2024-06

Automated Coffee Quality Prediction: A Study on Predicting Coffee Quality Attributes from Roasted Coffee Images using Image analysis and Machine Learning

Yamlaksira, Degu

Bahir Dar University

Bahir Dar Institute of Technology

School of Research and Postgraduate Studies

Faculty of Computing

Automated Coffee Quality Prediction: A Study on Predicting Coffee Quality Attributes from Roasted Coffee Images using Image analysis and Machine Learning

By:

Yamlaksira Degu

June, 2024

Bahir Dar, Ethiopia

Automated Coffee Quality Prediction: A Study on Predicting Coffee Quality Attributes from Roasted Coffee Images by using Image analysis and ML

By:

Yamlaksira Degu

A Thesis Submitted to the School of Research and Graduate Studies of Bahir Dar Institute of Technology, BDU in Partial Fulfilment of the requirements for the Degree of Master of Science in Computer Science in the Faculty of Computing.

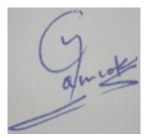
Advisor: Tesfa Tegegne (Assoc. Prof)

June, 2024

Bahir Dar, Ethiopia

DECLARATION

This is to certify that the thesis entitled "**Automated Coffee Quality Prediction: A Study on Predicting Coffee Quality Attributes from Roasted Coffee Images by Using Image Analysis and ML**", submitted in partial fulfillment of the requirements for the degree of Master of Science in Computer Science under Faculty of Computing, Bahir Dar Institute of Technology, is a record of original work carried out by me and has never been submitted to this or any other institution to get any other degree or certificates. The assistance and help I received during the course of this investigation have been duly acknowledged.



Yamlaksira Degu	______________	6/18/2024
Name of the candidate	signature	Date

BAHIR DAR UNIVERSITY

BAHIR DAR INSTITUTE OF TECHNOLOGY

SCHOOL OF GRADUATE STUDIES

COMPUTING FACULTY

Approval of thesis for defense result

I hereby confirm that the changes required by the examiners have been carried out and incorporated in the final thesis.

Name of Student: Yamlaksira Degu Signature ______ Date 23/09/2024

As members of the board of examiners, we examined this thesis entitled "Predicting Coffee Quality Attributes from Roasted Coffee Images using Image analysis and Machine Learning" by Yamlaksira Degu. We hereby certify that the thesis is accepted for fulfilling the requirements for the award of the degree of Masters of Science in Computer science.

Board of Examiners

Name of Advisor	Signature	Date
Dr. Tesfa Tegenge (Ph.D)	______	Sep23/09/2024
Name of External examiner	Signature	Date
Dr. Dereje Teferi (Ph.D)	______	September 2, 2024
Name of Internal Examiner	Signature	Date
Dr. Mekwanent (Ph.D)	______	Sep 30, 2024
Name of Chairperson	Signature	Date
Mr. Alemu Kumilachew	______	Sep 30, 2024
Name of Chair Holder	Signature	Date
Ms Aleumi Krumhrebow / Kel: St Mesheshe	______	Sep 30, 2024
Name of Faculty Dean	Signature	Date
Dr. Tesfa Tegenge (Ph.D)	______	Sep 30, 2024

iii | P a g e

Bahir Dar University Bahir Dar Institute of Technology Faculty of Computing

ACKNOWLEDGEMENTS

First, I would like to thank my god for providing me with the strength and inspiration to pursue this research. Secondly, I would like to express my sincere gratitude to my advisor, Dr. Tesfa Tegegne (Associate professor.), for his invaluable guidance and support throughout the development of this research, and also to coffee quality expert Mr. Getachew who gave a direction for getting data for the research. His expertise and insights have been instrumental in shaping the direction of this research. I am truly grateful for his patience, encouragement, and commitment to excellence. Thank you, Dr. Tesfa and Mr. Getachew, for your unwavering dedication and mentorship.

List of Abbreviations

AHE…………………..……. Adaptive Histogram Equalization

ANOVA………………….Analysis of variance

BBHE…………………… Brightness Preserving Bi-Histogram Equalization

CSV…………………..….Comma Separated Value

CLAHE………………….Contrast-Limited Adaptive Histogram Equalization

DNN……………………..Deep Neural Network

ECTA…………………....Ethiopian Coffee and Tea Authority

ECTACQICC…………. Ethiopian Coffee and Tea Authority Coffee Quality Inspection and Certification Center

ECX…………………….. Ethiopia Commodity Exchange

FNC…………………….. National Federation of Coffee Growers

FNC…………………….. Federation of Coffee Growers

FN………………………..False Negative

FP……………………...…False Positive

GLCM……………….…. Gray-Level Co-occurrence Matrix

HE……………………… Histogram Equalization

LLC……………………… Limited Liability Company

LTD ………………………Limited

MAE………………………Mean Absolute Error

ML………………………..Machine Learning

NIR……………………… Near Infrared Spectroscopy

ORP………………………Oxidation-Reduction Potential

PH………………………...Potential of hydrogen

PLS……………………….Partial least squares

RF………………...………Random Forest

RNN………………………Recurrent Neural Network

SIFT………………………Scale Invariant Feature Transform

SURF…………………..….Speeded-up Robust Features

SVM……………………...Support Vector Machine

TN………………………..True Negative

TP………………………...True Positive

Table of Contents

LIST OF FIGURE

LIST OF TABLE

LIST OF APPENDIX

ABSTRACT

Coffee is a popular and widely consumed beverage that people worldwide enjoy. It has a long history and cultural significance, often part of daily routines. Coffee quality assessment is vital in the industry, influencing consumer satisfaction and market value. Traditional assessment methods usually rely on subjective judgments, which can be time-consuming, labor-intensive, and costly. In this study, we develop a prediction model for coffee quality attributes by fusing multiple modalities. The model was trained on a comprehensive dataset of various coffee samples and their corresponding quality attribute values. We used machine learning and image analysis techniques to develop the model. We explore various algorithms like linear regression, random forests, and other regressions, along with feature selection techniques such as recursive feature elimination and polynomial featuring with scaling. We used HOG, GLCM, and color descriptors to extract color, texture, and shape features from roasted coffee beans. The model was trained and tested using a dataset divided into training (80%) and testing (20%) sets. To evaluate the model we use mean squared error (MSE), mean absolute error (MAE), and coefficient of determination (R^2 score). Our model demonstrated exceptional performance, achieving a high R^2 score of 0.98754. This indicates a strong correlation between the predicted and actual sensorial scores. The low MAE of 0.014088 and MSE of 0.00969 further validate the accuracy and reliability of our model's predictions. This automated approach may saves time, reduces labor costs, and enhances overall efficiency in the coffee production and distribution processes. Furthermore, the model may improve consumer satisfaction by offering high quality coffee products for the Ethiopian and international market.

Key Words: Coffee, Quality Attributes, Linear Regression, Machine learning, Feature Extraction and Feature Selection

CHAPTER ONE

INTRODUCTION

1.1 Background

Coffee is an immensely popular and widely consumed beverage, with a global market that continues to grow (Santos, 2020)**.** Ethiopia is the biggest producer of Coffee Arabica in Africa and the fifth-largest exporter worldwide. About 30-35% of Ethiopia's overall export revenues come from coffee, making it the country's top export revenue source. Ethiopia exclusively produces coffee in Arabica. The projected total production in 2023–2024 is 8.35 million 60 kg bags (“Review on Coffee Production and Marketing in Ethiopia,” 2020). In Ethiopia, there are some places that more commonly have high amounts of coffee products, commonly in the southern and some northern parts of Ethiopia.

Without question, the highest quality coffee for natural processing and washed coffee is Yirgacheffe (Toma et al., 2023). The greatest places to buy washed coffee are Sidama, Guji, and Limu. Both Jimma and Harrar work well with unwashed coffee (Muktar Bedaso et al., 2022). Specialty coffee is promised in all these areas. The secret to flavor is altitude (Ma et al., 2022). Sidama is the highest coffee-growing region in Ethiopia, situated 2,200 meters above sea level. For instance, Yirgacheffe has a light to medium body, flowery aromas, and a floral flavor. Sidama is a full-bodied wine coffee with a touch of acidity and spice (Worku et al., 2023)

The quality of coffee is a crucial aspect that not only affects consumer preferences but also plays a significant role in the reputation and success of coffee producers. Traditionally, the assessment of coffee quality has heavily relied on sensory evaluations performed by trained experts known as cuppers (Javier, 2020). These evaluations involve a complex and meticulous process, including aroma profiling, flavor analysis, and the assessment of various sensory attributes such as acidity, body, sweetness, and aftertaste (Javier, 2020; Muktar Bedaso et al., 2022; Kim, 2022). However, this manual evaluation process is time-consuming, labor-intensive, and subject to inter- and intra-rater variability. To this end, this type of evaluation is cumbersome to achieve consistent and objective assessments.

The color of roasted coffee beans is a crucial factor in determining the taste and smell of coffee. Light roasts have a more acidic and floral aroma with citrus, and herbal flavors, while medium roasts offer a well-balanced flavor with notes of caramel, chocolate, and nuts. Dark roasts have a smoky aroma with bold, rich flavors of bittersweet chocolate, toasted nuts, and smokiness. The roasting level influences the development of flavor compounds, acidity, sweetness, and body in coffee. Understanding the relationship between roasted coffee color and its sensory attributes can help coffee enthusiasts make informed choices (Ma et al., 2022).

In recent years, there have been remarkable advancements in the fields of computer vision, image analysis, and machine learning, which offer promising opportunities to revolutionize coffee quality assessment (Javier, 2020; Muktar Bedaso et al., 2022; Kim, 2022; Esteban-Díez et al., 2004; Petch Sajjacholapunt et al., 2022). These technologies provide the means to automate the assessment process, offering objective, consistent, and efficient methods for evaluating coffee quality. By analyzing visual information extracted from roasted coffee images, it is possible to identify unique patterns, colors, textures, and shapes that may be correlated with specific quality attributes.

The use of image analysis and machine learning algorithms in the context of coffee quality assessment offers several advantages. Firstly, it enables the processing of a large volume of data in a relatively short time, allowing for a more efficient evaluation of coffee samples (Muktar Bedaso et al., 2022). Instead of relying solely on human cuppers, who are limited in their capacity to assess countless coffee samples, an automated system can rapidly analyze numerous images and provide consistent results. Secondly, by reducing the reliance on human expertise, the automated system can minimize subjective biases that may influence sensory evaluations (Esteban-Díez et al., 2004). This leads to more objective and standardized assessments, enhancing the reliability and accuracy of coffee quality predictions. Thirdly, an automated system has the potential to provide real-time and continuous quality monitoring throughout the coffee production and supply chain, ensuring consistent quality control and enabling proactive intervention when necessary.

1.2 Motivations

Currently, Ethiopian coffee faces a challenge where other countries profit from its value-added processes, such as roasting and standardized washing, which generate substantial revenue. The cause of using traditional sensorial evaluation, the method heavily relies on subjective human judgment, creating a space for potential inconsistencies and biases. This study plays a vital role in addressing this issue. By automating the coffee quality evaluation process, we aim to enhance objectivity, efficiency, and reliability in assessing Ethiopian coffee. This, in turn, empowers coffee producers and exporters to add value to their products locally, ensuring that the economic benefits remain within the country. By improving the quality evaluation methods by considering factors like the degree of roasting and the impact of brewing methods, we can strengthen Ethiopia's position in the global coffee market and enable the country to maximize the financial benefits derived from its exceptional coffee beans.

1.3 Statement of the Problem

There are two systems used in Ethiopia for coffee grading. The liquor value or cup test and the raw quality or green analysis. Raw quality is computed out of 40% and liquor value is computed out of 60% which is a cup test (Muktar Bedaso et al., 2022). The quality of coffee in cup tests is primarily affected by the level of coffee roasting. Another factor to consider in evaluating coffee quality is its sensory attributes. Coffee quality assessment is a critical aspect of the coffee industry, impacting consumer satisfaction, market value, and purchasing decisions (Muktar Bedaso et al., 2022).

Traditional sensory evaluation methods in Ethiopia, conducted by expert cuppers, are subjective, time-consuming, and often require specialized training. While some researchers have explored using roasted coffee image color and shape for quality assessment (Santos, 2020), there is a gap in the literature regarding the fusion of multiple sensory modalities to enhance prediction accuracy and also they don’t consider different roasting profiles to predict sensorial attributes (Javier, 2020; Muktar Bedaso et al., 2022; Kim, 2022). Understanding these quality attributes is crucial for achieving the highest quality coffee. The main issue of this study is to develop an automated coffee quality assessment system that fuses roasted coffee image color and shape analysis and its sensorial score with machine learning techniques to predict coffee quality attributes.

By doing so, the study aims to address a specific problem. Traditional sensory evaluation methods in Ethiopia rely on subjective judgments, leading to potential biases and inconsistencies among cuppers. Previous research has examined roasted coffee quality using image color and shape analysis (Santos, 2020; Muktar Bedaso et al., 2022). However, some studies, such as Esteban-Díez et al., (2004), have utilized an unstandardized espresso brewing method that cannot be used for coffee quality standardization, this has its impact on coffee quality. This study aims to overcome subjectivity by fusing multiple sensory modalities. By incorporating additional information from visual features extracted from roasted coffee images, the proposed system can provide objective and consistent evaluations of coffee quality attributes.

In Ethiopia, traditional sensory evaluation methods for assessing large volumes of coffee samples can be costly, labor-intensive, and time-consuming (Tolessa et al., 2016). To overcome this problem, Putra et al., (2023) suggest automatic quality assessment using AI techniques will improve efficiency and reduce costs. The proposed system will enable faster decision-making, reduce the need for extensive human involvement, and lower the overall resources required for coffee quality assessment. The coffee industry encompasses a wide range of coffee origins, processing methods, and roast profiles, resulting in diverse coffee quality attributes. While previous studies have explored using individual sensory modalities, this study seeks to develop a scalable and standardized automated system that fuses multiple modalities.

The proposed system effectively integrates the complex aspects of color, texture, shape, and coffee quality attributes. Previous research has explored roasted coffee image color and shape for quality assessment, but limited studies have investigated the fusion of these modalities (Javier, 2020; Kim, 2022; Esteban-Díez et al., 2004; Petch Sajjacholapunt et al., 2022). By integrating color, texture, shape, and coffee quality attributes, the proposed system introduces a unique and comprehensive methodology that aims to revolutionize the way coffee quality is evaluated. Leveraging roasted coffee images and their corresponding sensorial scores, the system captures the visual cues and the impact of roasting protocols on coffee quality attributes. This integration allows for a more holistic analysis, facilitating the prediction of coffee quality attributes even in cases where standardized protocols may not have been followed. Finally, this study attempts to answer the following research questions.

RQ1. How Image analysis and ML can accurately predict roasted coffee quality attributes?

RQ2. What is the comparative effectiveness of the automated system in predicting coffee quality attributes compared to human cuppers utilizing traditional sensory evaluation methods?

RQ3. How does the fusion of roasted coffee image color and coffee quality attributes improve the prediction accuracy of coffee quality attributes?

1.4. Objectives of the Research

1.4.1. General Objective

The general objective of this study is to predict coffee quality attributes from roasted coffee and sensorial data using image analysis and machine learning techniques.

1.4.2. Specific objective

To accomplish the overall objective, the subsequent particular objectives are formulated:

- Identify the most important attributes to predict the coffee quality attributes.
- Collect and prepare a roasted coffee image dataset
- Develop a predictive model for roasted coffee quality attributes.
- Conduct a comparative analysis between the performance of the developed automated system and at least five human cuppers who utilize traditional sensory evaluation methods, and demonstrate that the automated system outperforms or achieves comparable performance to the human cuppers with a statistically significant difference.
- Use Explainability techniques and get insight from the models' predictions.

1.5 Scope and Limitation

1.5.1. Scope of the Study

The main focus of this study is predicting coffee quality attributes from roasted coffee images using image analysis and machine learning techniques. The data has already been collected from various Ethiopian regions coffee, such as Limu, Harrar, Sidamo, Yirgacheffe, Jimma, Kaffa, Nekemt, Wenbera, and Zegae coffees. Expert evaluators from the Ethiopian Coffee and Tea Authority Coffee Quality Inspection and Certification Center (ECTACQICC) have conducted the sensorial scoring of the roasted coffees. The attributes include 'Aroma', 'Flavor', 'Aftertaste', 'Acidity', 'Body', 'Balance', 'Uniformity', 'Sweetness', 'Cupper Points', 'Clean Cup'. In this study, coffee grading was performed by utilizing predicted values for the coffee quality attribute. The grading system encompassed three classes: grade 1, grade 2, and grade 3. It is worth noting that grades 4 and 5 were not included in the analysis due to the limited availability of sufficient data.

1.6 Significance of the study

One of the key benefits of this research is its potential to enhance the coffee quality assessment process. Currently, coffee quality assessment relies heavily on human judgment often subjective that can lead to bias and inconsistencies. This study aims to provide a more objective and standardized approach to evaluating coffee quality. This automated assessment method has several implications for the coffee industry. Firstly, it can help coffee producers and exporters guarantee the reliability and excellence of their products. By accurately predicting coffee quality attributes, such as flavor, aroma, acidity, and other attributes, this method can assist in identifying the best coffee batches and ensuring that only high-quality coffee reaches the market.

Moreover, this study holds immense potential for providing consumers with a more dependable and unbiased approach to assessing coffee quality. By offering such a method, consumers can make well-informed decisions when purchasing coffee, instilling in them a sense of assurance regarding the coffee's quality and its alignment with their particular preferences. Consequently, consumers can have greater confidence in the coffee they buy, leading to heightened satisfaction and enjoyment. The implementation of an automated system for predicting coffee quality attributes can have significant implications for producers and sellers alike. By leveraging a more accurate and consistent assessment method, they can confidently classify and market their coffee as premium-grade, thereby commanding higher prices in the market. This increased profitability not only serves as a reward for producers' efforts in cultivating high-quality coffee but also bolsters their competitiveness within the industry.

Furthermore, from a societal perspective, this research contributes to the advancement of image analysis and machine learning techniques. By applying these technologies to the coffee industry, this study showcases their potential in various fields beyond traditional applications. This can inspire further research and innovation in other industries, leading to advancements in automated quality assessment processes and improving overall product quality. In general, this study on automated coffee quality prediction holds significant significance in the following ways:

- Introduces an automated approach to coffee quality assessment, reducing reliance on subjective human judgment.
- By providing an objective and standardized method, it enhances the reliability and excellence of coffee products.

- Coffee producers and exporters can benefit from accurately predicting quality attributes, ensuring only high-quality batches reach the market.
- Consumers can make well-informed purchasing decisions based on reliable assessments, leading to increased satisfaction and enjoyment.
- Implementing an automated system allows producers and sellers to classify and market their coffee as premium-grade, commanding higher prices and improving profitability.
- The research contributes to the advancement of image analysis and machine learning techniques by giving an open data set.

1.7 Methodology of the Research

This study employs an experimental research approach to analyze the coffee quality thoroughly attributes prediction system. We can manipulate variables and systematically collect data by conducting controlled experiments. This approach allows for directly observing the system's performance and provides empirical evidence to support the findings. The experimental approach also enables hypothesis testing and validation against predefined criteria or benchmarks.

This research initiated a system for automated coffee quality prediction predicting coffee quality attributes. In this study, the coffee quality attributes were analyzed and evaluated to advance automatic prediction of roasted coffee. To maintain a clear vision throughout the study, various resources such as journal articles, conference papers, books, company reports, and online sources have been reviewed to achieve the research objective and address the research questions from the beginning of the study until its completion. The following methods are applied in this research.

1.7.1. Research Design

In this study, an experimental research approach is employed. Applying empirical or experimental research involves image acquisition, data pre-processing, feature extraction, feature selection training, and performance evaluation.

1.7.2. Data Acquisition and Dataset Preparation

An image can be described as a two-dimensional function, denoted as f(x, y), where (x, y) represents the coordinates in the image space, and ‘f’ represents the intensity or value at that specific coordinate (Mishra et al., 2017). Each position in the image is referred to as a pixel, which serves as the fundamental building block of the image, representing the smallest unit or picture element. This concept of pixels allows us to analyze and manipulate images by considering the individual values and their spatial arrangement within the image (Mishra et al., 2017).

In this study, we used a smartphone with a 50-megapixel camera for image acquisition 8288 x 6032 resolution images collected from Yirgacheffe, Limu, Harrar, Sidamo, Jimma, Kaffa, Lekemt, Wenbera, and Zegae coffees in Ethiopian Coffee and Tea Authority Coffee Quality Inspection and Certification Center (ECTACQICC) in Addis Ababa, Ethiopia. We capture the image by taking three Varieties of distances from the camera to the roasted coffee 100mm, 155mm, and 210mm. After testing each distance we used an average of the nearest and widest distance between the roasted coffee image and the camera which has a better quality. To reduce noise due to handshakes, we used a camera stand. In this research, we captured the images by roasting coffee beans at different time intervals at a constant voltage level (210°C). We have collected a total of 675 original roasted coffee images with their sensorial scores, and we used augmentation techniques, and the augmented data is 2700. In this study, six distinct cuppers used the cupping technique to determine the sensory scores. To evaluate the sensory qualities and quality of coffee samples, the coffee industry frequently uses cupping, a standard assessment procedure. The cuppers assess the roasted coffee's aroma, flavor, aftertaste, acidity, body, and overall quality using cupping. Smelling both wet and dry coffee grounds, slurping the brewed coffee to evaluate flavor, and then rating the coffee according to their sensory impressions are all part of the methodical procedure that the cuppers utilize. This cupping method yields sensorial scores that offer important insights into the sensory profile and qualities of roasted coffee, enabling a thorough assessment and comparison of various samples. In addition to that, the image size we used by taking an image sample was 124x124 pixels. Out of the total data set, 80% was used as a training set for creating the coffee quality attribute identification model and 20% for testing or evaluating the model's performance.

1.7.3. Design procedure

Our system comprises four main components. The first component handles image acquisition, capturing images of roasted coffee using a camera or similar devices with its corresponding sensorial score. The acquired images undergo pre-processing steps to enhance their quality and prepare them for feature extraction. These steps include histogram equalization for the balanced color distribution, noise reduction to eliminate unwanted artifacts, resizing for suitable dimensions, image enhancement to emphasize specific features and conversion to grayscale for simplified feature extraction.

Feature extraction is the subsequent step, where meaningful information is extracted from the pre-processed images. This is achieved through the utilization of a color histogram feature, which employs statistical analysis techniques to capture and represent the distribution of colors in an image and identify the significant difference between the two groups. By utilizing color histograms, significant characteristics and patterns can be extracted, enabling precise analysis and understanding of the coffee's color properties. It is used to recognize patterns and extract significant characteristics from the images, enabling accurate feature extraction.

By integrating image acquisition, pre-processing, and feature extraction using linear regression, our system effectively extracts relevant features from the acquired roasted coffee images. These features can then be utilized for further analysis, such as predicting the sensorial score of the coffee based on its image attributes. The Testing phase is responsible for evaluating the capabilities of the final system. Finally, the validation of the testing phase is applied after finding the accuracy using the testing dataset. The experimentation was conducted using a Computer with Core TM Intel® core™ i7-7500U CPU @ 2.7GHz and 8 GB of RAM, Hartanzah professional coffee roasting timer app, and coffee sample roaster equipment.

1.7.4. Evaluation

The developed system is evaluated to measure how well it supports a solution to the problem. To evaluate the system in a rational method, testing datasets fed into the developed model. After that, the system was assessed by comparing its output against the observed data using the confusion determinant, MSE, MAE, accuracy, precession, recall, and f1-score and also by analyzing variance between human cuppers and this model.

1.8. Organization of the Document

The remainder of this thesis is organized as follows. Chapter two explores the literature review of image processing, sensorial evaluation, feature extraction, and identifying gaps in existing literature. The system architecture for coffee quality attribute prediction, including pre-processing techniques, feature extraction methods, and prediction algorithms is presented in chapter three. Chapter four presents experimental results and discussions, evaluating the study's performance.

The final chapter concludes by summarizing findings, offering recommendations, and exploring future research possibilities.

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview

This chapter focuses on the analysis and review of literature related to coffee quality and processes. We examined various studies and research papers to gain insights into the concept of coffee quality. Additionally, we explored the literature on image processing, feature extraction, and prediction methods used in the overall activities of this study. By reviewing the relevant literature, we aimed to enhance our understanding of coffee quality prediction and improve the accuracy of our analysis.

2.2 Coffee Quality

Coffee is a popular beverage made from the roasted seeds of the coffee plant, known as coffee beans. The best-quality coffee beans have a deep, rich color and a glossy appearance (Javier, 2020). Coffee is enjoyed worldwide for its aroma, flavor, and stimulating effects. It can be consumed in various forms, such as brewed coffee, espresso, and more (Muktar Bedaso et al., 2022). Freshly roasted coffee offers superior taste and aroma compared to pre-ground or stale coffee. Coffee is not only used as a beverage but also adds flavor to culinary creations. It holds cultural significance and serves as a social lubricant, bringing people together. Coffee has become an integral part of daily routines, providing a moment of relaxation and enjoyment (Eron et al., 2024).

2.2.1 Coffee Quality in Ethiopia

Ethiopia is commonly regarded as the origin of Arabica coffee and is one of the world's largest coffee producers (Gebeyehu et al., 2020). The country has a long history of coffee production and is well-known for its high quality. Ethiopian coffee quality assessment is done by traditional sensorial evaluation techniques (Abebe et al., 2020; Girma Adugna, 2021). Ethiopian coffee quality is influenced by various factors, as highlighted in a study by Abebe et al. (2020). These factors include the diverse range of coffee varieties, the favorable growing conditions characterized by different microclimates and altitudes, the implementation of traditional farming methods that prioritize organic and sustainable practices, the unique processing methods like the natural and washed processes, and the emphasis on the specialty coffee industry (Abebe et al., 2020).

According to a study conducted by Fadri et al. in 2019, the temperature and duration of roasting significantly impact various quality attributes of Arabica coffee. The study found that these roasting parameters affect the yield, water content, color value, acidity, flavor, taste, and color of the coffee. The temperature and duration of the roasting process play a crucial role in determining the coffee's final characteristics and sensory profile, including its aroma, acidity, body, and overall flavor.

2.2.2 Sensorial Evaluation

The most important step in the preparation of coffee is roasting, which develops the desired flavor and aroma. The beans undergo several reactions during this process that alter their chemical composition and physical characteristics (Aliah, et al., 2015). Coffee quality assessment is a critical aspect of the coffee industry, impacting consumer satisfaction, market value, and purchasing decisions (Muktar Bedaso et al., 2022). When coffee is raw, it has no taste or unique aroma. The aroma and flavor that we know to as coffee flavor require some sort of processing. The assessment and determination of coffee quality, especially in the context of roasting, heavily relies on sensory evaluation (Tarigan et al., 2022).

To fully extract the taste of coffee, dry roasting of the coffee beans is necessary, and careful attention to detail is needed to ensure the right flavor. Flavor oils can be damaged and the final coffee flavor ruined by over- or under-roasting, as well as by roasting too fast or slowly. Getting the beans just right to maximize flavor is the art of roasting. An overdone roast produces a harsh, burnt flavor, whereas an underdone roast impedes the development of flavor. Experienced roasters can use visual clues to examine the texture and maturity of the beans, as much evaluation of roasted coffee is based on visual cues (Bolka & Emire, 2020).

A roasted coffee's color does not always reflect how developed the beans are; some coffees can be fully roasted at lighter shades than others. Furthermore, the bean's full development is indicated by a smooth texture free of cracks or balkiness, which the roaster must make sure the bean has. Through constant practice and experience, one gains this proficiency in visual evaluation (Wulandari et al., 2022). The process of roasting coffee beans over a dry heat source is required to generate the oils, compounds, and lipids that are edible and contribute to the flavor of real coffee. Before roasting the sample, it is best to keep the temperature of the roasting machine between

300°F – 464°F, or 150°C – 240°C. Ten to twelve grams of roasted coffee, blended with 250 milliliters of water, is the suggested amount for sensory evaluation (Santoso et al., 2021).

Figure 1: Samples Coffee roasting using a sample roaster machine (six-cylinder roaster).

2.2.3 Cup testing

Cup testing is a crucial method used in evaluating the quality of coffee. It involves tasting the coffee when it has cooled to a palatable temperature, neither too hot nor too cold. During cup testing, a spoonful of the liquid is taken to the back of the mouth, ensuring that air is incorporated to activate the taste buds (Henrique & Paula, 2023). Sipping the liquid without air does not allow the taste buds to fully perceive the flavors. The liquid is held in the mouth for a few seconds, typically around 10 seconds, allowing for the assessment of defects, acidity, body, and flavor characteristics (Santoso et al., 2021). It is important to set personal preferences aside during cup testing and focus on determining the quality based on industry standards.

Figure 2: Cup testing

In coffee cupping, key sensory attributes are evaluated. Acidity represents sharp, clear notes without being sour. Body refers to the thickness or weight of the tongue. The flavor is a combination of acidity and body. The aftertaste is the lingering pleasant flavors. Sweetness indicates agreeable flavor and can be influenced by carbohydrates. A clean cup lacks negative impressions during cupping (Tarigan et al., 2022). In general, the sensorial score of each sample is taken by Ethiopian Coffee and Tea Authority Coffee Quality Inspection and Certification Center (ECTACQICC) cuppers.

2.3. Applications on Coffee Quality

Techniques for image analysis and machine learning have proven to be useful in evaluating the quality of coffee. Through the application of these techniques, researchers have been able to predict cupping scores a measure of the sensory qualities of coffee accurately by examining the visual information taken from images of roasted coffee (Machado et al., 2019; Farré et al., 2020). These methods have also been used to categorize coffee roast levels, which helps with flavor and aroma characterization. Researchers have accurately classified roast levels, from light to dark roasts, by using machine learning models and assessing color data from images of roasted coffee (Machado et al., 2020; Jiang et al., 2021).

Furthermore, image analysis and machine learning have been used have been utilized in coffee grading to assess the size and uniformity of coffee beans. By analyzing visual features such as shape, color, and texture extracted from coffee bean images, machine learning models can accurately classify beans into different size categories, enabling precise grading based on size consistency (Bosilj et al., 2018; Pérez-Sánchez et al., 2020). These techniques have also been applied to assess coffee freshness by analyzing visual features such as color changes or surface texture variations in roasted coffee images, allowing for accurate classification of freshness levels (López-Ruiz et al., 2021; Doan et al., 2022). These applications highlight the potential of image analysis and machine learning in providing objective and efficient methods for evaluating coffee quality attributes, thereby contributing to enhanced decision-making processes and quality control in the coffee industry.

2.4. Image Processing

An image is a visual representation of an object, a person, or a scene. The image analysis is a mechanism to operate on the source image to become an improved image, which helps extract representing feature vectors from the input image (Yogesh Shankar Ghodake et al., 2021). It's a sort of signal processing in which the source is an image and the output is also an image or the image's characteristics/features. In today's digital life, digital images are everywhere around us (Tyagi, 2018).

2.4.1. Image Acquisition

The process or technique of obtaining an image from a source is known as image acquisition, and it usually comes first in the workflow sequence since processing can't be done without an image (Shi et al., 2020). The primary goal of image acquisition is to convert an optical picture (real-world data) into a range of numerical data that can be processed on a computer. However, before processing video or images, an image needs to be taken using a camera and transformed into a manageable entity (Mishra et al., 2017). The Image Acquisition process consists of three steps:-

1. An optical device that focuses energy
2. The energy that the interesting object reflects
3. A sensor for measuring energy levels.

A suitable camera allows for image acquisition (Xiu et al., 2018). Various cameras can be utilized for various purposes. Film or a camera with X-ray sensitivity can be used if an X-ray image is required. The use of cameras that are sensitive to infrared radiation is an option if we want an infrared image. We can employ cameras with visual spectrum sensitivity for regular shots, such as photographs of relatives. This study used a smartphone camera as an image acquisition mechanism because the smartphone is appropriate or suitable enough.

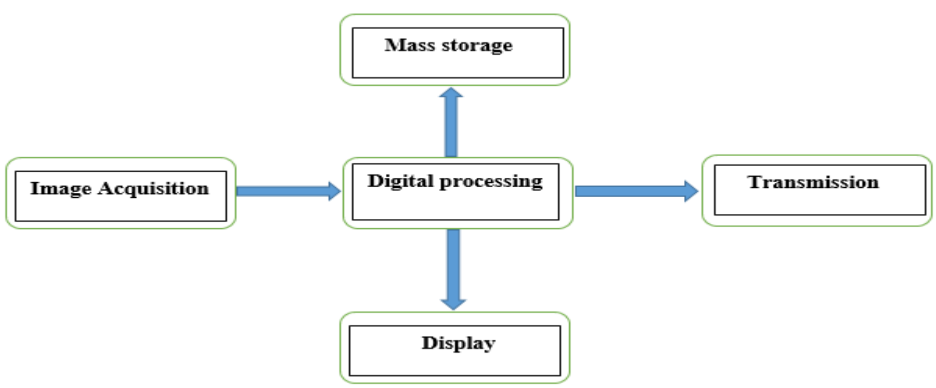


Figure 3: Image Processing

There are numerous methods for acquiring images, each with benefits and uses of its own. Among the noteworthy methods are near-infrared (NIR) imaging and hyperspectral imaging (HSI).

Hyperspectral imaging: it is a modality that combines imaging and spectroscopy. It often delivers continuous scanning imaging of tens or hundreds of spectral ranges at ultraviolet (UV), visible (VIS), infrared, and even mid-infrared wavelengths, covering a continuous portion of the spectrum (Cui et al., 2022).

Near-infrared imaging: Based on their wavelengths, infrared (IR) electromagnetic radiation is classified into three categories: Wavelengths between 0.78 and 2.5 μm are classified as near-infrared (NIR), 2.5 and 25 μm as mid-infrared (MIR), and 25 and 1000 μm as far infrared (FIR). NIR is a vibrational spectroscopy technique that uses photon energy (hv) in the 2.65 x 10-19 to 7.96 x 10-20 J energy range, which corresponds to the 750–2,500 nm wavelength range (Zhang et al., 2022).

2.4.2. Pre-processing

Several actions and methods performed on unprocessed data before its use in modeling or analysis are referred to as data preparation. To prepare the data for additional processing, it must be cleaned, transformed, and organized. Assisting with missing values, eliminating errors or inconsistencies, eliminating noise, and preparing the data in a format that works with the particular modeling or analysis methods being used are all objectives of data preprocessing (Mishra et al., 2020). Pre-processing is used to reveal clear information in digital image components, identify essential areas, and decrease image noise (S. Kumar et al., 2019).

Computer algorithms are used in the digital image processing process. With several benefits over analog image processing. It is a subfield of digital imaging that has a lot of benefits over traditional image processing. The purpose of computer image processing is to enhance some important picture attributes or remove superfluous distortions from image data to improve the data for our computer vision models. Generally speaking, the goal of this stage is to prepare your data for the machine-learning model so that it can be processed and interpreted more easily (Arias et al., 2020).

2.4.2.1 Image Enhancement

Image enhancement is a collection of methods and procedures used to raise an image's viscosity, clarity, and observability. Removing noise or artifacts, adjusting brightness and contrast, and improving overall image aesthetics are the objectives of image enhancement. In many different domains, such as computer vision, satellite imaging, medical imaging, and imaging, it is frequently used (Yogesh Shankar Ghodake et al., 2021).

In image enhancement techniques contrast and brightness are enhanced and the image does not lose its original information and the brightness is preserved (Raja Lakshmi & Annapurani, 2021). Histogram Equalization (HE), Adaptive Histogram Equalization (AHE), Brightness Preserving Bi-Histogram Equalization (BBHE), and Contrast-Limited Adaptive Histogram Equalization (CLAHE) are some of the image enhancement techniques.

- **Histogram Equalization (HE)**

Histogram equalization is an image processing technique that makes use of the histogram of the image to improve contrast. A popular technique in image processing to enhance contrast and look overall is histogram equalization (Xiong et al., 2021).

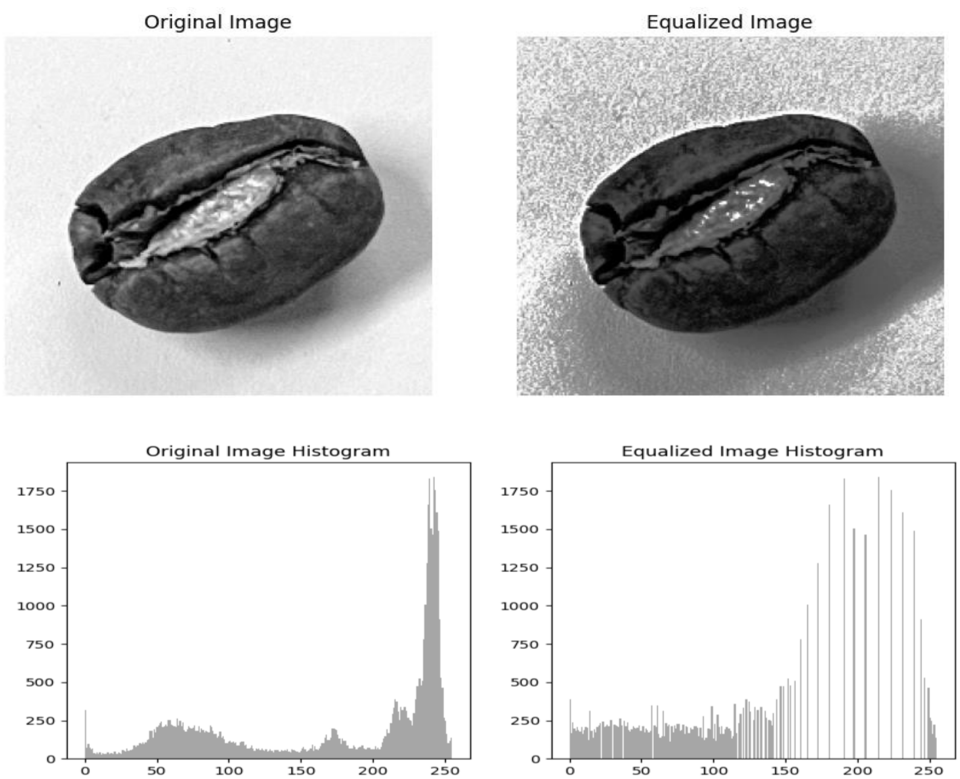


Figure 4: Comparing the impact of the algorithm for histogram equalization.

Histogram equalization (HE) can be implemented in different forms, each addressing specific requirements or limitations. Here are two common forms of histogram equalization:

Global Histogram Equalization: This is the basic form of histogram equalization, where the histogram of the entire image is computed and equalized. It redistributes the pixel intensities across the entire image, aiming to achieve a uniform histogram. Global histogram equalization is simple to implement and can enhance the overall contrast of an image. However, it does not consider local variations in contrast and may lead to over-enhancement or loss of detail in specific regions (Pandey et al., 2017).

Local Histogram Equalization: Also known as Adaptive Histogram Equalization, this form of histogram equalization takes into account the local characteristics of an image. Instead of equalizing the histogram of the entire image, it divides the image into smaller regions, or tiles, and performs histogram equalization on each tile independently (Chourasiya & Khare, 2019).

2.4.2.2 Image resizing

Image resizing is the process of changing the dimensions of an image to make it larger or smaller. It allows you to adjust the size of an image to meet specific requirements or constraints (Vahidi, 2022). Resizing can be done for various purposes, such as optimizing file size, adapting images for display on different devices, preparing images for printing, or creating thumbnails. It involves adjusting the pixel values using interpolation techniques to maintain visual quality and aspect ratio (Danon et al., 2021). It is performed with the help of various interpolation approaches including nearest neighbor, bi-cubic, and bilinear techniques.

- **Nearest neighbor interpolation**

Nearest neighbor interpolation uses the nearest pixel to fill the interpolated point, making it the simplest interpolation mechanism for resizing the image. It also takes less computing time than other interpolation approaches. These methods of interpolation impact the enlargement and reduced size of the image. It chooses a pixel's value that is very close to the intended interpolation point's surrounding coordinates. This interpolation technique is suitable for images with all edges vertical and images that have a high-resolution pixel. In most cases the edges of the image are random. So, the quality of the original image is better than the resized image (Manjunatha & Patil, 2018).

- **Bilinear interpolation**

The bilinear interpolation method creates an image that is smoother than the original and can be applied in both vertical and horizontal directions (Rukundo, 2019). To determine the proper color intensity values for a particular pixel, bilinear interpolation employs just the 4 nearest pixel values that are positioned in diagonal directions from that pixel. This method allows better results than nearest neighbor interpolation and takes less computational time compared to Bicubic interpolation (S. Parsania & V.Virparia, 2015).

- **Bi-cubic interpolation**

Bi-cubic interpolation is the most effective interpolation strategy out of all the interpolation methods. Averaging the pixels at various distances between known and unknown pixels yields the final interpolated value. A smoother image is produced by bi-cubic interpolation as opposed to bilinear and nearest neighbor interpolation (S. Parsania & V.Virparia, 2015). However, it takes more computational time.

2.4.2.3 Noise Removal

Noise is any unwanted information that entered the image during the capture process and produced pixel values that were not true to the scene's intensity. It is unnecessary information that degrades the visual quality of an image (Tania & Rowaida, 2016). Noise removal is significant because it removes the image's noise content, which is necessary. After all, noise causes inaccuracies in image processing techniques. Noise reduction determines the quality of the entire image processing cycle (Mohamed, 2015). It is the necessary stage of image preprocessing for removing the noise content of the image. To reduce noise from images, there are several noise removal techniques, such as Mean filters, Wiener filters, Median filters, Gabor filters, and Gaussian filters.

- ✓ **Mean filters: -** The mean filter reduces the amount of intensity variance between pixels, which is a basic, easy-to-understand, and uncomplicated way to reduce image noise (Tania & Rowaida, 2016). This filtering technique replaces each pixel value in a picture with the average or means of all of its neighbors, including itself.

- ✓ **Median filter: -** The median filter is a basic and effective nonlinear statistical filtering method for eliminating noise from images. The median filter performs better than the mean filter when it comes to maintaining meaningful detail in the image. The idea behind median filtering is to move each pixel one by one and replace its value with the neighborhood pixels' median value (Dhruv et al., 2018). It is used to reduce the amount of intensity variation between the pixels and it is better to avoid Poisson noise, impulse noise, salt, and pepper noise from the image while it is poor for avoiding Gaussian noise.

- ✓ **Gaussian filter: -** Gaussian filter is a linear noise reduction method that keeps clear of Gaussian noises and selects the smoothing function's weights based on the Gaussian function's shape (Padmaja et al., 2020). However, it is poor for removing salt and pepper noise. All of the image's edges and details are smoothed off with Gaussian filters.
- ✓ **Wiener filter**: - The Wiener filter is a component of the statistical filter methodology. This technique is primarily used to lower the mean square error between the intended process and the estimated random process by removing noise from the corrupted image. A Wiener filter is used to reduce speckle and Gaussian noise. However, the limitation is that Wiener works best when the noise is constant-power and it requires more computation time than linear filtering (Khudayer Jadwa, 2018).

2.4.2.4 Image Augmentation

Data augmentation is a common preprocessing technique used to enlarge the dataset size to avoid overfitting (Alzubaidi et al., 2021). The most common methods of generating new images or augmentation are: rotating at some degrees, copping randomly, flipping vertically or horizontally, scaling outward and inward, translating, and enhancing the learning capability.

2.4.2.5 Segmentation

Segmentation is the process of dividing an image into meaningful and distinct regions or segments. The goal of segmentation is to partition an image into homogeneous regions based on certain criteria such as color, texture, intensity, or other visual features (Song & Yan, 2018). The most common image segmentation techniques are threshold, region-based, edge-based, and cluster-based segmentation.

Threshold segmentation is a technique used in image processing and computer vision to separate objects or regions of interest from the background based on a specific threshold value. It is a simple and commonly used method for image segmentation, particularly when the foreground and background have distinct intensity or color characteristics (Niu & Li, 2019).

Region-based methods refer to a class of techniques used in computer vision and image processing for object detection and segmentation. These methods aim to identify and delineate specific objects or regions of interest within an image. Unlike pixel-based methods that operate on individual pixels, region-based methods focus on grouping pixels into coherent regions based on certain criteria (Ganatra & Patel, 2018).

Cluster-based segmentation is a method used in computer vision and image processing to divide an image into coherent areas based on pixel property similarities. To distinguish different areas or objects within the image, the objective is to cluster pixels that share similar properties, like color, texture, or intensity (Song & Yan, 2018).

2.4.2.6 Color Conversion

Color conversion, also known as color space conversion or color transformation, is the process of converting the representation of colors from one color space to another. Color spaces define the way colors are represented numerically, allowing for consistent and standardized color communication and manipulation (Bi & Cao, 2021). The following describes a few widely used color spaces and how color conversion works:

RGB (Red, Green, and Blue): RGB is an additive color model where colors are represented as combinations of red, green, and blue primary colors. In this color space, each color is defined by three numerical values representing the intensities of the red, green, and blue components. RGB is widely used in electronic displays and digital imaging devices (Guangwu Lv & Cao, 2020).

CMYK (Cyan, Magenta, Yellow, and Key/Black): CMYK is a subtractive color model used in printing and reproduction. It represents colors as combinations of cyan, magenta, yellow, and black inks. CMYK color space is used to achieve accurate color reproduction in print media, where the inks absorb light to create colors (Velastegui & Pedersen, 2021).

HSV (Hue, Saturation, and Value): HSV is a cylindrical color space that represents colors based on their hue, saturation, and value/lightness. Hue represents the dominant color, saturation represents the intensity or purity of the color, and value/lightness represents the brightness. HSV is often used for color selection, image editing, and color-based image analysis (Bi & Cao, 2021).

Several widely employed color conversion techniques are commonly used in various applications:

RGB image to Grayscale: RGB is widely used and relatively easy to utilize in practically all computer systems. It serves as a foundational color space for several uses (Kora & K. Thangadurai, 2016). The conversion process of RGB image to Grayscale is done using a function COLOR_RGB2GRAY ().

RGB to HSV Conversion: This conversion is used to transform colors from the RGB color space to the HSV color space (Kora & K. Thangadurai, 2016). This conversion allows for manipulating color attributes such as hue, saturation, and value independently, making it useful for tasks like color selection, image editing, and color-based.

RGB to CMYK Conversion: This technique is used to convert colors from the RGB color space (used in digital displays) to the CMYK color space (used in printing). The conversion involves transforming the RGB values to corresponding CMYK values, taking into account the color gamut and limitations of the CMYK color model (Bi & Cao, 2021).

Image Binarization: Image Binarization is the process of converting either a color image or a grayscale image, to a bi-level image through which each pixel is categorized either as a foreground or a background pixel. Image binarization or thresholding is used to distinguish the content from the background. During the process of binarization, image pixels are classified into two values one and zero, which represent foreground and background classes respectively (Goel et al., 2017). Binarization is done based on the image pixel value; for a threshold value T, if the value of the image pixel is greater than or equal to the threshold then it becomes one or white otherwise it is zero or black.

2.5 Feature extraction

Feature extraction refers to the process of transforming raw data or input into a reduced representation of relevant and distinctive features. It aims to extract meaningful and discriminative information from the data while reducing its dimensionality (Chavan et al., 2020). Feature extraction is widely used in various fields, including computer vision, pattern recognition, machine learning, and signal processing. In image processing, features are a quantitative description of an image, which is represented by feature vectors. Different techniques, such as handcrafted and deep feature extraction, are used to extract features from an image (Bansal et al., 2020)

Feature extraction algorithms are used to find features that better reflect an image by using fewer parameters. The image can be represented meaningfully with fewer parameters when the required components are used. Removing unimportant parameters requires a faster and more successful classification with fewer computational resources. When an algorithm's input data is huge to process and assumes that some of it is often redundant (data with little information), it is converted into a reduced representation of a group of features (called features vector). Features removal is the method of transforming input data into a group of elements (Öztürk & Akdemir, 2018).

There are two types of feature extraction techniques these are handcrafted feature extraction and Automatic Feature Extraction techniques. The design of handmade features involves finding the right balance between precision and machine performance. For example, the full-scale Invariant Feature Transform is well-known for its capability to rotate objects and its variations in size; however, this power comes with a high calculation cost. SIFT has been proposed using real-time applications and on devices with low computational power (Qasaimeh et al., 2019).

2.5.1 Handcrafted feature extraction

There are two types of features extracted from an image: generic and domain-specific features. Local and global features are the two types of general features. The global feature of an image is considered the whole image and helps to identify the target from the background. However, these features are only effective when there is no occlusion or lighting problem. Local features are considered the selected region of the image edge and shape. Domain-specific features are specific to any particular application and often a combination of low-level features of the specific domain. (Bansal et al., 2020) explained that handcrafted feature extraction algorithms grouped features into different categories including color-based, textured-based, and shape-based.

Feature extraction using HOG: It is a feature extraction technique that is used to determine the shape and appearance of the object. It also describes the object by counting the occurrence of gradient orientation (Alhindi et al., 2018). It is used for extracting local region features using the histogram orientation intensity (Yassin Kortli et al., 2018). During HOG first, calculate the gradient using the appropriate filter mask to extract edge gradient and orientations. In the end, a grid of histograms is produced using the detected gradients and orientation.

Feature extraction using a Gabor filter: It is a feature extraction algorithm used to extract the feature from the magnitude of the Gabor-filtered images. Gabor filter is used to interpret certain frequency content in the image in a specified direction in a small area surrounding the area of interest (Ganatra & Patel, 2018). This technique is a multi-scale and multi-resolution filter that is utilized to determine the spatial extent, direction, and spectral bandwidth. It is invariant to rotation changes and scale in the images. The most important properties of the Gabor feature are rotation, invariance to light, scaling, and translation. However, when the number of filters increases it requires high computational time.

Feature extraction using Scale Invariant Feature Transform: SIFT is a local feature descriptor that takes an image transforms it into N*N data and generates a set of features from it (Qasaimeh et al., 2019). In case features are extracted locally. Its output is invariant to rotation, camera viewpoint, light, and scale. The benefit of SIFT is invariant to scale, light, and rotation. However, it takes more time to compute feature vectors, mathematically complicated, and most of the time it does not give accurate results on the blurred image.

Feature extraction using Speeded-up Robust Features: SURF is a local feature descriptor that is invariant to scale, rotation, illumination, viewing direction, and noise. SURF is built based on the Scale Invariant Feature Transform (SIFT) descriptor (Yassin Kortli et al., 2018). It utilizes an intermediate image representation known as an integral image. This descriptor is designed first to detect local key points and then to produce a descriptor for these points.

2.5.2 Texture-based Algorithms

Texture refers to the visual and tactile surface characteristics of an object or image. It describes the patterns, arrangement, and variations of elements within an image, such as the roughness, smoothness, coarseness, or regularity of its surface. Texture plays a crucial role in image analysis, as it provides valuable information for tasks like object recognition, segmentation, and classification (Öztürk & Akdemir, 2018).

Texture feature extraction encompasses various techniques, including the widely used Gray-Level Co-occurrence Matrix (GLCM) and Gabor filters. GLCM calculates statistical measures based on the spatial relationships of pixel intensities, capturing texture properties like contrast, homogeneity, and entropy (Öztürk & Akdemir, 2018; Humeau-Heurtier, 2019)). On the other hand, Gabor filters employ a set of spatially localized filters tuned to different frequencies and orientations, enabling the extraction of texture features related to edges, corners, and texture orientation. These methods are integral components of texture feature extraction, providing valuable information for tasks such as texture analysis, classification, and recognition. These filters imitate certain aspects of the human visual system since they are based on multichannel filtering (Ganatra, 2020).

2.6 Machine Learning Algorithms

In the field of machine learning, a diversity of algorithms exists to tackle the complex task of prediction. Each of these algorithms has its special techniques and advantages. Linear regression is a basic and often-used approach for predicting continuous outcomes. It seeks to build a continuous target variable and input feature connection that is linear (Maulud & Abdulazeez, 2020). Linear regression aims to minimize the difference between the actual and anticipated values by fitting a best-fit line to the data. It is especially helpful in situations where a straight line can roughly represent the relationship between the variables. Multiple input features and polynomial relationships can be handled via linear regression, which is also interpretable and computationally efficient (Kumari & Yadav, 2018).

Decision trees are flexible algorithms that work with both categorical and numerical data. Based on feature values, they divide the input space into regions and then assign the majority class or mean value inside each zone to generate predictions. Nonlinear correlations and interactions between variables are captured by decision trees (Tsami et al., 2018). Since the resulting tree structure clearly outlines the decision-making process, they are simple to interpret. Decision trees are capable of handling irrelevant features, outliers, and missing values. Multiple decision trees are used in ensemble methods like random forests or gradient boosting to further improve performance (Tsami et al., 2018). A collection of unpruned classification or regression trees created from a random selection of training data samples is known as Random Forest, it is made up of several tree-structured classifiers and it was developed by Leo Breiman. In the induction process, features are chosen at random (Liu, 2019).

Vapnik first suggested SVM in 1995, and it was first known as "support-vector networks". The concept of mapping non-linear vectors to a very high-dimension feature space and creating a linear decision surface (hyperplane) in this feature space is implemented by the binary classification method time. Therefore, it has a positive impact on both separable and non-separable problem-solving (Wang et al, 2017). The other algorithm is XG-Boost. The XG-Boost system is a scalable end-to-end tree-boosting tool that is extensively utilized by data scientists to attain cutting-edge outcomes on various machine learning tasks such as classifications (Chen, 2019). Adaptive Boosting, or Ada-Boost, is a well-liked ensemble learning technique that builds a powerful classifier by combining several weaker ones. When used to enhance performance on binary classification problems, it works especially well (Sui & Ghosh, 2024). It works by passing these steps Initializing Weights, Building Weak Classifiers, Weighted Training, Weighted Voting, and Final Classification.

2.7 Model Performance Evaluation

2.7.1 Regression Performance Evaluation Metrics

Predictive models utilize various statistical methods to assess and measure the performance of the model, including the confusion matrix, accuracy, precision, recall, F1-score, mean absolute error (MAE), mean squared error (MSE), and R^2 score (confusion of determinant). These performance metrics provide valuable insights into the model's effectiveness and its ability to make accurate predictions. (M & M.N, 2015).

Scatter plot: a scatter plot is a type of plot that depicts the relationship or correlation between two numerical variables. It is made up of a series of points plotted on a two-dimensional graph, with each point reflecting a distinct combination of values from the two variables (Nguyen et al., 2020).

Mean Absolute Error (MAE): The mean difference between the regression model's values and their actual values is determined using the absolute error (MAE) measure. Utilizing this method, the average of the absolute values of the deviations between each actual value and its related prediction is calculated. MAE provides a straightforward and easy-to-understand method of determining the model's accuracy, with lower numbers indicating higher performance (Qi et al., 2020).

The Mean Absolute Error (MAE) is calculated using the following formula:

$$\mathrm{MAE} = \frac{1}{n} * \sum |\mathrm{a} - \mathrm{p}|^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. Equation\ 1$$

- Where: n is the total number of samples or observations.
- ‘a’ represents the actual values of the target variable.
- ‘p’ represents the predicted values of the target variable.

Mean Squared Error (MSE): A popular metric for assessing a prediction model's performance or estimating the output accuracy of an algorithm. It measures the average squared difference between a dataset's true values and anticipated values. The mean of the squared deviations between each predicted value and its matching true value is used to compute the MSE (Gorriz et al., 2024).

The formula for MSE is as follows:

$$\mathrm{MSE} = \frac{1}{n} * \sum |\mathrm{yi} - \hat{\mathrm{y}}_{\mathrm{i}}|^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. Equation\ 2$$

- Where: n is the number of data points in the dataset
- y_i represents the true value of i-th data point
- $\hat{y}_i$ represents the predicted value of i-th data point

R-squared (R^2) score: The coefficient of determination, or R-squared (R^2) score, is a statistical metric used to evaluate a regression model's goodness-of-fit. It shows how much of the variance in the dependent variable the model's independent variables can account for (M & M.N, 2015). The R^2 score ranges from 0 to 1, a value of 0 indicates that the model does not explain any of the variability in the dependent variable. But, a value of 1 indicates that the model perfectly predicts the dependent variable.

The R^2 score is calculated using the following formula:

$$\mathrm{R}^2 = 1 - \frac{\mathrm{res}}{\mathrm{tot}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..\ldots\ldots Equation\ 3$$

- Where: res is the sum of squares of residuals
- tot is the total sum of squares

2.7.2 Classification Performance Evaluation Metrics

Classification models employ many statistical techniques to assess the model's performance, including f1-scores, confusion matrix, accuracy, precision, and recall. These performance indicators are explained (M & M.N, 2015).

Accuracy: Accuracy is the number of correct predictions (i.e., true positives TP and true negatives TN) out of the total amount of instances evaluated (total positives P and total negatives N). The mathematical description of Accuracy is:

$$\mathrm{Accuracy} = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN})} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. Equation\ 4$$

Precision: The number of true positives, or accurately predicted out of all the positives in a class, is measured using precision. It gives a sense of how effectively a classifier categorizes each class accurately. The mathematical description of Precision is:

$$\text{Precision} = \frac{(\text{TP})}{(\text{TP} + \text{FP})} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . \mathit{Equation}\ 5$$

Recall: Recall, which is often referred to as sensitivity or true positive rate, quantifies a model's accuracy in identifying every positive instance. The ratio of true positives to the total of false negatives and true positives is calculated:

$$\text{Recall} = \frac{(\text{TP})}{(\text{TP} + \text{FN})} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . \mathit{Equation}\ 6$$

F1 scores are scores computed from the precision and the recall. It offers a single metric that strikes a balance between recall and precision. Finding the ideal balance between recall and precision is made easier with the help of the F1 score, especially in cases where the classes are unbalanced (M & M.N, 2015).

$$\text{F1 score} = \frac{(2 * (\text{precision} * \text{recall})\ \text{TP})}{(\text{TP} + (\text{precision} + \text{recall})\text{FN})} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . \mathit{Equation}\ 7$$

2.8 Related Work

Relevant literatures should be reviewed to gain a thorough understanding of the image analysis and machine learning techniques used to predict coffee quality attributes using roasted coffee images and its sensorial scores. To the best of our knowledge, only the acidity level and aroma were measured for quantifying coffee quality features from an image of roasted coffee beans.

Petch Sajjacholapunt et al., (2022) study the automatic measurement of acidity from roasted coffee beans images using efficient deep learning. They use the pH scale to measure the acidity of the roasted coffee beans. The pH scale is a degree of the acidity or basicity of a solution that starts, from 0-14, with 7 being neutral. The acidity level of the roasted coffee beans is quantitatively represented on the pH scale, and the actual acidity level is measured using a Bante (Bante920) pH/ORP meter. In particular, the work investigates how to train a model that can autonomously determine the acidity level from input images of roasted coffee beans using machine learning (Petch Sajjacholapunt et al., 2022). However, it is important to note that while image analysis alone may not be sufficient, there is a gap in exploring the integration of multimodal data, such as combining image data with additional sensory information (e.g., aroma profiles, and texture analysis). There is an opportunity to bridge this gap by incorporating additional data sources such as chemical analysis and sensorial scores. By considering the chemical composition and sensorial evaluation of coffee, it becomes feasible to predict acidity more accurately.

Caporaso et al., 2022 use spectroscopic methods to analyze coffee aroma. They demonstrate the effectiveness of hyperspectral imaging in inferring coffee aroma from roasted beans but highlight the need for further research with larger sample sizes and diverse coffee species. The study also emphasizes the importance of exploring how different brewing methods impact the final coffee product. However, sensory evaluation was not considered when correlating specific volatile chemicals with flavor qualities (Caporaso et al., 2022). While the study focused on predicting volatile compounds and sensory attributes from individual roasted beans, coffee aroma encompasses a broader range of chemical compounds. A more comprehensive analysis of these compounds could enhance our understanding of coffee aroma and improve predictive models. Overall, the study contributes valuable data for future research in coffee aroma analysis (Caporaso et al., 2022).

Esteban-Díez et al., (2004) investigate how roasted coffee samples can be used to predict the sensory attributes of espresso coffee using near-infrared spectroscopy. In addition to highlighting the shortcomings of conventional sensory analysis, the study offers a productive and non-destructive approach to quality control. The results show that NIR spectroscopy may be used to predict sensory qualities with accuracy, providing the coffee industry with a quicker and more objective method of sensory analysis, but the research was conducted on a small dataset and with no diversified data and also they don't consider the impact of various roasting profiles (level of roasting degree). The brewing method has a key impact on coffee quality assessment (Córdoba et al., 2021). The second gap is the study did not use the standardized brewing method, which is cupping.

Ferreira Lima dos Santos et al., (2020) employed computer vision and machine learning algorithms to detect coffee bean defects. The study discovered that color descriptors are crucial for classifying coffee beans based on defects (Ferreira Lima dos Santos et al., 2020). However, the study only focused on a limited number of defects and did not consider other factors that may affect coffee quality. Also, the study did not compare the performance of machine-learning algorithms to that of human experts. Further research is needed to address these limitations and expand the analysis.

Kim, (2022) employed machine learning algorithms to detect coffee bean defects. This study experimented with image processing techniques to grade coffee beans based on quality. 10,000 beans from various grades were used, along with 145 image data files. However, internal content and sensory evaluation are also necessary to accurately assess coffee quality. The dataset used for training the model had a small number of samples for certain coffee quality grades. The study highlighted altitude as a significant factor in predicting coffee quality (Kim, 2022). However, the importance of altitude varies across different geographic regions and countries. Future studies could consider examining country-specific datasets to account for such variability, and also they don't consider the effect of degree level of roasting.

A recent study conducted in collaboration with Almacafé, affiliated with Colombia's National Federation of Coffee Growers (FNC), used machine learning methods to evaluate the quality of coffee by analyzing almond and roasted coffee bean samples. With an average accuracy rate of 81% for a 10-fold stratified cross-validation, the classification results show superior accuracy using the Neural Network technique (Javier, 2020). However, the study does not consider the influence of roasting level on both beans and notes a gap in the analysis of the sensorial similarity of almonds and coffee.

Kassaye et al., (2016) utilized partial least square (PLS) regression analysis to predict the quality of specialty coffee cups based on near-infrared spectra of green coffee beans. The model was developed with Opus quant software and assessed through various parameters (Kassaye et al., 2016). However, the limited sample size used for model development and external validation may be a weakness. The authors attempted to assess raw coffee quality, but the big issue that affects the quality of coffee is the degree level of roasting coffee so it was not considered.

2.9 Summary

By making use of earlier studies in the areas of image analysis and machine learning, the study "Automated Coffee Quality Prediction: A Study on Predicting Coffee Quality Attributes from Roasted Coffee Images" advances the subject. Previous studies have shown that these methods are effective in several areas, such as agricultural product analysis and the evaluation of food quality.

The study intends to provide a strong framework for predicting coffee quality attributes by evaluating images of roasted coffee, utilizing this body of knowledge. By offering an automated, impartial, and standardized technique for evaluating coffee quality, this novel approach has the potential to completely transform the coffee business, which will benefit both growers and consumers.

Table 1: Summary of related work

Authors	Title	Method	Limitation
Javier, (2020)	Machine Learning for Cup Coffee Quality Prediction from Green and Roasted Coffee Beans Features	They used machine learning methods to evaluate the quality of coffee	The study does not consider the impact of roasting degree on both beans and identifies a hole in the research on the sensory similarities between coffee and almonds. They also don’t predict the coffee quality attributes.
Kim, (2022)	Coffee Beans Quality Prediction Using Machine Learning	They Used Random forest and image analysis methods to evaluate the quality of coffee	The dataset used for training the model had a small number of samples for certain coffee quality grades. The performance is too low, F1 score of 61.7%
Ferreira Lima dos Santos et al., (2020)	Quality assessment of coffee beans through computer vision and machine learning algorithms	They Used SVM, CNN, and Random Forest	The study only focused on a limited number of defects and did not consider other factors that may affect coffee quality. They classify coffee beans based on shape and color features extracted from images, they don’t consider sensorial score and texture.

Caporaso et al., (2022)	Prediction of coffee aroma from single roasted coffee beans by hyperspectral imaging	They used Partial least squares (PLS) regression	They don't consider the geographical variability in the volatile compounds of coffee beans. It varies from place to place.
Tolessa et al., (2016)	Prediction of specialty coffee cup quality based on near-infrared spectra of green coffee beans	They Used a partial least square regression analysis	The coffee quality defects like Quakers have the same spectral pattern at this time of using the hyperspectral method for an imaging system, so it is not good for coffee that has defects.
Petch Sajjacholapunt et al., (2022)	Automatic measurement of acidity from roasted coffee beans images using efficient deep learning	This study uses the pH scale to measure the acidity of the roasted coffee beans.	The data they used had a small dataset size and diversity, The dataset used in the study consists of images from 12 bags of coffee from four different brands and three roasted levels. Also, it is not enough to predict acidity using a coffee image without its chemical composition and sensorial score.

CHAPTER THREE

System Design

3.1 Overview

This study proposes an automated coffee quality prediction system that consists of four main steps: data collection, data preprocessing, feature extraction, and model development. The system aims to predict various coffee quality attributes based on diverse and representative coffee samples (roasted coffees). In the data collection phase, a comprehensive dataset of coffee samples with its sensorial score is gathered from different sources, covering various coffee varieties, and processing methods. During data preprocessing, techniques such as resizing, noise removal, segmentation, and color conversion are applied to enhance the quality of the coffee samples and prepare them for analysis. Generally, in this chapter, we should state a detailed description of the coffee quality attribute prediction system based on the roasted coffee images. Moreover, the material and methods for achieving the general objective and specific objective are also stated.

3.2 System Architecture

The proposed automated coffee quality attribute prediction system follows a system architecture that consists of four main tasks: image acquisition, image preprocessing, feature extraction, and coffee quality attribute prediction. Figure 5 illustrates the overall procedure employed to recognize the input coffee sample and predict its quality attributes accurately. The system architecture for the proposed automated coffee quality prediction research involves two key components: preprocessing for roasted coffee images and preprocessing for sensorial scores. The roasted coffee image preprocessing includes resizing, noise removal, segmentation, and histogram equalization to enhance image quality. The sensorial score preprocessing focuses on data cleaning, such as handling missing files and removing duplicates. After these two components are completed, the system proceeds to concatenate the preprocessed roasted coffee images and sensorial scores. At this stage, a common preprocessing step is applied, which includes techniques like data augmentation and normalization.

After preprocessing, the data undergoes feature extraction, which includes label encoding, Recursive Feature Elimination (RFE), and polynomial features to select relevant features. The preprocessed data is then split into training and testing sets for model development and evaluation.

In the case of applying random forest regression, linear regression, and other regressions, the training set is used to train the model by fitting the data to the algorithm. This allows the model to learn the relationships between the input features and the corresponding coffee quality attributes. The testing set is then used to evaluate the performance of the trained model by comparing its predictions against the actual values of the coffee quality attributes.

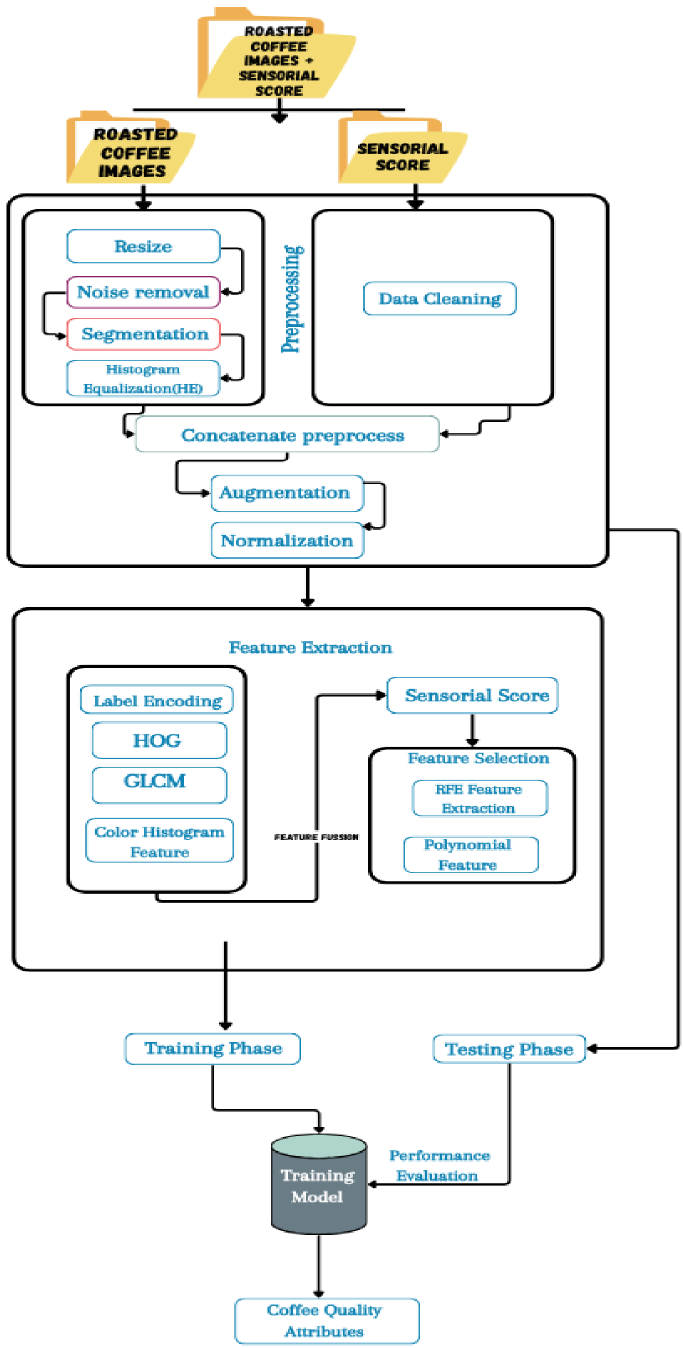


Figure 5: The proposed System Architecture for the coffee quality prediction model

3.3 Image Acquisition and Sensorial Evaluation

3.3.1 Data Collection and Preparation

The first activity of model development is data collection. To acquire roasted coffee images, the process begins by setting up the roasted coffee beans. The next step involves capturing the images using a camera positioned at three different distances: 100mm, 155mm, and 210mm from the roasted coffee. To avoid hand shaking while taking the image, we used a camera stand or a tripod. After testing each distance, we used an average of the nearest and furthest distance between the roasted coffee and the camera. Then put with their corresponding sensorial score. These selected images are then saved in PNG format and labeled with their corresponding sensorial scores. Each roasted coffee image has corresponding sensorial score. The sensorial score of each sample is taken by Ethiopian Coffee and Tea Authority Coffee Quality Inspection and Certification Center (ECTACQICC) cuppers.

We used a smartphone with a 50-megapixel camera for image acquisition 8288 x 6032 resolution images collected from Yirgacheffe, Limu, Harrar, Sidamo, Jimma, Kaffa, Nekemt, Wenbera, and Zegae coffees in Ethiopian Coffee and Tea Authority Coffee Quality Inspection and Certification Center (ECTACQICC). The coffee beans used for image acquisition undergo roasting at different time intervals, specifically 2, 4, 6, 8, 10, and 12 minutes. This time variation allows for capturing the changes in color and the associated differences in sensorial scores. We have a total of 675 dataset images and their corresponding sensorial score. To reduce noise due to shaking hands while taking the image, we used a camera stand. In this research, we also captured the images by making human sensorial evaluation scores. We captured the images by roasting coffee beans at different time intervals at a constant voltage level (210°C). We used augmentation techniques, and the augmented data is 2700.

In this study, six distinct cuppers used the cupping technique to determine the sensory scores. To evaluate the sensory attributes and characteristics of coffee samples, the coffee industry frequently uses cupping, as a standard assessment procedure (Henrique & Paula, 2023). The cuppers assess the roasted coffee's aroma, flavor, aftertaste, acidity, body, and overall quality attributes. Smelling both wet and dry coffee grounds, slurping the brewed coffee to evaluate flavor, and then rating the coffee according to the cupper’s sensory measurements. This cupping method yields sensorial scores that offer important insights into the sensory profile and qualities of the roasted coffee, enabling a thorough assessment and comparison of different samples. It is very difficult to get large datasets in coffee research institutions, coffee farmers' union, coffee exporters, and coffee quality controllers' labs, because the data is not recorded correspondingly to the coffee samples. The data was collected from Yirgacheffe, Limu, Harrar, Sidamo, Jimma, Kaffa, Lekemt, Wenbera, and Zegae coffee regions. This study passes through preprocessing, segmentation, image enhancement (histogram equalization), augmentation, feature extraction, training model, testing and finally predicting the coffee quality attributes. Then we organize the roasted coffee images into three groups (light, medium, and dark).

Table 2: Total samples of roasted coffee images and its sensorial score

Region	**Number of images + sensorial score**
Limu	70
Harrar	76
Sidamo	86
Yirgacheffe I, II, III	141
Jimma	99
Kaffa	85
Nekemt	42
Wenbera	26
Zegaie	55

Figure 6: Samples of roasted coffee images

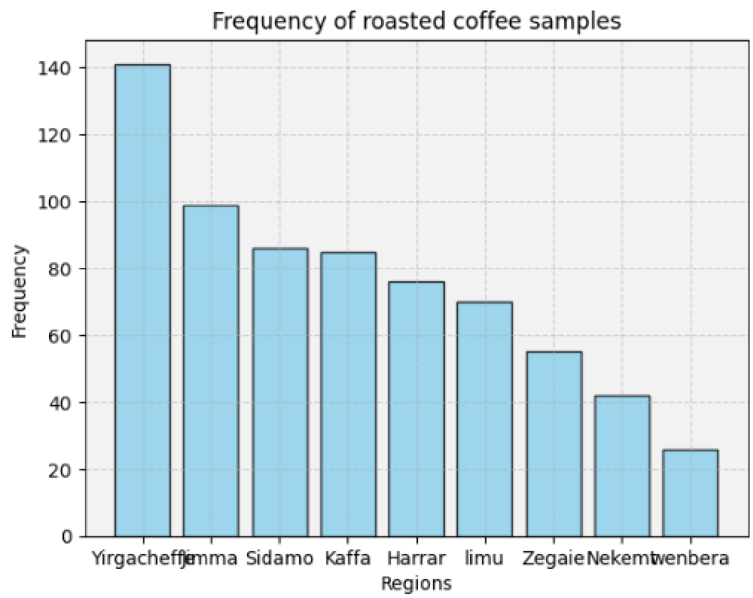


Figure 7: Number of samples (number of coffee beans) taken from each region

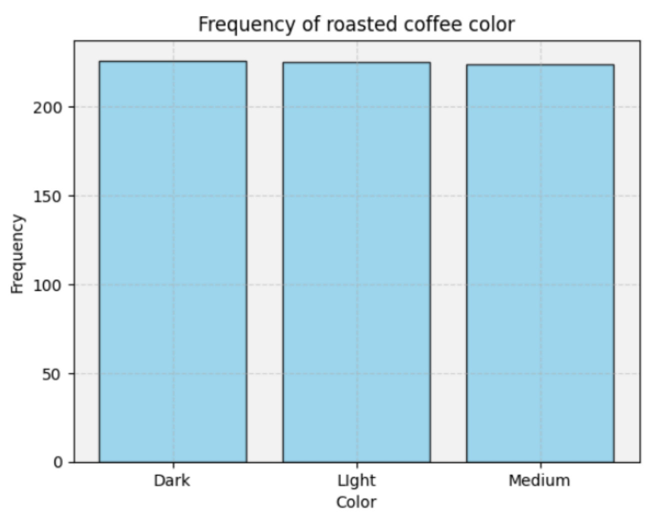


Figure 8: Roasted coffee color categories

As we can see in Figure 8 the color of the roasted coffees is categorized in to three colors (light, medium, and dark). Diversifying the color of roasted coffee image help the training model to capture the relationships of different color combinations with its corresponding sensorial score.

```
data.head(n=5)
```

image	Region	Aroma	Flavor	Aftertaste	Acidity	Body	Balance	Uniformity	Sweetness	Cupper.Points	Clean.Cup	Total.Cup.Points	M
/content/drive/MyDrive/SAMP/roasted coffee-bea...	Yirgacheffe	8.67	8.83	8.67	8.75	8.50	8.42	10.0	10.0	8.75	10.00	99.26	
/content/drive/MyDrive/SAMP/roasted coffee-bea...	Sidamo	7.58	7.50	7.42	7.58	7.50	7.50	10.0	10.0	7.67	10.00	90.33	
/content/drive/MyDrive/SAMP/roasted coffee-bea...	Jimma	7.33	7.17	7.17	7.25	7.42	7.33	10.0	10.0	7.25	10.00	88.25	
/content/drive/MyDrive/SAMP/roasted coffee-bea...	Kaffa	7.17	7.33	7.08	7.58	7.33	7.08	10.0	10.0	7.00	9.33	87.07	
/content/drive/MyDrive/SAMP/roasted coffee-bea...	Zegaie	7.25	7.08	7.00	7.17	7.17	7.08	10.0	10.0	7.00	10.00	87.00	

Figure 9: Dataset of the proposed system

3.3.2 Data Description

The purpose of the data description is to equip future researchers with the necessary knowledge to analyze and effectively utilize datasets while providing a comprehensive explanation of the collected attributes. It serves as a tool for understanding the characteristics and patterns exhibited by the data. In this study, the relevant data was obtained from the Ethiopian Coffee and Tea Authority Coffee Quality Inspection and Certification Center (ECTACQICC) and Kerch Anshe Trading PLC. The collected attributes include Image, Region, Aroma, Flavor, Acidity, Body, Sweetness, Balance, Aftertaste, Uniformity, Cupper Points, Clean Cup, Total Cup Points, and Moisture. Table 6 provides a comprehensive overview of the attributes present in our dataset, along with their respective descriptions and data types.

Table 3: The data type representation of each attribute

No	Attribute Name	Description	Data Types
1	Image	Image of roasted coffee	object
2	Region	The region of coffee grown	Categorical
3	Aroma	The aroma value of that roasted coffee	Numeric
4	Flavor	The Flavor value of that roasted coffee	Numeric
5	Acidity	The acidity value of that roasted coffee	Numeric
6	Body	The body value of that roasted coffee	Numeric
7	Sweetness	The sweet value of that roasted coffee	Numeric

8	Balance	The balance value of that roasted coffee	Numeric
9	Aftertaste	The aftertaste value of that roasted coffee	Numeric
10	Uniformity	The aroma value of that roasted coffee	Numeric
11	Cupper. Points	The Cuppers point value	Numeric
12	Clean. Cup	The cleanness value sampled roasted coffee	Numeric
13	Total.Cup.Points	The total cup point of the roasted coffee	Numeric
14	Moisture	The moisture content of coffee	Numeric
15	Color	The color category of roasted coffee	Categorical

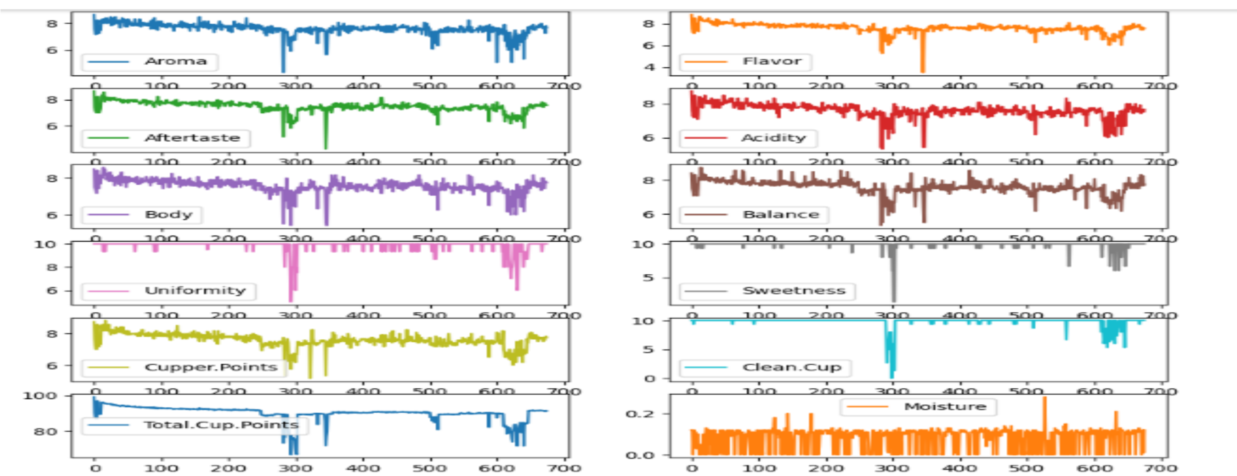


Figure 10: The data range representation of each attribute

Additionally, the sensorial scores are obtained using the cupping method, which is a standardized brewing technique for evaluating the sensory attributes of coffee According to Ethiopian Coffee and Tea Authority Coffee Quality Inspection and Certification Center protocol and guidelines for working a sensorial score, each coffee quality attributes have its way of giving the sensorial score. The sensorial score can be done in different point scales like 10 and 15 point scale. After that, by using the total cup point cuppers give the grade for each coffee sample. In this study, the cuppers used a 10-point scale. This method ensures consistency and reliability in the sensorial evaluation process.

Table 4: Sensorial score guideline of some coffee quality attributes (Training manual for coffee cuppers in Ethiopia, 2004)

Aroma		**Flavor**		**Acidity**		**Body**		**Clean cup**		**Grading**	
Quality	**Points**	**Quality**	**Points**	**Intensity**	**Points**	**Quality**	**Points**	**Quality**	**Points**	**Range**	**Grade**
Clean	10	Good	10	Pointed	10	Full	10	Clean	10	> =85	Excellent(Grade 1)
F. Clean	8	Fairly good	8	Medium	8	M. full	8	F. clean	8	75-84	V.good(Grade 2)
Trace	6	Average	6	M.pointed	6	Medium	6	1 CD	6	63-74	Good (Grade3)
Light	4	Fair	4	Light	4	Light	4	2 CD	4	47-62	Moderate (Grade 4)
Moderate	2	Commonish	2	Lacking	2	Thin	2	3 CD	2	31-46	Poor(Grade 5)
Strong	1	Not detected	1	Not detected	1	Not detected	1	>3 CD	1	UG(p) =15-30	Ungraded (p)
										UG(np) =15-30	Ungraded (np)

UG (p) refers to ungraded defects in coffee grading that are not individually counted. The range of 15-30 indicates that the coffee sample can have a maximum of 15 to 30 ungraded defects. UG (np) is a similar category where defects are not specified or counted individually, and "np" can represent "non-qualifying" or "non-specific" defects. However, the exact meaning of "np" depends on the grading standards used (Training manual for coffee cuppers in Ethiopia, 2004). Analyzing sensorial scores of coffee samples helps to evaluate quality attributes. By studying these scores, we can identify relationships and patterns between the coffee quality attributes. To see the relationship of coffee attributes we use person correlation (for detail see Figure 11).

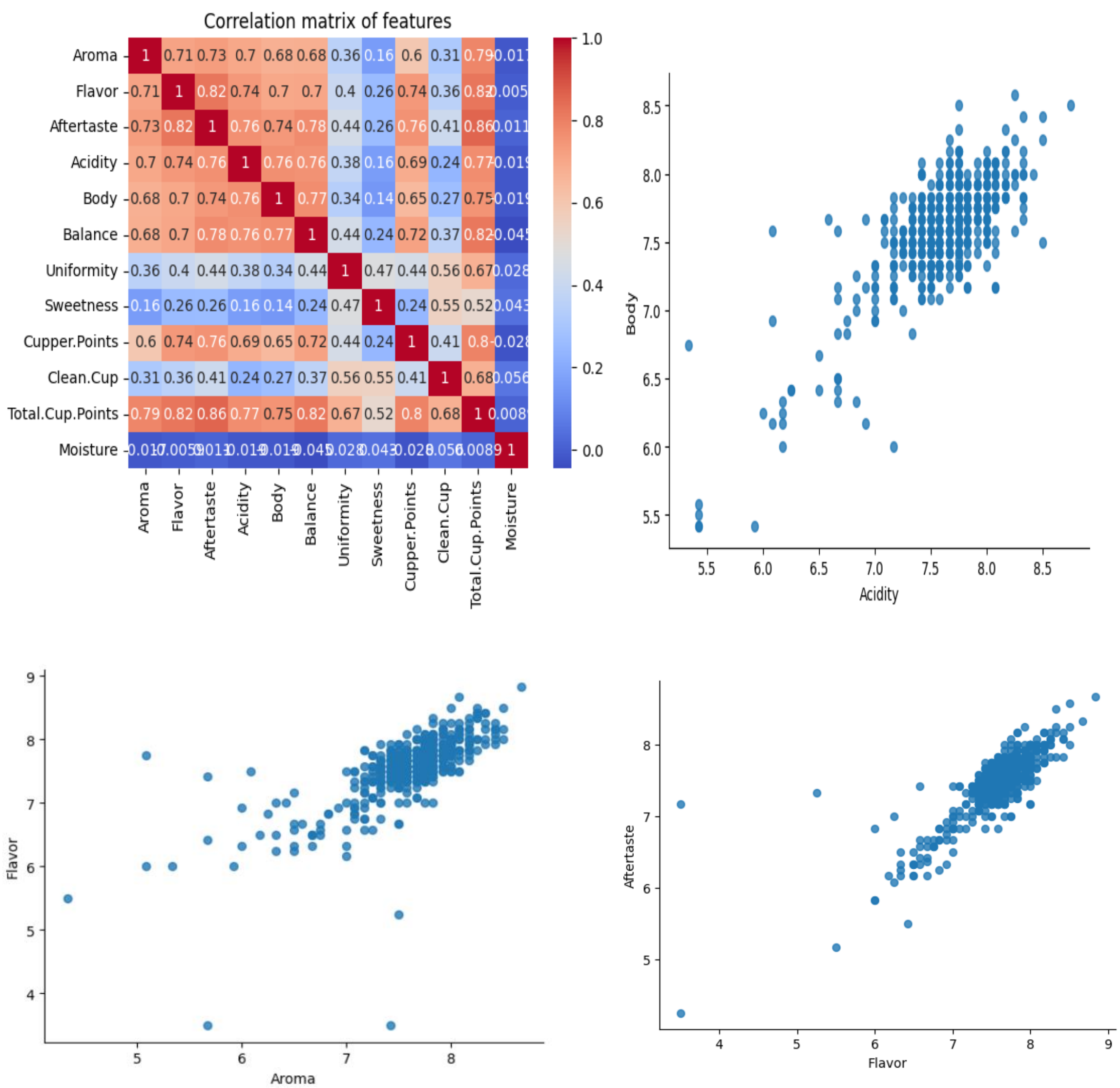


Figure 11: Co-relationship between coffee qualities attributes

3.4 Image preprocessing

3.4.1 Image Resizing

The collected images have different sizes so it must be resized. To resize the image we use an image resizing technique. In this study, we tested different roasted coffee image sizes such as 124x124, 256x256, and 512x512 pixels. We have used the OpenCV Python library. The technique used to handle aspect ratio during image resizing is by utilizing the resize () function from the PIL (Python Imaging Library) package, which automatically maintains the original aspect ratio of the image. To select the good one we considered the performance they achieve and computational time that has been used. Based on those considerations we have achieved better performance when testing the model in 124x124 roasted coffee image size.

```
import cv2
from PIL import Image
resized_images = []
output_file = 'preprocessed_data.csv'
image_size = (124, 124)
for image_path in image_paths:
    image = load image using cv2.imread and convert color space
to RGB
    resized_image = resize image using cv2.resize to image_size
    resized_image_pil = convert resized_image to PIL Image
format
    add resized_image_pil to resized_images
resized_images_array = convert resized_images to numpy array
for i, resized_image in enumerate(resized_images):
    save resized_image to file with a unique name or index i
```

Figure 12: Image resizing pseudocode

3.4.2 Noise Removal

Noise removal is significant because it removes the image's noise content, which is necessary. After all, noise causes inaccuracies in image processing techniques. Noise reduction determines the quality of the entire image processing cycle (Mohamed, 2015). It is unnecessary information that degrades the visual quality of an image (Tania & Rowaida, 2016). In this study, we apply a Gaussian and median filter for noise removal. The median filter gives high performance with less computational time. This is how the median filter removes noises in roasted coffee images, the image is loaded and resized to a desired size using the PIL library. Then, the adaptive histogram equalization function is called to perform noise removal. The median filter is applied to the image to reduce noise, and then adaptive histogram equalization is performed to enhance the image contrast. The basic process of how a median filter works is shown in Figure 13.

```
import cv2
from PIL import Image
from scipy.ndimage import median_filter
filtered_images = []
```

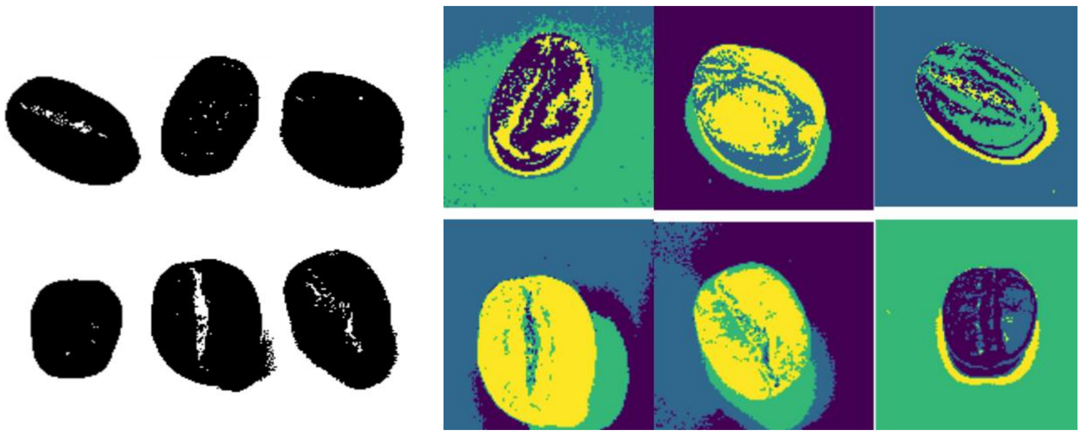
```
output_file = 'preprocessed_data.csv'
median_filter_size = 3
for image_path in image_paths:
    image = load image using cv2.imread and convert color space
to RGB
    image_pil = convert image to PIL Image format
    image_array = convert image_pil to numpy array
    filtered_image_array = apply median_filter to image_array
with size=median_filter_size
    filtered_image_pil = convert filtered_image_array to PIL
Image format
    add filtered_image_pil to filtered_images
filtered_images_array = convert filtered_images to numpy array
for i, filtered_image in enumerate(filtered_images):
    save filtered_image to file with a unique name or index i
```

Figure 13: Pseudocode of median filter

3.4.3 Segmentation

Segmentation is used to split digital images into a group of pixels, each representing homogeneous regions with similar features like shape, pattern, texture, and color. It is the process of dividing digital images into meaningful regions depending on certain criteria to extract the area in which the one is interested (Song & Yan, 2018). Image segmentation makes the feature extraction process of an image easier to understand by separating each feature of the image. In this study, we have tested different segmentation techniques; those are K-means (Figure 14 B) and threshold (Figure 14 A) segmentation techniques. However, the threshold segmentation techniques give better results with good computational time. Once the equalized image is obtained, thresholding is applied to segment the image. Using Otsu's method, the threshold_otsu function from skimage.filters is used to calculate an optimal threshold value. The resulting threshold is used to create a binary mask where pixel values above the threshold are considered foreground, and pixel values below the threshold are considered background.

The threshold_otsu function calculates the threshold value based on the histogram of the input image. It determines the threshold that minimizes the intra-class variance of the foreground and background pixels, effectively separating them. The calculated threshold is then used to create a binary segmented image. Finally, the segmented image is converted back to a PIL Image



A). Threshold segmentation technique B). K-means segmentation technique

Figure 14: Segmentation techniques

3.4.4 Color Conversion

Color space conversion or color transformation, is the process of converting the representation of colors from one color space to another. Color spaces define the way colors are represented numerically, allowing for consistent and standardized color communication and manipulation (Bi & Cao, 2021). In this study, we test an RGB image to HSV, RGB to CMYK Conversion, and BGR (Blue-Green-Red) to RGB (Red-Green-Blue) color conversion to enhance the image. In the end, the BGR to RGB color conversion gets good computational time and confusion determinant value.

A) BGR to RGB B) RGB to CMYK C) RGB to HSV

Figure 15: Color conversion

3.4.5 Augmentation

Data augmentation is a machine learning technique that applies several modifications to improve the quantity and diversity of training data. By introducing random modifications such as rotation, flipping, and scaling, augmentation helps improve model performance and generalization by exposing it to a wider range of variations and reducing overfitting (Alzubaidi et al., 2021). Data augmentation is typically applied to the input data (roasted coffee images) rather than the output data (attribute scores). Augmenting the attribute scores would involve modifying or altering the target labels, which is not a common practice in data augmentation. Attribute scores are typically fixed values that represent specific characteristics or qualities of the images. Augmenting the attribute scores could introduce noise or misalignment between the augmented images and their corresponding labels, making it harder for the model to learn accurately. Thus, we apply augmentation to the images, the attribute scores remain unchanged and are used as reference labels during the training process.

Table 5: Augmentation parameters

Parameters	Best parameters	Explanation
rotation_range	20	Rotate images randomly up to 20 degrees
width_shift_range	0.1	Shift the width of images by 0.1
height_shift_range	0.1	Shift the height of images by 0.1
horizontal_flip	True	Flip images horizontally
vertical_flip	False	Do not flip images vertically
fill_mode	'nearest'	Fill any empty pixels with the nearest available pixel

This augmentation technique (table 5) help to enhance the model's performance include random rotation of images up to 20 degrees, shifting the width and height of images by 0.1 to introduce positional variations, horizontal flipping to increase data diversity and improve left-right orientation recognition, and the use of 'nearest' fill mode to preserve image appearance and prevent artificial patterns or artifacts. These techniques collectively contribute to the model's robustness and ability to handle different angles, positions, and orientations, resulting in improved performance and generalization in real-world scenarios.

3.4.6 Normalization

Normalization is a data preprocessing technique used to rescale numeric data to a common scale or range. The goal of normalization is to bring different features or variables to a similar scale so that they can be compared and analyzed more effectively. There are different normalization methods, but two commonly used techniques are Min-Max scaling and Z-score normalization (Kalyani A. Sankpal, 2020). In this study we used the Min-Max scaling normalization technique because it is good for Intuitive interpretation, Preserves relationships and outliers, is Suitable for algorithms with specific input range requirements, and is Non-sensitive to small changes in data distribution.

```
from sklearn.preprocessing import MinMaxScaler
image_paths = convert data['image'] to list
attribute_scores = convert selected columns of data to numpy
array
output_file = 'preprocessed_data.csv'
preprocessed_attributes = []
scaler = MinMaxScaler()
preprocessed_attributes = apply Min-Max scaling to
attribute_scores using the scaler
preprocessed_df = create DataFrame from preprocessed_attributes
add columns with attribute names to preprocessed_df and assign
preprocessed_attributes as values
preprocessed_df[['Aroma', 'Flavor', 'Aftertaste', 'Acidity',
'Body', 'Balance', 'Uniformity', 'Sweetness', 'Cupper.Points',
'Clean.Cup', 'Total.Cup.Points','Moisture']] =
preprocessed_attributes
save preprocessed_df to CSV file named output_file without index
```

Figure 16: pseudocode of minimax normalization

3.5 Feature Extraction

Feature extraction is a technique for extracting color, texture, shape, and other descriptions from roasted coffee images and their sensorial score. Feature extraction is the process of involving the image feature in identifiable and distinguishable attributes in numerical value (Kabbai et al., 2019). In this study, the HOG, GLCM, and color histogram feature descriptors are considered.

3.5.1 Feature Extraction using HOG

Histogram of Oriented Gradients (HOG) descriptor is a well-liked technique for feature extraction in local areas for applications like object detection and recognition (Yassin Kortli et al., 2018). The HOG descriptor works by calculating the gradient image using filters to extract edge gradients and orientations. It then divides the local image into cells and creates histograms of gradient orientations within each cell. These histograms capture the distribution of edge orientations in the image, providing a representation of the local structure and texture (Alhindi et al., 2018). The HOG descriptor is known for its robustness to variations in lighting and background clutter, making it a valuable tool in computer vision applications (Yassin Kortli et al., 2018).

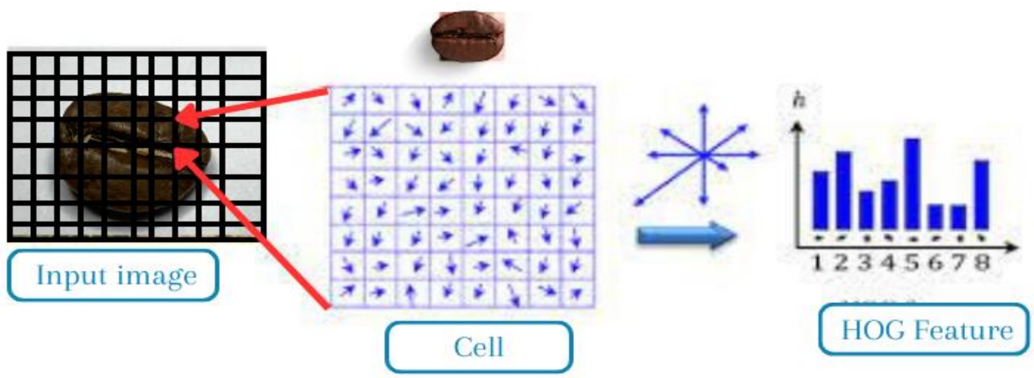


Figure 17: HOG Feature extraction process (Yassin Kortli et al., 2018).

The Histogram of Oriented Gradients (HOG) feature extraction process involves several key steps to capture the local structure and texture information from the roasted coffee image:

- Gradient Calculation: The first step in HOG feature extraction is to calculate the gradient image of the input image. This is typically done using gradient filters such as Sobel, Prewitt, or Laplacian filters. These filters help in identifying the edges and gradients in the image.
- Cell Division: The image is divided into small, non-overlapping cells. Each cell typically covers a small spatial area of the image.
- Orientation Histograms: Within each cell, orientation histograms are computed based on the gradient orientations of the pixels. The gradient magnitudes and orientations are used to cast a vote in the corresponding orientation bin of the histogram.

- Block Normalization: Normalization is applied to sets of cells referred to as blocks to increase the features' resistance to variations in light and contrast. The impacts of local variances are lessened with the aid of this normalization.
- Descriptor Formation: The final HOG descriptor is formed by concatenating the normalized histograms from all the cells within each block. This results in a feature vector that represents the distribution of gradient orientations in the image.

In this study in addition of HOG, the GLCM feature extraction technique was applied to extract the texture of roasted coffee images.

Table 6: HOG Parameters

Parameters	**Best Parameters**	**Explanation**
orientations	9	Number of gradient orientations
pixels_per_cell	(8, 8)	Size of each cell
cells_per_block	(2, 2)	Number of cells in each block
block_norm	'L2-Hys'	Block normalization method
transform_sqrt	False	Making apply power-law compression false

3.5.2 Grey Level Co-Occurrence Matrix (GLCM)

GLCM stands for Gray-Level Co-occurrence Matrix. It is a statistical method used in image processing to analyze the spatial relationship between pixels in an image based on their gray levels (P.S & V.S, 2016). GLCM quantifies how often different combinations of pixel intensity values occur in a specified spatial relationship within an image. By calculating GLCM, texture features such as contrast, correlation, dissimilarity, energy, and homogeneity can be extracted to characterize the texture properties of an image. The construction of the GLCM involves several steps. First, the image undergoes pre-processing techniques such as connected regions and morphological operations. These techniques enhance the image quality, ensuring optimal analysis results.

Once pre-processing is complete, the GLCM is formed as a matrix with dimensions corresponding to the number of gray levels present in the image. It captures the relationship between a reference pixel and its neighboring pixel in different orientations (0°, 45°, 90°, and 135°) within the image. Each element in the GLCM represents the frequency of occurrence of specific pairs of pixel intensity values of the reference pixel and its neighboring pixel at defined spatial relationships. These relationships are determined by the orientation and distance between the reference pixel and its neighboring pixel. The construction of the GLCM is crucial for extracting essential texture features such as Contrast, Correlation, Dissimilarity, Energy, and Homogeneity. These features play a significant role in image classification and analysis tasks. By capturing the statistical properties of pixel relationships, the GLCM provides valuable information for characterizing and distinguishing different textures within roasted coffee images.

3.6 Data Set Splitting

In this study, we have expanded our dataset through data augmentation techniques, resulting in a total of 2,700 images of roasted coffee. Each image is labeled with its corresponding sensorial score, which serves as the ground truth for our analysis. To develop a predictive model for assessing coffee quality attributes, we employed the Random Forest (RF) algorithm. The dataset was partitioned into training and testing sets, with 80% (2,160 images) used for training the RF model, linear regression, decision tree, and logistic regression the remaining 20% (540 images) reserved for evaluating the model's performance.

Training the RF model with a larger dataset allows for the extraction of more relevant features during the training process, leading to potentially improved predictive capabilities. Following the model development phase, we assessed the performance of the RF model using the testing dataset of 540 images. By comparing the predicted sensorial scores with the actual scores, we evaluated the accuracy and effectiveness of our model in predicting the quality attributes of roasted coffee based on the augmented image data.

3.7 Fusion of Image and Sensorial Scores

The fusion of multiple modalities in predicting coffee quality attributes from roasted coffee images involves the combination of different types of information or data sources to gain a comprehensive understanding. In this study, this fusion process consists of two main modalities: the image modality and the attribute modality. The image modality focuses on extracting visual information from the roasted coffee images. These visual features provide valuable insights into the physical characteristics of the coffee beans or grounds, enabling a detailed analysis of their visual properties (color, texture, shape). First of all the color histogram features are extracted from the preprocessed data by selecting the relevant columns. These features are then converted to a 2D array for further processing. The sensorial scores are encoded using a label encoder (LabelEncoder) to transform them into numerical values, facilitating their usage in regression models. The fusion of modalities would involve combining the visual features (HOG features, color histogram features, and texture) with the sensory attribute scores. This fusion is implicitly achieved by training a Random Forest Regression model using the HOG features (X_train) as input and the encoded sensory attribute scores (y_train) as the target variable.

In a linear regression model, a fusion process is applied to combine the color histogram features with sensorial scores. This fusion allows for the integration of both modalities, enhancing the model's understanding and predictive capabilities. Initially, the color histogram features are extracted from the preprocessed data and transformed into a 2D array. Then, the Histogram of Oriented Gradients (HOG) features are computed for each color histogram using the skimage library. These HOG features are stored in a list and subsequently converted into a data frame. Simultaneously, the sensorial scores are preprocessed and stored as the target attributes. The fusion occurs by merging the color histogram features and the HOG features into a single data frame using feature concatenation. The merged data is then split into training and testing sets to facilitate model training and evaluation. Linear regression models, including one with recursive feature elimination (RFE) and another with polynomial features and scaling, are applied to the fused data. Finally, model evaluation metrics such as mean absolute error (MAE) and mean squared error (MSE) are calculated to assess the performance of the fused data models.

3.8 Linear Regression

Linear regression is a statistical modeling technique used to establish a relationship between a dependent variable (also known as the target variable or response variable) and one or more independent variables (also known as predictor variables or features). The goal of linear regression is to find the best-fitting linear equation that represents the relationship between the variables (Ashiru & Oladele, 2023).

In the context of predicting coffee quality attributes, linear regression can be applied to estimate the values of attributes such as Aroma, Flavor, Aftertaste, Acidity, Body, Balance, Uniformity, Sweetness, Cupper.Points, and Clean.Cup. These attributes serve as the dependent variables, while specific features or characteristics of the coffee (e.g., color histogram) act as the independent variables. The color histogram has the color, texture, and shape of the roasted coffee image. In the context of predicting coffee quality attributes, linear regression uses a color histogram feature to predict attributes like Aroma, Flavor, Acidity, and others. The process involves preparing the data, splitting it into training and testing sets, training the model with the training data, evaluating its performance using metrics like MSE, MAE, and R^2 score, and making predictions on new data.

3.9 Random Forest

Random Forest is a potent and adaptable machine learning algorithm within the ensemble learning area. It is known for its high accuracy, robustness, and ability to handle complex data sets. As we can see in figure 17 random forest works by constructing a multitude of decision trees during the training phase and outputs the mode of the classes (classification) or the average prediction (regression) of the individual trees (Naghibi et al., 2017). One of the key features of Random Forest is its ability to reduce overfitting, a common issue in machine learning, by averaging the predictions of multiple decision trees. Each tree in the Random Forest is trained on a random subset of the training data and a random subset of features, adding an element of randomness that helps improve generalization and robustness.

Random Forest can handle large amounts of data with high dimensionality, categorical features, and missing values without the need for extensive data preprocessing. It is also capable of providing feature importance scores, which can help in understanding the underlying patterns in the data. It is a powerful ensemble learning algorithm that operates by constructing a multitude of decision trees during the training phase and combining their predictions to make accurate and robust predictions. To predict the attributes of coffee using Random Forest, the algorithm goes through two main phases: training and prediction. During the training phase, Random Forest randomly selects a subset of the training data. It creates individual decision trees using a random selection of color, shape, and texture features. These decision trees are built by making splits in the data based on these features. The goal is to minimize the difference between predicted and actual values, such as the sensorial scores associated with roasted coffee. The splitting process continues until a stopping criterion is met, like reaching a maximum tree depth or a minimum number of samples in a leaf node. Recursive Feature Elimination (RFE) is employed to select the most important features. RFE ranks and eliminates less significant features, resulting in a subset of selected features. RFE ranks and eliminates less significant features, resulting in a subset of selected features. The model is trained using only these selected features, and predictions are made on the testing set. Finally, evaluation metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE) are calculated to assess the model's performance.

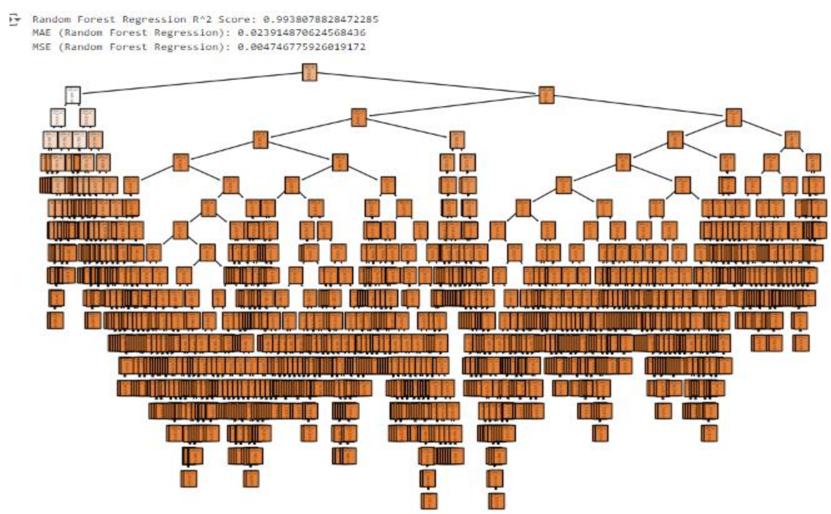


Figure 18: Decision trees in Random forest coffee quality attribute prediction model

3.10 Coffee Grading

In this research, we also work on coffee grading by taking the result of the coffee quality attribute prediction model. By summing up those results and taking them as a total cup point we develop a classifier model that grades each coffee. According to Ethiopian commodity exchange (ECX) coffee grading protocol, there are a threshold value to set the coffee grading. To do a classification we employed different machine learning algorithms such as Ada-Boost, XG-Boost, RF, and decision trees.

Table 7: Threshold values of coffee grading (Training manual for coffee cuppers in Ethiopia, 2004)

Threshold values	
Range	**Grade**
> =85	Excellent(Grade 1)
75-84	Very good(Grade 2)
63-74	Good (Grade3)
47-62	Moderate (Grade 4)
31-46	Poor(Grade 5)

The architecture of the classifier model looks like this:

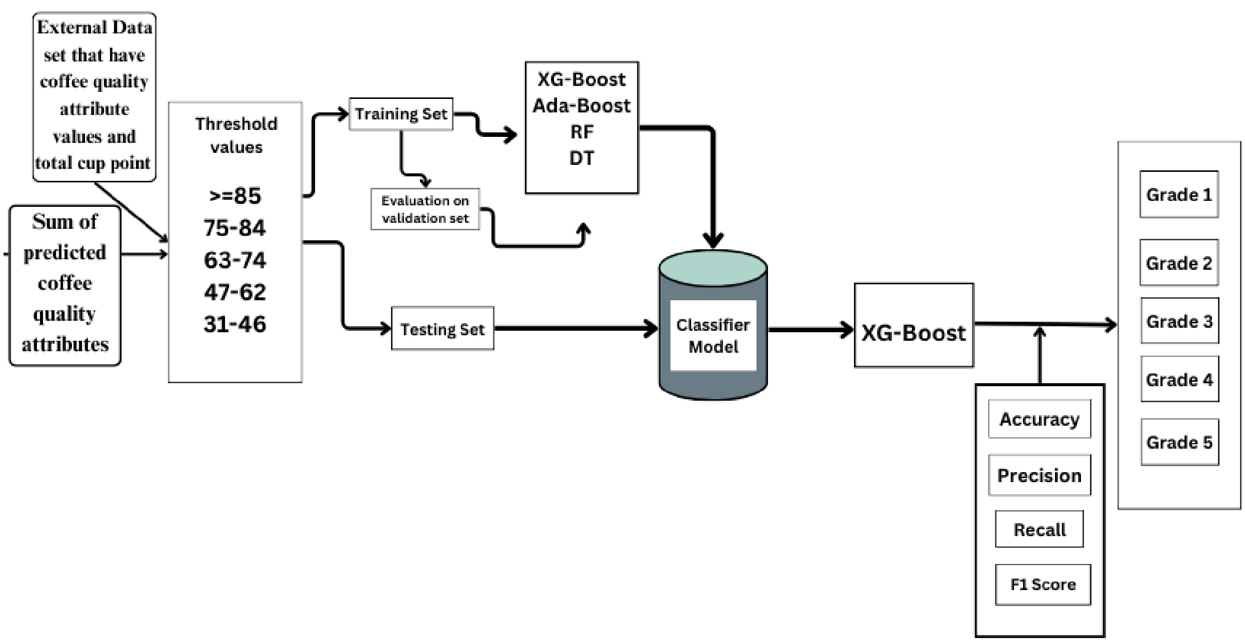


Figure 19: The architecture of coffee grading or classifier

3.10.1 Classifier Models

3.10.1.1 XG-Boost (Extreme Gradient Boosting)

XG-Boost is a sophisticated machine-learning method that belongs to the ensemble learning category. It is designed to efficiently handle regression and classification tasks and has grown in prominence as a result of its outstanding performance in different data science competitions. XG-Boost combines gradient boosting and regularization techniques to produce a highly accurate predictive model (Budholiya et al., 2022). We train the XG-Boost model and it gives better results than the other classifier.

3.10.1.2 Ada-Boost (Adaptive Boosting)

Ada-Boost is an ensemble learning technique that combines several weak classifiers to form a stronger classifier. The approach iteratively trains a series of weak classifiers using weighted representations of the training data. We also train this model by adding the predicted coffee quality attribute values and it give less performance than the XG-Boost classifier model.

3.11 Summary

In this chapter, we have explored the architecture of a predictive model for coffee quality attributes. The architecture encompasses various components, including preprocessing, feature extraction, training and testing stages. To extract meaningful features, we employed techniques such as Histogram of Oriented Gradients (HOG), Gray-Level Co-occurrence Matrix (GLCM), and color descriptors. These methods allowed us to capture relevant information from the coffee images. In addition, we applied feature selection techniques such as Recursive Feature Elimination (RFE), polynomial featuring, and scaling. RFE helped us identify the most important features, while polynomial features enabled us to capture non-linear relationships. Scaling ensured that the features were on a comparable scale, avoiding potential bias in the model. We used different regression models to train our model like random forest, linear regression, decision tree, and logistic regression. We also make a coffee grading by taking the result of the regression model and by sum up those coffee quality attribute values. We take the sum value as a total cup point, in the base of the ECX coffee grading protocol, we develop a model a classifier model that can classify based on the given threshold value of ECX.

CHAPTER FOUR

4. EXPERIMENTAL RESULT AND DISCUSSION

4.1 Overview

This chapter includes a discussion of the experimental results of the proposed model. In this section, the results of the developed model are explained and presented based on the training and test results of the proposed model on the given parameter considerations. Experimental evaluation approves the realization of the proposed model architecture. This chapter also includes the results of each research question.

4.2 Experimental Design

4.2.1 Development Environment

In this study, we have used tools such as Canva design and lucid chart for drawing the proposed architecture and flow charts. For implementation, we have used Google colab. We also use Keras, Tensor flow, Pillow (PIL), skimage, pandas, numpy, and OpenCV packages. At the time of coffee roasting Hartanzah professional coffee roasting timer app, and coffee sample roaster equipment are used. In our experiment, we utilized a regression Random Forest (RF) and linear regression model implemented in Keras with TensorFlow as the backend. The model is tested on an Intel(R) Core ™ i7-7600 CPU with 8 GB RAM and we use an external T4 GPU hardware accelerator in Google Colab. To evaluate the performance of the RF model, we employed several evaluation metrics, including the R^2 score, MAE, MSE, accuracy, precision, recall, and f1 score. Additionally, we visualized the results using scatter plots and confusion matrix to gain insights into the relationship between the predicted and actual sensorial scores. These metrics provide insights into how well the model fits the data and predicts the target variable. After this, we select the best model that gets high accuracy and low computational time.

4.2.2 Implementation

We have used an experimental design process strategy to produce research with a high level of internal validity. The implementation begins by importing necessary libraries and defining functions for different image processing operations, such as adaptive histogram equalization, noise removal, and feature extraction. We perform several data preprocessing tasks, including checking for missing values and removing any duplicate entries. It applies a series of image processing operations to each image, such as resizing, noise removal, histogram equalization, and thresholding, to enhance the image quality. Due to the scarcity size of the dataset, a data augmentation technique is applied to increase the dataset and to train a model. Data augmentation is applied using an Image Data Generator object, which performs operations like rotation, shifting, and flipping on the preprocessed images. The augmentation is applied to the input data (roasted coffee images) rather than the output data (attribute scores).

As a result, data augmentation is carried out for the image dataset, the attribute scores remain unchanged, and the sensorial scores are used as reference labels during the training process. The images are converted into grayscale that is converting all images from RGB color to one color channel. Noise removed by median filter and thresholding algorithm was used for image segmentation. We resize all images to 124x124 by using OpenCV's cv2.resize () function. Rescaling and rotating techniques are also handled during image preprocessing. The sensorial score of each attribute passed through the data cleaning process. Then we concatenate the preprocessed image with a cleaned sensorial score. These input image datasets should be converted into numerical arrays by using NumPy.

The preprocessed images and attributes are flattened and stored in arrays for further processing. The attribute scores are normalized using Min-Max scaling to ensure all features are within a consistent range. After this, the preprocessed data is stored in the newly created Data Frame as a CSV file. By reading the preprocessed data, color, texture, and shape features are extracted using the Gray-Level Co-occurrence Matrix (GLCM) approach. The shape and color of the roasted image are extracted using HOG and color histogram feature extraction techniques.

To ensure fair evaluation, we divided the available data into training, validation, and testing sets using a train-test splitter. This division allowed us to train the model on a portion of the data and assess its performance on the remaining unseen data. During the training process, we explored different parameter settings to optimize the random forest, linear regression, and other model's performance. The goal was to find the optimal combination of parameters that produced the best results based on the evaluation metrics.

After the attribute value is predicted the classifier model is trained by adding those attribute values. There is a threshold value that we use in this system for classification. The threshold value is at base of the ECX protocol. The regression model predicted values and external data is passed through preprocessing techniques such as filling missing values, data transformation, handling imbalanced data, and label encoding. By partitioning the data into training and testing data we train our classifier models.

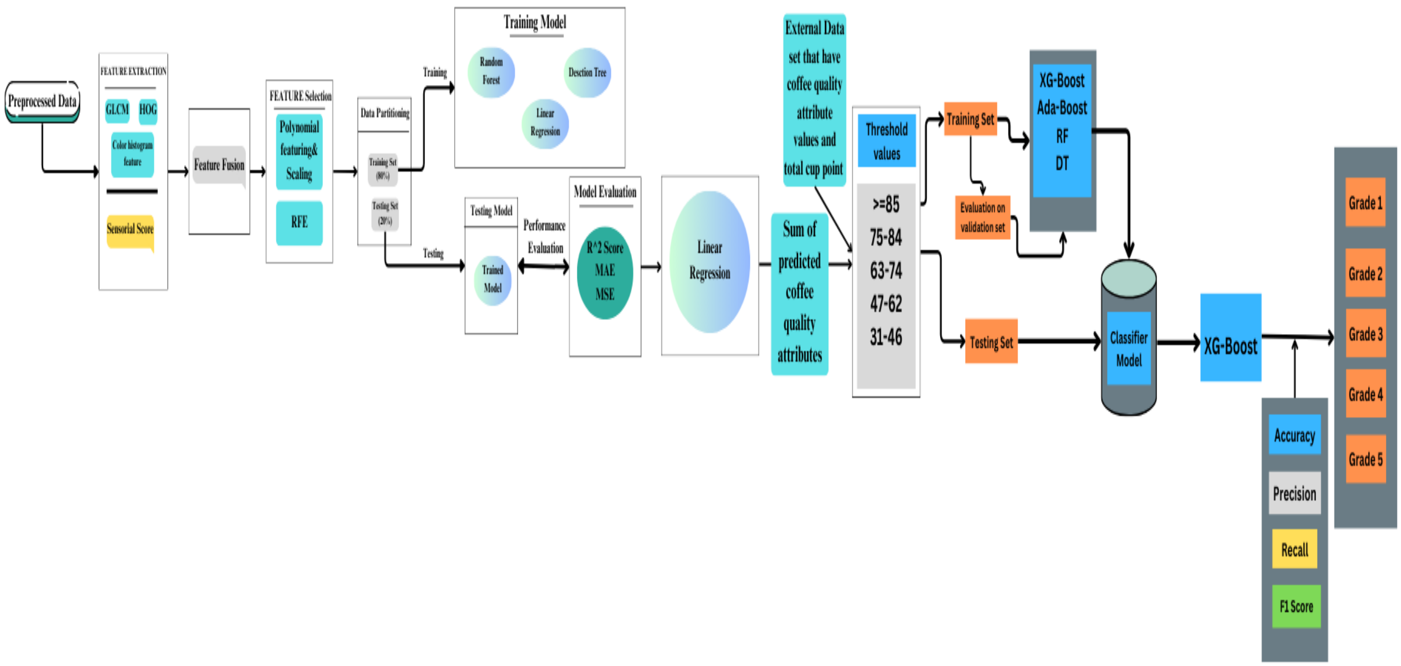


Figure 20: General Approaches of Coffee Quality Attribute Prediction and Coffee Grading

4.4 Experimental Results, Discussion, and Evaluation

4.4.1 Experimental with different image sizes

We tested different roasted coffee image sizes such as 124x124, 256x256, and 512x512 pixels. The first one is 512x512 pixels (figure 21), these images give lower confusion determinant values, MAE, and MSE than others with high computational time.

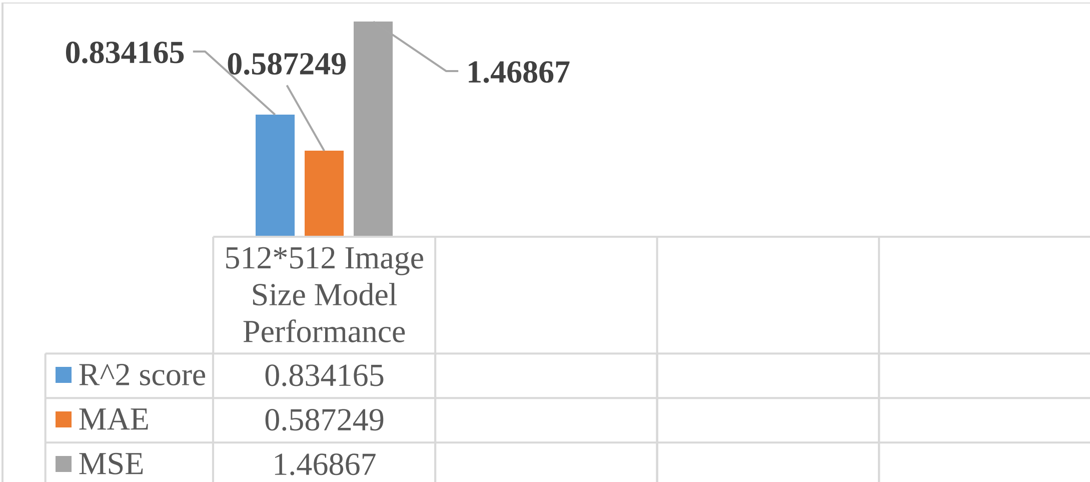


Figure 21: Model performance of 512*512 image size

We also tested the 256x256 pixels size, which yielded an R2 score of 0.8624, MSE of 1.254, and MAE of 0.352. It demonstrated superior performance compared to the 512x512 size. However, it incurred a higher computational time. The 124x124 (Figures 22) size images give better performance with smaller computational time. The reason the smaller image sizes have a better performance evaluation than others as they reduce complexity, smaller images may have less noise or irrelevant information compared to larger images, overfitting prevention, and model architecture compatibility. The 124x124 pixel size may have contained the optimal balance of information for the specific task. It is possible that this resolution captured the necessary details and features required for the analysis without excessive noise or loss of relevant information.

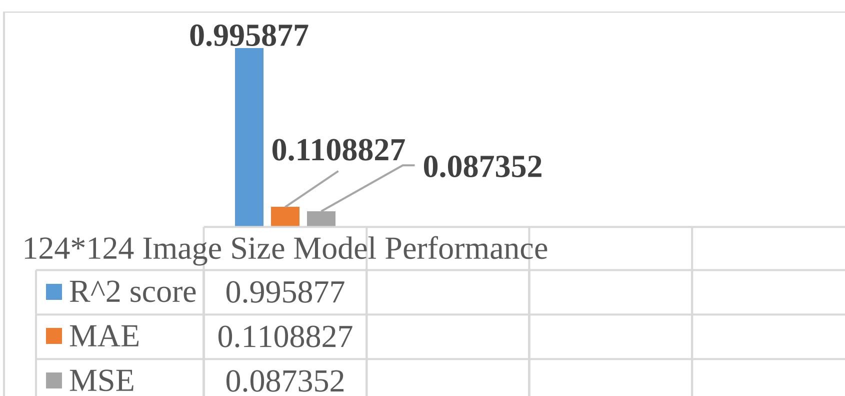


Figure 22: Model performance of 124*124 image size

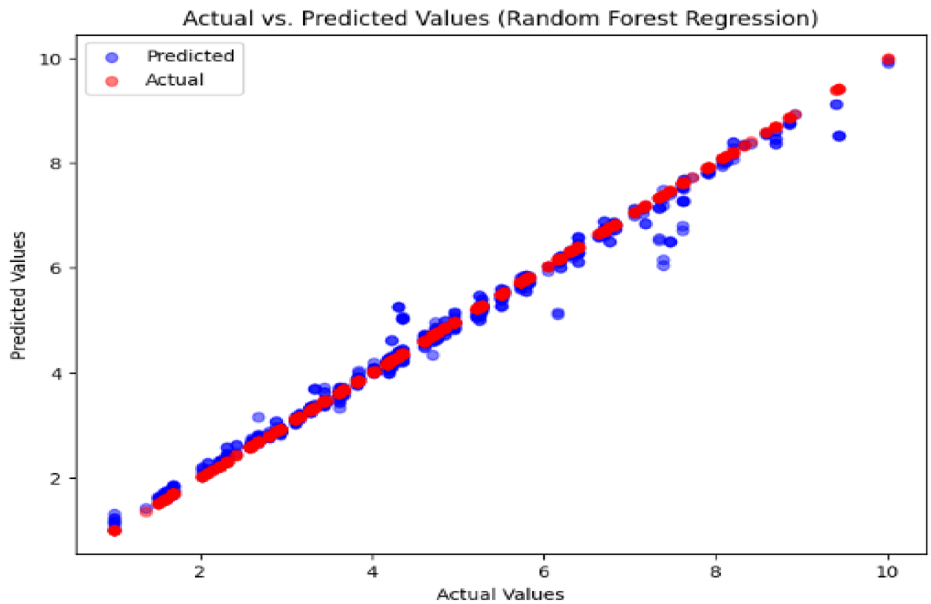


Figure 23: Scatter plot of RF model with 124*124 image size

The RF model we developed for predicting coffee quality attributes has demonstrated excellent performance. The R^2 score, which measures the proportion of variance explained by the model's predictions, is 0.995877. This high R^2 score indicates that the RF model can accurately capture the patterns and trends in the data, resulting in a strong fit. The Mean Squared Error (MSE) of 0.08735 measures the average squared difference between the predicted and actual scores. The lower the MSE, the better the model's predictive accuracy and ability to minimize large errors (Gorriz et al., 2024).

4.4.2 Experiment within and without HOG

HOG features provide an enriched representation of texture and shape information extracted from the images. By capturing detailed patterns and structures, the RF model gains a deeper understanding of the underlying characteristics, leading to more accurate predictions. This enhanced representation offers improved discriminative power, enabling the model to differentiate between subtle variations in texture and shape that may greatly influence the coffee quality attributes being predicted. The prediction value of the model using HOG is displayed in Figure 24 and 25

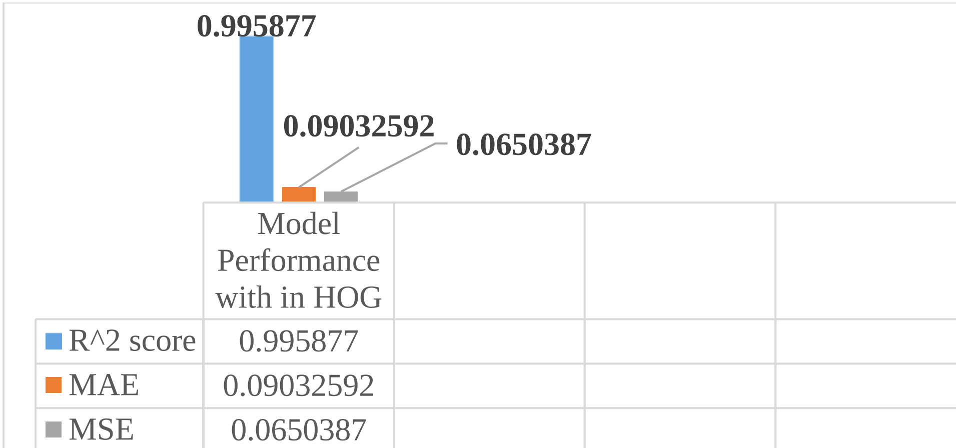


Figure 24: Model performance with HOG

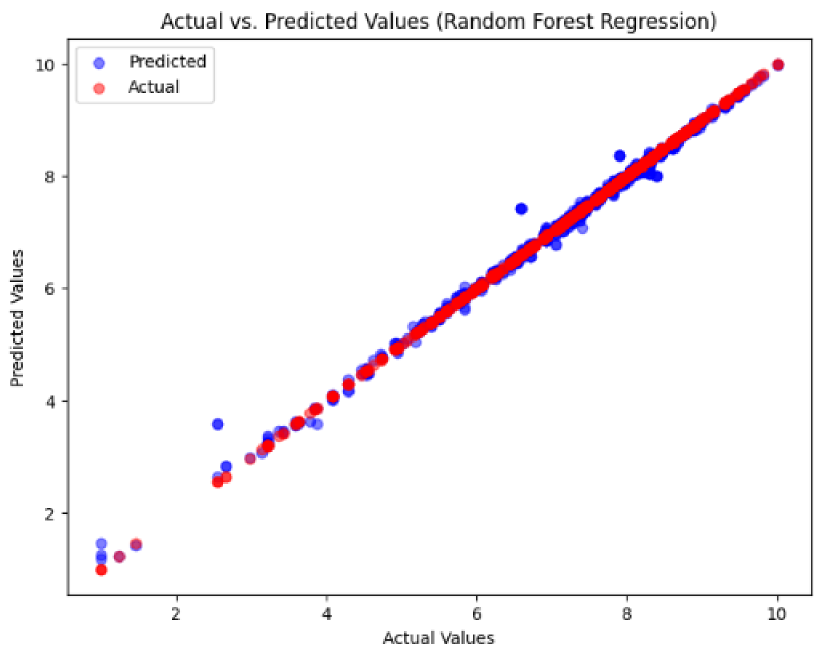


Figure 25: Scatter Plot of the model with HOG

During the development of a model without utilizing Histogram of Oriented Gradients (HOG) features, an important aspect to consider is the absence of shape and texture information captured by HOG. As a result, the model's performance is comparatively lower. The exclusion of HOG features limits the model's ability to comprehend the intricate texture and shape characteristics present in the coffee images, which can significantly impact its predictive accuracy.

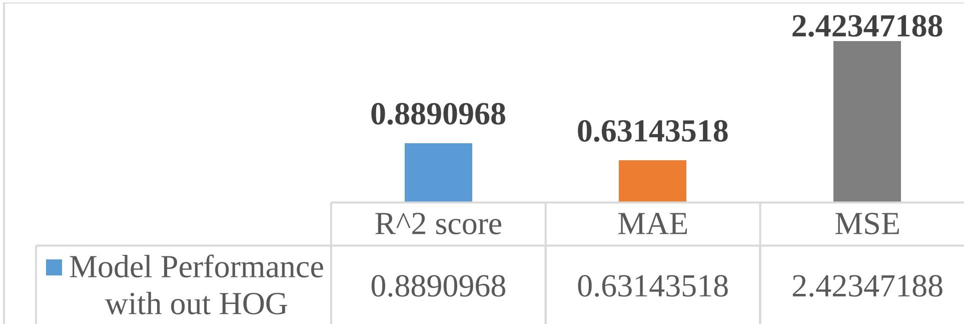


Figure 26: Model performance without HOG

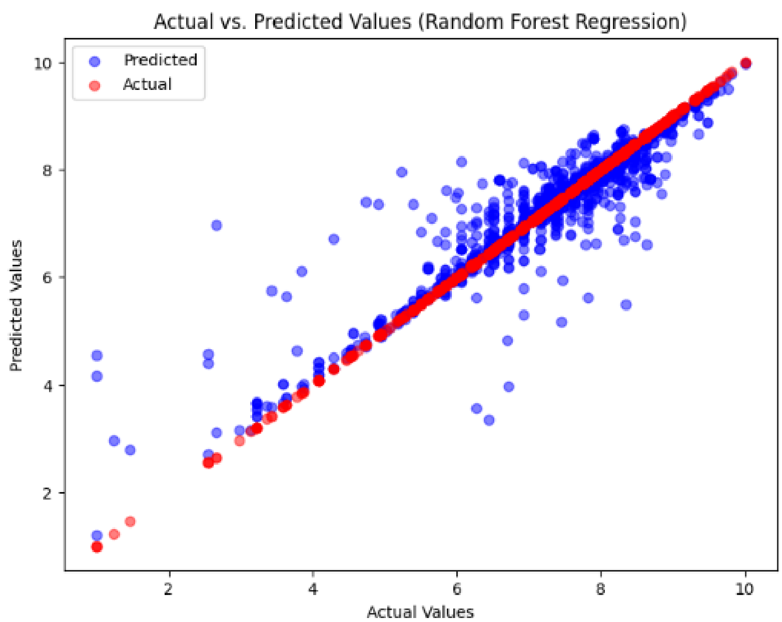


Figure 27: Scatter plot of the model without HOG

4.4.2.1 Feature Importance of HOG

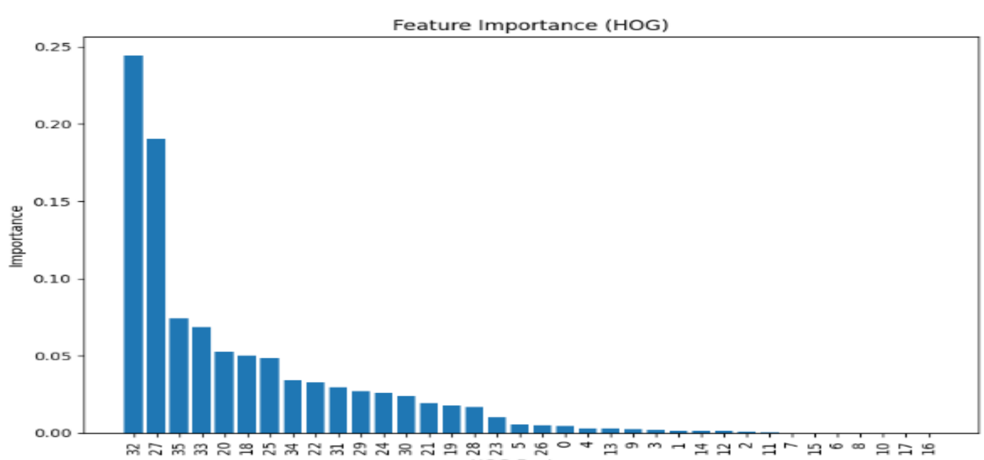


Figure 28: Feature Graph based on Mutual Information Score

	Feature	Importance
32	32	0.240673
27	27	0.191459
33	33	0.068297
35	35	0.064674
18	18	0.056355
20	20	0.051226
25	25	0.044819
31	31	0.037713
34	34	0.035908
22	22	0.035163
30	30	0.028390
24	24	0.024322
29	29	0.023825
21	21	0.017649
28	28	0.016529
19	19	0.014749
23	23	0.012025
3	3	0.010118
26	26	0.004784
9	9	0.004384
13	13	0.003081
4	4	0.003010
12	12	0.002650
5	5	0.001725
0	0	0.001696
2	2	0.001629
11	11	0.001420
14	14	0.000880
17	17	0.000146
15	15	0.000144
6	6	0.000143
8	8	0.000113
1	1	0.000106
10	10	0.000103
7	7	0.000070
16	16	0.000023

Figure 29: Mutual information score for each feature sorted with its feature number.

Without HOG (Figures 26 and 27), the model is unable to leverage the detailed information about texture and shape, leading to reduced performance in accurately predicting coffee quality attributes (Yassin Kortli et al., 2018). As we can see in Figures 28 and 29, the feature importance of HOG is shown. Feature number 32 has a greater feature importance value than the other, which indicates that the feature has a strong influence on the model's predictions. The feature is deemed more relevant in explaining the target variable compared to other features. The feature importance is displayed in descending order (Figure 29). Those feature importance values are better for feature selection, and interpretability, allowing us to select a subset of the most informative features and reduce computational complexity. To enhance clarity in illustrating the performance of each attribute, we utilize scatter plots that display both the actual and predicted values. By plotting the actual values on one axis and the corresponding predicted values on the other axis, we provide a visual representation of the accuracy and precision of the Random Forest (RF) algorithm for each attribute. This approach allows for a clear comparison between the expected and observed values, enabling a comprehensive assessment of RF's performance (R^2 score=0.995877) across different attributes.

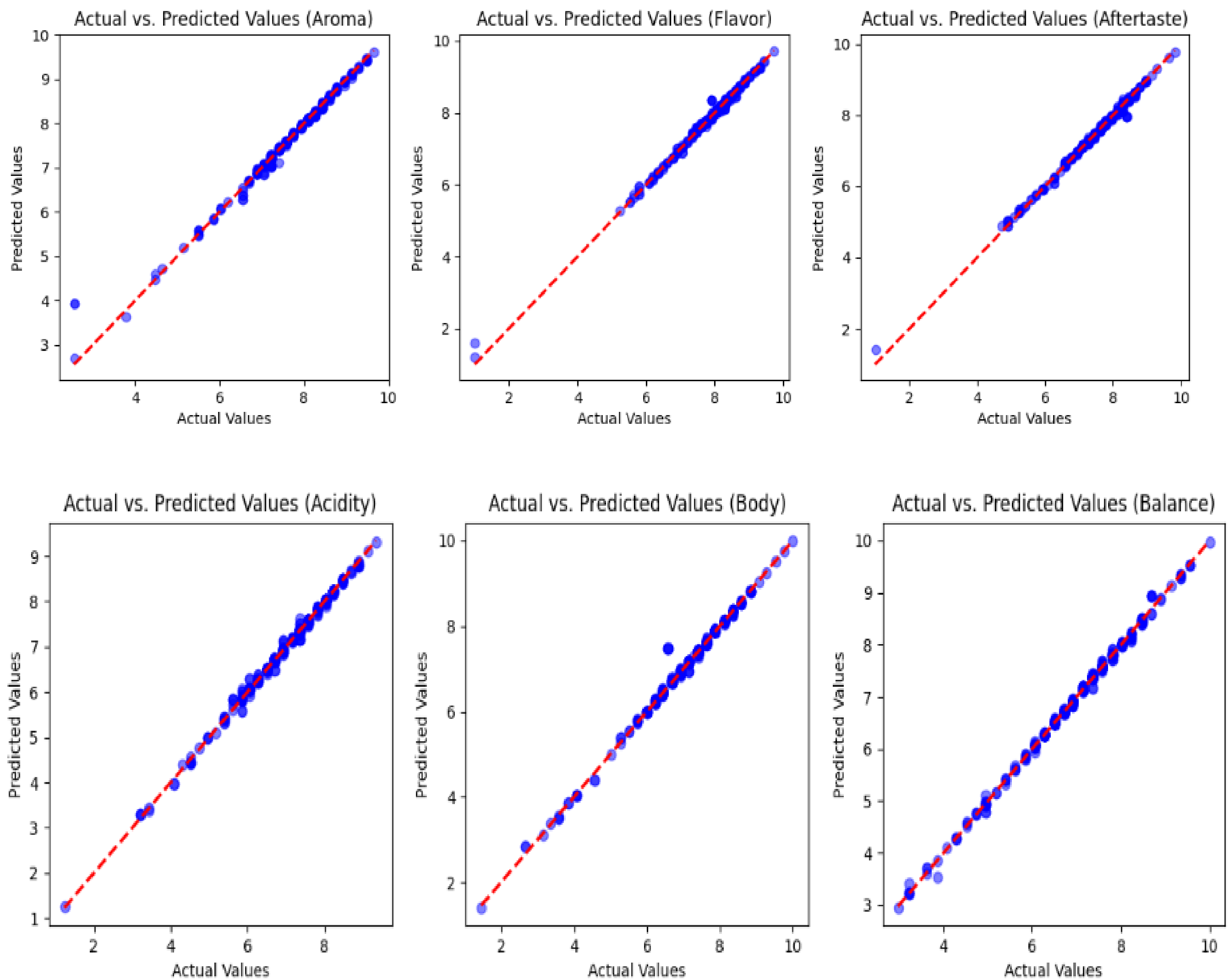


Figure 30: Scatter plot of each Coffee quality attribute of random forest

As we can see in Figure 30, each quality attribute provides better performance in capturing the relationship between the input and targeted attributes. Specifically, aroma, balance, and body exhibit high performance in accurately predicting the sensorial score of the attributes. Acidity, flavor, and aftertaste display a deviation in their actual values compared to the other attributes. This discrepancy occur due to the intricate or nonlinear relationship between the input variables and the target variable.

4.5 Model Comparison between Linear Regression and Random Forest

In this study, six models were evaluated: Random Forest with Recursive Feature Elimination (RFE), Linear Regression with RFE, Linear Regression with Polynomial Features and Scaling, logistic regression, decision tree, and deep neural network.

The Random Forest with RFE model demonstrated exceptional performance with an R^2 value of 0.9951, indicating a highly accurate. Random Forest is an ensemble learning approach, which combines multiple decision trees to make predictions and reduce overfitting. The model's use of Recursive Feature Elimination helps select the most informative features, reducing noise and improving predictive power. However, it had slightly larger errors, as reflected by the higher mean absolute error (MAE) of 0.09 and mean squared error (MSE) of 0.06. On the positive side, this model boasted the fastest computational time than Linear Regression with RFE, completing the evaluation in just 12.25 seconds.

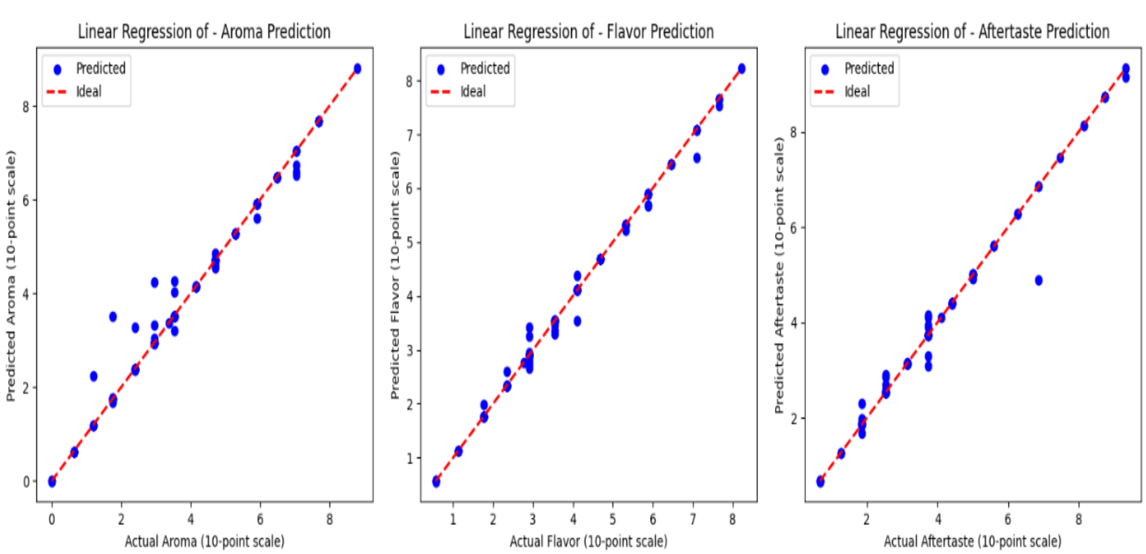
In contrast, the Linear Regression with the RFE model achieved a moderate R^2 score of 0.79764, suggesting a reasonable fit to the data. It exhibited relatively lower errors with an MAE of 0.1812 and MSE of 0.1563 compared to the Random Forest model. However, it took a slightly longer computation time 16.43 seconds, to complete the evaluation. The Linear Regression model with Polynomial Features and Scaling proved to be highly accurate, boasting an impressive R^2 score of 0.98754. It exhibited the lowest errors among the models, with an MAE of 0.0141 and MSE of 0.00969, and it has the fastest computational time (10.21 seconds).

In summary, the Random Forest with RFE model offered the fastest computational time but had slightly larger errors compared to the other models. The Linear Regression model with Polynomial Features and Scaling demonstrated excellent performance in terms of accuracy metrics, albeit with the fastest computational time. Considering both accuracy and computational time, the Linear Regression model with Polynomial Features and Scaling appears to be the best choice in this study. The reason why this model has the best performance is that RFE involves recursively eliminating less important features from the model. While this can be beneficial in reducing overfitting and improving model interpretability by focusing on the most relevant features, it may also result in the exclusion of potentially useful information. If important features are eliminated during the RFE process, it can lead to a decrease in model performance. By incorporating polynomial features into the linear regression model, the model gains the ability to capture non-linear relationships between the input variables and the target variable. This flexibility allows the model to fit the data more accurately, especially when the underlying relationship is non-linear. The overall result is presented in Table 8 and summarized in Figure 31.

Table 8: Summary of model evaluation

Model	Performance Evaluation Values	Computational Time
Random Forest with RFE	R^2=0.9951 MAE=0.09 MSE=0.06	12.25 Seconds
Linear Regression with RFE	R^2 Score: 0.79764 MAE: 0.1812 MSE: 0.1563	16.43 Seconds
Linear Regression with Polynomial Features and Scaling	R^2 Score: 0.98754 MAE:0.014088 MSE: 0.00969	10.21 Seconds

We present a sample prediction from the Linear Regression model with Polynomial Features and Scaling. This prediction provides an estimation of the target variable based on the input data, incorporating polynomial features and scaling techniques.



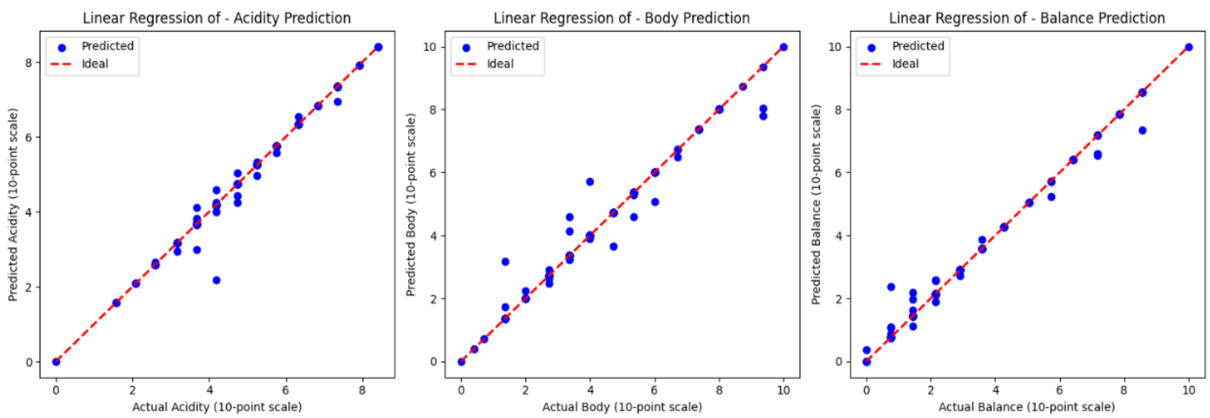


Figure 31: Attributes performance scatter plot of linear regression

```
Sample 3
Actual (10-point scale): Aroma          7.908451
Flavor          5.787234
Aftertaste      4.962687
Acidity         6.696203
Body            8.200000
Balance         4.230769
Uniformity      10.000000
Sweetness       10.000000
Cupper.Points   4.515625
Clean.Cup       10.000000
Name: 96, dtype: float64
Predicted (RFE) (10-point scale): [5.69651027 5.78723404 5.87734554 6.39195728 8.2       4.84485907
 9.31538754 9.39483294 4.515625   9.73200787]
Predicted (Poly & Scale) (10-point scale): [ 7.9084507   5.78723404  4.96268657  6.69620253  8.2        4.23076923
 10.         10.          4.515625   10.        ]
```

Figure 32: Linear regression with RFE, Poly & Scale Prediction Sample result

Based on the sample prediction observed above (Figure 32), the total actual cup point for the coffee quality attributes is 72.2988. In comparison, the total predicted value generated by the Linear Regression model with Polynomial Features and scaling amounts is 72.3003. This proximity between the total actual cup point and the predicted value suggests that the model can accurately estimate the overall coffee quality attributes. It indicates that the model captures the essential factors influencing the cup point and can provide reliable predictions for assessing the coffee's overall quality. We can observe the comparison of Random Forest and Linear Regression with other models in Figure 33. Additionally, the comparison between the Random Forest and Linear Regression models can also be seen in Figures 30 and 31.

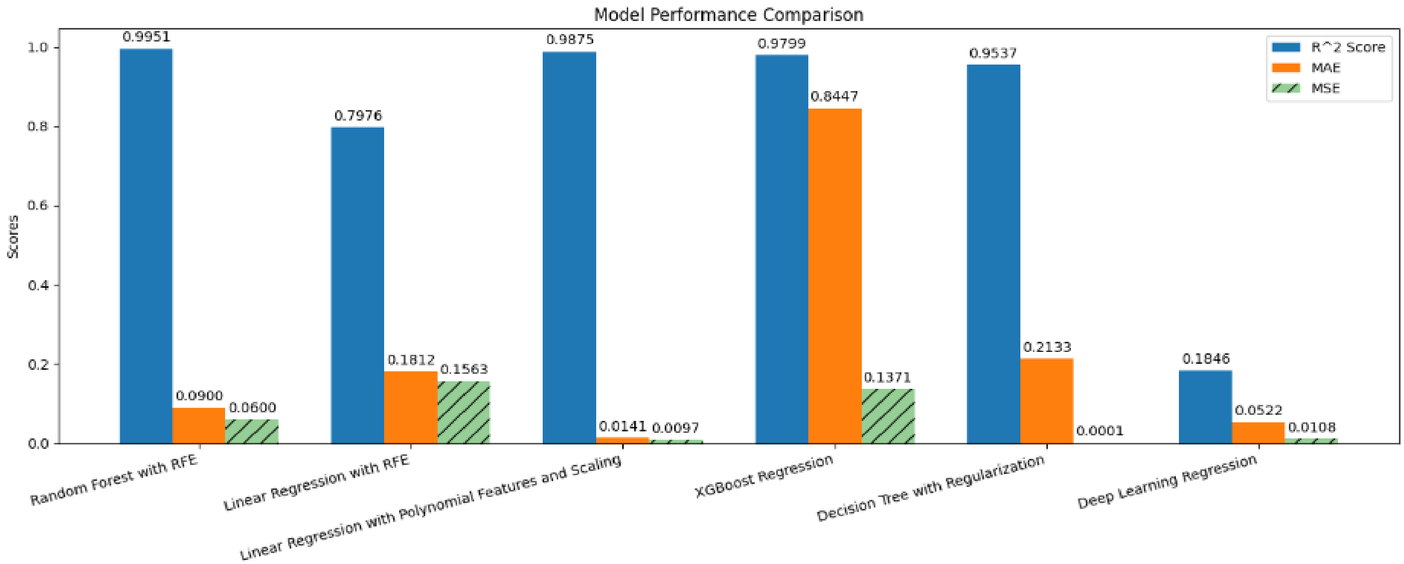


Figure 33: Model comparisons

4.6 Statistical Analysis of the model

ANOVA is a hypothesis test that determines whether there are significant differences between the means of two or more groups. It is commonly used to compare means across different groups or treatments and identify if there is evidence to support that the group means are statistically different (Nurrahma & Yusuf, 2020). We use ANOVA and t-test to make a statically analysis.

4.6.1 Statistical analysis of the distribution of colors

We experimented to examine how the degree of roasting, determined by varying the roasting time while keeping the voltage constant, affects the quality attributes of coffee. After roasting the coffee beans, we obtained sensory scores for each batch. We then conducted a statistical analysis to investigate the relationship between the color of the roasted coffee and its sensory score. The results of our analysis yielded promising outcomes, indicating a strong correlation between the roasted coffee color and its sensory attributes. This finding suggests that the duration of roasting significantly impacts the perceived quality of the coffee.

The formula of Independent samples t-test:

$$t - test = \frac{(mean_{group1} - mean_{group2})}{\sqrt{\left(\frac{s1^2}{n1}\right) + \left(\frac{s2^2}{n2}\right)}} \ldots\ldots\ldots\ldots\ldots\ldots.. Equation\ 4$$

Here, "mean_group1" and "mean_group2" represent the means of the two independent groups being compared, while "s1" and "s2" are the respective sample standard deviations, and "n1" and "n2" are the sizes of the two groups. The formula for calculating the standard deviation is as follows (Gerald, 2018):

For Group 1 (standard deviation denoted as s1):

$$s1 = \sqrt{\sum \frac{(xi - mean\_group1)2)}{(n1 - 1)}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots Equation\ 5$$

For Group 2 (standard deviation denoted as s2):

$$s2 = \sqrt{\sum \frac{(xi - mean\_group2)2)}{(n1 - 1)}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots Equation\ 6$$

In these formulas:

The variable "xi" represents an individual observation within the corresponding group. "mean_group1" represents the average value of the observations within Group 1, while "mean_group2" represents the average value of the observations within Group 2. The variable "n1" denotes the sample size of Group 1, indicating the number of observations within that group. Similarly, "n2" represents the sample size of Group 2, indicating the number of observations within that group.

Table 5: Statically analysis of roasted coffee color and its sensorial score

Features	**F-statistic**	**P-value**
Aroma	1.7232117448845765	0.011683655155484233
Flavor	1.768616173580991	0.008893105585327821
Aftertaste	2.848374746566604	0.0254381678948008e-05
Acidity	1.4240409973213652	0.06791439808235637
Body	1.8817255839595834	0.004503319496610052
Balance	1.5073086360926018	0.04207825213461935
Uniformity	4.408122872045971	0.0325340246883961e-08
Sweetness	1.8526026286273078	0.0053651813950203716

Cupper. Points	2.8980694856322478	0.005436735870309e-05
Clean. Cup	190.48915573343598	0.030153811694246e-44
Total.Cup.Points	2.538083091238056	0.03582212125653e-05
Moisture	0.9973132485816734	0.5297705321891792

The statistical analysis between roasted coffee image color and sensory scores reveals intriguing insights into the relationship between these variables. The findings indicate that color information in the images significantly impacts various sensory attributes. For instance, the analysis demonstrates that the image color influences the perceived aroma, as evidenced by a small yet statistically significant difference ($F = 1.72$, $p = 0.01$). Similarly, the image color shapes the perceived flavor, with a statistically significant difference observed ($F = 1.768$, $p = 0.008$). Additionally, the analysis reveals a strong association between image color and aftertaste, as indicated by a notable difference and a highly significant p-value ($F = 2.85$, $p < 0.02543$).

This statistical analysis shows the importance of color information in capturing and influencing sensory experiences. Moreover, attributes such as uniformity and cleanliness of the cup are also significantly related to the image color, reflecting the visual impact on these sensory aspects (Uniformity: $F = 4.41$, $p < 0.0325$; Clean Cup: $F = 190.49$, $p < 0.0301$).

4.6.2 Statistical analysis of sensorial scores

We compared automated coffee quality attribute prediction and human cuppers' evaluations. To carry out this comparison, we utilized statistical analysis techniques. Initially, we saved the predicted attribute values generated by the automated system in a CSV file. Subsequently, we collected sensory evaluations from human cuppers and recorded their scores alongside the automated predictions in the same CSV file. By performing a statistical analysis on these combined data, we aimed to assess the agreement or disparity between the automated predictions and the human cuppers' evaluations.

We make a statistical analysis between a sensorial score of automated aroma and cupper’s aroma, automated aftertaste and cupper’s aftertaste, automated body and cupper’s body, automated balance, and cupper’s balance, and automated flavor and cupper’s flavor. By taking a null hypothesis.

H0: There is no significant difference in the performance of the automated system and human cuppers in predicting coffee quality attributes using traditional sensory evaluation methods.

```
t-statistic: 0.17751339325679766
p-value: 0.8594715352713738
There is no statistically significant difference between the means of the two groups.
```

Figure 34: Statically difference between cuppers and automated model

Based on the obtained statistical analysis, the comparison between automated and human cuppers' performance in assessing coffee quality attributes resulted in a t-statistic of 0.17751339325679766 and a p-value of 0.8594715352713738. These results indicate that there is no statistically significant difference between the means of the two groups. In other words, the performance of the automated system is comparable to that of human cuppers in predicting the coffee quality attributes. This suggests that the automated system shows promise as a reliable and efficient alternative to human assessments in evaluating coffee quality.

4.7 Model Explainability

4.7.1 Random Forest Model Explainability Using LIME and SHAP

LIME is an interpretability technique that aims to explain the predictions of a model by generating a locally faithful, interpretable linear model specific to a given instance. This approach helps identify the crucial features or attributes of the instance that significantly contribute to the predicted class probabilities. By understanding the importance of these features, LIME provides insights into the inner workings of the model and offers suggestions for model enhancements or modifications to improve prediction accuracy. Such interpretability techniques play a vital role in building trust, understanding complex models, and facilitating model refinement.

In Figure 35, we observe a sample prediction of coffee quality attributes. To understand the individual contributions of each attribute to the prediction, we employ LIME and SHAP for model Explainability. For instance, if the value of the "Body" feature is equal to or less than 0.63, it contributes a score of 0.01 toward the attribute's prediction. Similarly, if the value of the feature with an ID of 46157 exceeds 0.74, it adds a score of 0.01 to the prediction.

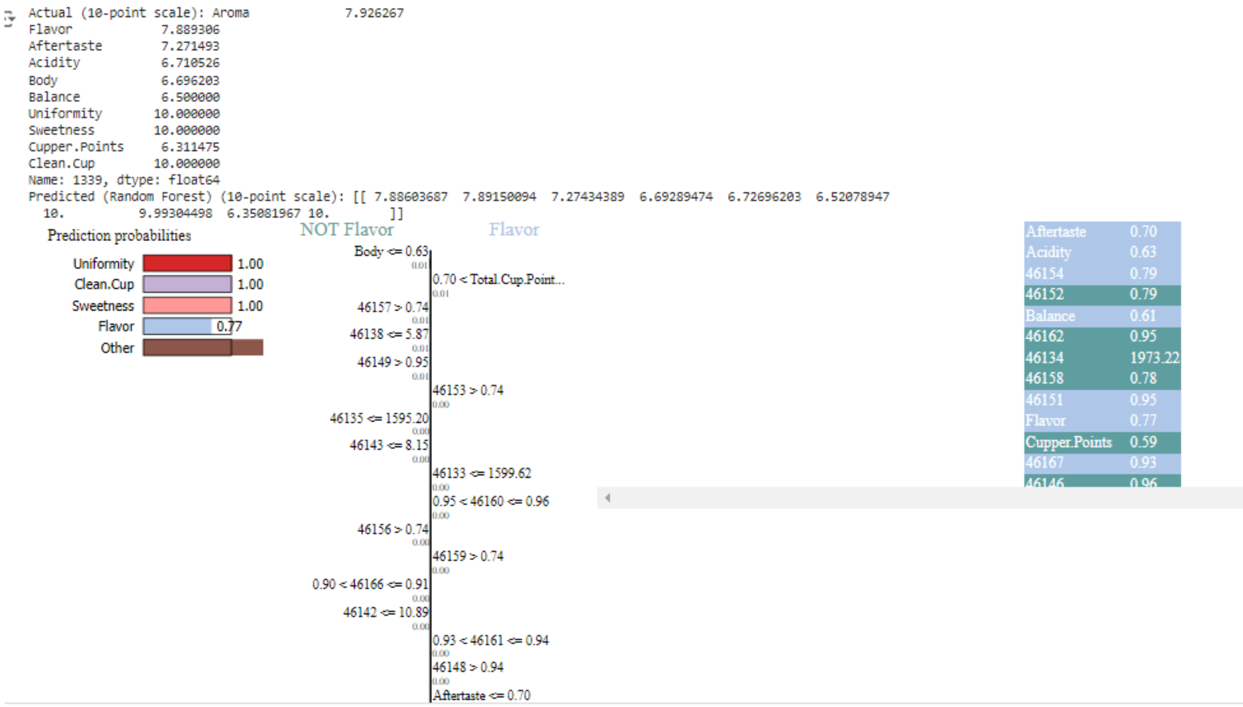


Figure 35: Random Forest Model Explainability using LIME.

To interpret machine learning model predictions, SHAP (Shapley Additive Explanations) is a framework and collection of algorithms. It tries to offer explicit feature attributions that clarify the reasoning behind a model's prediction for a given instance. The importance and influence of each feature on the model's predictions can be comprehended in a cohesive and mathematically sound manner with the help of SHAP values.

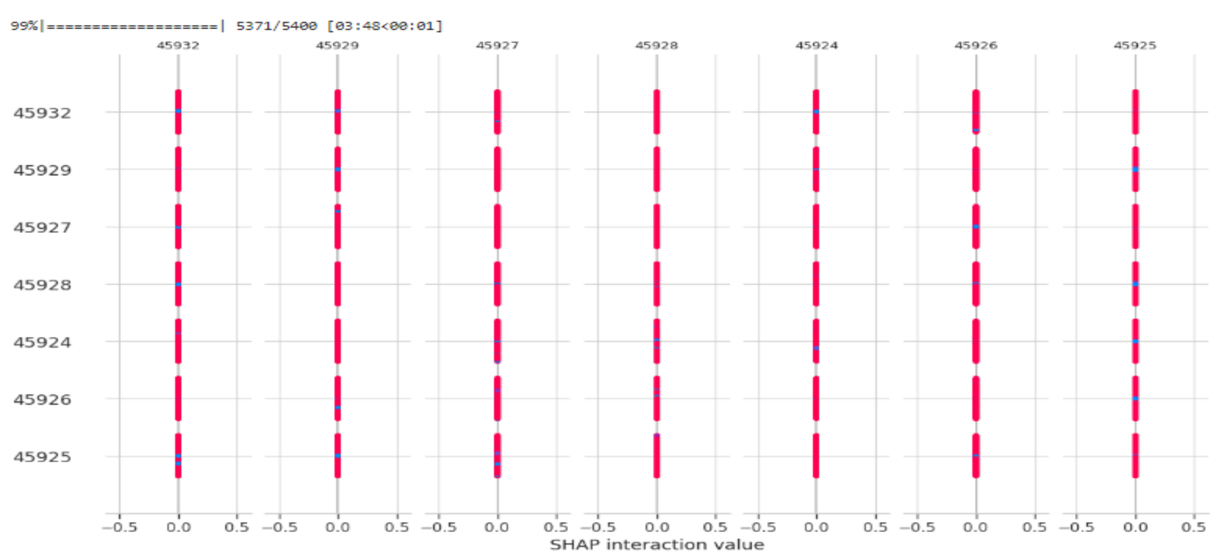


Figure 36: Random Forest Model Explainability using SHAP

In Figure 35 each feature is represented by a horizontal bar. The length of the bar indicates the magnitude of the SHAP values, representing the impact of the corresponding feature on the model's predictions. The color of the bar indicates the value of the feature, with red indicating high feature values and blue indicating low feature values. The vertical position of the feature bars represents the ordering of the features. Features at the top have the highest overall impact on predictions, while features at the bottom have the least impact. In Figure 35 SHAP helps to identify the role and importance of the color histogram features in the model's predictions.

4.8 Performance of Coffee Grading Model

In this study, our main focus is on predicting coffee quality attributes. Additionally, we have developed a comprehensive approach involving training a classifier model using total cup point and the outputs of a regression model that predicted coffee quality attributes, accurately predicting various coffee grading. We employed different ML algorithms to train classification models, like Ada-Boost, XG-Boost, random forest, and decision tree.

4.8.1 Performance of XG Boost Classifier

```
Accuracy: 0.8346883468834688
Precision: 0.8381160169778057
Recall: 0.8346883468834688
F1-score: 0.8349045186959618
Classification Report of XG Boost Classifier:
              precision    recall  f1-score   support

           0       0.92      0.95      0.93       129
           1       0.72      0.79      0.76       115
           2       0.86      0.76      0.81       125

    accuracy                           0.83       369
   macro avg       0.83      0.83      0.83       369
weighted avg       0.84      0.83      0.83       369

Cross-Validation Scores: [0.93984962 0.95488722 0.80451128 0.89393939 0.75757576]
```

Figure 37: Classification report of XG Boost classifier

The classification report reveals that class 0 had the highest precision, recall, and F1 score, indicating the model's exceptional ability to accurately identify instances within that class. Class 1 exhibited slightly lower performance, while class 2 demonstrated a favorable balance between precision and recall. The weighted average provides an overall assessment of the model's performance across all classes. Additionally, the cross-validation scores highlight performance variations across different data folds, demonstrating the model's generalization capabilities.

The reason behind achieving better performance can be attributed to several factors. The model has been trained on a diverse dataset, enabling it to learn intricate patterns and generalize effectively to unseen data. As Figure 38 shows, in RF 369 samples of data were used to test the model; among these test samples, 308 were correctly classified, and the remaining 61 were incorrectly classified.

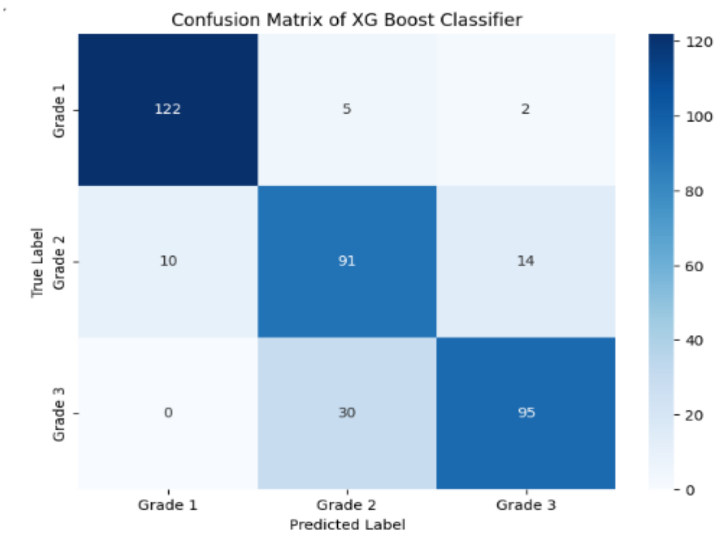


Figure 38: Confusion matrix for XG Boost

4.8.2 Performance of Ada Boost Classifier

```
Accuracy: 0.7994579945799458
Precision: 0.795901845825055
Recall: 0.7994579945799458
F1-score: 0.7969551773470087
Classification Report of Ada Boost:
              precision    recall  f1-score   support

           0       0.91      0.98      0.94       129
           1       0.70      0.70      0.70       115
           2       0.77      0.71      0.74       125

    accuracy                           0.80       369
   macro avg       0.79      0.79      0.79       369
weighted avg       0.80      0.80      0.80       369

Cross-Validation Scores: [0.92481203 0.95488722 0.06766917 0.90151515 0.87121212]
```

Figure 39: Classification report of Ada boost

This classifier gets less performance than other employed models because it is sensitive to outliers and imbalanced data, which leads to biased behavior. The algorithm may focus more on the majority class, leading to potential underperformance. As Figure 40 shows, in RF 369 samples of data were used to test the model; among these test samples, 295 were correctly classified, and the remaining 74 were incorrectly classified.

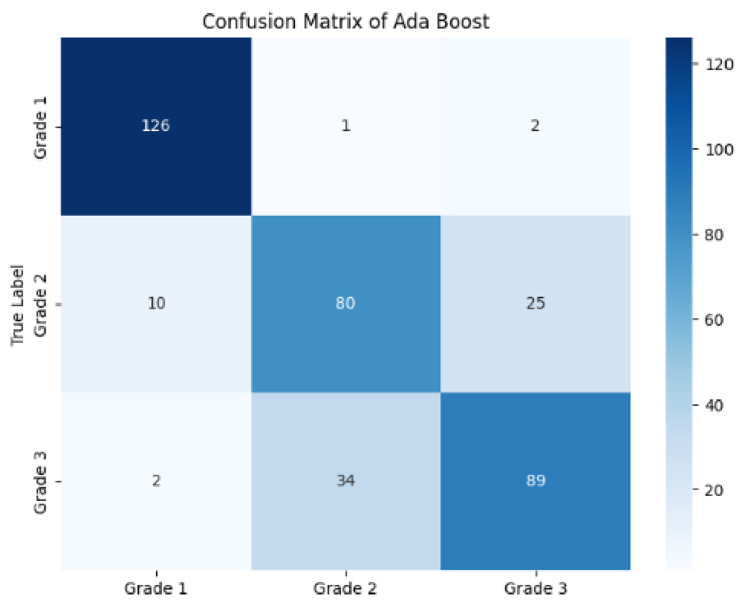


Figure 40: Confusion matrix of Ada boost classifier

4.8.3 Performance of Random Forest Classifier

```
Accuracy: 0.8157181571815718
Precision: 0.8237328407853444
Recall: 0.8157181571815718
F1-score: 0.8157831100578212
Classification Report of RandomForestClassifier:
              precision    recall  f1-score   support

           0       0.92      0.94      0.93       129
           1       0.69      0.81      0.74       115
           2       0.84      0.70      0.76       125

    accuracy                           0.82       369
   macro avg       0.82      0.81      0.81       369
weighted avg       0.82      0.82      0.82       369

Cross-Validation Scores: [0.91729323 0.95488722 0.80451128 0.89393939 0.68939394]
```

Figure 41: Classification report of random forest classifier

As we can see in the classification report of RF, a detailed analysis of the model's each class's performance is provided, including precision, recall, and F1-score values along with the support for each class. Observing the report, class 0 stands out with the highest precision, recall, and F1 score, indicating superior predictive performance compared to other classes. This suggests that the model excels in accurately identifying instances belonging to class 0. Figure 42 presents the weighted and macro average, considering the support and providing an overall measure of the model's performance in every class. Additionally, the cross-validation scores demonstrate variation in performance across different folds of the data, ranging from 0.6894 to 0.9549, which underscores the importance of assessing the model's generalizability. As Figure 42 shows, in RF 369 samples of data were used to test the model; among these test samples, 301 were correctly classified, and the remaining 68 were incorrectly classified.

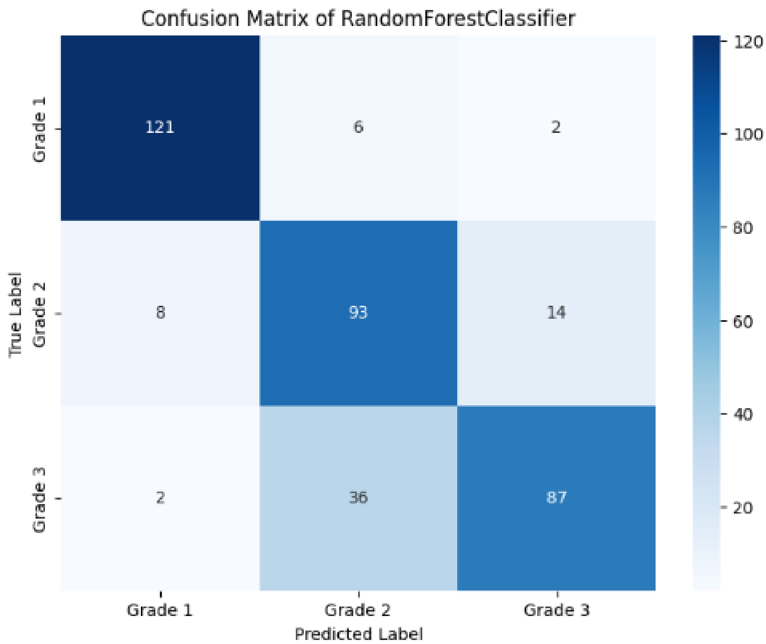


Figure 42: Confusion matrix of random forest classifier

4.8.4 Receiver Operating Characteristic (ROC) curve of classifiers

The Receiver Operating Characteristic (ROC) curve is a graphic representation and evaluation tool frequently used in binary classification applications. The area of the ROC curve (AUC-ROC) is a widely used metric for measuring a classifier's performance. It returns a single value indicating the model's total discriminating power. An AUC-ROC value of 0.5 suggests a random classifier, whereas a value of 1 implies a flawless classifier. ROC curve and AUC-ROC are useful tools for evaluating and comparing alternative categorization models, allowing researchers to choose the best model for their specific needs and performance criteria. As we can see in Figure 443

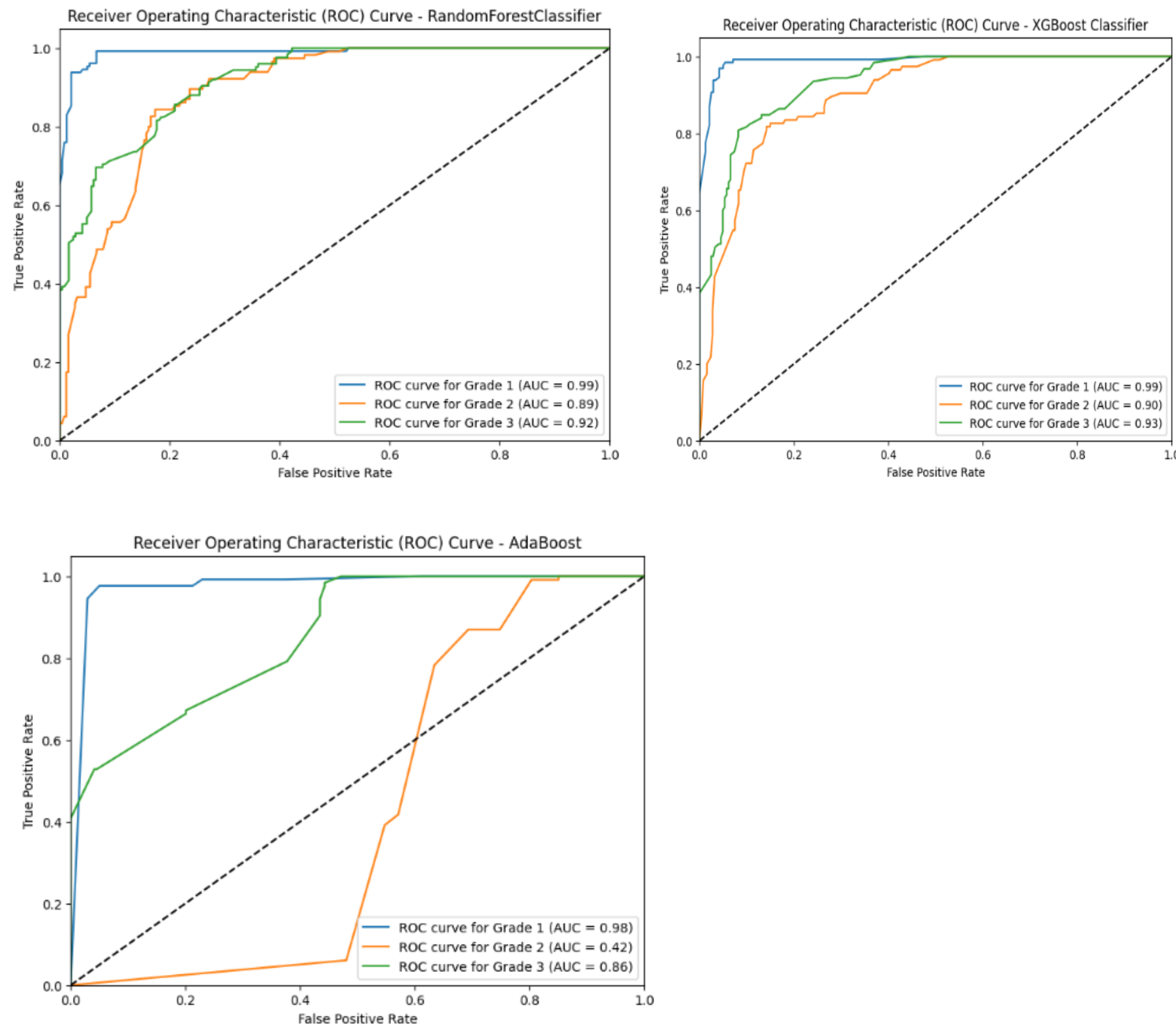


Figure 43: ROC curve of random forest, XG-Boost classifier, and Ada boost

4.8.5 Model Selections for Classifier Models

As we can see in confusion matric figures and ROC curve the XG-Boost classifier exhibits superior performance compared to other models due to its implementation of specific techniques. It incorporates regularization techniques, handles missing values, and utilizes a gradient-boosting algorithm that sequentially adds weak learners (decision trees) to rectify errors made by previous models. These attributes contribute to its enhanced performance and set it apart from other models.

4.9 Discussion of Result

The best finding of this research work is to combine GLCM, color descriptors, and HOG features with other modalities for coffee quality attribute prediction by using regression models. The experiments have been conducted to evaluate the performance of the coffee quality attribute prediction model. To enhance the performance of the model different image processing techniques on roasted coffee images like noise removal, image resizing, and others. We also applied different machine learning algorithms for both roasted coffee images and their sensorial score. The experiments have been performed with different image sizes to achieve better performance and minimize computational time by selecting optimal image sizes. So, we have taken 124x124, 256x256, and 512x512 image sizes. Based on the experimental results, we have selected a 124x124 image size for further analysis. This image size was chosen because it demonstrated performance that was better than the 256x256 and 512x512 sizes, with a minimum computational time. We evaluated six models: decision tree, Random Forest with Recursive Feature Elimination (RFE), Linear Regression with RFE, Linear Regression with Polynomial Features and Scaling, sequential deep learning model, and XG boost. The results of the evaluation revealed interesting findings. Among the six models, the Random Forest model demonstrated exceptional performance. It achieved an impressive R2 score of 0.9951, indicating a high degree of accuracy in predicting the target variable. Furthermore, the model exhibited a remarkably low mean absolute error (MAE) of 0.09 and mean squared error (MSE) of 0.06, indicating minimal deviation between predicted and actual values. Notably, the Random Forest model delivered these outstanding results within a short processing time of 12.25 seconds.

The Linear Regression model with RFE showcased a moderate level of performance. It attained an R2 score of 0.79764, implying a reasonable degree of predictive capability. However, compared to the Random Forest model, this model exhibited a higher MAE of 0.1812 and MSE of 0.1563, suggesting a slightly larger margin of error in its predictions. The processing time for this model was 16.43 seconds, which was longer than that of the Random Forest model. Lastly, the Linear Regression model with Polynomial Features and Scaling achieved a high R^2 score of 0.98754, indicating strong predictive accuracy. This model excelled in terms of the MAE, with an impressively low value of 0.01409, further highlighting its precision in predicting the target variable. Additionally, the MSE for this model was 0.00969, indicating minimal deviation between predicted and actual values. Remarkably, this model achieved these remarkable results within a relatively short processing time of 10.21 seconds. After predicting the quality attributes, the next step is to classify the grading of the coffee. However, due to limited data availability, we focus only on three classes instead of the full set of five classes that represent the overall coffee grading scheme. We assess the performance of the classifier model using evaluation metrics such as accuracy, precision, recall, and F1 score. Among the evaluated models, XG-Boost stands out as the best classifier with an accuracy of 83%.

4.10 Error Analysis

In this study, we conducted an error analysis to identify the sources of variation between the actual and predicted values. Two main factors contribute to these errors. Firstly, image quality and lighting conditions significantly impact the accuracy of predictions. Variations in lighting, image resolution, and noise can distort color representation and image clarity, leading to discrepancies between the actual and predicted values. Secondly, the variability in the roasting process introduces complexities that affect the model's ability to make accurate predictions. Differences in temperature, duration, and roasting techniques can result in variations between the actual and predicted values as the model may struggle to capture the intricacies introduced during roasting.

Additionally, the subjective nature of coffee quality assessment and the individual preferences of experts play a role in the errors observed. Limited training data and the selection of relevant features from the images also contribute to prediction errors. To improve the accuracy of coffee quality attribute predictions, future research should focus on enhancing image quality, accounting for roasting process variability, addressing subjectivity in assessment, expanding the training dataset, and refining feature extraction and selection techniques.

4.11 Summary

In this chapter, we conducted experiments to examine the impact of image size, Histogram of Oriented Gradients (HOG) features, and color descriptors on predicting coffee quality attributes based on image analysis. We systematically varied image dimensions to determine the optimal size that balances information preservation and computational efficiency. Additionally, we explored the effectiveness of HOG as a feature extraction technique, capturing texture and edge information within local image patches. We also investigated the relevance of color descriptors, particularly in relation to the degree of roasting. These experiments provided valuable insights into the role of these factors in predicting coffee quality attributes accurately. Furthermore, we compared the results generated by our automated system with the assessments provided by professional cuppers. Through rigorous statistical analysis, we quantified the level of agreement or discrepancy between the two sets of values. This comparative analysis allowed us to evaluate the accuracy and reliability of our image-based prediction model. By leveraging statistical techniques, we objectively assessed the performance of our automated system, identifying areas for improvement and gaining a deeper understanding of its strengths and limitations in predicting coffee quality attributes.

CHAPTER FIVE

CONCLUSION AND FUTURE WORK

5.1 Conclusion

In this study, we develop a model to predict the coffee quality attributes using a hybrid feature extraction (color, shape, and texture) by considering the most relevant features of the roasted coffee image like shape, color, textures, and relationships features of the roasted coffee image and its sensorial score. In this research firstly we collect the roasted coffee images with their sensorial score and then this input roasted coffee images and their sensorial score are preprocessed (resizing, color conversion, noise removal, and segmentation). The preprocessed datasets augmented to increase the dataset in number by horizontal and vertical flip, rotation range, width and height shift range, and flip mode. The augmented dataset segmented by using Otsu's thresholding and K-means. However, the threshold segmentation techniques give better results with good computational time. To extract meaningful features, we employed techniques such as Histogram of Oriented Gradients (HOG), Gray-Level Co-occurrence Matrix (GLCM), and color descriptors. These methods allowed us to capture relevant information from the coffee images. In addition, we applied feature selection techniques such as Recursive Feature Elimination (RFE), polynomial featuring, and scaling. Then we train our model by splitting the data into training and testing data. We evaluated three models: Random Forest with Recursive Feature Elimination (RFE), Linear Regression with RFE, and Linear Regression with Polynomial Features and Scaling. The results of the evaluation revealed interesting findings. Among the three models, the Random Forest model demonstrated exceptional performance.

5.1.2 Answering Research Questions

RQ1: The research question aims to investigate how image analysis and machine learning techniques can accurately predict roasted coffee quality attributes. By analyzing features such as color, texture, and shape from roasted coffee images, combined with machine learning algorithms, the study seeks to determine the effectiveness of this approach in predicting attributes related to coffee quality. Image analysis plays a crucial role in this process as it involves extracting meaningful visual features from the roasted coffee images. Color analysis enables quantification of the color properties exhibited by the roasted coffee, providing insights into its quality. Texture analysis focuses on capturing the textural characteristics of the coffee surface, which can be indicative of factors such as bean density or roast consistency. Shape analysis examines the geometric properties and contours of the coffee beans or grounds, which can offer valuable information about the bean structure and potential defects. These visual features serve as important cues for assessing coffee quality attributes.

Machine learning algorithms are then employed to build predictive models based on the extracted visual features. These algorithms learn from the relationships between the visual features and the associated sensory scores or quality attributes assigned by expert cuppers. By training on a diverse dataset, encompassing different origins, processing methods, and roast profiles, the models can capture the underlying patterns and correlations between the visual features and coffee quality attributes. The trained models can accurately predict roasted coffee quality attributes by leveraging the learned relationships. They can provide insights into various quality aspects such as flavor, aroma, acidity, body, and overall sensory experience. The models take advantage of the rich information present in the visual features extracted from the roasted coffee images and utilize this information to make accurate predictions.

To assess the accuracy of the predictions, the developed models are evaluated using appropriate performance metrics. These metrics include coefficient determinant value, mean absolute error, or mean squared error, depending on the specific attributes being predicted. Cross-validation techniques employed to evaluate the model's generalization ability and ensure robustness. The combination of image analysis and machine learning techniques offers a powerful approach to accurately predict roasted coffee quality attributes. By leveraging visual features and learning from the associations between these features and sensory scores, this approach provides an objective and consistent method for assessing coffee quality. It has the potential to enhance decision-making processes in the coffee industry, allowing for improved quality control, product development, and customer satisfaction

RQ2: This research question focuses on comparing the effectiveness of the proposed automated system in predicting coffee quality attributes with human cuppers who utilize traditional sensory evaluation methods. We utilized statistical analysis techniques. Initially, we saved the predicted attribute values generated by the automated system in a CSV file. Subsequently, we collected sensory evaluations from human cuppers and recorded their scores alongside the automated predictions in the same CSV file. By performing a statistical analysis on these combined data, we aimed to assess the agreement or disparity between the automated predictions and the human cuppers' evaluations.

We make a statistical analysis between a sensorial score of automated aroma and cupper's aroma, automated aftertaste and cupper's aftertaste, automated body and cupper's body, automated balance, and cupper's balance, and automated flavor and cupper's flavor. By taking a null hypothesis. The results indicate that there is no statistically significant difference between the means of the two groups. In other words, the performance of the automated system is comparable to that of human cuppers in predicting the coffee quality attributes. This suggests that the automated system shows promise as a reliable and efficient alternative to human assessments in evaluating coffee quality.

RQ3: This research question aims to explore how the fusion of roasted coffee image color and coffee quality attributes can enhance the prediction accuracy of coffee quality attributes. By combining visual features extracted from roasted coffee images with sensory scores, this study determine the integration of these modalities leads to improved prediction accuracy compared to using either modality in isolation. This investigation provide insights into the effectiveness of a comprehensive approach that considers both visual cues and the impact of roasting protocols on coffee quality attributes.

5.2 Contribution

This study shows the comparative analysis of regression algorithms using color, shape, and texture extraction techniques on the performance of coffee quality prediction, as well as the examination of how the degree of roasting affects coffee quality, determined by varying the roasting time while keeping the voltage constant, affects the quality attributes of coffee. By considering shape, texture, and color as the most relevant features of roasted coffee beans, the research offers a systematic model for predicting coffee quality attributes. Additionally, the study experimented to obtain sensory scores for each batch of coffee, enabling a comprehensive evaluation of its quality. A statistical analysis was then performed to investigate the relationship between the color of the roasted coffee and its sensory score. These findings provide valuable insights for coffee quality prediction by considering multiple features and optimizing the roasting process to achieve desired flavor profiles. We have identified the appropriate image size that is suitable for our model based on R^2, MAE, and MSE score and computational time. In general, the contributions of this study are:

Dataset Preparation: to the best of our investigation, this study is the first investigation of coffee quality attribute prediction from roasted coffee images in Ethiopia. We prepared our dataset for the study, which is suitable for system development.

Feature extraction: we extract hybrid features by fusing multiple modalities of roasted coffee images (HOG, GLCM, and color descriptors) and their sensorial score, also the study proposes a state-of-the-art machine learning approach to predict coffee quality attributes from roasted coffee images.

Static analysis: we experiment on the effect of roasting on coffee quality attributes by making a statically analysis between the color of roasted coffee images with its sensorial score value. We also make a comparison between automated and human cupper's performance in the prediction of coffee quality attributes.

Fusion multiple modalities: we use our data by fusing multiple modalities (image and sensorial score). It gives way to enhanced feature representation, complementary information, improved model robustness, improved decision-making, and increased interpretability.

Coffee grading: we work on coffee grading by taking the predicted value of coffee quality attributes.

5.3 Future work

In this study, we prepared fewer datasets, as the complexity of predicting coffee quality attribute prediction needs a large image dataset and we suggest increasing datasets without over-fitting and under-fitting problems. Also, it is better for the chemical analysis of coffee will be added to the data. In this research we focused on machine learning algorithms, to make them more reliable and sophisticated it may be better to apply deep learning techniques.

References

Abdulmajid, A. M. (2014). Sensory Evaluation of beverage characteristics and biochemical components of Coffee Genotypes. In Advances in Food Science and Technology (Vol. 2, Issue 12). www.internationalscholarsjournals.org

Abebe, Y., Juergen, B., Endashaw, B., Kitessa, H., & Heiner, G. (2020). The major factors influencing coffee quality in Ethiopia: The case of wild Arabica coffee (Coffea arabica L.) from its natural habitat of southwest and southeast afromontane rainforests. African Journal of Plant Science, 14(6), 213–230. https://doi.org/10.5897/ajps2020.1976

Adya Zizwan Putra, Chalvin Chalvin, Achmad Nurhadi, Andro Eriel Tambun, & Syahmir Defha. (2023). Coffee Quality Prediction with Light Gradient Boosting Machine Algorithm Through Data Science Approach. *Sinkron:Jurnal Dan Penelitian Teknik Informatika*, *8*(1), 563–573. https://doi.org/10.33395/sinkron.v8i1.12169

Alhindi, T. J., Kalra, S., Ng, K. H., Afrin, A., & Tizhoosh, H. R. (n.d.). Comparing LBP , HOG and Deep Features for Classification of Histopathology Images.

Aliah, N. (). A Review of Quality Coffee Roasting Degree Evaluation [Review of *A Review of Quality Coffee Roasting Degree Evaluation]*. *Journal of Applied Science and Agriculture*, *10*(7), 18–23.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Dujaili, A. Al, Duan, Y., Shamma, O. Al, Santamaría, J., Fadhel, M. A., Amidie, M. Al, & Farhan, L. (2021). Review of deep learning : concepts, CNN architectures, challenges, applications, future directions. In Journal of Big Data. Springer International Publishing. https://doi.org/10.1186/s40537-021-00444-8

Arias, F., Campo, D., Osiris, O., Villegas, V., Guadalupe, V., Sánchez, C., & Antonio, L. (2020). Influence of Image Pre-processing to Improve the Accuracy in a Convolutional Neural Network. 11(April), 88–96.

Ashiru, Y., & Oladele, L. (2023, July 7). *A Complete Guide to Linear Regression Model Interpretation*. Social Science Research Network. https://doi.org/10.2139/ssrn.4503197

Bansal, M., Kumar, M., & Kumar, M. (2020). 2D Object Recognition Techniques: State-of-the-Art Work. *Archives of Computational Methods in Engineering*, *28*(3), 1147–1161. https://doi.org/10.1007/s11831-020-09409-1

Bi, Z., & Cao, P. (2021). Color space conversion algorithm and comparison study. *Journal of Physics: Conference Series*, *1976*(1), 012008. https://doi.org/10.1088/1742-6596/1976/1/012008

Bolka, M., & Emire, S. (2020). Effects of coffee roasting technologies on cup quality and bioactive compounds of specialty coffee beans. *Food Science & Nutrition*. https://doi.org/10.1002/fsn3.1904

Bosilj, P., Hribar, U., & Likar, B. (2018). Detection and classification of coffee bean defects using computer vision and machine learning. Journal of Food Engineering, 228, 47-57.

Budholiya, K., Shrivastava, S. K., & Sharma, V. (2022). An optimized XGBoost based diagnostic system for effective prediction of heart disease. *Journal of King Saud University - Computer and Information Sciences*, *34*(7), 4514–4523. https://doi.org/10.1016/j.jksuci.2020.10.013

Caporaso, N., Whitworth, M. B., & Fisk, I. D. (2022). Prediction of coffee aroma from single roasted coffee beans by hyperspectral imaging. *Food Chemistry*, *371*, 131159. https://doi.org/10.1016/j.foodchem.2021.131159

Chen, T. (2019). *XGBoost: A Scalable Tree Boosting System* [Review of *XGBoost: A Scalable Tree Boosting System*].

Color Stability of Different Aesthetic Resin Composite Materials: A Digital Image Analysis. (2020). *Indian Journal of Forensic Medicine & Toxicology*. https://doi.org/10.37506/ijfmt.v14i1.239

Córdoba, N., Moreno, F. L., Osorio, C., Velásquez, S., Fernandez-Alduenda, M., & Ruiz-Pardo, Y. (2021). Specialty and regular coffee bean quality for cold and hot brewing: Evaluation of sensory profile and physicochemical characteristics. LWT, 145, 111363. https://doi.org/10.1016/j.lwt.2021.111363

Cui, R., Yu, H., Xu, T., Xing, X., Cao, X., Yan, K., & Chen, J. (2022). Deep Learning in Medical Hyperspectral Images: A Review. *Sensors*, *22*(24), 9790. https://doi.org/10.3390/s22249790

Dhruv, B., Mittal, N., & Modi, M. (2017). Analysis of different filters for noise reduction in images. *2017 Recent Developments in Control, Automation & Power Engineering (RDCAPE)*. https://doi.org/10.1109/rdcape.2017.8358306

Eron, F., Noman, M., Ricon, R., & Chalfun-Junior, A. (2024). Computer Vision-Aided Intelligent Monitoring of Coffee: Towards Sustainable Coffee Production. *Scientia Horticulturae*, *327*, 112847–112847. https://doi.org/10.1016/j.scienta.2024.112847

Esteban-Díez, I., González-Sáiz, J. M., & Pizarro, C. (2004). Prediction of sensory properties of espresso from roasted coffee samples by near-infrared spectroscopy. *Analytica Chimica Acta*, *525*(2), 171–182. https://doi.org/10.1016/j.aca.2004.08.057

Fadri, R. A., Sayuti, K., Nazir, N., & Suliansyah, I. (2019). The Effect of Temperature and Roasting Duration on Physical Characteristics and Sensory Quality Of Singgalang Arabica Coffee (Coffea arabica) Agam Regency. Journal of Applied Agricultural Science and Technology, 3(2), 189–201. https://doi.org/10.32530/jaast.v3i2.117

Farré, E., Blasco, J., & Cubero, S. (2020). Deep learning for coffee quality assessment. Computers and Electronics in Agriculture, 173, 105386.

Ganatra, N., & Patel, A. (2018). A Survey on Diseases Detection and Classification of Agriculture Products using Image Processing and Machine Learning. *International Journal of Computer Applications*, *180*(13), 7–12. https://doi.org/10.5120/ijca2018916249

Ganesan, K., Baskaran, R., & Raguram, A. (2020). Automated classification system for coffee bean defects using image processing and machine learning techniques. Computers and Electronics in Agriculture, 174, 105494.

Gebeyehu, G. M., Feleke, D. G., Molla, M. D., & Admasu, T. D. (2020). Effect of habitual consumption of Ethiopian Arabica coffee on the risk of cardiovascular diseases among non-diabetic healthy adults. Heliyon, 6(9). https://doi.org/10.1016/j.heliyon.2020.e04886

Gerald, B. (2018). A brief review of independent, dependent and one sample t-test. *International Journal of Applied Mathematics and Theoretical Physics*, *4*(2), 50–54. https://doi.org/10.11648/j.ijamtp.20180402.13

Ginting, M. B. (2022). The Influence of Servicescape and Brand Image on Consumer Loyalty with Consumer Satisfaction as An Intervening Variable at The Coffee Crowd Medan

[Review of *The Influence of Servicescape and Brand Image on Consumer Loyalty with Consumer Satisfaction as An Intervening Variable at The Coffee Crowd Medan*]. *International Journal of Research and Review*, *9*(8), 483–503. https://doi.org/10.52403/ijrr.20220841

Goel, V., Singhal, S., Jain, T., & Kole, S. (2017). Specific Color Detection in Images using RGB Modelling in MATLAB. *International Journal of Computer Applications*, *161*(8), 38–42. https://doi.org/10.5120/ijca2017913254

Guangwu Lv, & Cao, P. (2020). *Color Space Conversion Technology and Comparative Research*. https://doi.org/10.1145/3446999.3447021

Henrique, G., & Paula, A. (2023). Coffee Roasting, Blending, and Grinding: Nutritional, Sensorial and Sustainable Aspects. *Agriculture*, *13*(11), 2116–2116. https://doi.org/10.3390/agriculture13112116

Jansen, D., & Warren, K. (2020, June). *What is research methodology?* Grad Coach. https://gradcoach.com/what-is-research-methodology/

Javier, A. (2020). Machine Learning for Cup Coffee Quality Prediction from Green and Roasted Coffee Beans Features [Review of Machine Learning for Cup Coffee Quality Prediction from Green and Roasted Coffee Beans Features]. Research Gate, 5(6).

Jiang, J., Li, C., Fan, J., & Zhang, D. (2021). Deep learning-based coffee roast level classification using images. Computers and Electronics in Agriculture, 185, 106000.

Kabbai, L., Abdellaoui, M., & Douik, A. (2019). Image classification by combining local and global features. Visual Computer, 35(5), 679–693. https://doi.org/10.1007/s00371-018- 1503-0

Kalyani A Sankpal. (2020). A Review on Data Normalization Techniques. *International Journal of Engineering Research And*, *V9*(06). https://doi.org/10.17577/ijertv9is060915

Khudayer Jadwa, Dr. S. (2018). Wiener Filter based Medical Image De-noising. *International Journal of Science and Engineering Applications*, *7*(9), 318–323. https://doi.org/10.7753/ijsea0709.1014

Kim, J. Y. (2022). Coffee Beans Quality Prediction Using Machine Learning. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4024785.

Kora, P., & K. Thangadurai. (2016). *Implementation of RGB and Grayscale Images in Plant Leaves Disease Detection – Comparative Study*. *9*(6). https://doi.org/10.17485/ijst/2016/v9i6/77739

Kumar, S., Abid, I., Garg, S., Singh, A. K., & Jain, V. (2019). Brain tumor detection using image processing. 11(1), 20–25.

Liu, Y. (2019). New Machine Learning Algorithm: Random Forest [Review of *New Machine Learning Algorithm: Random Forest*]. *International Conference on Information Computing and Applications*, 246–252.

López-Ruiz, N., Rios-Solis, L., & Barbin, D. (2021). Image analysis coupled with machine learning for coffee quality assessment. Journal of Food Quality, 2021, 5593576.

Machado, R. M. A., Barbin, D. F., & Poppi, R. J. (2019). Prediction of cupping scores of coffee samples by color analysis using chemometric tools. Food Chemistry, 288, 273-279.

M, H., & M.N, S. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, *5*(2), 01-11. https://doi.org/10.5121/ijdkp.2015.5201

Ma, J., Li, J., He, H., Jin, X., Cesarino, I., Zeng, W., & Li, Z. (2022). Characterization of sensory properties of Yunnan coffee. *Current Research in Food Science*, *5*, 1205–1215. https://doi.org/10.1016/j.crfs.2022.07.010

Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, *1*(4), 140–147. https://doi.org/10.38094/jastt1457

Manjunatha, S., & Patil, M. M. (2018). Interpolation techniques in Image Resampling. International Journal of Engineering and Technology (UAE), 7(3.34 Special Issue 34), 74 567–570. https://doi.org/10.14419/ijet.v7i3.2.14591

Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. TrAC Trends in Analytical Chemistry, 132, 116045. https://doi.org/10.1016/j.trac.2020.116045

Mishra, V. K., Kumar, S., & Shukla, N. (2017). Image Acquisition and Techniques to Perform Image Acquisition. *SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology*, *9*(01). https://doi.org/10.18090/samriddhi.v9i01.8333

Mohamed, A. (2015). Image Noise Reduction and Filtering Techniques [Review of *Image Noise Reduction and Filtering Techniques*]. *International Journal of Science and Research (IJSR)*, 2319–7064.

Münchow, M., Alstrup, J., Steen, I., & Giacalone, D. (2020). Roasting Conditions and Coffee Flavor: A Multi-Study Empirical Investigation. *Beverages*, *6*(2), 29. https://doi.org/10.3390/beverages6020029

Muktar Bedaso, Meshesha, M., & Chala Diriba. (2022). Grading Ethiopian Coffee Raw Quality Using Image Processing Techniques. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-1980632/v1

Naghibi, S. A., Ahmadi, K., & Daneshi, A. (2017). Application of Support Vector Machine, Random Forest, and Genetic Algorithm Optimized Random Forest Models in Groundwater Potential Mapping. *Water Resources Management*, *31*(9), 2761–2775. https://doi.org/10.1007/s11269-017-1660-3

Nguyen, Q. V., Miller, N., Arness, D., Huang, W., Huang, M. L., & Simoff, S. (2020). Evaluation on interactive visualization data with scatterplots. *Visual Informatics*, *4*(4), 1–10. https://doi.org/10.1016/j.visinf.2020.09.004

Niu, Z., & Li, H. (2019). Research and analysis of threshold segmentation algorithms in image processing. *Journal of Physics: Conference Series*, *1237*(2), 022122. https://doi.org/10.1088/1742-6596/1237/2/022122

Nurrahma, & Yusuf, R. (2020). *Comparing Different Supervised Machine Learning Accuracy on Analyzing COVID-19 Data using ANOVA Test*. https://doi.org/10.1109/icidm51048.2020.9339676

Öztürk, Ş., & Akdemir, B. (2018). Application of Feature Extraction and Classification Methods for Histopathological Image using GLCM, LBP, LBGLCM, GLRLM and SFTA. *Procedia Computer Science*, *132*, 40–46. https://doi.org/10.1016/j.procs.2018.05.057

Pandey, A. K., Sharma, P. D., Dheer, P., Parida, G. K., Goyal, H., Patel, C., Bal, C., & Kumar, R. (2017). Investigating the Role of Global Histogram Equalization Technique for 99mTechnetium-Methylene diphosphonate Bone Scan Image Enhancement. *Indian*

Journal of Nuclear Medicine: IJNM: The Official Journal of the Society of Nuclear Medicine, India, *32*(4), 283–288. https://doi.org/10.4103/ijnm.IJNM_61_17

Petch Sajjacholapunt, Akara Supratak, & Suppawong Tuarob. (2022). Automatic measurement of acidity from roasted coffee beans images using efficient deep learning. *Journal of Food Process Engineering*, *45*(11). https://doi.org/10.1111/jfpe.14147

Pérez-Sánchez, H., Pallottino, F., & Tisseyre, B. (2020). Machine vision system for real-time image analysis and size grading of coffee beans. Journal of Food Engineering, 283, 110023.

P.S, S. K., & V.S, D. (2016). Extraction of Texture Features using GLCM and Shape Features using Connected Regions. *International Journal of Engineering and Technology*, *8*(6), 2926–2930. https://doi.org/10.21817/ijet/2016/v8i6/160806254

Rukundo, O. (2019). Nearest Neighbor Value Interpolation [Review of *Nearest Neighbor Value Interpolation*]. *International Journal of Advanced Computer Science and Applications*, *03*(04).

Qasaimeh, M., Saleh, H., Mohammad, B., Tekeste, T., & Ismail, M. (2019). A novel SIFT architecture and ASIC implementation for real time SOC application. Analog Integrated Circuits and Signal Processing, 99(2), 325–338. https://doi.org/10.1007/s10470-019-01419- 9

Qi, J., Du, J., Siniscalchi, S. M., Ma, X., & Lee, C.-H. (2020). On Mean Absolute Error for Deep Neural Network Based Vector-to-Vector Regression. *IEEE Signal Processing Letters*, *27*, 1485–1489. https://doi.org/10.1109/lsp.2020.3016837

Rajalakshmi, M., & Annapurani, K. (2021). Enhancement of Vascular Patterns in Palm Images Using Various Image Enhancement Techniques for Person Identification. International Journal of Image and Graphics, 2250032. https://doi.org/10.1142/s0219467822500322

Review on Coffee Production and Marketing in Ethiopia. (2020). Journal of Marketing and Consumer Research. https://doi.org/10.7176/jmcr/67-02

Santos, F., Rosas, J., Martins, R., Araújo, G., Viana, L., & Gonçalves, J. (2020). Quality assessment of coffee beans through computer vision and machine learning algorithms. Coffee Science, 15, 1–9. https://doi.org/10.25186/.v15i.1752

Santoso, I., Mustaniroh, S. A., & Choirun, A. (2021). Methods for quality coffee roasting degree evaluation: a literature review on risk perspective. *IOP Conference Series: Earth and Environmental Science*, *924*(1), 012058. https://doi.org/10.1088/1755-1315/924/1/012058

S. Parsania, P. P., & V.Virparia, D. P. (2015). A Review: Image Interpolation Techniques for Image Scaling. International Journal of Innovative Research in Computer and Communication Engineering, 02(12), 7409–7414. https://doi.org/10.15680/ijircce.2014.0212024

Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y., & Shen, D. (2020). Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation and Diagnosis for COVID-19. IEEE Reviews in Biomedical Engineering, 14, 1–1. https://doi.org/10.1109/RBME.2020.2987975

Song, Y., & Yan, H. (2018). Image Segmentation Techniques Overview. AMS 2017 - Asia Modelling Symposium 2017 and 11th International Conference on Mathematical Modelling and Computer Simulation, 103–107. https://doi.org/10.1109/AMS.2017.24

Sui, Q., & Ghosh, S. K. (2024). Active Learning for Stacking and AdaBoost-Related Models. *Stats*, *7*(1), 110–137. https://doi.org/10.3390/stats7010008

Tania, S., & Rowaida, R. (2016). A Comparative Study of Various Image Filtering Techniques for Removing Various Noisy Pixels in Aerial Image. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, *9*(3), 113–124. https://doi.org/10.14257/ijsip.2016.9.3.10

Tarigan, E. B., Wardiana, E., Hilmi, Y. S., & Komarudin, N. A. (2022). The changes in chemical properties of coffee during roasting: A review. IOP Conference Series: Earth and Environmental Science, 974(1), 012115. https://doi.org/10.1088/1755-1315/974/1/012115

Tolessa, K., Rademaker, M., De Baets, B., & Boeckx, P. (2016). Prediction of specialty coffee cup quality based on near infrared spectra of green coffee beans. *Talanta*, *150*, 367–374. https://doi.org/10.1016/j.talanta.2015.12.039

Toma, T. (n.d.). Analysis of Coffee Value Chain: The Case of Yirgachefe District, Gedeo Zone, Southern Nation Nationalities and Peoples Regional State, Ethiopia [Review of *Analysis of Coffee Value Chain: The Case of Yirgachefe District, Gedeo Zone, Southern Nation Nationalities and Peoples Regional State, Ethiopia*]. *Asian Journal of Plant Science and Research*, *13*(3).

Tsami, M., Adamos, G., Nathanail, E., Budiloviča, E. B., Jackiva, I. Y., & Magginas, V. (2018). A Decision Tree Approach for Achieving High Customer Satisfaction at Urban Interchanges. *Transport and Telecommunication Journal*, *19*(3), 194–202. https://doi.org/10.2478/ttj-2018-0016

Tyagi, V. (2018). Understanding Digital Image Processing. Understanding Digital Image Processing, September. https://doi.org/10.1201/9781315123905

Vahidi, S. (2022). *A New Method for Resizing the Images* [Review of *A New Method for Resizing the Images*].

Velastegui, R., & Pedersen, M. (2021). CMYK-CIELAB Color Space Transformation Using Machine Learning Techniques. *London Imaging Meeting*, *2021*(1), 73–77. https://doi.org/10.2352/issn.2694-118x.2021.lim-73

Wang, H. (2017). Research Survey on Support Vector Machine [Review of *Research Survey on Support Vector Machine*]. *Chongqing, People's Republic of China*, *10th*.

Worku, M., Astatkie, T., & Boeckx, P. (2023). Quality and biochemical composition of Ethiopian coffee varied with growing region and locality. *Journal of Food Composition and Analysis*, *115*, 105015. https://doi.org/10.1016/j.jfca.2022.105015

Wulandari, S., Ainuri, M., & Sukartiko, A. C. (2022). Sensory Evaluation of Robusta Coffee under Various Postharvest and Processing. *Proceedings of the 2nd International Conference on Smart and Innovative Agriculture (ICoSIA 2021)*. https://doi.org/10.2991/absr.k.220305.061

Xiong, J., Yu, D., Wang, Q., Shu, L., Cen, J., Liang, Q., Chen, H., & Sun, B. (2021). Application of Histogram Equalization for Image Enhancement in Corrosion Areas. *Shock and Vibration*, *2021*, e8883571. https://doi.org/10.1155/2021/8883571

Xiu, L., Ma, B., Zhu, K., & Zhang, L. (2018). Implementation and optimization of image acquisition with smartphones in computer vision. International Conference on Information Networking, 2018-Janua, 261–266. https://doi.org/10.1109/ICOIN.2018.8343121

Yassin Kortli, Maher Jridi, Ayman Al Falou, & Atri, M. (2018). A comparative study of CFs, LBP, HOG, SIFT, SURF, and BRIEF techniques for face recognition. *HAL (Le Centre Pour La Communication Scientifique Directe)*. https://doi.org/10.1117/12.2309454

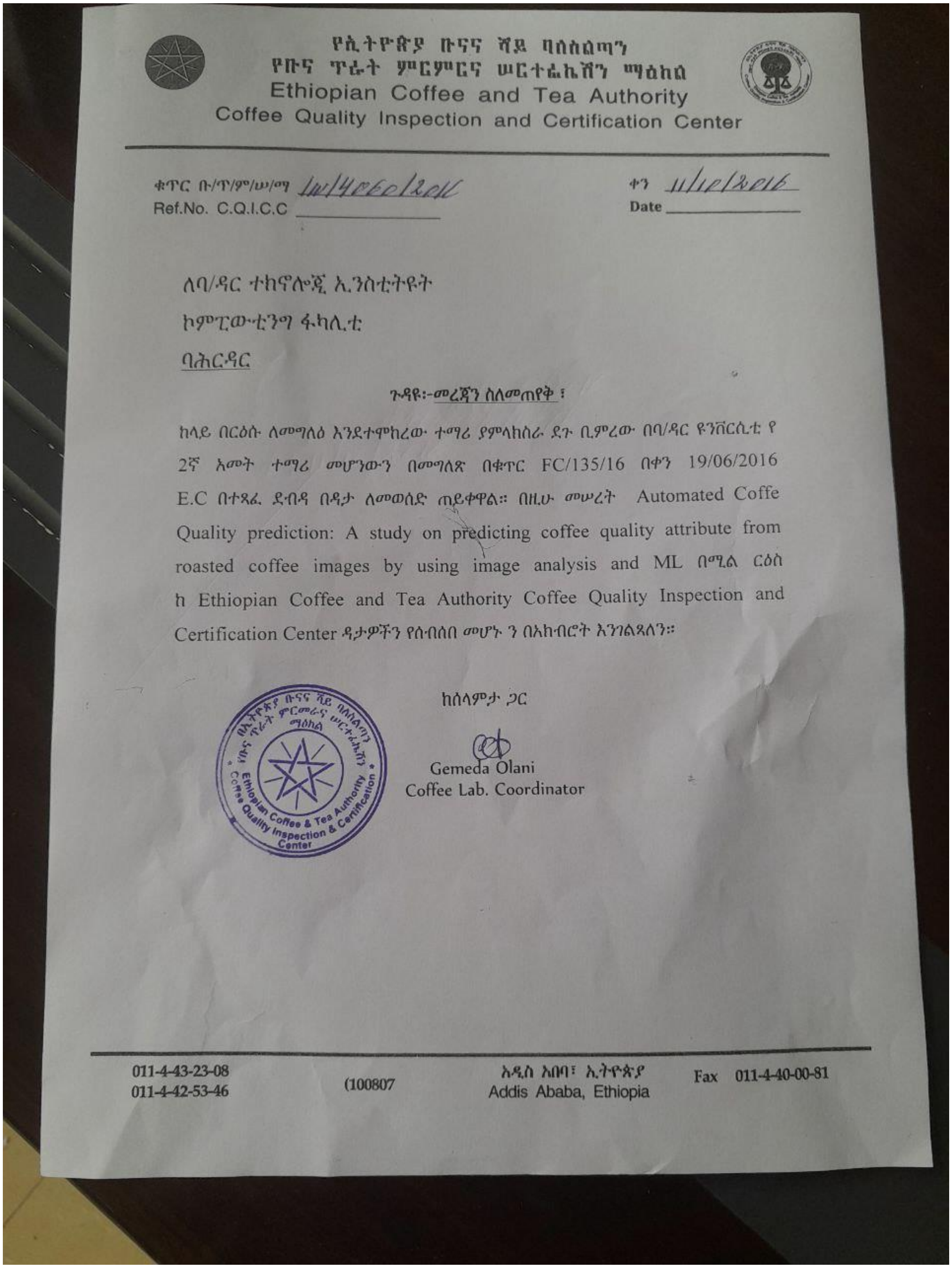
Yogesh Shankar Ghodake, Bhosale, D. V., & Kulkarni, S. S. (2021). Image Processing: Methods, Techniques, Applications Review. *AIJR Proceedings*. https://doi.org/10.21467/proceedings.118.30

Zhang, S., Wu, Y., & Chang, J. (2020). Survey of Image Recognition Algorithms. *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. https://doi.org/10.1109/itnec48623.2020.9084972

Zhang, W., Kasun, L. C., Wang, Q. J., Zheng, Y., & Lin, Z. (2022). A Review of Machine Learning for Near-Infrared Spectroscopy. *Sensors*, *22*(24), 9764. https://doi.org/10.3390/s22249764

Appendix

Appendix 1: Data collection approval letter.



የኢትዮጵያ ቡናና ሻይ ባለስልጣን
የቡና ጥራት ምርምርና ሠርተፊኬሽን ማዕከል
Ethiopian Coffee and Tea Authority
Coffee Quality Inspection and Certification Center

ቁጥር ቡ/ጥ/ም/ሠ/ማ [illegible]
Ref.No. C.Q.I.C.C

ቀን 11/12/2016
Date

ለባ/ዳር ተክኖሎጂ ኢንስቲትዩት

ኮምፒውቲንግ ፋካሊቲ

ባሕርዳር

ጉዳዩ፡-መረጃን ስለመጠየቅ፣

ከላይ በርዕሱ ለመግለፅ እንደተሞከረው ተማሪ ያምላከስራ ደጉ ቢምረው በባ/ዳር ዩንቨርሲቲ የ 2ኛ አመት ተማሪ መሆንውን በመግለጽ በቁጥር FC/135/16 በቀን 19/06/2016 E.C በተጻፈ ደብዳ በዳታ ለመወሰድ ጠይቀዋል። በዚሁ መሠረት Automated Coffe Quality prediction: A study on predicting coffee quality attribute from roasted coffee images by using image analysis and ML በሚል ርዕስ ከ Ethiopian Coffee and Tea Authority Coffee Quality Inspection and Certification Center ዳታዎችን የሰብሰበ መሆኑ ን በአክብሮት እንገልጻለን።

ከሰላምታ ጋር

Gemeda Olani
Coffee Lab. Coordinator

011-4-43-23-08
011-4-42-53-46

(100807

አዲስ አበባ፣ ኢትዮጵያ
Addis Ababa, Ethiopia

Fax 011-4-40-00-81