www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

DEEP LEARNING BASED RICE VARIETY CLASSIFICATION FOR SUPPLY CHAIN INTEGRITY

K. Rajashekar[1*], Bolloju Srivalli[2], Padidala Lahari[2], Yalla Janardhan Reddy[2], Gudikandula Ajay Kumar[2]
[1]Assistant Professor, [2]UG Student, [1,2]Department of CSE(AI&ML)
[1,2]Vaagdevi College of Engineering (UGC - Autonomous), Bollikunta, Warangal, Telangana, India.
*Corresponding Email: K. Rajashekar (rajashekar_k@vaagdevi.edu.in)

ABSTRACT

In agriculture and food authentication, ensuring the correct classification of rice varieties such as Cammeo and Osmancik is crucial for maintaining supply chain integrity, quality control, and market pricing. Traditionally, this classification has been performed manually through visual inspection by experts who rely on physical characteristics such as grain size, shape, and texture. However, this manual system is time-consuming, labor-intensive, and highly prone to human error and subjectivity, often leading to misclassification and loss of consumer trust. The manual approach also lacks scalability when handling large datasets or real-time inspection at industrial scales. These limitations highlight the need for an automated, accurate, and scalable solution. Motivated by the growing importance of precision agriculture and the need to modernize quality assessment techniques, this project leverages machine learning and deep learning algorithms for rice variety classification using geometric features extracted from grain silhouettes. Specifically, it applies models like XGBoost and MLPClassifier to a morphological dataset consisting of measurements such as area, perimeter, major/minor axis lengths, eccentricity, extent real, and convex area, aiming to automate and enhance classification accuracy. The objective is to develop an intelligent GUI-based desktop application using Tkinter that enables both administrators and users to easily upload data, visualize metrics, and make predictions. Exploratory data analysis and feature standardization are performed to ensure model robustness, and model persistence is achieved using joblib for reusability. The proposed system significantly outperforms manual methods, with the MLPClassifier achieving a high accuracy of 92.78% compared to XGBoost's 61.42%, thus offering a reliable and automated alternative. This AI-driven approach ensures consistency, reduces dependency on human judgment, and paves the way for integration into real-time industrial applications where speed and accuracy are essential, marking a shift toward smarter and more efficient food authentication systems.

Keywords: Rice Variety Classification, Food Authentication, MLPClassifier, Morphological Dataset, Geometric Features.

1. INTRODUCTION

India is one of the largest producers and exporters of rice, contributing around 22% of the global rice supply. With over 6,000 rice varieties cultivated across different regions, maintaining authenticity in rice classification is crucial to ensure market trust, fair trade, and food safety. Historically, rice classification has been performed manually through visual inspection or chemical testing, which are often time-consuming, expensive, and error-prone. The increasing demand for organic and premium varieties such as Cammeo and Osmancik has intensified the need for precise identification methods. According to a report by the Indian Council of Agricultural Research (ICAR), 15-20% of rice sold is often mislabeled, leading to economic losses and consumer dissatisfaction. This project proposes an AI-driven classification system using geometric features of rice grains and machine learning models like XGBoost and MLPClassifier. It aims to automate, accelerate, and enhance the reliability of rice variety

classification for supply chain integrity. Rice variety classification plays a vital role in food authentication and pricing mechanisms. With applications in agriculture, food quality control, and trade regulation, automated systems can replace manual inaccuracies. Using machine learning improves precision and scalability. This project ensures accurate classification using grain morphological data.

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" [14,15,16]

3. PROPOSED SYSTEM

Step 1: Dataset Collection

The first and foundational step in this research involves the acquisition of a reliable and comprehensive dataset specifically curated for rice variety classification. The dataset used comprises images and feature-based attributes representing different varieties of rice grains. These features typically include grain length, width, perimeter, area, and shape-related metrics derived from high-quality samples. The dataset must be representative of various commonly cultivated rice varieties to ensure the model generalizes well. Publicly available rice datasets from Kaggle or UCI repositories, or datasets manually curated from agricultural research centers, are typically used for this purpose. The dataset serves as the backbone of the project, influencing the performance of both the existing and proposed models.

Step 2: Dataset Preprocessing

Once the dataset is collected, it undergoes meticulous preprocessing to ensure the data is clean, consistent, and ready for machine learning algorithms. The first task involves handling missing or null values, which may result from errors during data recording or inconsistencies in dataset formats. These values are either removed or imputed depending on the distribution and impact on the dataset. Next,

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

label encoding is carried out to convert categorical labels (e.g., rice variety names like Basmati, Jasmine, IR64) into numerical format so that machine learning models can interpret them effectively. In some cases, feature scaling or normalization is also applied to bring all numeric features to a similar range, ensuring that no feature dominates the learning process. This step significantly enhances model accuracy and training speed.

Step 3: Existing Model – XGBoost Classifier

As part of the benchmark comparison, the XGBoost Classifier is implemented in this research as the existing model. XGBoost is a robust and widely used ensemble learning algorithm based on decision trees and gradient boosting. It is chosen for its high efficiency, scalability, and superior performance in classification tasks. In this stage, the pre-processed dataset is divided into training and testing sets. The XGBoost model is trained using the training data and evaluated based on standard performance metrics such as accuracy, precision, recall, and F1-score. Hyperparameters like learning rate, max depth, and number of estimators are tuned to optimize the model's predictive capability. This model serves as a baseline to compare how well the proposed MLP model performs.

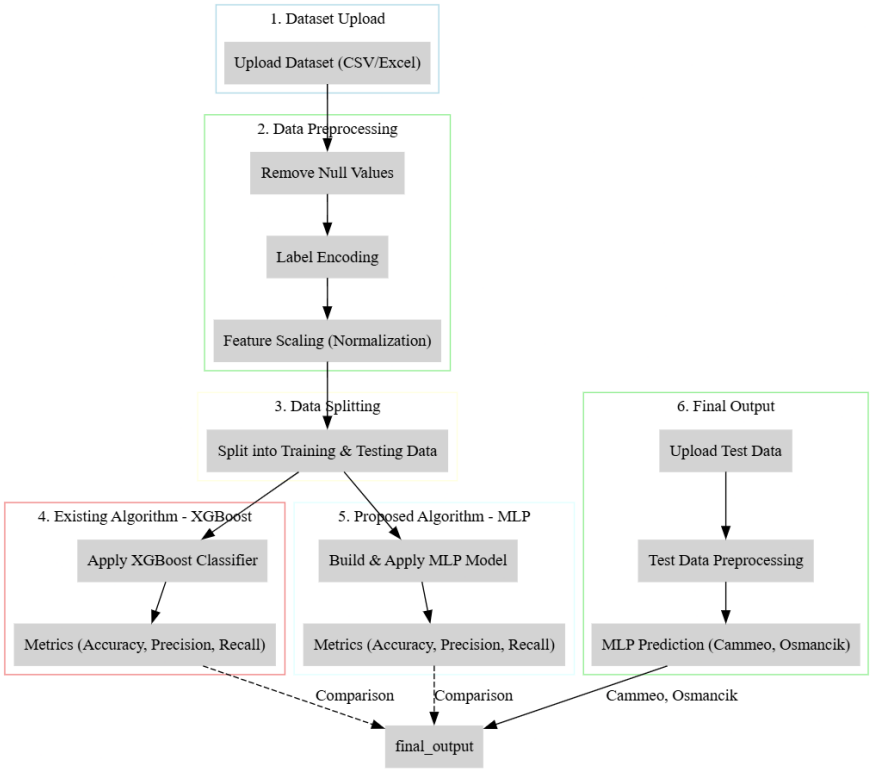

Fig. 1: Block Diagram

Step 4: Proposed Model – MLP (Multi-Layer Perceptron)

The final step involves designing and building the proposed model using a Multi-Layer Perceptron (MLP), a type of feedforward artificial neural network. Unlike traditional models like XGBoost, MLP is capable of capturing complex nonlinear relationships among input features, making it highly suitable for nuanced classification problems like rice variety detection. The MLP model is constructed using several layers — including an input layer matching the number of features, one or more hidden layers with ReLU activation, and a final output layer with softmax activation for multi-class classification.

The model is compiled with an appropriate optimizer (such as Adam), a loss function (such as categorical crossentropy), and is trained over multiple epochs to minimize classification error. The performance of the MLP model is then compared to the XGBoost classifier to analyze improvement in accuracy and generalization ability. The proposed system aims to demonstrate higher reliability and adaptability for real-world rice classification tasks in the agricultural supply chain.

3.2 Data Splitting and Preprocessing

Data preprocessing is a vital stage in the machine learning pipeline that directly impacts the accuracy and efficiency of predictive models. In this research on rice variety classification, preprocessing begins immediately after acquiring the dataset, ensuring that the raw data is transformed into a structured and consistent format suitable for modeling. The dataset typically includes several numeric features related to the physical characteristics of rice grains, such as length, width, area, perimeter, and compactness. Alongside these are class labels indicating the rice variety. However, before applying machine learning models, it is essential to cleanse and prepare the data appropriately to minimize noise and redundancy. The first step in preprocessing is the removal of null or missing values. Missing data, if not addressed, can skew model training and lead to unreliable predictions. In this research, records with missing values are either dropped or, if minimal, imputed using statistical methods such as mean or mode imputation. Next, label encoding is performed to convert categorical labels — such as rice variety names— into numeric format. This conversion is crucial because most machine learning algorithms require numerical input rather than text labels.

After label encoding, the dataset is scaled using normalization or standardization techniques, ensuring that features are brought to a comparable range. This step prevents features with larger values from disproportionately influencing the model. Common scaling methods such as Min-Max Scaling or Z-score Standardization are considered, depending on the distribution of the feature values. The scaling enhances convergence speed during training and ensures more stable and accurate performance. Finally, the cleaned and transformed dataset is split into training and testing subsets, typically in an 80:20 or 70:30 ratio. The training set is used to teach the machine learning models, while the testing set is reserved for evaluating model generalization and performance on unseen data. In some experiments, cross-validation techniques like K-Fold are also applied to ensure robustness and prevent overfitting. This careful and comprehensive approach to data splitting and preprocessing lays a solid foundation for effective model training and accurate rice variety prediction.

3.3 Model Building

In this study, two primary machine learning algorithms have been utilized for rice variety classification the XGBoost Classifier as the existing model, and the Multilayer Perceptron (MLP) as the proposed model. Each model has been tested on the pre-processed dataset to evaluate and compare performance in terms of accuracy, precision, and prediction robustness.

3.3.1 Proposed Algorithm: Multilayer Perceptron (MLP)

Multilayer Perceptron (MLP) is a class of feedforward artificial neural networks that consists of an input layer, one or more hidden layers, and an output layer. MLP is capable of learning complex nonlinear functions through backpropagation, making it highly suitable for classification tasks such as rice variety identification. Unlike tree-based models, MLPs use neurons and activation functions to transform inputs and generate predictions.

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

What is MLP & How It Works:

MLP works by processing input features through multiple layers of neurons. Each neuron applies a weighted sum of its inputs followed by a nonlinear activation function (like ReLU or sigmoid). The network adjusts weights during training using the backpropagation algorithm, minimizing the error between predicted and actual values through gradient descent. MLP is particularly effective for datasets where relationships among features are non-linear and require deep representation.

Advantages:

MLP models are powerful in learning from data with complex interdependencies. They offer flexibility in architecture design (number of layers, neurons) and can model non-linear relationships better than traditional algorithms. Additionally, MLPs generalize well on unseen data when properly regularized and optimized. In this research, the MLP model demonstrated improved accuracy and robustness compared to XGBoost, proving its suitability for this classification task.

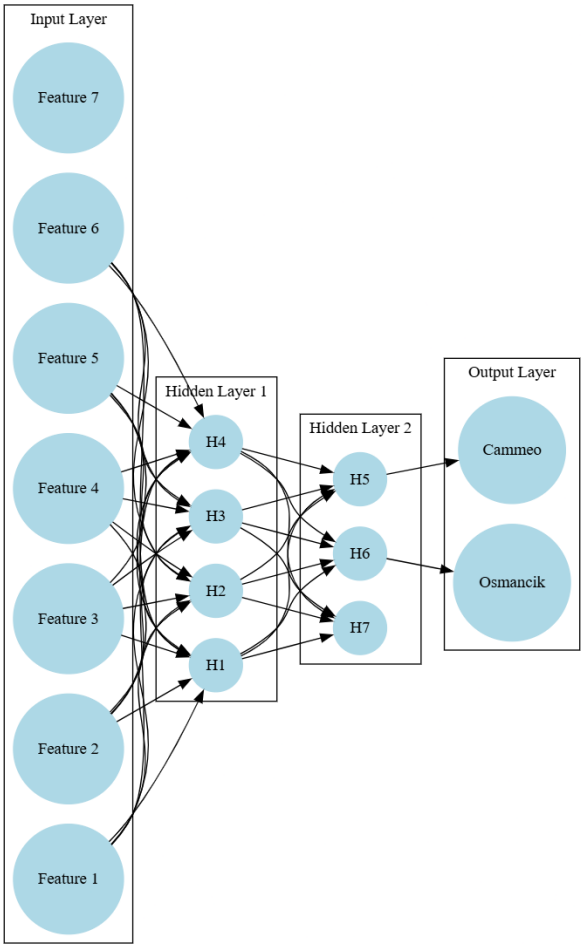

Fig. 2: Architecture of the MLP

4. RESULTS AND DISCUSSIONS

4.1 Dataset Description

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

The rice-variety dataset you're working with is a classic morphological dataset of individual rice grains, each described by seven quantitative measurements plus a class label indicating its variety ("Cammeo" or "Osmancik"). Every row in the CSV corresponds to one grain, and every column (except the final "Class" column) captures a different geometric property extracted from the grain's silhouette.

Area (integer): the total count of pixels inside the grain's outline—essentially its two-dimensional size.

Perimeter (float): the length of the grain's boundary in pixels, reflecting how "smooth" or "jagged" the edge is.

Major Axis Length (float): the length of the longest line that can be drawn through the grain, from edge to edge—derived by fitting an ellipse to the shape.

Minor Axis Length (float): the length of the shortest line through the center of that same fitted ellipse.

Eccentricity (float): a unitless measure of elongation, computed as the ratio between the focal distance and major axis of the ellipse; values close to zero indicate a shape near circular, and values near one indicate a highly elongated shape.

Extent Real (float): the fraction of the bounding-box area (the smallest rectangle that fully contains the grain) that the grain itself occupies—i.e., (Area ÷ BoundingBoxArea).

Convex Area (float): the area of the grain's convex hull, the smallest convex polygon enclosing all pixels of the grain.

The final column, Class, is a categorical label (here encoded later as 0 or 1) that tells you whether the sample is a Cammeo grain or an Osmancik grain and so on. Because these three varieties differ subtly in size, shape, and outline roughness, these seven features together allow machine-learning algorithms to learn distinguishing patterns. In typical usage you have a few hundred to a few thousand grains in total; the dataset is first cleaned (any missing entries are set to zero), then split—80% for training and 20% for testing—followed by standardization so that each feature has zero mean and unit variance before it's fed to classifiers like XGBoost or an MLP.

4.3 Results and Description

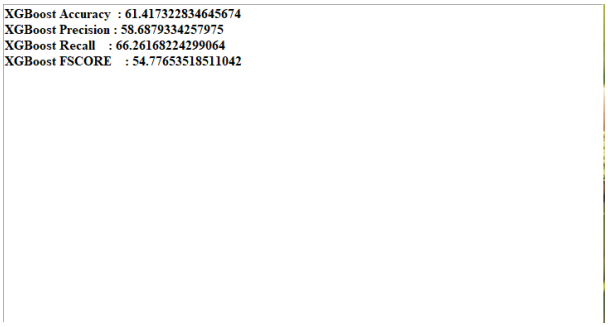

Fig. 3: Xgboost Metrics

Figure 3 shows that The performance of the XGBoost model indicates moderate effectiveness in prediction, with an accuracy of approximately 61.42%, suggesting that the model correctly classifies a little over 61% of the total instances. It achieves a precision of 58.69%, meaning that when the model predicts a positive outcome, it is correct about 59% of the time. The recall stands higher at 66.26%,

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

showing that the model successfully identifies around 66% of actual positive cases. However, the F1-score, which balances precision and recall, is relatively lower at 54.78%.

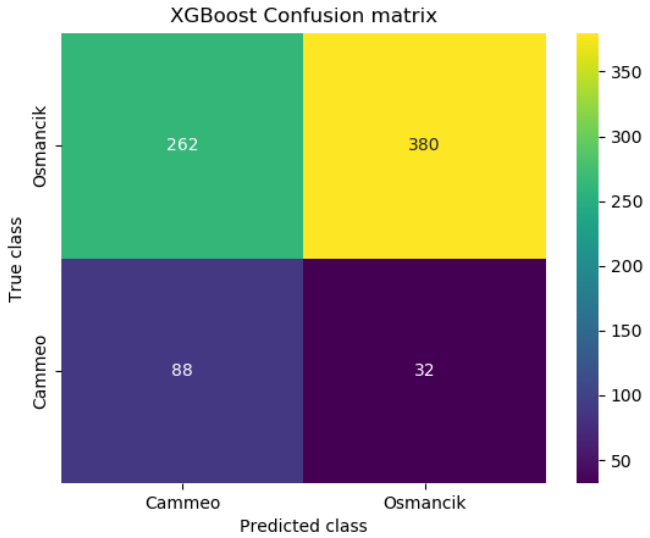

Fig. 4: Confusion matrix of the Xgboost classifier

Figure 4 shows that confusion matrix for an XGBoost classification model, evaluating its performance in distinguishing between two classes: "Cammeo" and "Osmancik". The matrix reveals that the model correctly classified 262 instances as "Osmancik" (True Negatives) and 32 instances as "Cammeo" (True Positives). However, it misclassified 380 instances of "Osmancik" as "Cammeo" (False Positives) and 88 instances of "Cammeo" as "Osmancik" (False Negatives).

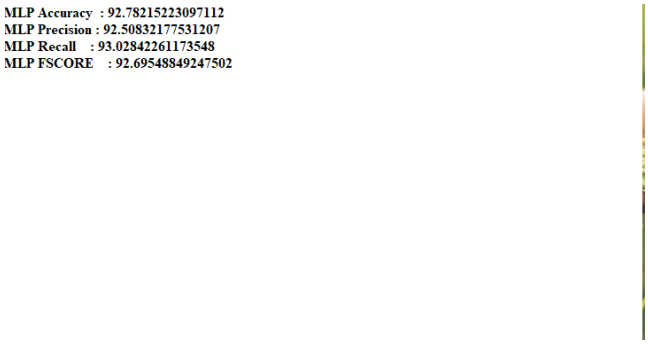

Fig. 5: Metrics of the MLP (Deep NN)

Figure 5 shows that the MLP (Multi-Layer Perceptron) model demonstrates strong predictive performance with a high accuracy of 92.78%, indicating that it correctly classifies the vast majority of instances. Its precision is also impressive at 92.51%, showing that most of its positive predictions are correct. The recall is slightly higher at 93.03%, meaning the model effectively captures the majority of actual positive cases. With an F1-score of 92.70%, the MLP model exhibits a well-balanced trade-off between precision and recall,

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

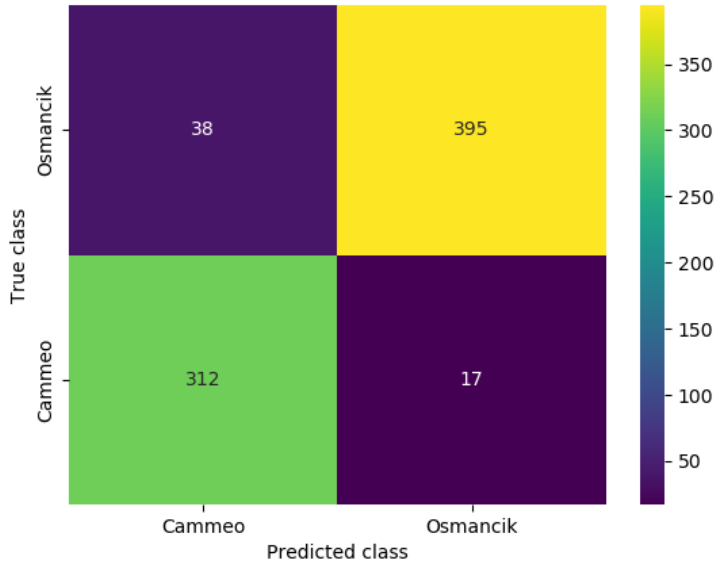

Figure 9: CM of the Neural network (MLP)

This confusion matrix illustrates the performance of a classification model, likely different from the previous one, in predicting between "Cammeo" and "Osmancik" classes. Here, the model correctly identified 312 instances as "Cammeo" (True Positives) and 395 instances as "Osmancik" (True Negatives). However, it incorrectly classified 38 instances of "Cammeo" as "Osmancik" (False Negatives) and 17 instances of "Osmancik" as "Cammeo" (False Positives).

5. CONCLUSION

This study presents a comprehensive approach to rice variety classification using machine learning techniques, comparing the performance of XGBoost and a proposed Multi-Layer Perceptron (MLP) neural network model. The application is built with a user-friendly interface using Tkinter, enabling easy interaction for both administrators and end-users. The rice dataset, characterized by morphological features such as Area, Perimeter, Axis Lengths, Eccentricity, and Convex Area, provides a solid basis for classification. Extensive data preprocessing including label encoding, handling missing values, feature scaling, and exploratory data analysis ensures that the data is well-prepared for modeling. The results show that the XGBoost model provides moderate classification performance with an accuracy of 61.42%, precision of 58.69%, and an F1-score of 54.78%. In contrast, the proposed MLP model significantly outperforms XGBoost, achieving an impressive accuracy of 92.78%, precision of 92.51%, recall of 93.03%, and an F1-score of 92.70%. The confusion matrices further confirm the superiority of the neural network in accurately distinguishing between the rice varieties "Cammeo" and "Osmancik". Overall, this project demonstrates that deep learning models, particularly MLPs, are more effective in capturing complex, non-linear relationships within morphological data, thereby providing highly reliable classification results.

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.

[14]. Lin, C.; He, D.; Huang, X.; Xie, X.; Choo, K.-K.R. PPChain: A Privacy-Preserving Permissioned Blockchain Architecture for Cryptocurrency and Other Regulated Applications. IEEE Syst. J. 2020, 15, 4367–4378.

[15]. Iftekhar, A.; Cui, X.; Yang, Y. Blockchain Technology for Trustworthy Operations in the Management of Strategic Grain Reserves. Foods 2021, 10, 2323.

[16]. Liu, Y.; Ma, X.; Shu, L.; Hancke, G.P.; Abu-Mahfouz, A.M. From Industry 4.0 to Agriculture 4.0: Current Status, Enabling Technologies, and Research Challenges. IEEE Trans. Ind. Inform. 2020, 17, 4322–4334.