Study of Coffee Bean Classification Using Neural Networks

G Goncalves[1], D V Magalhaes[1], C Arcanjo[1], J P M Galdames[1], and V S Bagnato[1,2]

[1]*Física, São Carlos Institute of Physics, São Carlos, Brazil*
[2]*Hagler Fellow, Department of Biomedical Engineering, , Texas A&M University, College Station, USA*
Contact Email: goncalves.engmec@gmail.com

The classification of raw coffee beans is currently performed manually, requiring about 40-80 minutes for a trained evaluator to assess 300 g samples containing approximately 6,500 beans. This process is prone to variations and inconsistencies due to the subjective nature of visual inspection by a human evaluator, creating bottlenecks in production and negatively impacting the coffee supply chain, as this analysis conditions the transport of these coffee lots. Therefore, this research aims to develop an automated computer vision system based on neural networks and cloud data processing to classify raw coffee beans in terms of impurities, defects, and distinguishing the sample beans in a single image obtained through a prototype capable of adequately capturing the sample image. The system was developed by creating a reference database containing around 14,000 images of coffee beans, categorized according to intrinsic and extrinsic defect types defined by MAPA (2003) (Ministry of Agriculture, Livestock, and Supply, 2003 regulation). The images were acquired with a resolution of 35 megapixels per bean unit. This megapixel information was used to train a neural network on the AWS (Amazon Web Services) platform. Initially, contour and defect recognition with individual bean identification were performed using an algorithm in Python. Various defect configurations were evaluated, and the image resolution was gradually reduced to the smallest possible size for each type of defect to determine the minimum effective resolution for identification. The results show that both the neural network and the developed prototype were able to classify coffee beans with high accuracy, using lower resolution images, reducing the average manual classification time from an average of 40 minutes in the traditional method to 1 to 10 seconds with automatic evaluation. The automated system offers a robust and impartial alternative to the traditional manual method, improving the reproducibility and reliability of the classification and potentially adding traceability to the grain classification process. This research presents an innovative solution for coffee classification and establishes a valuable image database for future technical scientific research and commercial applications, significantly contributing to the coffee supply chain.