DEFECT GREEN COFFEE BEAN DETECTION USING IMAGE RECOGNITION AND SUPERVISED LEARNING

SHAFIAN IZAN BIN SOFIAN

FACULTY OF COMPUTING AND INFORMATICS

UNIVERSITY OF MALAYSIA SABAH

2022

DEFECT GREEN COFFEE BEAN DETECTION USING IMAGE RECOGNITION AND SUPERVISED LEARNING

SHAFIAN IZAN BIN SOFIAN

THESIS SUBMITTED IN PARTIAL FULFILLMENT
FOR THE BACHELOR OF COMPUTER SCIENCE
(NETWORK ENGINEERING)

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITY OF MALAYSIA SABAH
2022

NAME	:	**SHAFIAN IZAN BIN SOFIAN**
MATRIC NUMBER	:	**BI18110271**
TITLE	:	**DEFECT GREEN COFFEE BEAN DETECTION USING IMAGE RECOGNITION AND SUPERVISED LEARNING**
BACHELOR	:	**BACHELOR OF COMPUTER SCIENCE (NETWORK ENGINEERING)**
DATE	:	**11th February 2022**

VERIFIED BY;

1. **SUPERVISOR** Signature
   DR. FARASHAZILLAH BIN YAHYA

   ____

2. **EXAMINER**
   DR. ERVIN GUBIN MOUNG

   ____

   ASSOC. PROF. DR. NG GIAP WENG

   Edmund
   ____

DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, excepts equations summaries and reference, which have been duly acknowledged.

....

11th February 2022

SHAFIAN IZAN BIN SOFIAN

BI18110271

ACKNOWLEDGEMENT

First and foremost, I would like to be thankful and grateful to Allah that made me to be able going through many hardships and obstacles during my studies in University of Malaysia Sabah, be it educationally, mentally, or physically. The amount of sheer will, and strength given to me is the prove that I was helped throughout these past years.

I would like to express my gratitude to my mother, my father, my brother, and those who surrounded me with words of encouragement, faith, directly and indirectly helped me up to this point my supervisor. Nevertheless, my recognition also goes towards Dr. Farashazillah Binti Yahya for the constant guidance for my Final Year Project assessment along with valuable criticism from both my examiners – Dr. Ervin Gubin Moung and Assoc. Prof. Dr. Ng Giap Weng. It is without a doubt that these role models inspired me to complete my Final Year Project.

. Apart from that, I would also like to extend my appreciation to my fellow friends, and lecturers in University of Malaysia Sabah for the supports and cares I had been received for many years. Last but not least, these group of people are the one that make everything possible for me to write this acknowledgement and complete my Final Year Project.

....

SHAFIAN IZAN BIN SOFIAN

11th February 2022

ABSTRACT

Addressing the quality of green coffee bean is an important process to define its quality and market price for any industry that processing it. Normally, the evaluation that is carried out in determining the quality of green coffee is by visual inspection where it has limitations, and it is prone to error. Therefore, in this research project, the process will be conducted by using an image classifier with the model of a machine learning algorithm which the candidates comprise of Support Vector Machine, k-Nearest Neighbour and Decision Tree. k-nearest neighbour has the highest F1-score (0.51) than the other two algorithms (Support Vector Machine: 0.50, and Decision Tree: 0.48). The model was integrated as web application with Flask where user can upload the image and the system will return result with precision and prediction. This integrated web application is tested with functionality test and integration test which it succeeded both successfully fulfilling each criterion tested.

ABSTRAK

KECACATAN PENGESAHAN BIJI KOPI HIJAU MENGGUNAKAN PENGIKTIRAFAN IMEJ DAN PEMBELAJARAN TERSELIAAN

Menangani kualiti biji kopi hijau adalah proses penting untuk menentukan kualiti dan harga pasarannya untuk mana-mana industry yang memprosesny, Biasanya, penilaian yang dijalankan dalam menentukan kualiti kopi hijau adalah dengan pemeriksaan visual di mana ia mempunyai had, dan ia terdedah kepada kesilapan. Oleh itu, dalam projek penyelidikan ini, prosese akan dijalankan dengan menggunakan pengelas imej dengan model algoritma pembelajaran mesin yang mana calon-calonnya terdiri daripada Support Vector Machine, k-nearest neighbour dan Decision Tree. k-nearest neighbourmempunyai skor F1 tertinggi (0.51) berbanding dua algoritma lain (Support Vector Machine: 0.50 Dan Decision Tree: 0.48). Oleh itu, model ini disepadukan sebagai aplikasi web dengan Flask di mana pengguna boleh memuat naik imej dan system akan mengembalikan hasil dengan ketepatan dan ramalan. Aplikasi web bersepadu ini diuji dengan ujian kefungsian dan ujian integrasi yang mana ia berjaya kedua-duanya memenuhi kriteria yang diuji.

TABLE OF CONTENTS

LIST OF FIGURES

PAGE

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Chapter Overview

Chapter 1 contains the insight about the research project. This chapter also summarizes the problem background, problem statement, project objectives, and organization of this research project.

1.2 Problem Background

In coffee mass production, the problem lies in the post-harvest of the manufacturing process of coffee production. At this stage coffee beans are carefully selected and sorted which known as visual inspection that is conducted by human inspectors. This is one of the crucial stages in the manufacturing process of coffee production since the healthy and defected coffee beans is clearly separated to prevent the spoilage in taste and quality of the production.

This tedious and labour extensive process is the main source problem of this manufacturing process. If this problem does not being address, it will impact on labours and the process itself. As mentioned, human inspectors which is the labours are recruited to conduct the visual inspection for defect detection of post-harvesting.

As human energy is limited and taking consideration that process is a mass production this will lead to their physical and mental state such as fatigue, and bias choice. The second impact mentioned is the process where products yield is degraded in terms of its quality, bad condition coffee beans is mixed with healthy ones, and it takes more human labours to speed up the production.

There is an alternative to solve this problem which is using the machinery that can sort beans by density and size, as well as eliminating any sticks, pebbles, nails or other debris that may have gotten mixed in with the coffee while drying. The beans are blown into the air by the first machines; the heaviest and largest beans fall into the bins nearest to the air source, while the lightest and most likely to be defected beans are blown into the farthest bin. Other machines sort the beans by size by shaking them through a succession of sieves. Next, a gravity separator rattles the sized beans on a titled surface, causing the heaviest, densest, best to vibrate to one side and the lightest to the other. There is one that needs to be noted that colour sorting is not included in this alternative. Colour sorting is the most difficult and crucial phase in the sorting and cleaning process. Colour sorting is done in the most basic method which is by hand to yield most high-quality coffees. This redirects towards the problem where the alternative for this problem is proven unsuccessful.

1.3 Problem Statement

As mentioned earlier, the main party that is to be concerned here is the human labours who works countless of hours to do visual inspection for the coffee bean. This problem arose in the process called cleaning and sorting, specifically to check the sizes and colour sorting of the coffee bean. According to (Ayitenfsu, 2014), visual inspection done by human inspectors check the size of the coffee bean which exist in 3 forms, which are, full, half, and broken. A significant number of bruises or half and broken states of the green coffee bean can affect the flavour and aroma which overall as its beverage qualities. Human visual inspectors' method is a process that is efficient, yet the possibility of the error made is high. If the problem is not being address there are several possibilities that could impact the mass manufacturing process. For the first point, labours' fatigue is serious matter as the company relies on them to do the visual inspection while the process is repeated twice or even thrice to product high quality coffee production. In relation to the first point, companies that are responsible to mass produce high quality bean will be affected because of its quality and probability of their labour quitting jobs corresponding to find more labours which is cost extensive. Solving this problem by using proposed system in this research project can help to replace the human labour method mostly if not eliminating the method completely. The solution to this problem can advance one's knowledge on how to take the advantage of our current technology, in this case machine learning technology that is proven to be helpful to mankind in many field of applications. This research paper also targets to be the base foundation for other research paper in the same research topic.

1.4 Project Objectives

The main aim of this research project is to eliminate the visual inspection that is carried out by human which is most likely to be a prone error. Any slight mistakes will make the quality production of sorting green coffee bean affecting the image of Malaysia as one of many countries that exports coffee such as Liberica coffee which has only 1% population worldwide.

Furthermore, correlating the sorting factor that most visual inspection carried out in the process of defect detection of green coffee bean in Malaysia, this research project has the potential in contributing the service of defect detection of green coffee bean which enhances the productivity of the concerned parties. The University of Malaysia Sabah has the Faculty of Sustainable Agriculture that specialized in crop production. This faculty can add coffee plantation to their list or offer the service to do the defect detection of the green coffee bean by using the algorithm from this research paper. Hence, this will make the University of Malaysia Sabah acknowledged across the coffee exporters and importers alike. According to (Institute, 2019), the pattern for Consumer Price Index between 2010 – 2018 is growing. Following this pattern, Malaysian are spending coffee is growing throughout the year.

The objectives of this research project are listed below, where these objectives are contributing to building a system of defect detection green coffee bean by using image recognition with the proposed machine learning technique (Support Vector Machine, K-Nearest Neighbour, Decision Tree).

i. To investigate and analyse the effectiveness of the Support Vector Machine, the K-Nearest Neighbour, and the Decision Tree model algorithm based on confusion matrix, and F1-score.
ii. To develop a prototype system for Defect Detection of Green Coffee Bean using the chosen machine learning from (i).
iii. To perform testing in terms of interaction between modules and actions within function using integration testing and unit testing.

1.5 Organization of The Project

The final report will be containing the following chapters for this research project:-

Chapter 1: Introduction

This chapter briefly explains the nature of this research project by relating it with real world problem that can be solved using machine learning as in the title per se for this paper. It also outlines the objectives for this research project to achieve through research study and implementation later.

Chapter 2: Literature Review

This chapter explains in detail regarding past years research/books/journals/articles that is related to this research paper's topic. There will be comparison of research among the other related work.

Chapter 3: Methodology

This chapter discusses the model approach on how this research paper conducts including its software and hardware requirement listed. Furthermore, this chapter also illustrate the sub process components for the model used in the web application later.

Chapter 4: System Analysis and Design

This chapter describes the system analysis of the system to show the comprehensive understanding for the system and how does the system design accordingly.

Chapter 5: Implementation

This chapter depicts two phases for this research project which are 1) preliminary implementation for the comparison of two algorithms, and 2) system implementation where the integration of the chosen model algorithm being integrated with the designated web application.

Chapter 6: Testing

This chapter describes how the web application's software modules interacting action taken by user. The tests are functionality and integration testing which is to detect problems, obtain confidence and information about the system's quality level, and prevent defects from occurring.

Chapter 7: Conclusion

This chapter concludes summarization of this research project with synopsize of each chapter previously and adding the future work remarks.

CHAPTER 2

LITERATURE REVIEW

2.1 Chapter Overview

This chapter explains in detail regarding past years research/books/journals/articles that is related to this research paper's topic. There will be comparison of research among the other related work.

2.2 Green Coffee Bean

The coffee bean is widely used in the world of culinary where mostly is used as beverages such as the base drinks from *Starbucks*. The coffee plant is the source of the coffee bean, but it does not become like that without undergoing a certain process. When the times come, the fruit will ripe, and it is most likely that the handpicked method will be used. There will be always two ways of handpicking, one of them is selective picking where only ripe fruit is removed from the branches and the other one is strip picking where in this method, all the fruit is removed from the branches.

Furthermore, the next stage is where coffee cherries undergoing the process known as wet or washed. This makes the cherries' flesh is separated from the seed and at the same time fermenting the seed itself which is soaked in the water for about two days. Mucilage surrounding the seed will become soften over time and then the layer will wash off with water. The third stage of the process is known as dry processing where the fruit is spread out on the concrete to be exposed to the sunlight where it will eventually be dried and left for about 2 to 3 weeks. These processes will then yield green coffee bean where it is an unroasted mature or immature coffee beans.

2.2.1 Healthy Green Coffee Bean

Healthy green coffee bean can be defined in perfect condition before they were roasted or used as green coffee. Usually, the shape and size of the bean itself do not have any imperfection to it such as broken. Figure 2.1 are the example images of healthy green coffee bean:

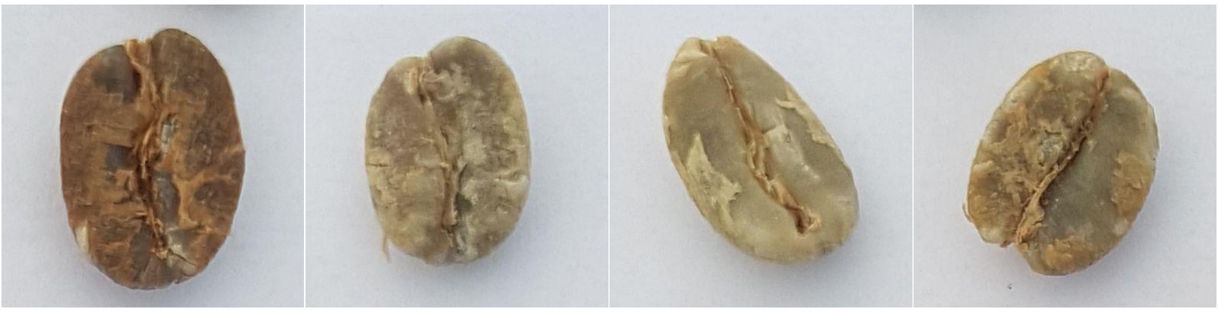

Figure 2.1: Healthy Green Coffee Bean sample images from the dataset

2.2.2 Defect Green Coffee Bean

Defect green coffee bean can be seen in Figure 2.2, where the defect can be defined as broken, large and medium stone, large and medium sticks, and husk.

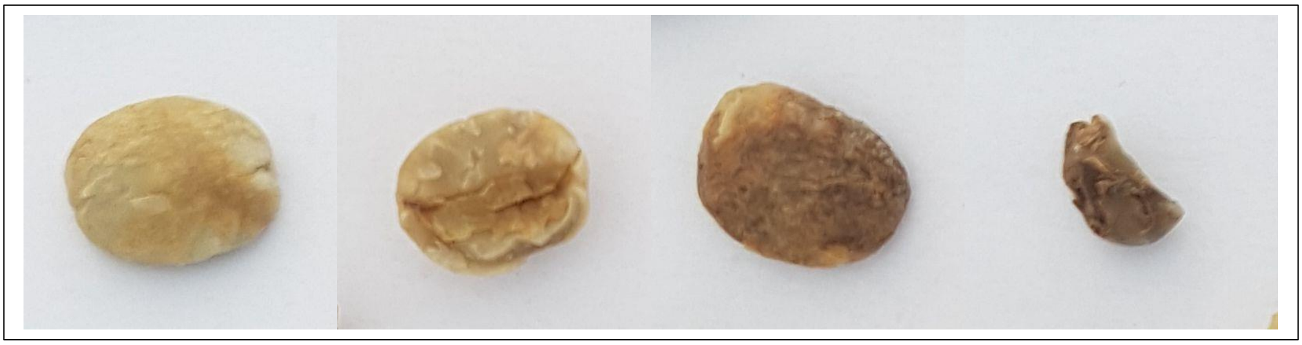

Figure 2.2: Defect Green Coffee Bean sample images from the dataset

2.3 Image Recognition

Image recognition is one of the most pursued and active area by the scientist in imaging sciences and engineering. The reason is evident (Javidi, 2002) that this field has the potential to replace human visual capabilities with a machine and its extensive application diversity. The main idea of image recognition is to examine an image to obtain valuable information from it from the sensors. This will substantially reduce human workload and improve the accuracy of decision making. For an instance, a journal article (Hjelmås & Low, 2001), where image recognition is applied gives the ability to recognize facial from an image. The research can be applied to nowadays smartphone security feature where access is granted after the scanning of the holder's is done.

2.3.1 Feature Extraction

Feature extraction is a vital procedure in the field of image processing. Feature extraction occurs after the image undergo various image pre-processing techniques such as binarization, thresholding, resizing and normalization (Kumar & Bhatia, 2014). This process will extract any valuable information that is going to be used in classifying and recognition of images. Various image processing application such as character recognition is done by feature extraction. The feature extraction process had been done, classified features revealing the behaviour of the image, where it defines the place in terms of storage taken, time consumption and classification efficiency.

2.3.2 Image Processing

Image processing is any form of signal processing that receives input as an image such as photographs. Images undergo image processing will be an image or parameter related to the image or a set of characteristic as an output for the process. Image processing techniques mostly treat its input as a two-dimensional signal and standard signal-processing is applied to it (Ng & Goldberger, 2010). Usually, this process is referring to digital image processing, but optical and analogue image processing is also possible. There are many examples of image processing operations, one of them is Euclidean geometry transformation that involves enlargement, reduction and rotation of an image.

2.4 Support Vector Machine

Support Vector Machine started as a theory created by Vladimir Vapnik and Alexey Chervonenski in the study of computational learning theory that is directly related to statistical learning theory from distribution-free data (Vapnik, 1995). The theory known as VC Theory became the foundation of what we know today as Support Vector Machine – a supervised learning technique for classification and regression. In this algorithm, there is an *n*-dimensional space that is determined by the number of the data item as a point of reference where *n* is the number of features. *n*-dimensional space will be the coordinate system for the data plotted. The main objective of this algorithm is to find a hyperplane within those boundaries of n-dimensional space that segregates and classifies the data points.