Robusta Coffee Beans Classification Based on Roast Levels Using Vision Transformer

Dhafindra Razaqa Stefano
Faculty of Informatics
Telkom University
Bandung, Indonesia
dhafindrarazaqa@student.telkomuniversity.ac.id

Ema Rachmawati
Faculty of Informatics
Telkom University
Bandung, Indonesia
emarachmawati@telkomuniversity.ac.id

***Abstract*—The accurate classification of coffee bean roast levels is crucial for delivering high-quality coffee drinks tailored to consumer preferences. Roast levels significantly impact the flavor profile of coffee, with different roasting stages catering to specific brewing methods and taste expectations. Despite the importance of precise classification, many coffee shops still rely on manual processes that are time-consuming and prone to human error. This study presents an automated approach using computer vision and deep learning techniques to classify coffee beans into eight distinct roast levels: Extremely Light, Very Light, Light, Medium Light, Medium, Moderately Dark, Dark, and Very Dark. A Vision Transformer (ViT) model was employed due to its state-of-the-art performance in image classification tasks. The model was trained on a custom dataset containing 3,600 images, evenly distributed across the eight classes. To ensure robust performance, preprocessing techniques, including histogram matching and normalization, were applied. The ViT achieved exceptional results, with a testing accuracy of 0.9778, precision of 0.9791, recall of 0.9778, and F1-score of 0.9777. These findings demonstrate the ViT's effectiveness in distinguishing subtle visual differences between roast levels. This approach offers a scalable and cost-effective solution for automating coffee bean classification, enhancing efficiency and reducing operational errors in the coffee industry.**

***Index Terms*—roast levels, coffee beans classification, computer vision, vision transformer**

I. Introduction

Coffee is one of the most popular beverages worldwide, with its demand and cultural significance growing steadily each year [1]. In response, an increasing number of coffee shops have emerged, each striving to differentiate themselves by offering diverse coffee beverages that cater to a wide range of customer preferences. A key factor influencing coffee flavor profiles is the roast level of the beans, as different coffee beverages require different roast levels to achieve the desired flavor [2]. There are eight recognized levels of coffee bean roast, as described in the book by Massey [3], Extremely Light, Very Light, Light, Medium Light, Medium, Moderately Dark, Dark, and Very Dark, each offering distinct flavor characteristics. Roasting green coffee beans induces chemical changes as the beans are heated, unlocking the aromas and flavors within [4]. However, darker roast levels do not necessarily mean better-tasting coffee. On the contrary, darker roasts increase bitterness while reducing acidity and sweetness [2]. Therefore, each roast level offers unique flavors and has specific applications [2].

Tools like the DiFluid Omni help determine roast levels with high accuracy by analyzing the color of coffee beans using advanced optical sensors. However, the Omni is expensive and limited to small-scale applications, making it impractical for larger operations. To address these challenges, this study proposes a cost-effective alternative for classifying coffee beans based on their roast levels by using computer vision and deep learning. The Vision Transformer (ViT), a state-of-the-art deep learning model [5], was applied for its ability to classify eight coffee bean roast levels, aiming to achieve high accuracy while ensuring scalability and affordability.

The benefit of this research is that it allows coffee shops to reduce costs in coffee production, while at the same time ensuring a more consistent coffee flavors. As a result, not only does the coffee drinks taste better, it is also more affordable, allowing a broader audience to enjoy them. Additionally, the cost savings could enable the coffee shops to achieve profitability more quickly, potentially leading to business expansion by way of opening new branches, which leads to more job opportunities.

This paper is organized into several sections. Section II reviews previous studies on coffee bean roast level classification and the application of machine learning techniques. Section III explains the dataset preparation, preprocessing steps, and the architecture of the Vision Transformer model used in this study. Section IV presents the experimental results, performance metrics, confusion matrices, and a comparative analysis of the two models. Section V summarizes the study's findings and proposes directions for future research.

II. Related Works

Several studies have explored the classification of coffee bean roast levels and the application of machine learning in this domain. Sarino et al. [6] utilized an Artificial Neural Network (ANN) to classify three coffee bean roast levels, namely light, medium and very dark roasts. The study achieved an excellent accuracy of 97.2%. However, the study lacks information regarding the dataset, such as the quantity and the visualization of the images in the dataset

On the other hand, Suryana and Raharja [7] improved upon the study by Sarino et al. [6] by extending the classification to four roast levels instead of three, namely green, light,