment, all the pixels forming the background were made black in order to resemble it. Hence, any pixel intensity p below the T value was set to 0, and any pixel intensity p above the T value was set to 255. Equation 3.1. below summarizes the applied operations. This led to the formation of a mask, which was later applied to the original image to blacken the background completely, remaining only with the object of interest.

$$dst(x,y) = \begin{Bmatrix} maxval & if\ src(x,y) > thresh \\ 0 & otherwise \end{Bmatrix} \qquad \textbf{(3.1)}$$

3.4. Feature Extraction and Feature Selection

3.4.1. Feature Extraction

The aim of this experiment was to use the colour feature in the classification of cashew kernels into white and scorched grades. For this reason, only the colour features were extracted from the images. The features were extracted from images in two different colour models i.e., RGB colour space and HSV (Hue, Saturation, Value) colour space. There are various colour features that can be extracted from an image. But in this experiment, some colour moments and the excess red, excess green and excess blue features were extracted. An image is composed of various colours distributed in different combinations to form an image. Just like how the central moments uniquely define a probability distribution, so do the colour moments, which are measures that describe how the colours in an image are distributed. Also, colour moments encode information about the colour and shape of an image. This means that from colour moments we can find the difference or similarities between images. Hence, making them the best option for our classification problem. In this experiment, only the low-order moments were used because they are the ones that contain most of the colour distribution information. Moreover, higher-order colour moments were not preferred because more data was required in order to accurately estimate their values. The colour moments extracted from the images in this experiment include; Mean, Standard deviation and Skewness. These colour moments were computed individually for each channel for both colour spaces used.

3.4.1.1. Mean

Mean is a fundamental concept in statistics. Mean is a central value of a finite set of numbers. It is computed by taking the sum of the values and dividing it by the number of values. It is one of the measures of the central tendency of a probability distribution in statistics. Unlike other central tendency measures like median and mode, the mean is computed by using values in the dataset. Hence, there is no information loss. Another advantage of the mean is its simplicity in understanding and computation. The mean for the colour channels in this experiment was computed using Equation 3.2.

$$Mean(\mu) = \sum_{j=1}^{N} \frac{1}{N} P_{ij} \quad \textbf{(3.2)}$$

3.4.1.2. Standard Deviation

Statistically, the standard deviation is the measure of variation or dispersion of a set of values in relation to the mean. Low standard deviation implies that the data values tend to be close to the mean, which is also referred to as the expected value of the set, whereas a high standard deviation implies that the values are scattered out over a wider range.

$$Standard\ deviations\ (\sigma) = \sqrt{\left(\frac{1}{N}\sum_{j=1}^{N}(P_{ij} - E_i)^2\right)} \quad \textbf{(3.3)}$$

3.4.1.3. Skewness

The measure of the symmetry of distribution in statistics is called Skewness, where the mode is the highest point of distribution. When the tail on one side of the distribution is longer than the other, then the distribution is said to be skewed, meaning it is asymmetrical. This usually occurs when there are a lot of extreme values, which distort the normal distribution. A positive skew implies that the tail is longer on the right side than the left side and a negative skew implies that the tail is longer on the left side than the right side. The skewness of the colour channels was calculated using Equation 3.4.

$$skewness\ (\gamma) = \sqrt[3]{\left(\frac{1}{N}\sum_{j=1}^{N}(P_{ij} - E_i)^3\right)} \quad \textbf{(3.4)}$$

3.4.2. Feature Selection

After extracting the features from the images, sometimes some features may be irrelevant. And using irrelevant features would affect the performance of the algorithms. So, in order to improve the efficiency of a classification model and decrease the running time, it is highly recommended to use only the important features and get rid of irrelevant features. This process is called feature selection. In this experiment, feature selection was performed in Python using the Boruta library, which uses the wrapper approach to select the relevant features by building a random forest classifier.

3.5. Manipulation of Data

3.5.1. Checking for Outliers

In a random sample of a population, the points that lie at an irregular distance from other values are what we refer to as outliers. In a sense, this definition of outliers leaves it up to a consensus process (or the analyst) to determine what will be considered irregular. This means that there are no stringent statistical rules for definitively pinpointing outliers. Outliers are challenging for many statistical analyses because they can distort the final results or make the tests miss significant findings.

In this experiment, the outliers were identified by using outlier fences. The outlier fences were created from the interquartile range (IQR). The IQR is the range of values between the third quartile and the first quartile (Q3 – Q1). The created outlier fences include; upper outer, upper inner, lower inner and lower outer. The values inside the two inner fences were kept, while the ones outside were removed as outliers.

3.5.2. Normalization of Features

The values of different types of features extracted from images differ considerably. The features ranged from a very small to a very large positive number. Some classifiers accomplish better results with the inputs values lying between 0.0 and 1.0. For example, the ones using the sigmoid function with a gain equal to 1.0. For this reason, the extracted features were normalized so that they could be properly used as input to different classifiers. The following procedure was followed for normalizing the feature values. When a feature, e.g., mean, of consecutive N cashew kernels from a class were arranged in sequential order, they formed a (N x 1) vector, V. Each element vi of V was normalized using:

$$V_i^{\ n} = \frac{V_i - V_s}{V_l - V_s}\ ,\ i = 1{,}2.\ldots.N \qquad \textbf{(3.5)}$$

Where V: is the normalized value of the element i, and V_l and V_s represent the largest and the smallest components of the feature vector, respectively. The equation above transformed the original vectors to a new set of vectors whose elements have a value between 0 and 1. The normalization process was performed for each of the individual features.

3.6. Classification Models

3.6.1. Logistic Regression

Despite bearing the word "regression", Logistic Regression is one of the classification models in ML. Logistic Regression models are models that contain a certain fixed number of parameters depending on various input features. And the output is a categorical prediction. For instance, in our case, is the cashew kernel is white or scorched.

Generally speaking, the theory behind both Linear Regression and Logistic Regression is the same. The only difference is that instead of fitting the data to a straight line like the way it is done in Linear Regression, the Logistic Regression approach fits the data into an S-shaped curve called Sigmoid. Equation 3.6 shows how the sigmoid is calculated.

The training data were fitted into a Logistic Regression algorithm to train the model. Then, the trained model was used to make predictions on new data.

$$sigmoid(x) = \frac{1}{1+e^{-x}} \tag{3.6}$$

3.6.2. Decision Tree

The second algorithm to be tested was the Decision Tree algorithm. This model uses a predictive model (a decision tree) to go from observations about a particular problem, which are usually represented in the branches, to conclusions about the problem's target value, which is usually represented in the leaves. The model's accuracy is heavily affected by the decision to make strategic splits in the tree branches. Multiple algorithms can be used to decide to split the node into sub-nodes. Figure 3.3. represents the split of a decision tree from the root node to the terminal nodes.

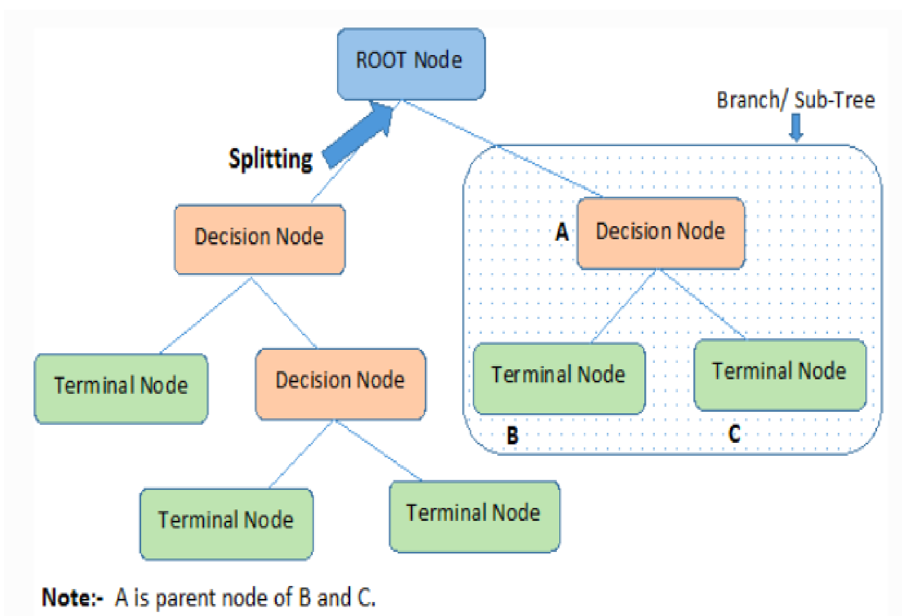

Figure 3.3. Flow diagram of a Decision Tree

Depending on the intended purpose, whether classification or regression, the criteria for decision may vary. Various criteria may include; Entropy, Information gain, Gini index, Gain Ratio, Reduction in Variance, Chi-Square. In this experiment, the Gini index was used as the classification criteria. The Gini index was calculated using the equation 3.7.

$$Gini = 1 - \sum_{i=1}^{c}(Pi)^2 \quad \textbf{(3.7)}$$

3.6.3. Random Forest

Random Forest is a classification algorithm comprised of several decision trees. This approach comes as an improvement for the Decision Tree. Random Forest makes more accurate decisions by utilizing the forest's predictive power. It uses randomness to construct individual trees to promote uncorrelated forests. The Random Forest applied in this experiment used the Gini index as a decision criterion for its trees. Two different approaches were employed in this experiment. First, a forest of a collection of 10 trees was trained and tested to find the classification performance. Later, the number of trees was increased to 100 trees.

3.6.4. Support Vector Machine

Support Vector Machine algorithm was also tested. This algorithm is based on the concept of finding a hyperplane that best splits a dataset into two classes, as demonstrated in the image below (Figure 3.4).

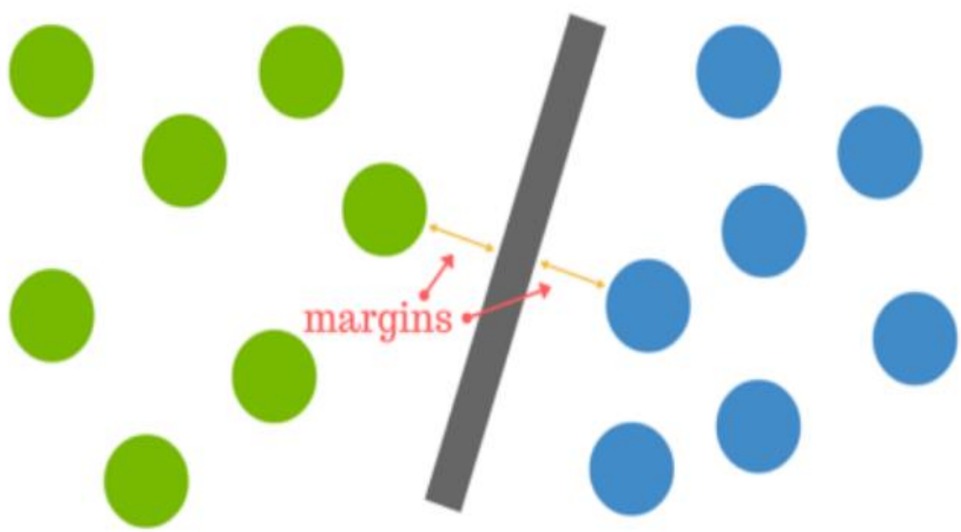

Figure 3.4. Illustration diagram of data split in a Support Vector Machine

We can imagine a hyperplane as a line that linearly splits and classifies a set of data. After the dataset is split, the distance between the nearest data point from either set and the hyperplane is known as the margin. In this approach, the goal was to determine a hyperplane with the maximum possible margin between the hyperplane and any point within the training set, providing a more significant chance of new data being correctly classified.

3.6.5. K-Nearest Neighbour

The last model to be tested in this experiment was the KNN algorithm. Here various values of K were tested to find the best one. First, the value of K was set to 7, and the model was trained and tested for classification accuracy. Later, the value of K was lowered to 5 and finally 3.

CHAPTER FOUR

4. RESULTS AND DISCUSSION

4.1. Introduction

The detailed results of the performed methods in the experiment from image acquisition, feature extraction, feature selection and classification are explained in this chapter. Images of the captured cashew kernels are presented, and the list of all features extracted from these images is given. Later, the selected features that will be used in the classification algorithm are outlined. Various algorithms are discussed in this chapter, and at the end, a general discussion about the features selected and the classification algorithms is presented.

4.2. Captured Images

An example of the captured white and scorched images is shown in (Figure 4.1) below.

Figure 4.1. Sample of the captured cashew kernels scorched nut on the left, white nut on the right

4.3. Image Segmentation

After testing different threshold values, the threshold value of 60 was found to give the best performance in segmenting the foreground from the background. An example of raw and segmented images is shown in Figure 4.2. below.

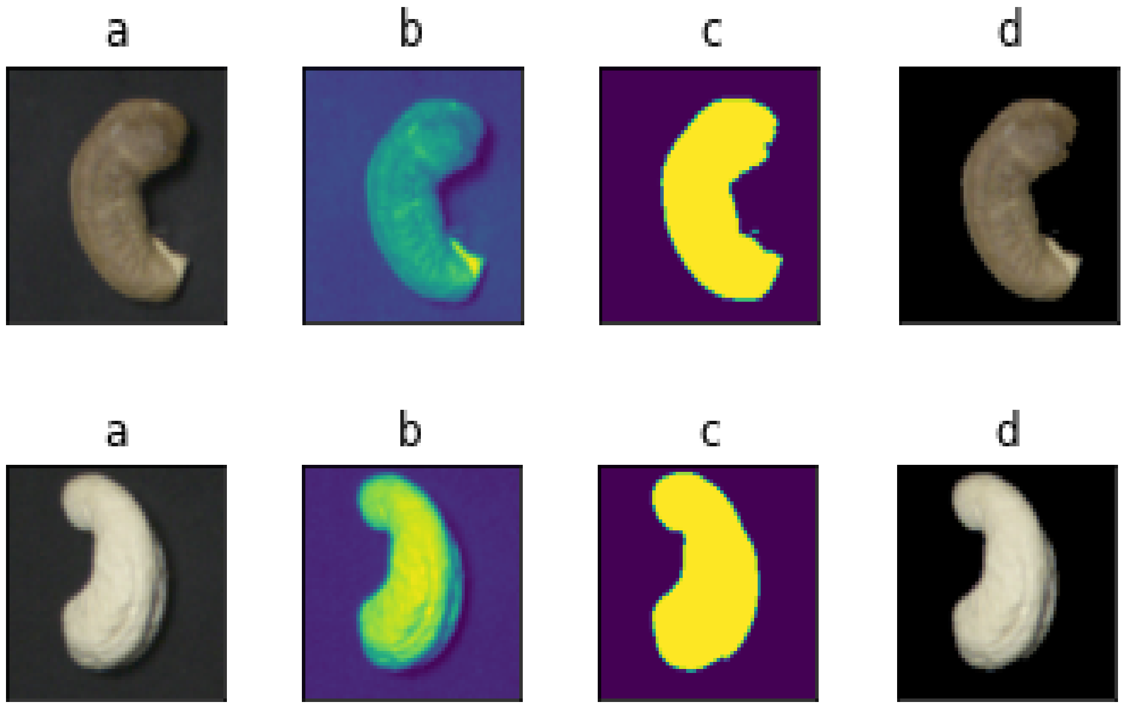

Figure 4.2. Raw image segmentation

4.4. Feature Selection and Extraction

This experiment used the colour features of the cashew kernels to classify them into white or scorched categories. 21 colour features were extracted and after evaluating the effectiveness of all features in predicting the category of the cashew kernels, only 14 features were found to have a significant effect and the other irrelevant 7 features were left out. Table 4.1 below shows the list of features used and those left out.

Table 4.1. List of accepted and rejected features

No.	Feature	Status	No.	Feature	Status
1.	Red Mean	Rejected	12.	Excess Blue	Rejected
2.	Green Mean	Rejected	13.	Hue Mean	Accepted
3.	Blue Mean	Accepted	14.	Saturated Mean	Rejected
4.	Red Std	Accepted	15.	Value Mean	Rejected
5.	Green Std	Accepted	16.	Hue Std	Accepted
6.	Blue Std	Accepted	17.	Saturated Std	Accepted
7.	Red Skewness	Accepted	18.	Value Std	Accepted
8.	Green Skewness	Accepted	19.	Hue Skewness	Accepted
9.	Blue Skewness	Accepted	20.	Saturated Skewness	Accepted
10.	Excess Red	Rejected	21.	Value Skewness	Accepted
11.	Excess Green	Rejected			

4.5. Manipulation of Data

The dataset of the selected features was checked for outliers. All the features were found to have clean data except the "Saturated Standard Deviation" and the "Saturated Skewness" features. The figure below shows the data before the outliers were removed (Figure 4.3). By using the method explained in chapter three above, the outliers were removed from the dataset.

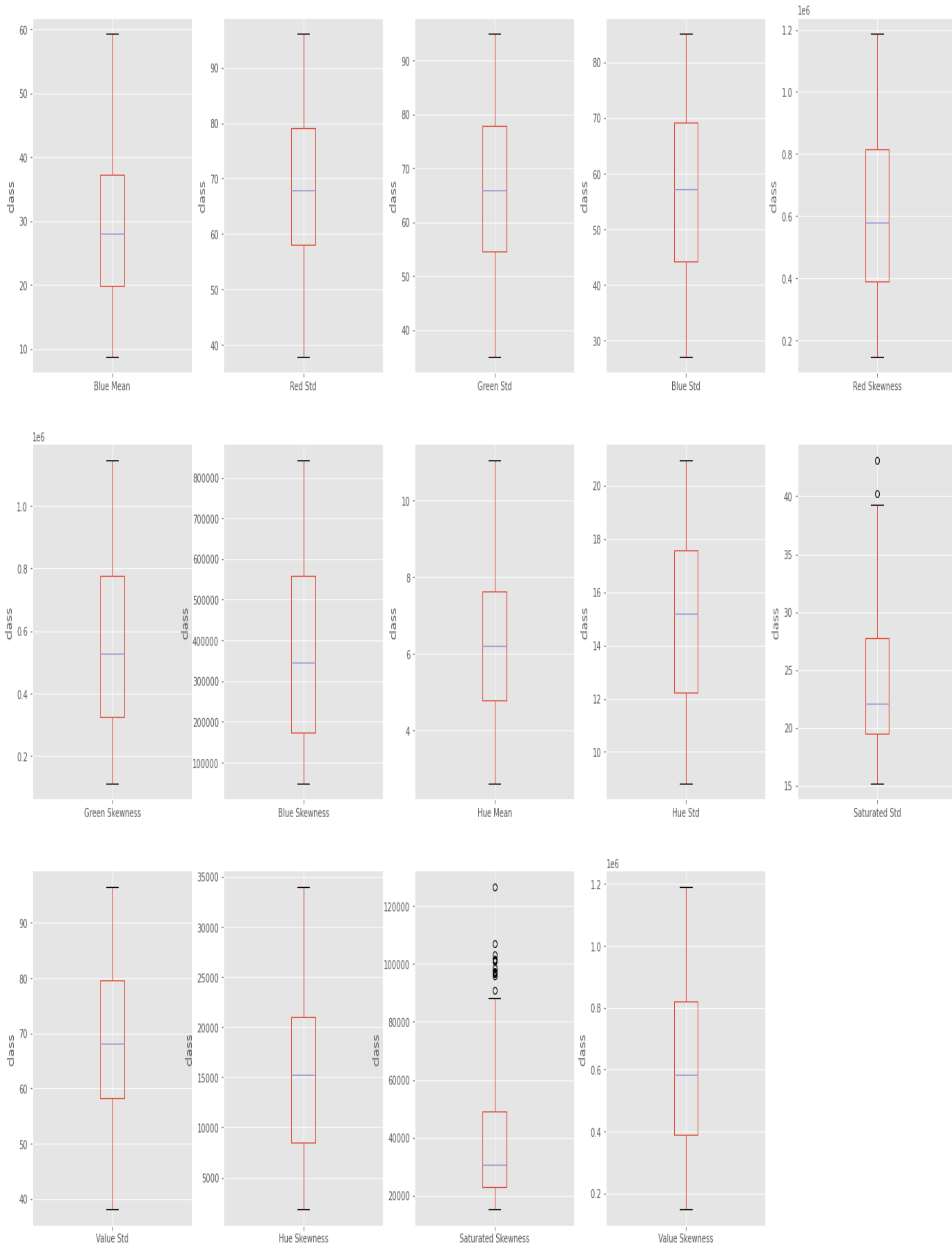

Figure 4.3. Dataset before removing the outliers

4.6. Classifiers

Five traditional ML algorithms were tested in this experiment. The first algorithm was the Logistic Regression. Later a Decision Tree model was tested followed by Random Forest, SVM, and finally, the KNN algorithm. The algorithms were compiled and trained properly. Then, a set of validation data was used to check the performance of the algorithms and determine the classification accuracy.

Evaluating the efficiency of the model by using the classification accuracy alone can sometimes be misleading if there is an unequal number of observations in the classes. That is why it is worth using the confusion matrix as well. In simple words, the confusion matrix can be defined as a summary of prediction results of a particular classification problem. This means that calculating a confusion matrix provides a better idea of what the classification model is getting right and the types of errors that are being made. The confusion matrix for each model is presented in this section. More details about each algorithm are given in the sections that follow below.

4.6.1. Logistic Regression

The first algorithm to be tested was the Logistic Regression. The efficiency of this model was found to be 99.4%. The confusion matrix is shown in Figure 4.4.

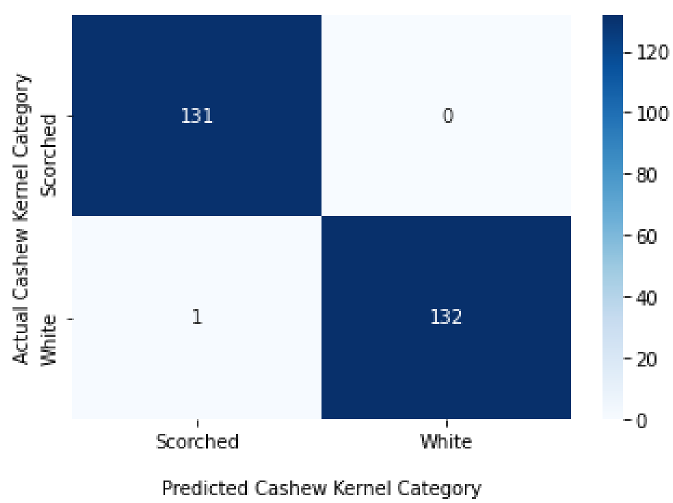

Figure 4.4. Logistic Regression confusion matrix

4.6.2. Decision Tree

After testing data on the logistic regression, they were tested on the decision tree. The tree used in this model started to create nodes in the data from the "Blue Skewness" feature and went all the way down to the leaf nodes. The Gini value was used as a criterion for decisions at the nodes. The tree started with the Gini value of 0.5. The accuracy of the decision tree algorithm was 98.4%, the confusion matrix for the decision tree algorithm is shown in Figure 4.5. and Figure 4.6. below represents the decision tree used in this model.

The decision tree recorded the lowest performance in this experiment. Despite seeming to be simple and easy to compute, the decision tree classifier has got some challenges. One of which is the identification of which attributes should be considered as rood nodes at each level. And, it is because of this issue that it may have a low classification performance. This can also be seen in a study that was performed by Thakkar, Bhatt, & Bhensdadia (2011) to evaluate the performance of different classification techniques in classifying cashew nuts. In this study, the physical properties of the cashew kernels were extracted and used as attributes for the classification model. Various classification algorithms were tested, including; ML-Perceptron KNN, Naive Bayes SVM and Decision Tree. The Decision Tree algorithm had a classification accuracy of 79%, which was among the lowest performance.

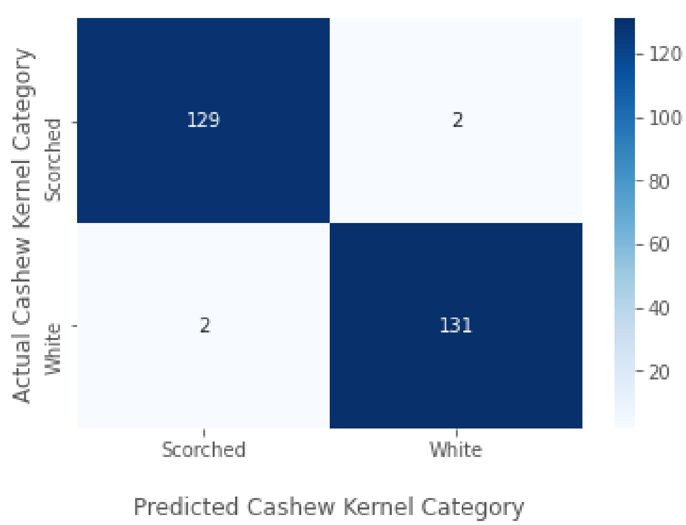

Figure 4.5. Decision Tree confusion matrix

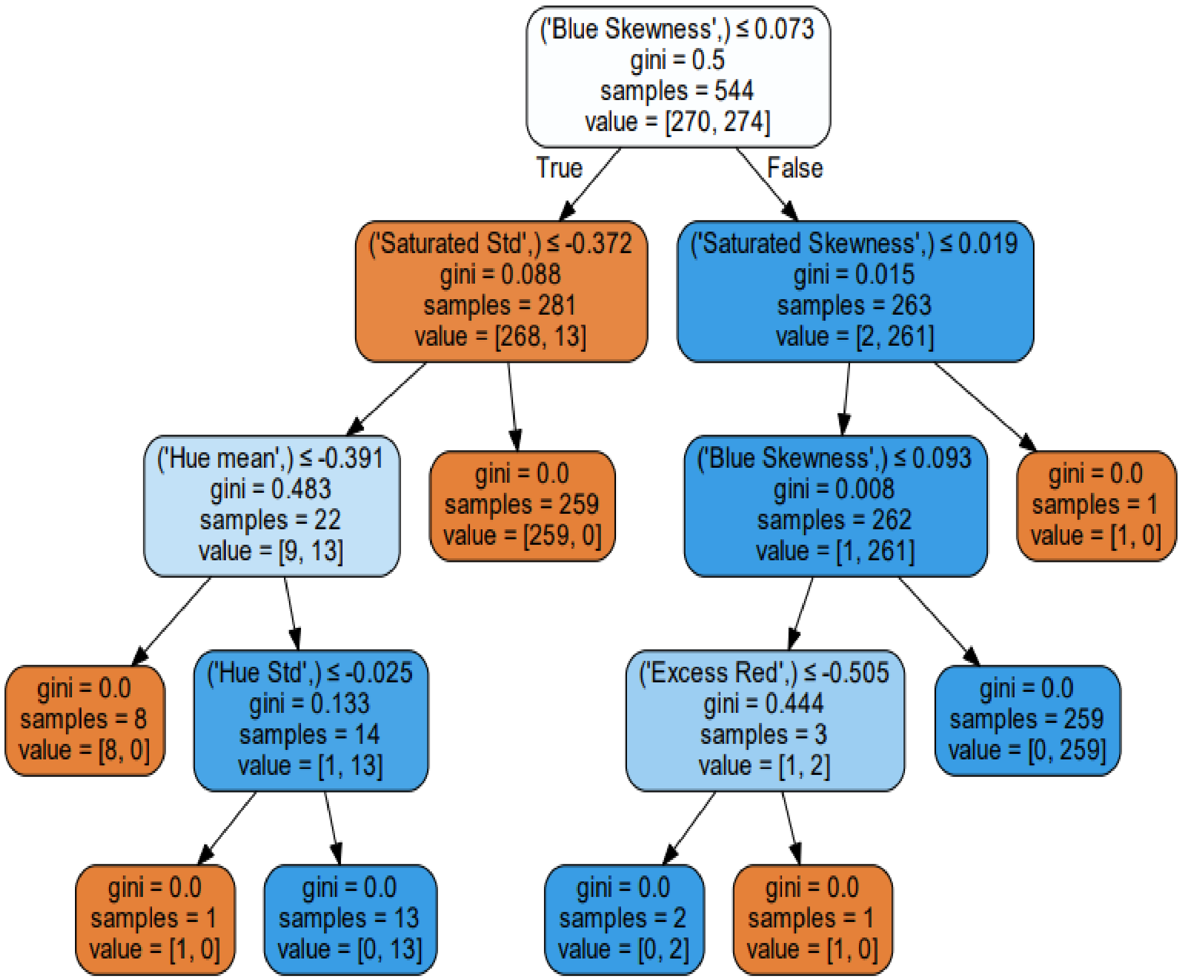

Figure 4.6. Illustration of the Decision Tree model

4.6.3. Random Forest

A Random Forest is a collection of trees. In this experiment, two kinds of forests were used. Firstly, a random forest made of 10 trees was used. The accuracy of this algorithm was 99%. Upon the increase in the number of trees, the classification accuracy increased as well. Thus, when the number of trees was increased up to 100, the accuracy was 99.8%. This was the maximum accuracy recorded from all the tested algorithms. Figure 4.8. below represents one of the trees from the forest with the highest classification accuracy.

Random Forest model has proved to be more efficient in cashew classification in various studies as well. Another study was performed by Arora & Devi (2019) to classify the cashew kernels into various grades by using a combination of different features i.e., morphological, texture and colour features. In this study also, the Random Forest model out performed other models with the highest classification efficiency of 94.28%.

Apart from studies done in cashew nuts classification, the high performance of the Random Forest algorithm can be seen in other types of nuts as well. For instance, a study performed by Koc et al. (2020) to classify various hazelnut cultivars. In this study, three algorithms were tested and compared to each other. The tested algorithms are; Feed-forward DL, Gradient Boosting and Random Forest. The results revealed that the Feed-Forward DL and the Gradient Boosting algorithms had the same performance accuracy of 94%. And the highest classification accuracy of 100% was obtained from the Random Forest algorithm. From these examples, it is clear that the Random Forest algorithm is an ideal choice for the classification problems of nuts, particularly cashew nuts in our case.

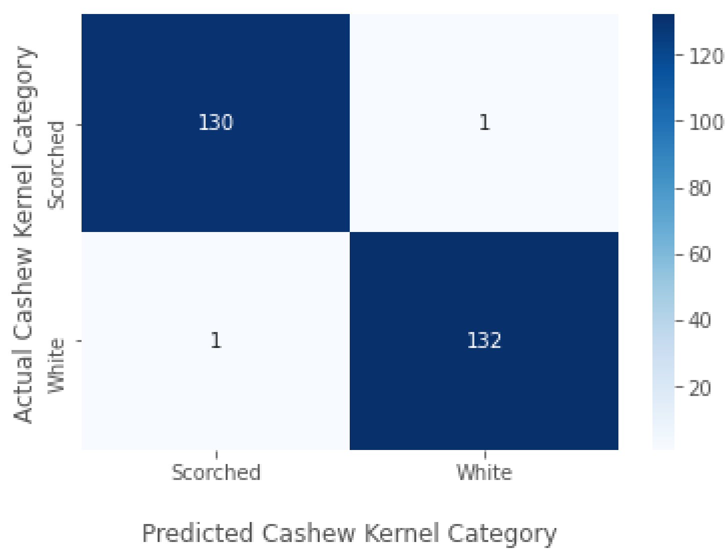

Figure 4.7. Random Forest confusion matrix

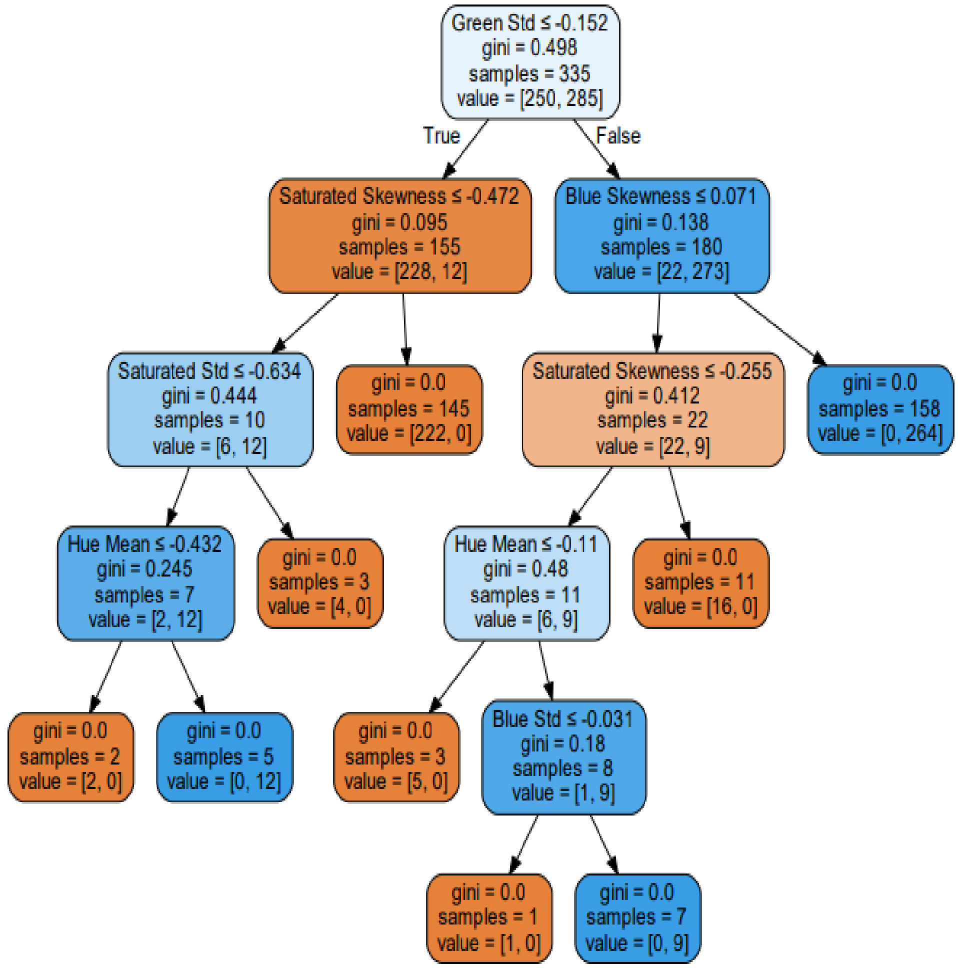

Figure 4.8. Illustration of the Random Forest model with 100 trees

4.6.4. Support Vector Machine

Another algorithm to be tested was the SVM. The classification accuracy of this algorithm was 99.6%. As it can be seen from the confusion matrix in (Figure 4.9) below, only one cashew kernel was mistakenly classified.

SVM was also used in a study performed by Arora & Devi (2019) to classify cashew kernels into various grades. In this study, a combination of morphological, texture and colour features were used. Two different kinds of SVM algorithms were tested in this study. The first one was the One-vs-All algorithm, which had a classification accuracy of 85%. And the second one was the One-vs-One, which had a more improved classification accuracy of 90.6%.

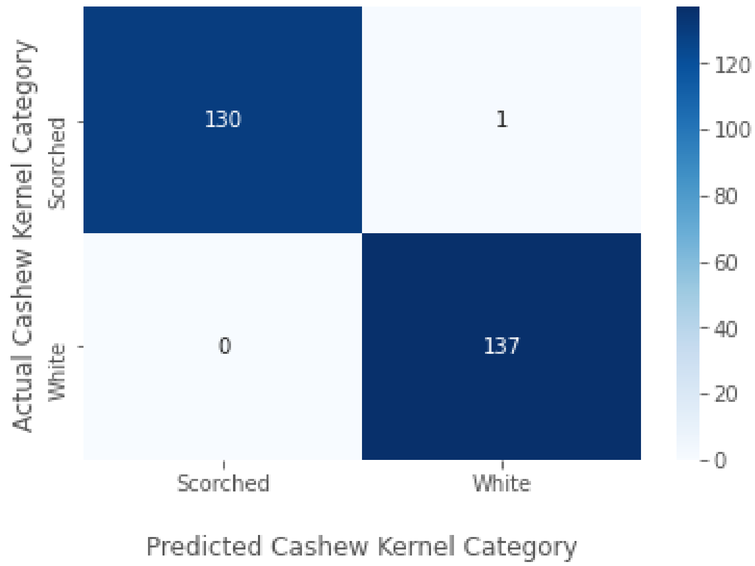

Figure 4.9. Support Vector Machine confusion matrix

4.6.5. K-Nearest Neighbour

Last but not least, the KNN algorithm was tested. Three different values of K were tested. At first, the value of K was set to 7, and the classification accuracy of 98.8% was obtained under these conditions. Upon decreasing the value of K to 5 the accuracy increased to 99.5%. The maximum accuracy was obtained after lowering the value of K down to 3, where the classification accuracy was 99.7%.

KNN is regarded as one of the oldest nonparametric classification algorithms that are used in classification problems with promising results. For example, let's consider another study that was performed by Özkan, Köklü, & Saraçoğlu (2021). In this study, the KNN algorithm was used to create a computer vision system to distinguish two different varieties of pistachio nuts. A high-resolution camera was used to capture 2148 images of the two varieties of pistachio nuts. 16 features were extracted from these images and used as attributes for the classification model. The system achieved a classification accuracy of 94.18%.

CHAPTER FIVE

5. CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

Cashew kernel classification is still a very big problem in many places where manual labour is used extensively. This means that still there is a need to apply computer vision-based classification techniques to help improve the efficiency of the classification process as well as lower the overall operation costs. This study aimed to evaluate the effectiveness of colour features in the classification of cashew kernels into white and scorched categories. From the results obtained, the colour features have proved to be effective in the classification of the cashew kernels into these two categories. All the tested traditional machine learning techniques have given promising results in this classification. Random Forests are usually easy to construct, and with the accuracy obtained in this study, it would be more practical to prefer Random Forest for this classification.

5.2. Recommendations

I. The development and use of multi-camera systems are important for research on agricultural products. This is due to the fact that sometimes it may be impossible to control the product's position under the camera. Hence the use of multi-camera systems would be of great help.

II. In this experiment, the images of the cashew kernels were captured in a stationary position. A prototype for sorting the cashew kernels should be developed which captures and analyses the images of the nuts in motion (e.g., while sliding down a channel or falling in an air column).

III. For developing a high-speed and more efficient grading machine, the behaviour of the cashew nuts under motion and/or their aerodynamic behaviour should be studied.

IV. Studies on machine vision classification of cashew nuts should be extended to include the detection of degrading factors such as foreign material, shell stain, discolouration, and contaminated nuts.

REFERENCES

Abdul, W. (2019). Region Based Segmentation Techniques for Digital Images. *Journal of Computational and Theoretical Nanoscience, 16*, 3792-3801. doi:10.1166/jctn.2019.8252

Ahmadabadi, H. N,. Omid, M,. Mohtasebi, S. S,, Firouz, M. S. (2017). Design, development and evaluation of an online grading system for peeled pistachios equipped with machine vision technology and support vector machine. *Information Processing in Agriculture* 4(4): 333-341

Ajith, S., Stephen, P., Kumari, C. P., & Potty, V. (2014). Effect of storage temperatures and humidity on proximate composition, peroxide value and iodine value of raw cashew nuts. *Journal of Food Science and Technology, 52*. doi:10.1007/s13197-014-1476-6

Anonymous. (2021a). The History of Nuts. Retrieved from http://www.nutcrackermuseum.com/history_nuts.htm

Anonymous. (2021b). Nuts & Dried Fruits Statistical Yearbook 2020/2021. Retrieved 2/25/2022, from International Nut and Dried Fruit Council https://www.nutfruit.org/industry/technical-resources?category=statistical-yearbooks

Aran, M., G Nath, A., & Shyna, A. (2016). Automated cashew kernel grading using machine vision. Paper presented at the 2016 International Conference on Next Generation Intelligent Systems (ICNGIS).

Arora, M., & Devi, D. V. (2019). A machine vision based approach to cashew kernel grading for efficient industry grade application. *International Journal of Advance Research, Ideas and Innovations in Technology, 4*(6), 865-871.

Azam-Ali, S., H, & Judge, E., C (2001). *Small-scale cashew nut processing*. Retrieved from Warwickshire, UK: https://www.fao.org/3/ac306e/ac306e.pdf

Babu, C. S., Thota, L. S., Rao, A. A., Hanuman, T., Rambabu, P., Y.V, S., . . . Hanif, A. K. A. (2012). Intelligent model to classify cashew kernels. *International Journal of Engineering and Innovative Technology (IJEIT), 2*(6).

Balasubramanian, B., Sherfudeen, K., Kaliannan, S., & Murugesan, K. (2016). Cashew nut shell liquid poisoning. *Indian Journal of Critical Care Medicine, 20*, 57. doi:10.4103/0972-5229.173696

Banks, N. H. (2009). Chapter 1 - Postharvest Handling: A Discipline that Connects Commercial, Social, Natural and Scientific Systems. In W. J. Florkowski, R. L. Shewfelt, B. Brueckner, & S. E. Prussia (Eds.), *Postharvest Handling (Second Edition)* (pp. 1-8). San Diego: Academic Press.

Bao, Y., Han, J., Hu, F. B., Giovannucci, E. L., Stampfer, M. J., Willett, W. C., & Fuchs, C. S. (2013). Association of nut consumption with total and cause-specific mortality. *New England Journal of Medicine, 369*(21), 2001-2011. doi:10.1056/NEJMoa1307352

Bhuwad, A., & Pawaskar, S. (2019). Storage study of cashew apples. *Multilogic in Science Journal for Applied Science, 6*(19), 176-182.

Britannica, T. E. o. E. (2020). cashew: Encyclopedia Britannica.

Catarino, L., Menezes, Y., & Sardinha, R. (2015). Cashew cultivation in Guinea-Bissau – risks and challenges of the success of a cash crop. *Scientia Agricola, 72*, 459-467. doi:10.1590/0103-9016-2014-0369

Demant, C., Garnica, C., & Streicher-Abel, B. (2013). Overview: Image preprocessing *Industrial Image Processing Visual Quality Control in Manufacturing* (2nd edition ed., pp. 25-63).

Donkoh, A., Attoh-Kotoku, V., Kwame, R., & Gascar, R. (2012). Evaluation of nutritional quality of dried cashew nut testa using laboratory rat as a model for pigs. *TheScientificWorldJournal, 2012*, 984249. doi:10.1100/2012/984249

Drugman, T., Huybrechts, G., Klimkov, V., & Moinet, A. (2018). Traditional machine learning for pitch detection. *IEEE Signal Processing Letters, PP*, 1-1. doi:10.1109/LSP.2018.2874155

Eang, R., & Tippayawong, N. (2018). Optimization of process variables for drying of cashew nuts by superheated steam. *Cogent Engineering, 5*. doi:10.1080/23311916.2018.1531457

Fahey, F., Harkness, B., Keyes, J., Madsen, M., Battisti, C., & Zito, V. (1992). Sensitivity, resolution and image quality with a multi-head SPECT camera. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine, 33*, 1859-1863.

FAO. (2022). Production share of Cashew nuts, with shell by region. Retrieved 7/03/2022, from FAOSTAT https://www.fao.org/faostat/en/#data/QCL/visualize

Fernando, D. (2012). Plants: An International Scientific Open Access Journal to Publish All Facets of Plants, Their Functions and Interactions with the Environment and Other Living Organisms. *Plants, 1*. doi:10.3390/plants1010001

Ganganagowdar, N. V., & Siddaramappa, H. K. (2011). Cashew kernels classification using colour features. *International Journal of Machine Intelligence, 3*(2), 52-57.

Ganganagowdar, N. V., & Siddaramappa, H. K. (2016). Recognition and classification of White Wholes (WW) grade cashew kernel using artificial neural networks. *Acta Scientiarum. Agronomy, 38*, 145. doi:10.4025/actasciagron.v38i2.27861

Gone, R., & Vp, C. (2017). Estimation of initial moisture content of agricultural products by an ASTM Experimental Procedure. *Indian Journal of Natural Sciences, 8*(43).

Gordon, J. (1974). Ujamaa in Tanzania: Problems and Prospects. *A Current Bibliography on African Affairs, 7*, 360-377. doi:10.1177/001132557400700404

Huddar, V. B., & Kamoji , M. A. (2019). Feasibility of introducing solar air dryer for drying process in cashew industries. *International Journal of Recent Technology and Engineering (IJRTE), 7*(5S3).

Jain, P., & Kumar, S. (1998). Automatic feature extraction in PRIZCAPP. *Int. J. Computer Integrated Manufacturing, 11*, 500-512. doi:10.1080/095119298130534

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets, 31*. doi:10.1007/s12525-021-00475-2

Jindal, A., Joshi, S., Jangwal, R., Rathi, A., & Jain, R. (2021). Image segmentation *Proceedings of Second International Conference on Smart Energy and Communication* (pp. 247-259): Springer, Singapore.

Jøker, D. (2003). *Information About Cashew Nuts (Anacardium Occidentale)* Retrieved from

Karthikeyan, J., Hie, T., Jin, D., & Adgaonkar, S. (2022). Artificial Intelligence *Learning Outcomes Of Classroom Research* (pp. 236-242): L ORDINE NUOVO PUBLICATIONS.

Kilanko, O., Ojolo, S., Inegbenebor, A., Ilori, T. A., Leramo, O., Babalola, P., . . . Onwordi, P. (2019). Development of cashew nut shelling machine. *Journal of Physics: Conference Series, 1378*, 042091. doi:10.1088/1742-6596/1378/4/042091

Kilanko, O., Ojolo, S., J , Leramo, R., O , Ilor, T., A, Oyedepo, S., O, Babalola, P., O , . . . Ekwere, A. (2020). Dataset on physical properties of raw and roasted cashew nuts. *Data in Brief*.

Kim, K.-B., Nam, J.-H., & Suk, O. (2004). Data reduction in human vision system *Parallel and Distributed Computing: Applications and Technologies* (pp. 153-168).

Kluczkovski, A. M., & Martins, M. (2016). Cashew Nuts. *Encyclopedia of Food and Health*. doi:10.1016/B978-0-12-384947-2.00123-9

Koc, C., Gerdan, D., Eminoğlu, M. B., Yegül, U., Koc, B., & Vatandaş, M. (2020). Classification of hazelnut cultivars: comparison of DL4J and ensemble learning algorithms. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca, 48*(4), 2316-2327. doi:https://doi.org/10.15835/nbha48412041

Kosoko, S., Sanni, L., Adebowale, A.-R., Sokabi, A., & Oyelakin, M. (2009). Effect of period of steaming and drying temperature on chemical properties of cashew nut. *Afr. J. Food Sci., 3*.

Krepl, V., Kment, P., Rajdlová, G., & Kapila, P. (2016). African Countries' Agricultural Trade Value Chain Assessment Case study: Tanzania (Cashew nut exports). *Agris on-line Papers in Economics and Informatics, VIII*, 45-55. doi:10.7160/aol.2016.080105

Kubat, M. (2021). Deep Learning *An Introduction to Machine Learning* (2 ed., pp. 327-351): Springer, Cham.

Labudzki, R., Legutko, S., & Raos, P. (2014). The essence and applications of machine vision. *Tehnicki Vjesnik, 21*, 903-909.

Lin Hu, F., Hong, Y., Hong, J., Jeong, J., Han, Y., Kwon, I., & Kim, Y. (2017). Effects of cashew nut testa levels as an alternative to wheat bran in gestating sow diets. *Asian-Australasian Journal of Animal Sciences, 31*. doi:10.5713/ajas.17.0600

Liu, H., Zhou, M., & Liu, Q. (2019). An embedded feature selection method for imbalanced data classification. *IEEE/CAA Journal of Automatica Sinica, PP*, 1-13. doi:10.1109/JAS.2019.1911447

Mahajan, S. (2011). White Gold:An Experience of Cashew. *Golden Research Thoughts, 1*, 1-4.

Makundi, R., Massawe, A., & Laswai, H. (2006). Storage and protection of durable food crops and their products in Tanzania *Management of Selected Crop Pests in Tanzania* (1 ed., pp. 185-204). Dare es Salaam, Tanzania: Tanzania Publishing House (THP).

Masawe, P., & Z, Z. (2014). *Diseases and Insect Pests of Cashew in Tanzania*.

Mascarenhas, A. C., Bryceson, D. F., Ingham, K., & Chiteji, F. M. (2021). Jamhuri ya Mwungano wa Tanzania, United Republic of Tanzania. In E. Britannica (Ed.).

Masters, T. (2018). Image Preprocessing *Deep Belief Nets in C++ and CUDA C* (Vol. 1, pp. 85-101): Apress, Berkeley, CA.

Maurice, S. (2021). Evolution and opportunities for transactional Machine Learning in almost every industry *Transactional Machine Learning with Data Streams and AutoML: Build Frictionless and Elastic Machine Learning Solutions with Apache Kafka in the Cloud Using Python* (pp. 223-241). Toronto, ON, Canada.

Meshram, V., Patil, K., Meshram, V., Hanchate, D., & Ramkteke, S. D. (2021). Machine learning in agriculture domain: A state-of-art survey. *Artificial Intelligence in the Life Sciences, 1*, 100010. doi:https://doi.org/10.1016/j.ailsci.2021.100010

Mhd Zafer al Hosny, M. (2017). *basic description of CMOS camera sensor and color precnible*.

Mich, L. (2020). Artificial Intelligence and Machine Learning. In Z. Xiang, M. Fuchs, U. Gretzel, & W. Höpken (Eds.), *Handbook of E-Tourism* (pp. 1-21): Springer, Cham.

Mihayo, I., & Swai, R. (2019). Green economy in Tanzania: Is it foreseeable? *Journal of Applied and Advanced Research, 4*. doi:10.21839/jaar.2019.v4i4.299

Mohod, A., Khandetod, Y. P., Sengar, S. H., & Shrirame, H. Y. (2012). Energy Analysis of Baby Boiler for Steaming of Raw Cashew Nut Seeds. *ISRN Renewable Energy, 2012*, 476702. doi:10.5402/2012/476702

Mugnier, C. (2008). United Republic of Tanzania. *Photogrammetric Engineering and Remote Sensing, 74*, 149-150.

Nagpure, V., Joshi, K. (2016). Grading of cashew nuts on the bases of texture, color and size. *International Journal on Recent and Innovation Trends in Computing and Communication* 4(4): 171–173.

Ngatunga, E., Cools, N., Dondeyne, S., & Deckers, J. (2001). Soil suitability for cashew production in southeast Tanzania. *The Land, 5,* 3-16.

Nyein, A., & Khaung Tin, D. H. H. (2021). *Handwritten Myanmar Character Recognition System using the Otsu's Binarization Algorithm*.

Ojolo, S., Damisa, O., Orisaleye, J., & Ogbonnaya, C. (2010). Design and development of cashew nut shelling machine. *Journal of Engineering, Design and Technology, 8,* 146-157. doi:10.1108/17260531011062528

Opara, U., & Mditshwa, A. (2013). A review on the role of packaging in securing food system: Adding value to food products and reducing losses and waste. *African journal of agricultural research, 8,* 2621-2630. doi:10.5897/AJAR2013.6931

Özkan, İ. A., Köklü, M., & Saraçoğlu, R. (2021). Classification of pistachio species using improved K-NN classifier. *Journal of Nutrition and Internal Medicine, 23*(2). doi:https://doi.org/10.23751/pn.v23i2.9686

Paulsen, R., & Moeslund, T. (2020). Image Acquisition *Introduction to Medical Image Analysis* (pp. 7-25). Switzerland: Springer Nature Switzerland AG.

Pavlidis, T. (1978). A review of algorithms for shape analysis. *Computer Graphics and Image Processing, 7*(2), 243-258. doi:https://doi.org/10.1016/0146-664X(78)90115-6

Pratt, C. (1980). Beyond Ujamaa in Tanzania: Underdevelopment and an uncaptured peasantry. *African Affairs, 79*. doi:10.1093/oxfordjournals.afraf.a097269

Qingge, L., Zheng, R., Zhao, X., Wei, S., & Yang, P. (2020). An improved Otsu threshold segmentation algorithm. *International Journal of Computational Science and Engineering, 22*, 146. doi:10.1504/IJCSE.2020.10029225

Reddy, V., & Swathi, K. (2017). Investigation of Effectiveness of Simple Thresholding for Accurate Yawn Detection. *Advances in Intelligent Systems and Computing, 553*, 81-89. doi:10.1007/978-981-10-3770-2_8

Riad, R., Al-Taie, R., Basma, J., Saleh, L., & Abdalhasan, S. (2021). Image Edge-Segmentation Techniques : A Review. *International Journal of Scientific Research in Science, Engineering and Technology, 8*, 252-257. doi:10.32628/IJSRSET218528

Ros, E. (2010). Health Benefits of Nut Consumption *Nutrients, 2*. doi:10.3390/nu2070652

Rwahwire, S., Tomkova, B., Periyasamy, A. P., & Kale, B. M. (2019). Chapter 3 - Green thermoset reinforced biocomposites. In G. Koronis & A. Silva (Eds.), *Green Composites for Automotive Applications* (pp. 61-80): Woodhead Publishing.

Sarkar, N. (2017). Machine Vision for Quality Control in the Food Industry *Instrumental Methods for Quality Assurance in Foods* (pp. 167-187).

Savadi, S., Bm, M., & Preethi, P. (2020). Advances in genomics of cashew tree: molecular tools and strategies for accelerated breeding. *Tree Genetics & Genomes, 16*. doi:10.1007/s11295-020-01453-z

Savoy, J. (2020). Machine Learning Models *Machine Learning Methods for Stylometry* (1 ed., pp. 109-151): Springer International Publishing;Springer.

Semmlow, J., & Griffel, B. (2021). Image Segmentation *Biosignal and Medical Image Processing* (3rd edition ed., pp. 451-481): Taylor & Francis.

Silva, J., Phillips, H., & M, R. (2009). Post-Harvest Handling and Processing *Freshwater Prawns: Biology and Farming* (pp. 400-419).

Singh, R., Gehlot, A., Prajapat, M., & Singh, B. (2021). Machine Learning *Artificial Intelligence in Agriculture* (pp. 41-49).

Sivaranjani, A., Senthilrani, S., Ashokumar, B., & Murugan, A. (2018). An improvised algorithm for computer vision based cashew grading system using deep cnn. Paper presented at the 2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN).

Smith, N. J. H., Williams, J. T., Talbot, J. P., & Plucknett, D. L. (1992). *Tropical Forests and Their Crops*: Comstock Pub. Associates.

Steger, C., & Ulrich, M. (2021). A camera model for line-scan cameras with telecentric lenses. *International Journal of Computer Vision, 129*. doi:10.1007/s11263-020-01358-3

Subbiah, P. (2013). *Experimental Investigation Of Quality Characteristics In Macro And Micro Machining Processes Using Intelligent Techniques.* (Doctor of Philosophy), Anna University.

Sun, Y., Wang, Y., Chen, L., Du, W., Cao, Z., & Zhou, Y. (2013). Improved filter feature selection methods by samples localization. *Journal of Bionanoscience, 7*, 371-377. doi:10.1166/jbns.2013.1150

Sunoj, S., Igathinathane, C., & Jenicka, S. (2018). Cashews whole and splits classification using a novel machine vision approach. *Postharvest Biology and Technology, 138*, 19-30. doi:https://doi.org/10.1016/j.postharvbio.2017.12.006

Taiwo, E. A. (2015). Cashew Nut Shell Oil-A Renewable and Reliable Petrochemical Feedstock. In V. Patel (Ed.), Advances in Petrochemicals (pp. 152). Retrieved from https://www.intechopen.com/chapters/48869. doi:10.5772/61096

Thakkar, M., Bhatt, M., & Bhensdadia, C. K. (2011). Performance evaluation of classification techniques for computer vision based cashew grading system. *International Journal of Computer Applications, 18*(6), 9-12.

Tran, S. D., Uchiyama, N., Kirita, S., Yamanaka, N., & Ho, P. M. (2011). Towards the automation of cashew shelling operation. *Proceedings of the 16th International Symposium on Artificial Life and Robotics, AROB 16th'11*, 912-915.

Usman, A., Abubakar, S., Alaku, C., & Nnadi, O. (2014). Plant: A Necessity of Life. *International Letters of Natural Sciences*, *20*, 151-159. doi:10.18052/www.scipress.com/ILNS.20.151

Vidyarthi, S., Singh, S., Tiwari, R., & Rai, R. (2020). Classification of first quality fancy cashew kernels using four deep convolutional neural network models. *Journal of Food Process Engineering, 43*. doi:10.1111/jfpe.13552

Walsh, J., O' Mahony, N., Campbell, S., Carvalho, A., Krpalkova, L., Velasco-Hernandez, G., . . . Riordan, D. (2019). *Deep Learning vs. Traditional Computer Vision*. Paper presented at the Computer Vision Conference (CVC) 2019, Las Vegas, Nevada, United States.

Wilkinson, J. (2005). *Nut grower's guide : the complete handbook for producers and hobbyists / Jennifer Wilkinson*. Collingwood, Vic: CSIRO Publishing.

Woitha, D., & Janich, D. (2004). Adaptive Threshold *Reading and Learning: Adaptive Content Recognition* (Vol. 2956, pp. 14-25).

Woodroof, J. G. (1979). *Tree Nuts : Production, Processing, Products* (2 ed.): AVI Publishing Company, Incorporated.

Zhou, W., Du, X., & Wang, S. (2021). Techniques for Image Segmentation Based on Edge Detection. Paper presented at the 2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI).

Zorita, E., & Tilya, F. F. (2002). Rainfall variability in Northern Tanzania in the March-May season (long rains) and its links to large-scale climate forcing. *Climate Research - CLIMATE RES, 20*, 31-40. doi:10.3354/cr020031

CURRICULUM VITEA

My name is Geofrey Prudence BAITU, I graduated from Kwiro High School in 2014 and later obtained a bachelor's degree of science in Agricultural Engineering from the Faculty of Agriculture, Sokoine University of Agriculture (SUA) in 2018. In 2019 I was enrolled in the Agricultural Machinery and Technologies Engineering Master's program at Ondokuz Mayis University in Turkey. As part of my master's degree requirements, I have prepared a dissertation with the title " Machine-Vision-Based Classification of Cashew Nuts Using Colour Features".

Contact information:

ORCID ID: 0000-0002-3243-3252

Skills:

- Microsoft skills:
  Microsoft Office, Google Docs, Microsoft Word, Microsoft Excel.
- Research skills:
  Research and analytical skills, Strategic Planning, Organizational and planning skills
- Computer skills:
  Computer vision (image processing, image restoration), Programming language PYTHON, Visual Basic for Application (VBA Excel Access), Basic knowledge in machine operating.

Languages:

- Mother tongue(s): HAYA | SWAHILI
- Other languages: ENGLISH - Bilingual Proficiency
  TURKISH - Full Professional Proficiency
  SPANISH - Elementary Proficiency

Conferences and Seminars:

1. 07/10/2021 – 10/10/2021 – Jahorina mountain, Bosnia and Herzegovina
   XII International Scientific Agriculture Symposium "AGROSYM 2021"

   Title: AGRICULTURAL PRODUCTS GRADING METHODS AND THEIR ASSOCIATED CHALLENGES
2. 14/01/2022 – 14/01/2022 – Ondokuz Mayiz University, Samsun Turkey
   Master's Seminar

   Title: NEW METHODS FOR SORTING AND GRADING OF NUTS