Machine Learning-Based Rice Variety Classification for Authenticity and Quality Control

Dr. T. Venkata Ramana[1], M Varaprasad Rao[2], Dr. Srikanth Lakumarapu[3], Dr.Banoth Samya[3]

[1]Associate Professor, Department of CSE (AIML), CVR College of Engineering, Hyderabad, India.
meetramana1204@gmail.com
[2]Department of CSE(Data Science),CVR College of Engineering, Hyderabad, India.
varam78@gmail.com
[3]Associate Professor,Department of CSE,CVR College of Engineering, Hyderabad, India.
dr.srikanthcse@gmail.com&samyabanoth@gmail.com

ABSTRACT
Rice serves as a staple food for more than half of the global population, making its authenticity and quality vital to consumer trust and food security. With numerous rice varieties available in the market, accurately identifying and classifying these varieties is crucial for maintaining supply chain integrity. Traditional methods of rice variety classification often rely on manual inspection and laboratory testing, which are time-consuming, resource-intensive, and prone to errors.This research explores the potential of advanced computer vision techniques and machine learning algorithms to automate and enhance rice variety classification. By training models on extensive datasets containing rice images and genetic information, the proposed system aims to autonomously and accurately identify rice varieties. The integration of machine learning enables the extraction of subtle visual and genetic features that are difficult to discern through traditional methods, ensuring precise and reliable classification.The adoption of this AI-driven approach has significant implications for the rice supply chain. It not only ensures consumers receive the quality and variety of rice they expect but also supports stakeholders in maintaining transparency and traceability. This work demonstrates how cutting-edge technologies can revolutionize rice classification processes, contributing to greater efficiency, authenticity, and trust across the global food supply chain.

Keywords: Rice variety, Supply chain management, Predictive analytics, Data analytics, Machine learning.

1. INTRODUCTION

Rice is a fundamental dietary staple for over 50% of the global population, serving as a primary source of calories and nutrition, particularly in Asia, Africa, and Latin America. According to the Food and Agriculture Organization (FAO), global rice production reached approximately 515 million metric tons in 2022, with Asia accounting for nearly 90% of this total. As one of the most widely consumed grains, ensuring the authenticity and quality of rice is essential for maintaining consumer trust, food security, and the economic stability of the agricultural sector.The global rice industry features a vast diversity of varieties, each with distinct physical, genetic, and aromatic traits. However, distinguishing between these varieties is a complex process that has traditionally relied on manual inspection and laboratory-based testing methods. These conventional approaches are time-intensive, require significant expertise, and often fail to detect subtle differences, leading to inaccuracies in classification. Misclassification not only impacts consumer satisfaction but also threatens supply chain integrity, resulting in financial losses and compromised food quality.

With advancements in technology, computer vision and machine learning have emerged as transformative tools in agricultural classification tasks. By analyzing visual and genetic data, these techniques can identify patterns and features with remarkable precision. For example, studies have shown that machine learning algorithms achieve classification accuracies exceeding 95% for various agricultural products, highlighting their potential to revolutionize rice variety identification.This research leverages advanced computer vision techniques and machine learning algorithms to automate the classification of rice varieties. By training models on extensive datasets of rice images and genetic information, this system aims to enhance the efficiency and accuracy of rice variety classification. Such an innovation not only ensures consumers receive the desired rice variety but also strengthens traceability and transparency throughout the supply chain. This work underscores the pivotal role of AI-driven technologies in meeting the growing demand for food authenticity and quality assurance in an increasingly globalized market.

2. LITERATURE SURVEY

[1,2,3]. In recent years, the combination of blockchain technology with artificial intelligence, big data, 5G, and the industrial internet have been explored by researchers to strengthen regulatory capabilities, which has been mainly reflected in the following aspects [4,5,6]. Firstly, artificial intelligence (AI) and smart contracts were combined to solve the problem of redundancy of blockchain information and improved supervision efficiency [7,8]. Secondly, blockchain technology and big data technology were combined to unify different data sources and realize unified data supervision [9.10]. Thirdly, blockchain technology and 5G technology were combined to solve the problem of slow real-time data transmission [11,12]. Fourthly, the blockchain was combined with the industrial internet, and the precise traceability of regulatory information was achieved through identification analysis [13]. Compared with the traditional agricultural and food supply chain supervision model, the "blockchain+" model can ensure the safety and credibility of the data in the agricultural and food supply chain. The credible traceability and precise accountability of the agricultural products and food data can be realized, thereby improving the supervision of the agricultural and food supply chain efficiency and authenticity.

The rice supply chain is characterized by complex links, diverse data types, and long-life cycles. The application of the blockchain and smart contracts has promoted the digitization and intelligence of the rice supply chain, and the supervision of the rice supply chain by the regulatory authorities has been improved to a certain extent. However, as the amount of data has increased, the application of a blockchain and smart contracts in the supervision of the rice supply chain has encountered the following shortcomings.The research on blockchains in the rice supply chain is mostly on single link blockchains such as the "production blockchain", "processing blockchain", and "storage blockchain".

3. PROPOSED SYSTEM

TheProject overview for a food authentication project specifically focused on the classification of rice varieties. The script uses machine learning and deep learning techniques to build and evaluate classification models. Below is a description of the key components and steps in the code:

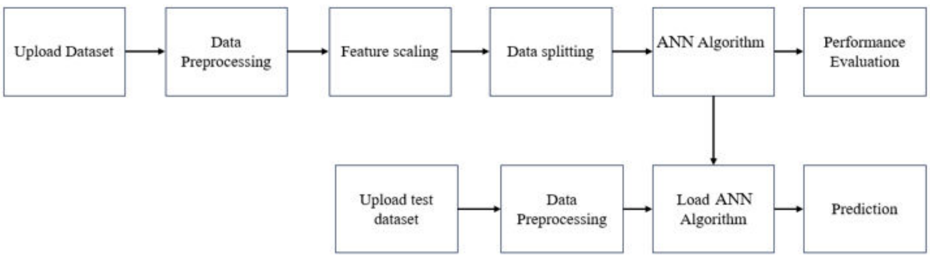

Figure 1: Block diagram of proposed ANN Model.

Although today the Perceptron is widely recognized as an algorithm, it was initially intended as an image recognition machine. It gets its name from performing the human-like function of perception, seeing, and recognizing images. Interest has been centered on the idea of a machine which would be capable of conceptualizing inputs impinging directly from the physical environment of light, sound, temperature, etc. — the "phenomenal world" with which we are all familiar — rather than requiring the intervention of a human agent to digest and code the necessary information. Rosenblatt's perceptron machine relied on a basic unit of computation, the neuron. Just like in previous models, each neuron has a cell that receives a series of pairs of inputs and weights. The major difference in Rosenblatt's model is that inputs are combined in a weighted sum and, if the weighted sum exceeds a predefined threshold, the neuron fires and produces an output.

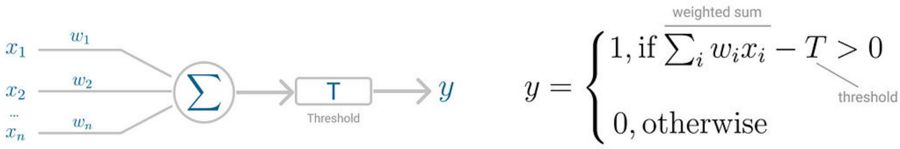

Fig. 2: Perceptron neuron model (left) and threshold logic (right).

Threshold T represents the activation function. If the weighted sum of the inputs is greater than zero the neuron outputs the value 1, otherwise the output value is zero.

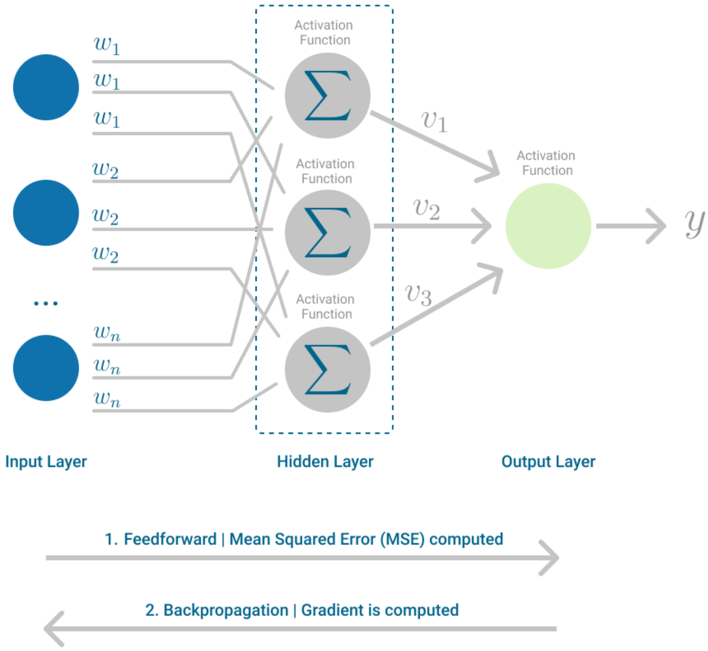

Fig. 3: Artificial neural network, highlighting the Artificial neural network and Backpropagation steps.

4. RESULTS AND DISCUSSION

Figure 4represents the dataset used for rice variety classification. It includes visual representations of rice images or relevant features used for the classification task.Figure 5 showcases an image displaying comprehensive information about the complete dataset used for rice variety classification. It includes details such as the number of samples, features, and other dataset characteristics.Figure 6 presents a count plot depicting the distribution of the target column in the dataset. It provides insights into the balance or imbalance of different classes related to rice varieties.Figure 7 illustrates the features of the dataset after undergoing preprocessing steps. Preprocessing includes tasks such as cleaning, normalization, or feature engineering to prepare the data for model training.Figure 8 shows the confusion matrix for a model trained using XGBoost. The confusion matrix provides a detailed breakdown of true positive, true negative, false positive, and false negative predictions, aiding in the assessment of model behavior.

	Area Integer	Perimeter Real	Major_Axis_Length Real	Minor_Axis_Length Real	EccentricityReal	Convex_Area Integer	Extent Real	Class
0	15231	525.578979	229.749878	85.093788	0.928882	15617	0.572896	Cammeo
1	14656	494.311005	206.020065	91.730972	0.895405	15072	0.615436	Cammeo
2	14634	501.122009	214.106781	87.768288	0.912118	14954	0.693259	Cammeo
3	13176	458.342987	193.337387	87.448395	0.891861	13368	0.640669	Cammeo
4	14688	507.166992	211.743378	89.312454	0.906691	15262	0.646024	Cammeo
...	...	...	...	...	...	...	...	...
3805	11441	415.858002	170.486771	85.756592	0.864280	11628	0.681012	Osmancik
3806	11625	421.390015	167.714798	89.462570	0.845850	11904	0.694279	Osmancik
3807	12437	442.498993	183.572922	86.801979	0.881144	12645	0.626739	Osmancik
3808	9882	392.296997	161.193985	78.210480	0.874406	10097	0.659064	Osmancik
3809	11434	404.709992	161.079269	90.868195	0.825692	11591	0.802949	Osmancik

3810 rows × 8 columns

Figure 4: Image shows the dataset used for rice variety.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3810 entries, 0 to 3809
Data columns (total 8 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   Area Integer            3810 non-null   int64  
 1   Perimeter Real          3810 non-null   float64
 2   Major_Axis_Length Real  3810 non-null   float64
 3   Minor_Axis_Length Real  3810 non-null   float64
 4   EccentricityReal        3810 non-null   float64
 5   Convex_Area Integer     3810 non-null   int64  
 6   Extent Real             3810 non-null   float64
 7   Class                   3810 non-null   object 
dtypes: float64(5), int64(2), object(1)
memory usage: 238.2+ KB
```

Figure 5: Displays the information of complete dataset used for Rice variety classification.

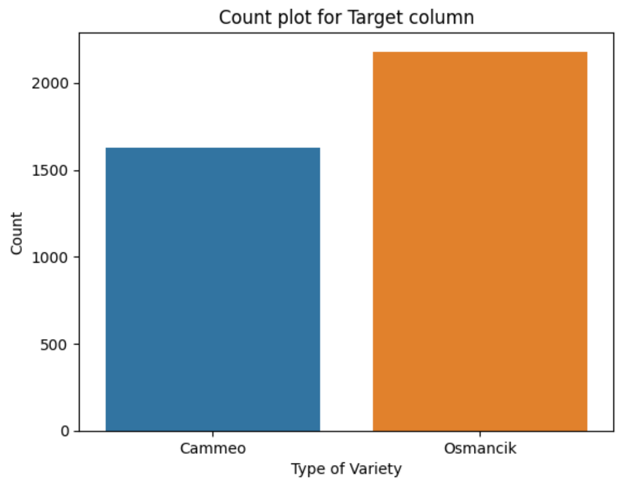

Figure 6: Count plot of Target column of a dataset.

```
array([[ 1.47982953,  2.0043543 ,  2.34854657, ...,  2.01833746,
         1.49965944, -1.15292093],
       [ 1.14787029,  1.12585309,  0.98839042, ...,  0.41001813,
         1.19291767, -0.60207876],
       [ 1.13516924,  1.31721425,  1.45190846, ...,  1.21295648,
         1.12650386,  0.405611  ],
       ...,
       [-0.13320373, -0.32985087, -0.29824512, ..., -0.27509915,
        -0.17306812, -0.45573108],
       [-1.60825742, -1.74032002, -1.58097116, ..., -0.59882135,
        -1.60715621, -0.03716757],
       [-0.71225612, -1.39156604, -1.58754648, ..., -2.93916012,
        -0.76628981,  1.82594693]])
```

Figure 7: Features of a dataset after preprocessing.

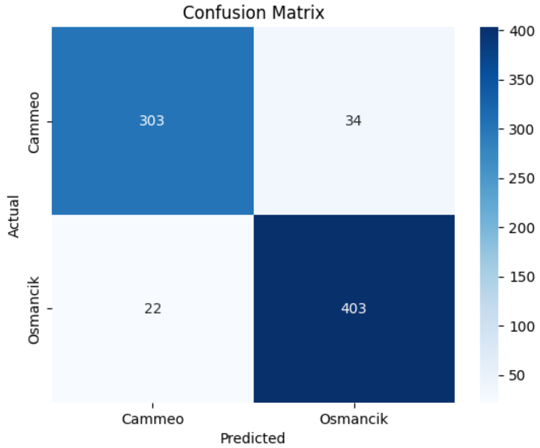

Figure 8: Confusion matrix of XGBOOST model.

XGBoost Classifier Classification_report:

	precision	recall	f1-score	support
Cammeo	0.93	0.90	0.92	337
Osmancik	0.92	0.95	0.94	425
accuracy			0.93	762
macro avg	0.93	0.92	0.93	762
weighted avg	0.93	0.93	0.93	762

Figure 9: Classification report of XGBOOST classifier.

Figure 9showcases the classification report generated for a classifier using the XGBoost algorithm. The classification report typically includes metrics such as precision, recall, and F1-score for each class, providing a comprehensive evaluation of the classifier's performance.Figure 10displays the confusion matrix for Artificial Neural network. It provides a detailed breakdown of predictions, helping assess the neural network's performance.Figure 11presents the classification report generated for anArtificial Neural network. It includes metrics such as precision, recall, and F1-score for each class, offering insights into the neural network's performance.

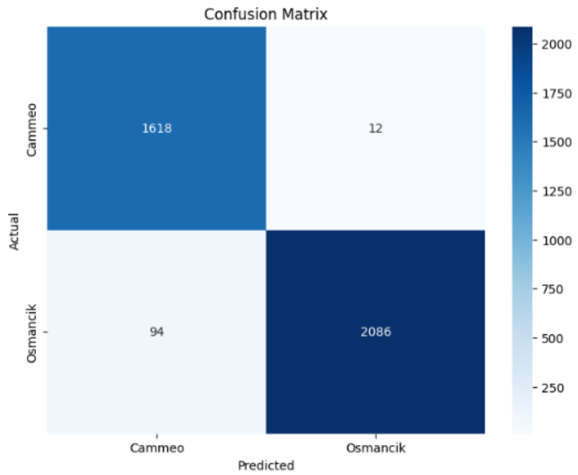

Figure 10: Confusion matrix of Artificial neural network

	precision	recall	f1-score	support
Cammeo	0.95	0.99	0.97	1630
Osmancik	0.99	0.96	0.98	2180
accuracy			0.97	3810
macro avg	0.97	0.97	0.97	3810
weighted avg	0.97	0.97	0.97	3810

Figure 11: classification report of Artificial neural network

Table 1: Performance comparison of quality metrics obtained using XGBoost and Artificial Neural network model.

Model	XGBoost	Artificial Neural Network
Accuracy (%)	93	97
Precision (%)	92	97
Recall (%)	93	97
F1-score (%)	93	97

5. CONCLUSION

In conclusion, thiswork addressed a critical need in the global rice market. With rice being a staple for a significant portion of the world's population, ensuring the authenticity and quality of rice varieties is paramount. The traditional methods of identification, though accurate, are often inefficient for large-scale supply chains. The project proposes a solution by leveraging advanced computer vision techniques and machine learning algorithms to enhance the accuracy and efficiency of rice variety classification.By combining extensive datasets of rice images and genetic information, the project aims to create a system capable of autonomously and accurately identifying different rice varieties. The integration of machine learning enables the extraction of subtle visual and genetic features, overcoming challenges associated with manual inspection and limited genetic testing. This advancement not only ensures the integrity of the rice supply chain but also contributes to consumer confidence, fraud prevention, and overall food security.

REFERENCES

[1] Egala, B.S.; Pradhan, A.K.; Badarla, V.; Mohanty, S.P. Fortified-Chain: A Blockchain-Based Framework for Security and Privacy-Assured Internet of Medical Things with Effective Access Control. IEEE Internet Things J. 2021, 8, 11717–11731.
[2] Li, W.; Feng, C.; Zhang, L.; Xu, H.; Cao, B.; Imran, M.A. A Scalable Multi-Layer PBFT Consensus for Blockchain. IEEE Trans. Parallel Distrib. Syst. 2020, 32, 1146–1160.
[3] Agyekum, K.O.-B.O.; Xia, Q.; Sifah, E.B.; Cobblah, C.N.A.; Xia, H.; Gao, J. A Proxy Re-Encryption Approach to Secure Data Sharing in the Internet of Things Based on Blockchain. IEEE Syst. J. 2021, 16, 1685–1696.
[4] Peng, S.; Hu, X.; Zhang, J.; Xie, X.; Long, C.; Tian, Z.; Jiang, H. An Efficient Double-Layer Blockchain Method for Vaccine Production Supervision. IEEE Trans. NanoBioscience 2020, 19, 579–587.
[5] Rachakonda, L.; Bapatla, A.K.; Mohanty, S.P.; Kougianos, E. SaYoPillow: Blockchain-Integrated Privacy-Assured IoMT Framework for Stress Management Considering Sleeping Habits. IEEE Trans. Consum. Electron. 2020, 67, 20–29.
[6] Yanez, W.; Mahmud, R.; Bahsoon, R.; Zhang, Y.; Buyya, R. Data Allocation Mechanism for Internet-of-Things Systems with Blockchain. IEEE Internet Things J. 2020, 7, 3509–3522.
[7] Huang, C.; Wang, Z.; Chen, H.; Hu, Q.; Zhang, Q.; Wang, W.; Guan, X. RepChain: A Reputation-Based Secure, Fast, and High Incentive Blockchain System via Sharding. IEEE Internet Things J. 2021, 8, 4291–4304.
[8] Wang, L.; He, Y.; Wu, Z. Design of a Blockchain-Enabled Traceability System Framework for Food Supply Chains. Foods 2022, 11, 744.

[9] Giraldo, F.D.; Barbosa Milton, C.; Gamboa, C.E. Electronic Voting Using Blockchain and Smart Contracts: Proof of Concept. IEEE Lat. Am. Trans. 2020, 18, 1743–1751.
[10] Kshetri, N.; DeFranco, J. The Economics Behind Food Supply Blockchains. Computer 2020, 53, 106–110.
[11] Katsikouli, P.; Wilde, A.S.; Dragoni, N.; Høgh-Jensen, H. On the benefits and challenges of blockchains for managing food supply chains. J. Sci. Food Agric. 2020, 101, 2175–2181.
[12] Li, X.; Huang, D. Research on Value Integration Mode of Agricultural E-Commerce Industry Chain Based on Internet of Things and Blockchain Technology. Wirel. Commun. Mob. Comput. 2020, 2020, 8889148.
[13] Shahid, A.; Almogren, A.; Javaid, N.; Al-Zahrani, F.A.; Zuair, M.; Alam, M. Blockchain-Based Agri-Food Supply Chain: A Complete Solution. IEEE Access 2020, 8, 69230–69243.