



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 14, Issue 4, April 2025**

# An Big Data for Informed Decision Making using Convolutional Neural Networks for Coffee Trading

K.Rajapriya, S.Subhashini

Assistant Professor, Department of Computer Science and Engineering, Mrk Institute of Technology, Nattarmangalam,  
Kattumannarkoil, Cuddalore District, Tamil Nadu, India

Student, Department of Computer Science and Engineering, Mrk Institute of Technology, Nattarmangalam,  
Kattumannarkoil, Cuddalore District, Tamil Nadu, India

**ABSTRACT:** Coffee, the second-largest global soft commodity, can take advantage of a comprehensive mining of daily and historical market data for more effective informed trading decisions. The paper presents a design and implementation of a coffee commodity trading Big Data warehouse supporting informed decision-making for trading coffee of small and mid-size enterprises (SMEs), business households, farmers, and others concerned through the visualization analysis dashboards. Specifically, we collected data from different sources, including the coffee market data from the New York Market of Intercontinental Exchange (ICE) from the portal Yahoo Finance, the Commitment of Traders (COT) report, and the Commodity Index Trader (CIT) report from January 2000 to October 2022. The Extract, transform, and load (ETL) process is adopted to ingest coffee futures trading crawled data into the 3 layers data warehouse. Finally, the analytical system will extract and visualize selected key dimensions that influence coffee futures prices within different observation windows and perspectives. Our research focused our visualization on four main parameters: market price, volume, trailing over 250 days, and the highest/lowest 40 days. As a result, we implement a prototype of a coffee trading data warehouse on the crawled data and visualize trends in coffee futures prices.

**KEYWORDS:** artificial intelligence; deep learning; convolutional neural networks (CNNs); coffee and coffee bean

## I. INTRODUCTION

Arabica coffee is a species of coffee tree with small beans and belongs to the Rubiaceae family. This unique species characterized by the features of an evergreen tree, which, through proper breeding in the zone of tropical areas, allows it to maintain its green leafy state almost throughout the year. Coffee beans are used in the preparation of coffee beverage, which is the second most consumed beverage in the world. However, among the 100 species of coffee, it is Arabica coffee that has the greatest economic importance due to its flavor, aroma, low caffeine content, cultivation requirements and product price, among other reasons.

Arabica coffee is one of many group of food products that exhibit beneficial effects on the human body. This is related to its content of key bioactive compounds, such as caffeine, lipids, polyphenols, vitamins, minerals and aroma compounds. In the literature, the consumption of Arabica coffee has been shown to help neutralize free radicals in the body, reducing the risk of chronic diseases including cardiovascular disease, Alzheimer's and Parkinson's disease, and liver disease.

The coffee bean also represents a valuable dietary source due to, among other things, the chlorogenic acids (CGAs) they contain. This is a group of polyphenols found naturally in coffee beans. They are also a powerful antioxidant, which, as mentioned, in addition to inhibiting certain diseases, can promote weight loss by, among other things, affecting metabolism and glucose absorption. It seems crucial to assess the quality of coffee beans due to the impact of sourcing the right beans during the coffee roasting process, where the quantity and quality of the CGA they contain can degrade rapidly. Therefore, the process of sourcing and selecting the right beans is important in analyzing the health properties of coffee. The interest in health properties has attracted the attention of scientists as well as healthy eaters.



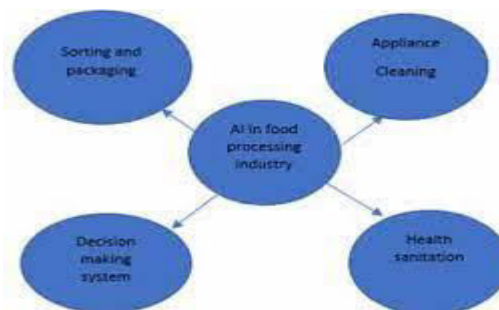


Fig 1: Major application of AI in food industry

Artificial intelligence methods have gained great importance in the food production and distribution process through process optimization, crop monitoring and management, quality assessment [19,20] and food prediction, among others.

One of the artificial intelligence methods includes deep learning, which is a technique that uses artificial neural network models containing, among other things, two or more layers of hidden neurons. Deep learning models allow for data analysis without the researcher knowing the exact rules. This network topology processes information and is a tool that offers the mapping of the complex relationships occurring between the analyzed data. Actions such as extracting the shape, color or texture of an analyzed object in a bitmap image requires a large amount of data, which translates into longer processing time and statistical analysis of that object. The application of deep learning methods makes it possible to automate some of the activities, among others, related to image data processing at various stages of statistical analysis.

## II. LITERATURE SURVEY

The integration of fog computing, big data analysis, and optimization techniques can enhance decision-making in coffee commodity trading. Studies have shown that smart ant colony optimization can improve task offloading efficiency (Kishor & Chakarbarty, 2022), while big data analysis can extract valuable insights from large datasets (Le, 2022). Effective data management using ETL tools (Singh, 2022) and reliable data sources such as World Bank Commodities Price Data (World Bank, 2023) and International Coffee Organization reports (ICO, 2022) support accurate market forecasting. These findings imply improved efficiency, reduced costs, and better risk management in coffee trading.

Future research directions include integrating AI, blockchain, and real-time monitoring systems to further enhance transparency and accuracy in coffee commodity trading. The international coffee market involves complex supply chains, and risk management strategies mitigate market volatility. Market forecasting models predict price fluctuations, and sustainable practices ensure longterm coffee supply. Research addresses global coffee market dynamics, regional trends, and stakeholder collaboration. Innovative technologies transform coffee trading, enhancing supply chain efficiency and sustainability. By exploring these themes, researchers and practitioners can optimize coffee commodity trading, ensuring a resilient and prosperous coffee industry.

The convolutional networks, also called convolutional neural networks (CNNs), are a case of deep learning that analyze individual areas of the input dataset and generate new tensors based on them. By means of convolution operations, CNNs use a kernel function and involve increasing the ordinality of copies of the tensor with which operations are performed without increasing their dimensionality. Convolutional networks use an algorithm to classify images due to the presence of repeated, recognizable and similar shapes in the image. The input tensor shape of the network accounts for the sample (pattern), image height, image width, and depth as the number of dimensions of the RGB color space.

In the literature, from the point of view of the problem posed, many researchers have tried to apply CNNs using digital images to assess the quality of green coffee beans during harvesting, for example, by the team of Huang et al. (2020), among others, who achieved a coffee grading rate of 93%. On the other hand, the team of Pinto et al. (2020), based on

green coffee beans, similarly made a quality assessment according to defects, obtaining a grading rate of min. 72.4%. Research was also conducted by Wang et al. (2021) to effectively build an intelligent system based on deep learning and computer vision to assist in the detection of defects, including mold, fermentation, insect bites and crushed coffee beans, obtaining neural models with recognition quality rates of 78% and 93%.

In view of a number of research works aimed at improving the evaluation of the quality of harvested coffee beans, the current technological solutions in the process of coffee harvesting and roasting are still being sought to be made more effective.

In this work, as mentioned above, the focus was on recognizing three quality classes of coffee beans because of the coffee roasting process. The assumption of the study was to assess the authenticity based on the observation of the color properties of selected Arabica coffee samples obtained through the processing of digital images. The utilitarian purpose was to develop convolutional networks to assist in the identification of quality classes of Arabica coffee beans. The application of modern methods using the Python 3 environment seems to be a key aspect. This is due to the fact that it is easier to implement this type of solution in the food industry than traditional statistical methods. The designed convolutional models, according to the machine learning criterion, have high scalability, which makes it possible to adapt to applications in the food industry.

### III. METHODS

Future work will be needed to collect data from the other markets and compare the correlation between different markets. We also want to define more parameters that can affect the coffee price. We tend to build an Android app based on cloud computing and use model machine learning, model computer vision, or natural language processing to build a real-time prediction application. The market for derivatives develops in parallel with the need to manage price changes of critical commodities. According to modern financial theory, introducing a derivative market in a spot market helps to complete the process of market-based pricing and prevent the risk of price fluctuations in the spot market at a low cost. In research authors synthesized a complete case study of commodity futures and agricultural option markets regarding the inter-time price relationship of stocks, speculation, price behavior, and institutional problems. The authors raise controversial issues and challenges for the studies focusing on risk management and market strategy, price and volatility behavior, e-transactions and trading funds, and international framework. In the coffee derivatives market, these problems will be affected by similar effects and problems. In the future, we will focus on data analysis and discovery and build effective prediction models for the coffee derivative market based on deep learning or machine learning algorithms.

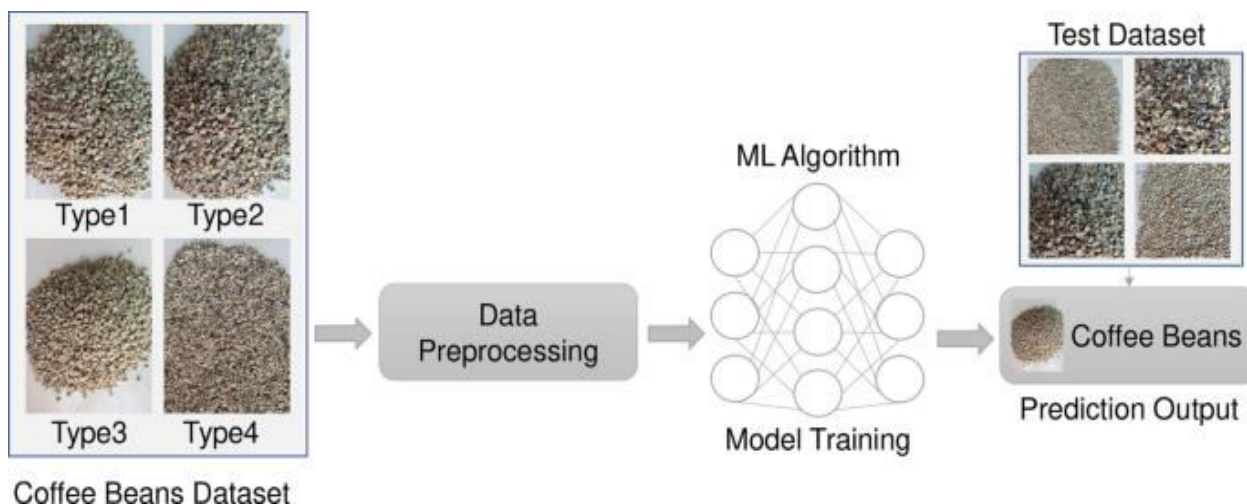


Fig 2: Machine learning techniques for coffee classification

The future scope of AI-driven coffee commodity trading holds immense potential for growth and innovation. Integration with emerging technologies such as blockchain, Internet of Things (IoT), and natural language processing (NLP) can further enhance transparency, security, and decision-making capabilities. Expanding the system to incorporate real-time weather forecasting, soil health monitoring, and farm-level data can improve predictive accuracy. Additionally, exploring applications in other commodity markets, such as cocoa and sugar, can broaden the system's impact. Advancements in explainable AI (XAI) and transfer learning can also improve model interpretability and adaptability. Furthermore, integrating with trading platforms and exchanges can enable seamless execution of AI-generated trading recommendations, revolutionizing the coffee commodity trading landscape.

#### IV. RESULTS AND DISCUSSION

The knowledge that the RGB model is based on the light reflected in a fixed environment may prompt an appropriate representation of these colors for a food product in reality. In having an accurate representation of these colors using the RGB model for a food product, there is greater quality control and it leads to maintaining the authenticity of that product. Currently, techniques are being used to adulterate food products due to their color to meet consumer expectations and at the same time extend their shelf life. The RGB color space model, together with machine learning, may soon become a key aspect for determining the quality grade and checking the authenticity of food products. As a result of image acquisition, the dependence on illumination has been established, and this allows for the adjustment of the visual presentation with the help of the RGB model, minimizing the differences between the digital image and the actual visual impression. Nowadays, through increasing consumer awareness, it is also required that a food product contain information as described for authenticity in the production and the distribution of products. By the same means, the RGB color space model, by distinguishing more precisely between different classes of coffee beans than the  $L^*a^*b^*$  model, can be of greater benefit in assessing the quality of these products.

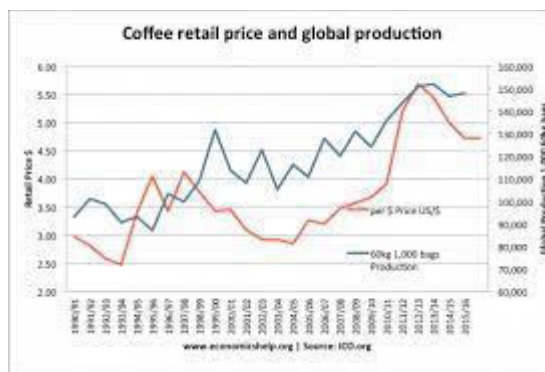


Fig 3: Are Coffee Prices Fair

The designed Sequential model contained eight input layers, including four convolution and four MaxPooling layers, one global average layer and one output layer. The convolution window in the convolutional layers was shifted by two units instead of one, which resulted in a faster dimensionality reduction in large tensors in the network layers. After a series of convolution and reduction operations, the feature tensor was transformed to flat link to the dense layer for the final classifier. For the output layer, a sigmoid function was used, as it gave better results than the more commonly used softmax function in convolutional networks. For the input layers, the Rectified Linear Unit (ReLU) function was used, which has emerged as the more effective activation function in this issue as in other studies. This is due to the fact that the ReLU algorithm relative to other functions, including sigmoidal and hyperbolic tangent (tanh), prevents gradient from fading when learning convolutional models as well as other types of neural networks.

#### V. CONCLUSIONS

The statistical analysis by ANOVA for color helped distinguish exactly three classes of coffee beans in the RGB color space model, which translated into the efficient recognition of coffee defects obtained during the coffee roasting process. The  $L^*a^*b^*$  color model demonstrated the difficulty in distinguishing between the quality classes of under-

roasted and properly roasted coffee beans. Nevertheless, the  $L^*a^*b^*$  model successfully distinguished between the class of over-roasted coffee. Analyzing the properties of coffee color parameters as an indicator of authenticity can make valuable contributions to food quality and safety. More challenging was the classification based on digital images with a resolution of  $678 \times 678$  by, among other things, the multidimensionality and complexity of the information contained in the acquired image as a result of image acquisition. The application of deep artificial neural networks made it possible to classify coffee bean samples using computer vision based on image pixels. As a result, an adequate CNN model was acquired, which was characterized by a high fit factor for the test set of 0.81. The quality assessment of coffee on the basis of representative characteristics (in this case, color) significantly affects both producers and consumers. The effective evaluation of coffee quality will maintain a high product standard in the industrial coffee-roasting process. Consumers will be inclined to select and consume the right kind of coffee through taste and satisfaction with the obtained high-quality food product.

## REFERENCES

1. Yu, Y.; Cao, Y.; Lai, Q.; Zhao, Q.; Sun, Z.; Zhou, S.; Song, D. Design and Operation Parameters of Vibrating Harvester for *Coffea arabica* L. *Agriculture* **2023**, 13, 700. [Google Scholar] [CrossRef]
2. Batista, L.R.; Chalfoun de Souza, S.M.; Silva e Batista, C.F.; Schwan, R.F. Coffee: Types and Production. In *Encyclopedia of Food and Health*; Elsevier Ltd.: Amsterdam, The Netherlands, 2015. [Google Scholar]
3. Bolka, M.; Emire, S. Effects of Coffee Roasting Technologies on Cup Quality and Bioactive Compounds of Specialty Coffee Beans. *Food Sci. Nutr.* **2020**, 8, 6120–6130. [Google Scholar] [CrossRef] [PubMed]
4. Rusinek, R.; Dobrzański, B.; Oniszcuk, A.; Gawrysiak-Witulska, M.; Siger, A.; Karami, H.; Ptasińska, A.A.; Żytek, A.; Kapela, K.; Gancarz, M. How to Identify Roast Defects in Coffee Beans Based on the Volatile Compound Profile. *Molecules* **2022**, 27, 8530. [Google Scholar] [CrossRef] [PubMed]
5. Liczbiński, P.; Bukowska, B. Tea and Coffee Polyphenols and Their Biological Properties Based on the Latest in Vitro Investigations. *Ind. Crops Prod.* **2022**, 175, 114265. [Google Scholar] [CrossRef]
6. Yesil, A.; Yilmaz, Y. Review Article: Coffee Consumption, the Metabolic Syndrome and Non-Alcoholic Fatty Liver Disease. *Aliment. Pharmacol. Ther.* **2013**, 38, 1038–1044. [Google Scholar] [CrossRef]
7. Tsai, C.F.; Jioe, I.P.J. The Analysis of Chlorogenic Acid and Caffeine Content and Its Correlation with Coffee Bean Color under Different Roasting Degree and Sources of Coffee (*Coffea arabica* Typica). *Processes* **2021**, 9, 2040. [Google Scholar] [CrossRef]
8. Mitiku, H.; Kim, T.Y.; Kang, H.; Apostolidis, E.; Lee, J.Y.; Kwon, Y.I. Selected Coffee (*Coffea arabica* L.) Extracts Inhibit Intestinal  $\alpha$ -Glucosidases Activities in-Vitro and Postprandial Hyperglycemia in SD Rats. *BMC Complement. Med. Ther.* **2022**, 22, 249. [Google Scholar] [CrossRef]
9. Ascherio, A.; Zhang, S.M.; Hernán, M.A.; Kawachi, I.; Colditz, G.A.; Speizer, F.E.; Willett, W.C. Prospective Study of Caffeine Consumption and Risk of Parkinson's Disease in Men and Women. *Ann. Neurol.* **2001**, 50, 56–63. [Google Scholar] [CrossRef]
10. Camandola, S.; Plick, N.; Mattson, M.P. Impact of Coffee and Cacao Purine Metabolites on Neuroplasticity and Neurodegenerative Disease. *Neurochem. Res.* **2019**, 44, 214–227. [Google Scholar] [CrossRef]
11. Qi, H.; Li, S. Dose-Response Meta-Analysis on Coffee, Tea and Caffeine Consumption with Risk of Parkinson's Disease. *Geriatr. Gerontol. Int.* **2014**, 14, 430–439. [Google Scholar] [CrossRef]
12. Hu, G.; Bidel, S.; Jousilahti, P.; Antikainen, R.; Tuomilehto, J. Coffee and Tea Consumption and the Risk of Parkinson's Disease. *Mov. Disord.* **2007**, 22, 2242–2248. [Google Scholar] [CrossRef] [PubMed]
13. Król, K.; Gantner, M.; Tatarak, A.; Hallmann, E. The Content of Polyphenols in Coffee Beans as Roasting, Origin and Storage Effect. *Eur. Food Res. Technol.* **2020**, 246, 33–39. [Google Scholar] [CrossRef]
14. Tajik, N.; Tajik, M.; Mack, I.; Enck, P. The Potential Effects of Chlorogenic Acid, the Main Phenolic Components in Coffee, on Health: A Comprehensive Review of the Literature. *Eur. J. Nutr.* **2017**, 56, 2215–2244. [Google Scholar] [CrossRef] [PubMed]
15. Reis, C.E.G.; Dórea, J.G.; da Costa, T.H.M. Effects of Coffee Consumption on Glucose Metabolism: A Systematic Review of Clinical Trials. *J. Tradit. Complement. Med.* **2019**, 9, 184–191. [Google Scholar] [CrossRef]





INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details