OmniaScience

Journal of Industrial Engineering and Management
JIEM, 2024 – 17(1): 151-167 – Online ISSN: 2013-0953 – Print ISSN: 2013-8423
https://doi.org/10.3926/jiem.6256

Design of a Traceability System for a Coffee Supply Chain Based on Blockchain and Machine Learning

Bonang Ligar[1*], Sarifuddin Madenda[1], Sutrisno Mardjan[2], Tubagus Kusuma[1]

[1]*Gunadarma University (Indonesia)*
[2]*IPB University (Indonesia)*

[*]*Corresponding author: bonangligar@staff.gunadarma.ac.id*
sarif@staff.gunadarma.ac.id, trisno406@apps.ipb.ac.id, mkusuma@staff.gunadarma.ac.id

Received: July 2023
Accepted: February 2024

Abstract:

Purpose: This paper aims to develop a coffee supply chain traceability system based on Blockchain (BC) and Machine Learning (ML) with the aim of ensuring the quality of coffee beans production. BC functions to ensure supply chain performance, while the ML model ensures product quality.

Design/methodology/approach: Smart Contracts will be built on the Ethereum Virtual Machine BC network based on Ethereum. The ML model to identify good and bad green coffee beans will be built using different YOLO algorithms, which will go through training and validation stages, namely using the k-fold cross validation method. The ML model algorithm is based on Convolutional Neural Network (CNN) using YOLOv5m, YOLOv6m and YOLOv7. The best model will be chosen based on the results of cross-validation with test data in the form of coffee image data that the model has never seen (unseen data). The whole process of building the ML model is done on the Google Collab Pro+ Virtual Machine.

Findings: YOLOv5m outperformed the other models in both non-augmented and augmented training datasets, highlighting the proficiency of YOLOv5m in managing compact datasets and its resilience in the face of data augmentation, positioning it as a prime selection for quality discernment tasks within the realm of green coffee beans. The smart contracts offer an all-encompassing approach for user management, monitoring product status, and presenting traceability data within the framework of coffee plantation administration.

Originality/value: This research contributes to the development of blockchain network as a solution to implement traceability systems along coffee supply chains in Indonesia. Moreover, it shows that while blockchain can ensure the process of production along the coffee supply to follow certain guidelines, machine learning can verify whether the product that was produced by utilizing BC is of high quality/acceptable.

Keywords: smart contracts, coffee beans, quality identification, blockchain, machine learning, traceability

To cite this article:

Ligar, B., Madenda, S., Mardjan, S., & Kusuma, T. (2024). Design of a traceability system for a coffee supply chain based on blockchain and machine learning. *Journal of Industrial Engineering and Management*, 17(1), 151-167. https://doi.org/10.3926/jiem.6256

1. Introduction

One of the most potential agri-food commodities to be traded in the world is coffee, and Indonesia is a leading player in the global market (Ligar, Madenda, Mardjan & Kusuma, 2023). According to the Indonesian Central Statistics Bureau, in the period between January-October 2022, Indonesia managed to export 348.6 thousand tons of coffee, with a transaction value of 918.3 million USD (BPS RI, 2022). The transaction value is the highest among all exported Indonesian agricultural commodities, contributing 23% of the total export value of agricultural commodities (BPS RI, 2022). In 2019, Indonesia ranked 4th in the world out of a total of 55 coffee exporting countries, accounting for 5.2% of world exports (ICO - International Coffee Organization, 2020).

Priangan Coffee or Java Preanger Arabica Coffee (JPAC), is a variant of coffee originating from West Java, more specifically from Garut, Bandung Regency, West Bandung and Sumedang area in Indonesia (Dasipah, Sukmawati & Faturachman, 2021). These areas produce specialty-type coffee which has excellent taste test results (Djuwendah, 2019) and has received a cupping test certificate in the very good category (Djuwendah, Karyani, Sadeli & Hapsari, 2019). JPAC has good quality, liked by domestic and foreign consumers, and until now it has been exported a lot, including to Morocco, South Korea, Australia and Germany (Dasipah et al., 2021; Ginanjar, Apiatno & Amanda, 2020).

However, the coffee supply chain in Indonesia still faces several problems such as low coffee production (Ibnu, 2019; Randy & Marianti, 2019), low quality (Jamil, 2019; Kudus, Widayat & Abubakar, 2019; Yustisar, 2018), as well as the low value of Indonesian coffee exports (Ginting & Kartiasih, 2019; Jamil, 2019).

Blockchain technology (BC), as a decentralized electronic ledger technology that ensures transparency, traceability and security, demonstrated the potential to mitigate several worldwide supply chain problems (Saberi, Kouhizadeh, Sarkis & Shen, 2019). Given the need for reliable suppliers in the industry and customers who are concerned about food safety (Chen, Brahma, Mackay, Cao & Aliakbarian, 2020), it is important for decision makers to consider using BC to help improve integration and overall supply chain performance, also significantly decrease the risks in losing credibility (Saberi et al., 2019).

Integration of BC with the supply chain is useful for more end-to-end tracking, transparency, accuracy, increasing trust between producers and consumers, increasing visibility and conformity of products with international standards, reducing or even eliminating fraud and counterfeit products, facilitating traceability of product origin, and enabling problematic products recall efficiently (Azzi, Chamoun & Sokhn, 2019). But can BC really help supply chains produce standardized quality coffee beans?

Coffee is valued by consumers for its aroma, taste, and refreshing impact (Yang, Li, Mei, Liu, Liu, Chen et al., 2021). These properties depend on the quality of the coffee beans which also have a significant effect on its economic value. However, not all coffee beans grown by coffee producers are of high quality. Damage to coffee beans can be caused by varied factors such as pest attacks or post-harvest processes which cause some or all the coffee beans to be damaged so that their quality decreases (Nasution, Rumansa & Lukman-Adlin, 2019).

Visual inspection of coffee beans by humans has been the traditional approach to spot defects and stone chips when selecting the highest quality green coffee beans; however, non-uniform selection can easily arise because of excessive working hours, insufficient training, and individual employee attitudes (García, Candelo-Becerra & Hoyos, 2019; Wang, Tseng, Chen & Hsia, 2021). The machine learning (ML) algorithm in computer vision (CV) systems is superior for classifying and selecting food products because it automatically extracts and evaluates important information from certain visual items (García et al., 2019). ML algorithms can not only classify and locate objects, but also differentiate them based on factors such as color, appearance, shape and size (Jian & Pan, 2022). Therefore, the integration between BC and ML is very significant in ensuring the quality of products produced by a supply chain. The production process must be reliable and flexible, for that it is better to use an intelligent control system that considers changes automatically while reducing human intervention as well as associated costs (Dhatterwal, Kaswan & Ojha, 2023).

In 2015, researcher Joseph Redmon and fellow researchers built the YOLO algorithm, an object identification system that, for the first time, completes all the steps required to recognize an item using a single neural network called You Only Look Once (YOLO). From the pixels in an image then generating class probabilities and bounding box coordinates, YOLO reframes object recognition as a single regression problem (Thuan, 2021). While other ML algorithms may be able to recognize items in an image after a number of runs, YOLO can do so after only one forward propagation through a neural network, making it ideal for use in real-time situations (Nepal & Eslamiat, 2022). YOLO algorithm became famous among other ML algorithms because of this feature. This integrated model predicts the bounding box as well as the class probabilities for the items that will be covered by the box. The YOLO algorithm has reached an extraordinary specification, which at the time of publication outperformed the top algorithms in terms of speed and accuracy for locating and identifying object locations (Redmon, Divvala, Girshick & Farhadi, 2016).

Very few studies have integrated traceability, BC, and ML as a single entity; specifically, to identify the quality level of coffee beans along the supply chain for export needs. Therefore, this study aims to develop a traceability system based on BC and ML to identify the quality of coffee beans in the Priangan coffee supply chain. Traceability systems integration with BC is expected to guarantee activities, interactions, and transactions between actors in the coffee supply chain network so that they continue to meet international needs. The algorithms used to build the green coffee bean quality identification model will be YOLOv5m, YOLOv6m and YOLOv7. They will be compared and the algorithm with the best performance will be chosen for identifying green coffee beans quality.

2. Literature Review

As science and technology advances as well as rapid changes in market demand, understanding the latest research and developments in supply chain management plays an important role in studying supply chains in more depth. Researchers at home and abroad have also made a lot of discussions and research on how to combine traceability, BC, and ML systems to improve the current supply chain management situation.

Wang, Yue and Zhou developed a traceability system to evaluate the quality of pork products along the supply chain (Wang, Yue & Zhou, 2017). The fuzzy classification method is used to evaluate food quality at each stage of the supply chain, while artificial neural networks (ANN) are adopted to obtain the final determination of food quality level based on the quality evaluation of all stages. The fuzzy classification algorithm used is k-NN. At each stage of the pork product supply chain, a quality assessment is given by several experts, whose results will be compared with the k-NN classification results. The results of the k-NN classification will be input to the ANN to obtain the final value for the quality level of pork products. The developed ST has an information processing subsystem that can carry out forward tracing and backward tracing, as well as quality evaluation.

Thiruchelvam, Mughisha, Shahpasand and Bamiah (2018), conducted a quantitative analysis research to conclude about the need for technology to help solve the problem of sustainability of coffee production (Thiruchelvam et al., 2018). The questionnaire was distributed using Google Doc with a sample size of 66 respondents representing the Burundian coffee industry. It can be seen from 98% of respondents' responses that awareness of coffee distribution and coffee origin traceability is still lacking. Where, 97.96% of respondents stated the importance of transparency to ensure the reliability of information shared by actors in the coffee supply chain. In addition, the visibility factor was considered by 93.88% of respondents to be important in the coffee supply chain to maintain sustainability and control. Furthermore, 97.96% of respondents agreed that technology can help make the coffee industry more sustainable. Respondents agreed that technology can help solve current coffee sustainability problems. Thiruchelvam et al., proposed that BC technology could be applied to the coffee supply chain.

Landinez, Rodriguez, and Gomez proposed the design and implementation of a software for traceability in coffee bean processing from harvest to drying (Landínez, Rodríguez & Gómez, 2019). The software makes it possible to support coffee producers in the process of obtaining certification of origin of their products in accordance with the regulations established for trade in agricultural products by the European Union since January 2005. There are six basic functions provided by the software, namely: (1) harvest records; (2) quality estimation records; (3) exfoliation machine entry records; (4) fermentation process records; (5) records of the coffee washing process; and (6) records of the coffee drying process. The developed software can be accessed via web pages and mobile devices.

Pradana et al developed a BC-based ST in the coffee supply chain with the main aim of knowing where a coffee product comes from (origin) (Pradana, Djatna & Hermadi, 2020). The system developed makes it possible to accommodate actor activities and interactions between actors in the supply chain, as well as storing all information formed along the supply chain. Transactions that occur between actors will be recorded in BC through a smart contract that has been designed according to transaction needs and system objectives. The collected information will be provided to customers with the aim of gaining customer trust. The quality of the coffee beans is also analyzed based on the coffee characteristics that customers want. The proposed system is aimed at the national market and is not based on export needs. One of the outputs is a coffee traceability software prototype. The traceability system developed does not use ML technology to determine the quality level of green coffee beans produced by the supply chain.

3. Research Methodology

The research is divided into three stages (Figure 1). The first stage involves identifying problems that exist in the coffee supply chain, analyze each existing business procedure, identify stakeholders and interactions between stakeholders, collect coffee supply chain data, regulations, and standards in the coffee chain, as well as factors which affect the quality of coffee products. Then the modeling of the Indonesian coffee supply chain will be carried out. At this stage, entity identification and modeling of the coffee traceability system entities will also be carried out. The results of this first stage will be used as a basis for designing a traceability system. The data collected will refer to international standards for Indonesian coffee importing countries and will be traced upstream to the farmer level. Based on these data, coffee quality identification will be carried out. Observations and interviews with the actors involved provide the study's primary data. Secondary data refers to information that has been analyzed and compiled by researchers based on primary data. Table 1 lists data related to the stakeholders and actors involved as well as the types and sources of data used in the research.

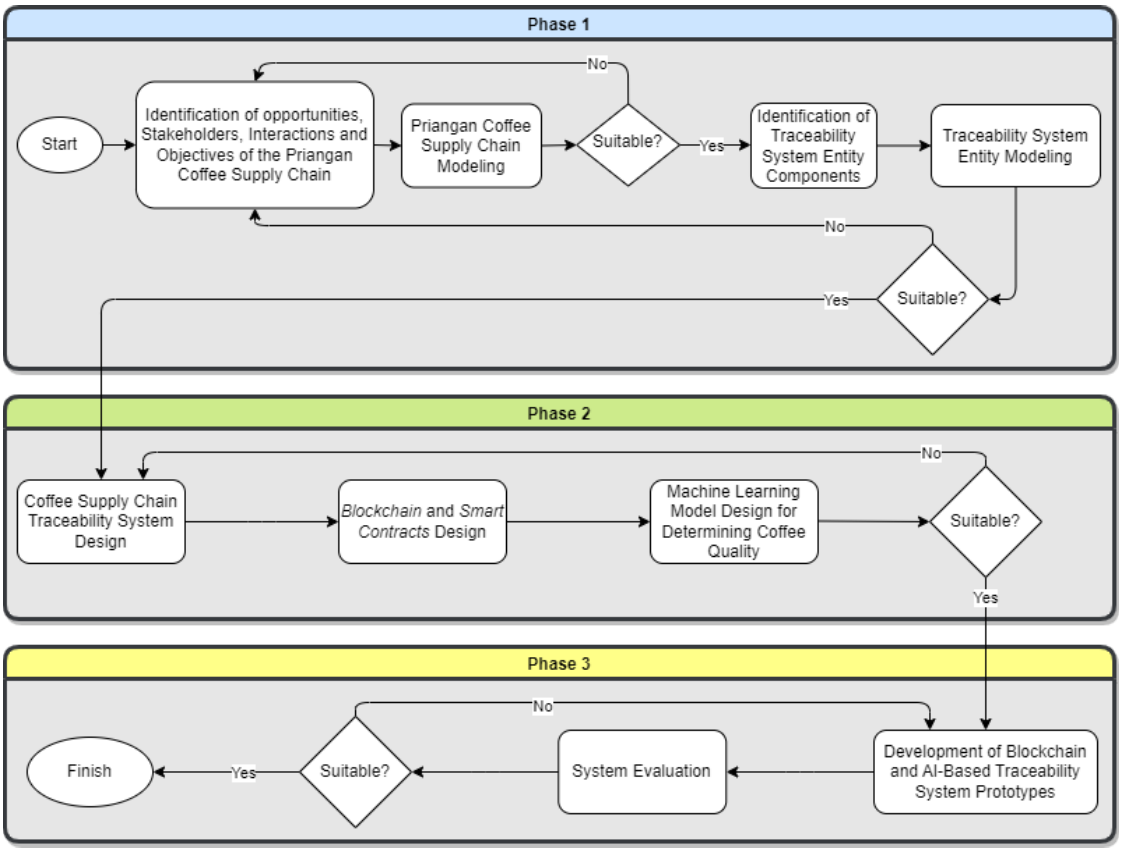

Figure 1. Stages of research in general

No.	Aspects	Data	Data Source
1.	Supply chain models	– Actors in the supply chain	Field observation and interview (head office and factory)
2.	JPAC agroindustry supply chain activities	– Activities of supply chain actors include: cultivation, processing, and export. – Company internal management.	Direct field observations and interviews with each section head.
3.	Traceability	– Analysis of production activity – Traceability data	Field observations, interviews, and literature study
4.	Coffee bean quality identification model	– Images of green coffee bean – ML algorithm	Sampling in the field (factory) and literature.
5.	Traceability system	– Verification, validation and testing of system implementation	Discussion and literature study

Table 1. Data and research data sources

The second stage is carried out based on the results of the analysis in the first stage whose purpose is to determine how the system operates. This stage will be divided into 3 parts, namely the design of the coffee supply chain traceability system, the design of smart contracts to record activities between actors along the supply chain, and the design of the ML model algorithm to determine coffee quality. The development stage of the ML model will go through the training and validation stages, namely using the k-fold cross validation method. In the context of k-fold cross-validation, the term "k" signifies the quantity of distinct subsets into which the dataset is segregated. In this study the number of k used is 5-folds. The next step is to test the model based on the results of cross validation with test data in the form of coffee image data that the model has never seen (unseen data). and designing the ML model algorithm based on Convolutional Neural Network (CNN) using YOLOv5m, YOLOv6m and YOLOv7.

The third stage will be the implementation of a prototype system. Testing will be carried out on prototypes to see if the system functionality is as expected, and to look for errors that may be experienced by the system. Testing will be carried out using the black box testing method. Black box testing attempts to find errors in several categories, including wrong or missing functions, interface errors, errors in data structures or external database access, performance errors, errors when starting and stopping the system (Siswanto, 2019).

3.1. Traceability System Design

An illustration of a blockchain (BC) and machine learning (ML)-based traceability system framework implemented in the software can be seen in Figure 2. In the initial transaction, the user enters data into a specially crafted web browser, which is then authorized by the system. In the process of developing an Ethereum-based design, testing is the most important part of the program life cycle. Data entered into BC is distributed using a smart contract code or application that is protected from possible data changes. A smart contract can be signed when two parties agree to all the terms and conditions. The input data recorded by the computer is used to draw up the contract, which is then converted into bytecode and added to the BC.

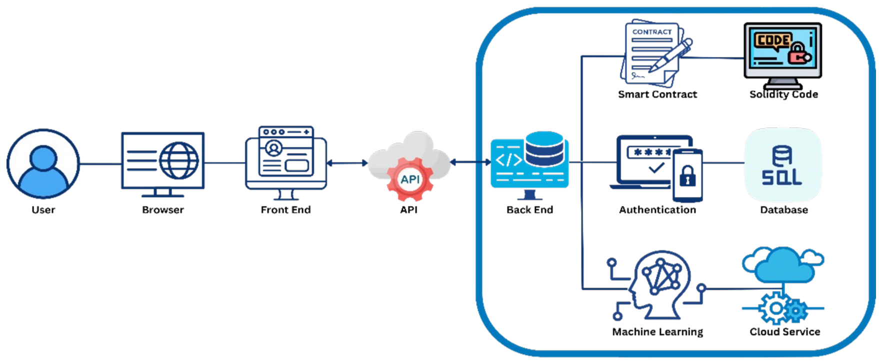

Figure 2. Blockchain and machine learning-based traceability system workflows

3.2. Smart Contract Design

Building a local private distributed ledger or BC network based on Ethereum, utilizing tools such as Ganache, which will generate BC's with ten accounts set up, with 100 ether each, is a common way to implement smart contracts. Then, smart contracts can be deployed on this blockchain by sending them to the Ethereum Virtual Machine via a single user account. Therefore, the steps for setting up, and implementing a smart contract are as follows (Kirli, Couraud, Robu, Salgado-Bravo, Norbu, Andoni et al., 2022):

1. Using Ganache, configure local BC with nodes (virtual machines) and accounts.
2. Create a smart contract in a specific programming language (e.g., Solidity or Vyper).
3. Use a language compiler to compile the smart contract code.
4. Using Python or JavaScript Web3 libraries to apply compiled code (byte code) in BC.
5. Interact with the smart contract (and BC) using python or JavaScript commands sent to the smart contract address via the local BC node.

3.3. Designing Machine Learning Model

This research proposes a ML model to identify the quality of coffee products along the supply chain. In this study, we utilized the YOLO (You Only Look Once) object detection algorithm, a machine learning method that employs a convolutional neural network for rapid and accurate object detection. YOLO stands out among other machine learning algorithms for its ability to detect objects in a single pass, making it both fast and efficient, this is where the YOLO algorithm shines with its ability to identify objects in a single forward pass through a neural network, making it ideal for real-time applications. This characteristic has propelled YOLO to the forefront of machine learning algorithms YOLO development has now reached the 7th version which has better accuracy than previous versions. YOLO offers the flexibility to adapt to the available computing resources, ensuring optimal performance on the device at hand.

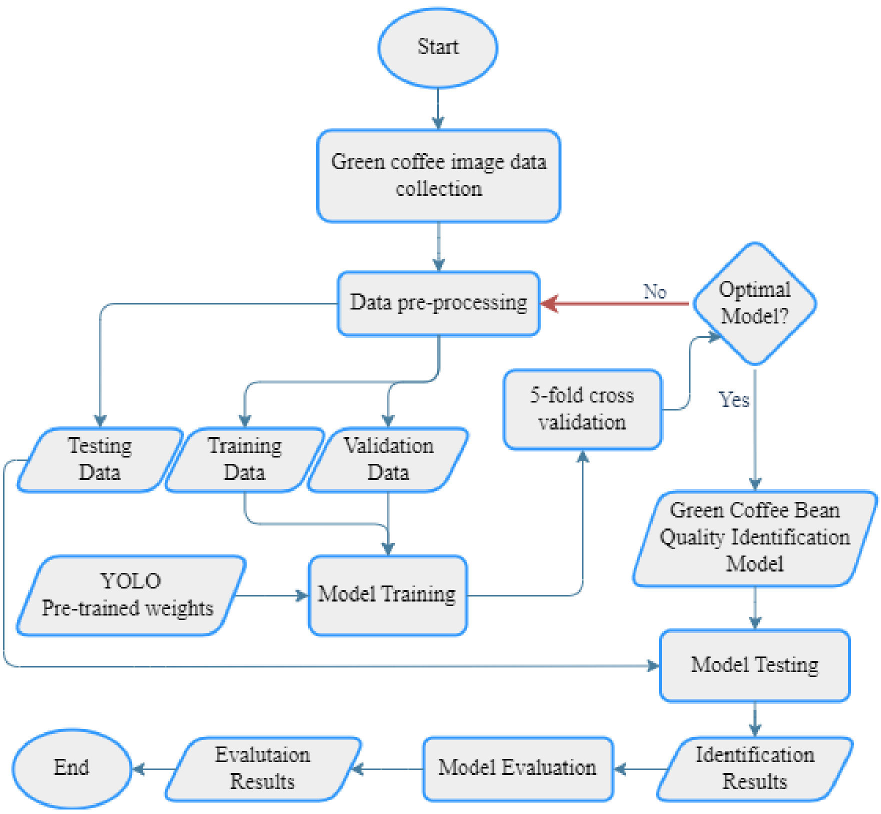

Figure 3. YOLO model determination flowchart

In this study, it will be determined which of the YOLOv5m, YOLOv6M, and YOLOv7 models is the best for the green coffee bean dataset. Figure 3 shows the flowchart for determining the YOLO model using training 5-fold cross-validation method. The results of cross-validation will be evaluated, then will be tested using test data (unseen data). Evaluation of model testing will be seen from the aspects of precision, recall, mean average precision (mAP), accuracy and classification time. Precision is the ratio of correct predictions to the total number of predictions. Recall calculates the ratio of correct predictions to the total number of objects in an image. The mAP value is used to measure the quality of a model, the greater the mAP value, the better the average detection accuracy and performance. Accuracy can be defined as the extent to which predicted values and observed values match each other. The model with the best test results will be used for green coffee bean quality qualification.

3.3.1. Acquisition of Green Coffee Bean Image Data

The coffee data image collection in this study was taken from samples of coffee beans produced at Java Frinsa Estate. There are 2 types of coffee beans taken as samples, namely (1) good quality green coffee beans; (2) poor quality green beans. Both types of coffee beans will be used as training data for the proposed computer vision model. 600 good bean images and 600 bad bean images were taken with a total of 1200 data. A digital camera (Sony ZV-1) with a resolution of 3648×3648 pixels was used to take the images. All copy images are taken in a mini photo box with a white background, with LED lighting.

3.3.2. Data Preprocessing

1. All the images collected were then cropped to a size of 640×640 pixels for each individual bean. A total of 1400 green coffee bean images were collected, with 1200 as training and validation data sets, while 200 images as test data (unseen data).
2. Perform labeling (annotation) on the training and validation data with two class labels, namely good bean and bad bean classes using Roboflow. The composition of the training data is 80% (960 images) and validation is 20% (240).
3. Perform data augmentation. The data augmentation technique used in this study is based on the geometric transformation technique. Geometric transformation is a data augmentation method based on basic image manipulation. The augmentation of 1200 coffee bean images resulted in a total of 10000 coffee bean images divided into 5000 good beans and 5000 bad beans. The augmented data will be used as training data and validation.

3.3.3. Development of Machine Learning Model

The YOLOv5, YOLOv6, and YOLOv7 algorithm models use the Pytorch framework. The best model for identifying JPAC beans will be determined by comparing the three models. The network input size (width × height) used is: 640 × 640 pixels. Google Collab Pro + (GPU A100-SXM4-40GB) environment was used for developing (training, validation, and testing) the coffee bean quality identification model.

4. Results and Discussion

4.4. Traceability System Architecture

Design development begins with building the functional architecture of the system. As shown in Figure 4, the development of this functional architecture produces a coffee traceability model based on BC and ML. This proposed architecture takes into account quality standards based on SCA and Indonesia National Standard for Java Preanger Arabica Coffee (JPAC). The resulting model describes the physical flow and information that flows from the initial actor to the final actor. The physical flow in the model involves supply chain actors and JPAC post-harvest products. While the flow of information consists of establishing traceability information, digital data and the BC network.

The proposed traceability system architecture is also designed based on two main obstacles found from observations and interviews at Frinsa, namely: (1) All records are still carried out conventionally using paper, this is often complained of by staff as one of the obstacles. It is necessary to find a solution and this BC-based traceability system is expected to provide a solution; (2) Green bean quality inspections that are often carried out repeatedly increase production time and costs. This repetition is due to the level of accuracy of the color sorter machine which is still inadequate. Repetition at this stage also increases the risk of the coffee beans breaking while in processing, so that the number of defects also increases. This problem can be solved by using ML technology, so it is proposed to develop a computer vision model to overcome the problems of quality inspection of green beans.

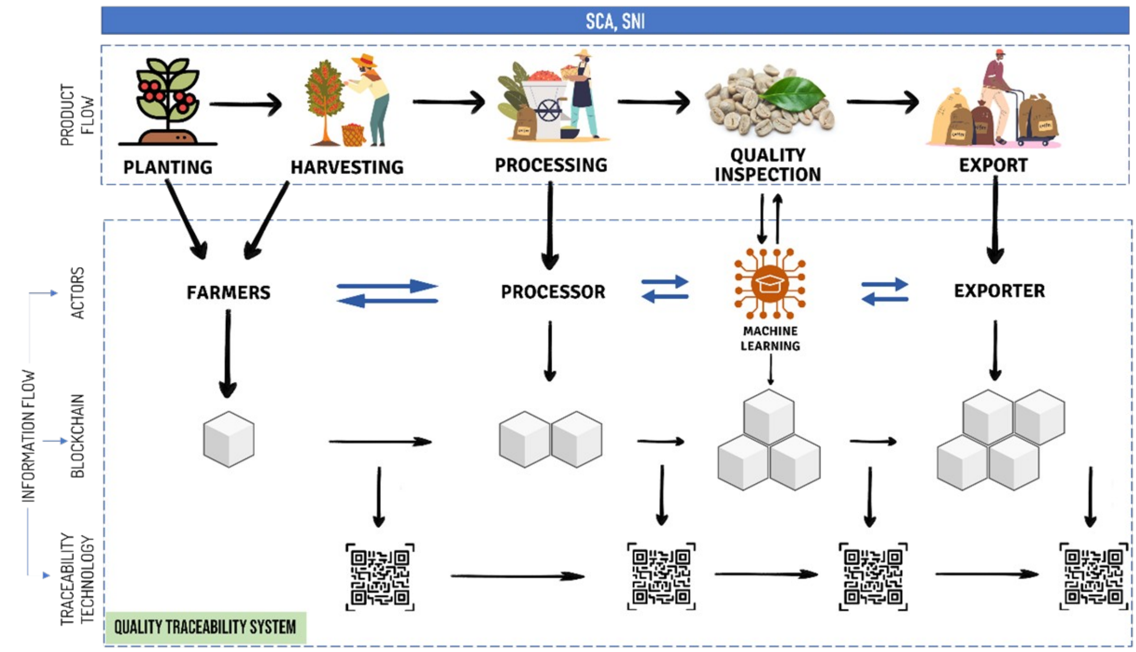

Figure 4. Proposed Blockchain and Machine Learning-Based Traceability System

Internal and external traceability that is formed due to transactions/activities involving two actors is transformed into a unique QR code which will be stored on the blockchain network. Every time there is a new transaction/activity, a new QR code will be generated which includes the previous code as additional information. This QR code will be stored as a block in the BC network, and each block will be linked from the first to the last created, making the traceability process easier. By utilizing this BC technology, it is easy for each actor in the supply chain to trace which activity might be the cause if there is a decline in the quality of the coffee beans.

4.5. Smart Contract Design

The design of the smart contracts was tested on the BC Ethereum network. The solidity programming language version 0.6.0 was used in the remix IDE to develop smart contracts that define the necessary tracing mechanisms. Table 2 shows the pseudocode of the Farmer added contract, which is used to manage users who have the role of farmer. The contract also functions to store data regarding farmers and the coffee plantations they own. After the contract is successfully written, it will be tested by running it on the test network, which provides a test environment similar to the actual BC network.

The next smart contract that is built is traceability.sol, which contains functions including harvesting products, selling farmer products, processing products, as well as functions that can trace the history of products, namely the function of traceability *infoketertelusuran1* and *infoketertelusuran2*. This contract also forms the structure of the products produced by actors in the supply chain, which can be seen in the pseudocode in Table 3.

The product structure also defines a Stock Keeping Unit (SKU) to manage the number of products, the product batch code (PBC) which will be used as a product marker and used as input to generate a QR Code and include it on the packaging. This PBC will be used as the main search key to display traceability information. Product ID is an identifier generated from SKU + PBC whose function is to indicate a new product which is a modification or the result of a particular processing activity.

Input:
{FarmerID, farmerName, plantationName, plantationPIC, plantationArea, coffeeVariety, plantationAltitude, Longitude, Latitude}
Output:
Farmer Information
begin
Input the necessary farmer data *Run the addFarmer function* *Input ID_Petani to return farmer info* *Execute the seeFarmer function* *Return farmer information*
end

Table 2. Pseudocode for addFarmer Smart Contract

Input:
{sku (stock keeping unit), product batch code (PBC), ownerID, FarmerID, farmerName, producerInfo, plantationLongitude, plantationLatitude, productID, productQuality, productPrice, productStatus, processorID, processorName, exporterID, exporterName, consumerID}
Output:
Product Information
begin
Input the required product data *Save product information*
end

Table 3. Product structure on smart contract traceability.sol

Input:
{FarmerID, farmerName, producerInfo, plantationLongitude, plantationLatitude, productQuality}
Output:
Product status has been harvested
begin
Input the required product harvest data *Product has been harvested status declaration*
end

Table 4. Harvest Product function on traceability.sol

Input:
{pbc}
Output:
Traceability info 1
begin
Input product batch code (pbc) *Run the traceability function 1* *Return traceability information {productSKU, product batch code (PBC), ownerID, FarmerID, farmerName, producerInfo, plantationLongitude, plantationLatitude}*
end

Table 5. Traceability info function 1 on contract traceability.sol

Input: *{kbp}* *Output:* *Traceability info 2*
begin *Input product batch code (pbc)* *Run the traceability function 1* *Return traceability information {productSKU, product batch code (PBC), IDproduk, productID, productQuality, productPrice, productStatus, processorID, exporterID, consumerID}* ***end***

Table 6. Traceability info function 2 on contract traceability.sol

Table 4 shows the product harvest function which aims to record info related to harvesting activities conducted by farmers. Products that have been harvested by farmers will be entered through a smart contract and recorded on the BC network, which will then change the product status according to the activities that have been carried out on the product. There are several product statuses in this contract along with their codes, namely Harvest (0), Processed (1), Sell (2), Sold (3), Sent (4), Purchase (5), Purchased (6), and Received (7). Table 5 shows the pseudocode of the traceabilityinfo1 function which is useful for finding and displaying traceability info such as product SKU, product PBC, OwnerID, farmer_id, farmer name, producer info, plantation latitude, and plantation longitude. Table 6 shows the pseudocode of the traceabilityinfo2 function.

4.6. Machine Learning Model

The neural network is trained at the initial stage using the original dataset of 1200 images which have not been augmented. This was done to see the effect of the number of samples on the YOLO model training results, whether a small amount of data produces good accuracy, or a dataset with a larger number and variations that affect the accuracy of the model. Table 7 shows the results of a comparison of three different YOLO algorithms that were trained using the original dataset in Google Collab Pro+ environment. The YOLOv5m had the best performance with all evaluation parameters above 99%, while YOLOv6m and YOLOv7 had parameters that resulted between 50% - 70%. YOLOv5 outperformed all other models using non-augmented training datasets.

Google Collab Pro+ (GCP+)			
Evaluation Parameters	**YOLOv5m**	**YOLOv6m**	**YOLOv7**
Precision	0.994 (99,4%)	0.536 (53,6%)	0.54 (54%)
Recall	1 (100%)	0.975 (97,5%)	0.922 (92,2%)
mAP@0.5	0.995 (99,5%)	0.685 (68,5%)	0.643 (64,3%)
mAP@0.5 :0.95	0.994 (99,4%)	0.679 (67,4%)	0.636 (63,6%)
Training time	7 m 8 s	16 m 34 s	17 m 34 s

Table 7. YOLO algorithm training results with original training set

In Table 7, the results of YOLOv5m training shows that with little training data it has insignificant effect on the accuracy level of the model. This is proven by the results of training with precision, recall, and mAP above 99%. mAP@0.5 means that it is the mAP calculated at IOU threshold 0.5, and mAP@0.5 :0.95 is the mAP calculated at IOU threshold between 0.5-0.95. In contrast to YOLOv6m and YOLOv7 which resulted with low accuracy in training with this original data. Proven by a mAP of around 60%, and a precision of around 54%.

After training and analyzing the results of the training using the original dataset of 1200 images, training will then be carried out using the augmented data of 10000 images. Determination of the best model for the green coffee dataset uses the 5-fold cross validation method. This validation method is used to avoid overfitting of a certain model. Overfitting is where a model obtains great performance in training, but bad results when tested with unseen data.

4.6.1. YOLOv5m

The results of the YOLOv5m model training with augmented data can be seen in Table 8. Training is carried out with 55 times iteration (epoch), with batch size = 60. Results show the fold with the best performance was k=0 for YOLOv5m model, with all evaluation parameters above 98%. This training with augmented data resulted with an overall lower performance in all evaluation parameters. Even though the results were lower, there were not significant differences, only about 1% difference. This means that the performance of the YOLOv5m in this study case of green coffee beans performed well.

Google Collab Pro+						
YOLOv5m	**Precision**	**Recall**	**F1 Score**	**mAP@0.5**	**mAP@0.5 :0.95**	**Training Time (hours)**
k=0	0.983	0.983	0.983	0.992	0.992	0,988
k=1	0.978	0.979	0.978	0.999	0.994	0,987
k=2	0.962	0.966	0.964	0.994	0.994	0,987
k=3	0.962	0.98	0.971	0.992	0.989	0,978
k=4	0.968	0.964	0.966	0.99	0.986	0,989
Mean	0.971	0.974	0.972	0.993	0.991	0,986
Std. Dev	0.01	0.009	0.008	0.004	0.003	-

Table 8. YOLOv5m algorithm training results using 10000 images of coffee data

4.6.2. YOLOv6m

The results of the YOLOv6m model training with augmented data can be seen in Table 9. The average values of Precision, Recall, F1-score, mAP@0.5, and mAP@0.5 :0.95 resulted with values of 96,7%, 97,2%, 97%, 99,4% and 99,2% respectively. When compared with the non-augmented training data, performance significantly increased. When trained with non-augmented data, YOLOv6m the values were Precision (53,6%), Recall (97,5%), mAP@0.5 (68,5%), and mAP@0.5 :0.95 (67,4%). The increase was between 30% - 40%, which shows that using greater numbers of data, also more variations of data for training, significantly increased he performance of the YOLOv6m model. With the best fold shown in k=1 with an average of 99% for all evaluation parameters, with the shortest training time.

Google Collab Pro+						
YOLOv6m	**Precision**	**Recall**	**F1 Score**	**mAP@0.5**	**mAP@0.5 :0.95**	**Training Time (hours)**
k=0	0.955	0.959	0.957	0.992	0.99	1,693
k=1	0.99	0.989	0.989	0.996	0.996	1,648
k=2	0.984	0.991	0.987	0.997	0.996	1,716
k=3	0.971	0.977	0.974	0.996	0.995	1,726
k=4	0.936	0.945	0.940	0.99	0.983	1,676
Mean	0.967	0.972	0.970	0.994	0.992	1,692
Std. Dev	0.02	0.019	0.020	0.003	0.005	-

Table 9. YOLOv6m algorithm training results using 10000 images of coffee data

4.6.3. YOLOv7

The results of the YOLOv7 model training with 10000 augmented data of coffee images can be seen in Table 10. The average values of Precision, Recall, F1-score, mAP@0.5, and mAP@0.5 resulted with values of 65,7%, 97,1%, 77,4%, 69%, and 68,8% respectively. When compared with the non-augmented training data, performance increased but not in a significant manner. The increase in performance was only under 10%. The YOLOv7 in this study case did not perform greatly.

Google Collab Pro+						
YOLOv7	Precision	Recall	F1 Score	mAP@0.5	mAP@0.5 :0.95	Training Time (hours)
k=0	0.971	0.966	0.968	0.992	0.992	2,084
k=1	0.5	1	0.667	0.532	0.532	2,051
k=2	0.952	0.949	0.950	0.989	0.989	2,046
k=3	0.501	0.997	0.667	0.525	0.524	2,073
k=4	0.528	0.945	0.677	0.569	0.565	2,089
Mean	0.657	0.971	0.774	0.690	0.688	2,069
Std. Dev	0.23	0.025	0.149	0.232	0.233	-

Table 10. YOLOv7 algorithm training results using 10000 images of coffee data

The model with the best overall performance was YOLOv5m whether using non-augmented training datasets or with augmented datasets. With this result, YOLOv5 was chosen as the model used to identify green coffee beans quality.

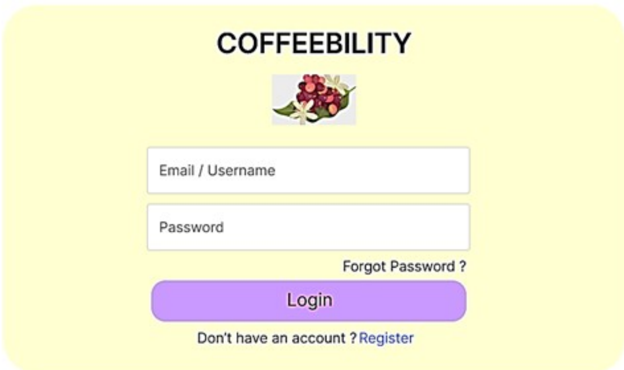

Figure 5. Application login page

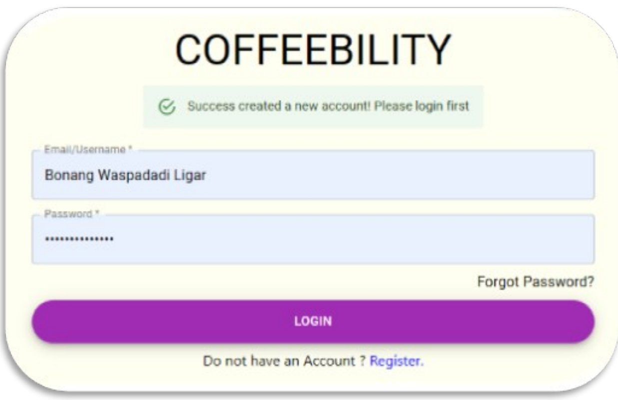

Figure 6. Successful account registration

4.7. Prototype Implementation

To demonstrate the system prototype development, user interface was displayed. Based on the results of the analysis and design of the previous system, this prototype uses the Figma program to create a user interface. Everyone using this program must log in with an email address and password. Users who do not yet have an account can create accounts based on their roles in the supply chain, including farmers, processors, exporters, and end consumers. Users must have a metamask account to register on the system, otherwise the homepage will not be accessible. Metamask is digital wallet used to store digital currency to be used for transactions in the blockchain network. Figure 5 shows the homepage of the prototype, in which users can login or register an account. Figure 6 shows notification of successful account registration in the system.

Figure 7 and 8 show the dashboard of the farmer account. In this page, a farmer can add activities related with cultivation. They can also see the history of certain activities that have been conducted before with all the details.

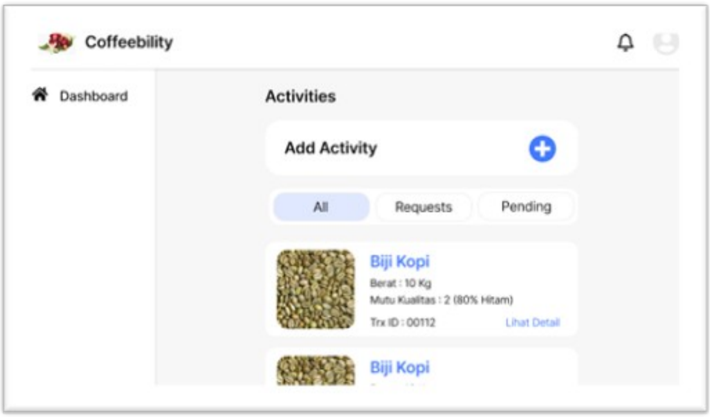

Figure 7. Farmer Dashboard

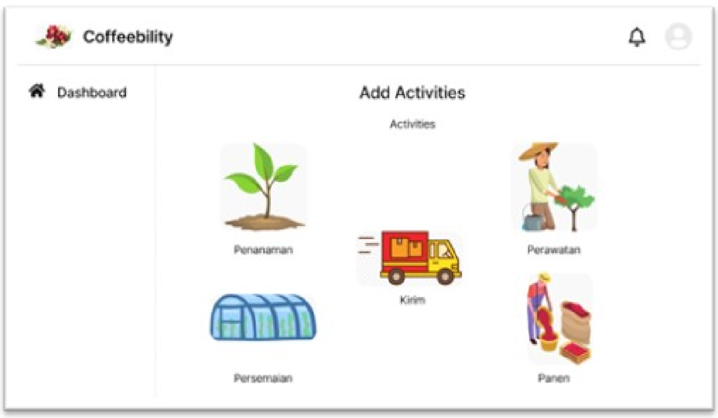

Figure 8. Farmer add activity page

Figure 9 shows the page which users can run the Machine Learning model to identify green coffee beans quality. Figure 10 shows the result of running the machine learning model. The application will show the percentage of good beans and bad beans.

End consumers who have registered in the application can access the home page (Figure 11). On the consumer's home page, there is only the option to enter the *id_transaksi* and also scan the barcode. Both functions retrieve and display traceability information on products that consumers have purchased (Figure 12). The prototype shows that the functions of the system planned in the analysis stage have successfully been implemented. Actors can use the application according to their roles in the supply chain.

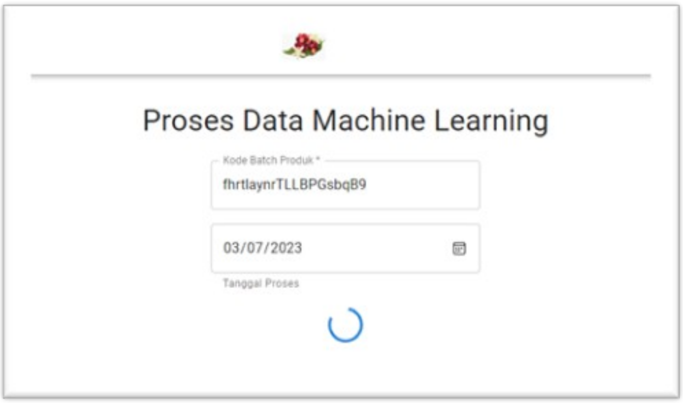

Figure 9. ML model execution

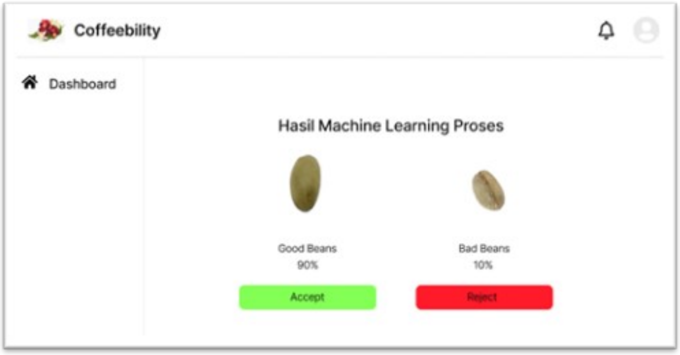

Figure 10. Result of quality identification

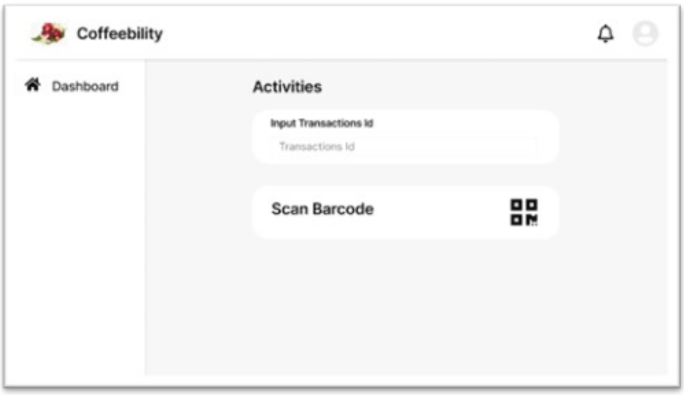

Figure 11. End consumer dashboard

4.8. Advantages of The Proposed System

Some of the studies shown in the literature review have not integrated traceability systems, blockchain and machine learning into one system, particularly for the green coffee supply chain. Those studies only developed traceability systems independently, combine traceability systems with either BC or ML, or only use BC as a standalone system. In our study these solutions are combined as one novel system, combining the advantages of these technologies. Incorporating the tracking and tracing prowess of traceability systems, the untampered and transparent nature of blockchain, with the logic, perception, and reasoning capabilities of machine learning algorithms. Blockchain and traceability systems alone don't have the capability to ensure the quality of the end products, this is where the integration with machine learning becomes significant. ML will verify whether the product manufactured using a blockchain based traceability system can fulfill the claim in improving the quality of end products (green coffee beans).

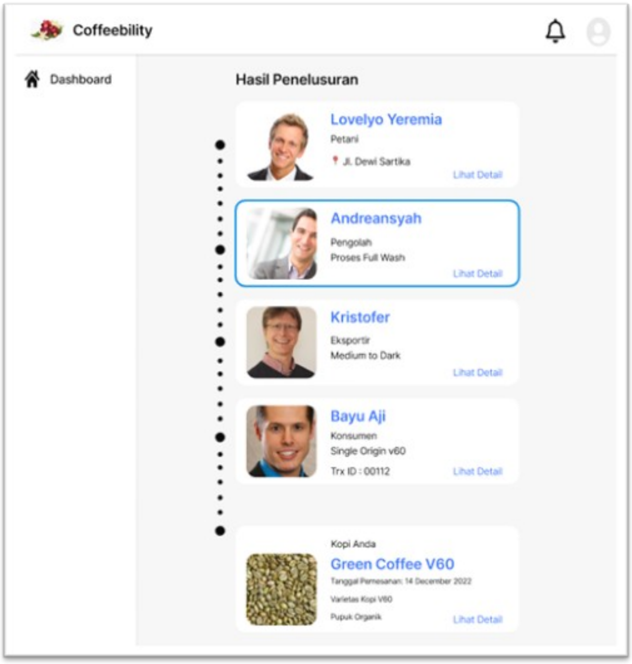

Figure 12. Traceability Information

5. Conclusion

This research aims to design a supply chain traceability system for JPAC based on blockchain and machine learning. Based on this, it can be concluded as follows:

1. Smart contracts provide a holistic and integrated solution for the administration of users, particularly those with the role of farmers in the context of coffee plantation management. These contracts are not only designed to manage users but also to meticulously track the status of the product at each stage of its lifecycle. Furthermore, they are equipped with the capability to present traceability information, which is crucial for maintaining transparency and accountability in the supply chain. This comprehensive approach provided by the smart contracts ensures that all aspects of coffee plantation management, from user management to product tracking and traceability, are efficiently handled, thereby enhancing the overall efficiency and effectiveness of the management process.
2. This research emphasizes the remarkable capability of the YOLOv5m algorithm in efficiently processing small datasets. This efficiency is not compromised even when the algorithm encounters augmented data, demonstrating its robustness and adaptability. These attributes of YOLOv5m, combined with its high performance, make it a highly suitable choice for tasks that require the identification of quality, particularly in scenarios dealing with green coffee beans. Therefore, in the realm of quality control and assurance for green coffee beans, YOLOv5m stands out as an optimal solution due to its effectiveness and resilience.
3. This research has produced a prototype of a web-based coffee supply chain traceability system, which allows users to carry out activities related to their respective roles in the JPAC supply chain.

For future research, it is crucial to delve into the study of hyperparameters within YOLO models. One must question whether the fine-tuning of these hyperparameters, specifically tailored to datasets -in this instance, images of green coffee beans- would yield enhanced performance. Furthermore, the integration of YOLO with alternative algorithms to augment performance warrants additional investigation.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work is supported by Research Program Kemenristek Indonesia in Hibah Penelitian Disertasi Doktor Contract Number: 155/E5/PG.02.00.PT/2022 dated 10 Mei 2022 and Yayasan Pendidikan Gunadarma Jakarta Indonesia, Contract Number: 11.11/LP/UG/VI/2022, 20 June 2022.

References

Azzi, R., Chamoun, R.K., & Sokhn, M. (2019). The power of a blockchain-based supply chain. Computers and Industrial Engineering, 135, 582-592. https://doi.org/10.1016/j.cie.2019.06.042

BPS RI (2022). *Buletin Statistik Perdagangan Luar Negeri Ekspor Menurut Kelompok Komoditi dan Negara, Juni 2022*. In BPS-Statistics Indonesia. Available at: https://www.bps.go.id/publication/2020/10/29/8be61162ee1931c11740da7a/buletin-statistik-perdagangan-luar-negeri-ekspor-menurut-kelompok-komoditi-dan-negara-agustus-2020.html

Chen, S., Brahma, S., Mackay, J., Cao, C., & Aliakbarian, B. (2020). The role of smart packaging system in food supply chain. *Journal of Food Science,* 85, 517-525. https://doi.org/10.1111/1750-3841.15046

Dasipah, E., Sukmawati, D., & Faturachman, D.P. (2021). Faktor Kelembagaan, Sosial Ekonomi Dan Penerapan (Adopsi) Teknologi Terhadap Keberhasilan Usahatani Kopi Arabika Java Preanger. *Paspalum: Jurnal Ilmiah Pertanian,* 9(1), 93-103. https://doi.org/10.35138/paspalum.v9i1.282

Dhatterwal, J.S., Kaswan, K.S., & Ojha, R.P. (2023). The Role of Multiagent System in Industry 4.0. In Chatterjee, J.M., Garg, H., & Thakur, R.N. (Eds.), *A Roadmap for Enabling Industry 4.0 by Artificial Intelligence* (227-246). Scrivener Publishing LLC. https://doi.org/10.1002/9781119905141.ch13

Djuwendah, E. (2019). Penguatan Kelembagaan Koperasi Produsen Kopi Java Preanger Dalam Upaya Meningkatkan Kinerja Bisnis Dan Pendapatan Petani Kopi. *Dharmakarya,* 7(4), 228-233. https://doi.org/10.24198/dharmakarya.v7i4.9773

Djuwendah, E., Karyani, T., Sadeli, A.H., & Hapsari, H. (2019). Study on the Roles and Benefits of Cooperative Producers of Coffee Margamulya in Java Preanger Coffee Agribusiness. *IOP Conference Series: Earth and Environmental Science,* 334(1), 1-7. https://doi.org/10.1088/1755-1315/334/1/012055

García, M., Candelo-Becerra, J.E., & Hoyos, F.E. (2019). Quality and defect inspection of green coffee beans using a computer vision system. *Applied Sciences (Switzerland),* 9(19), 1-18. https://doi.org/10.3390/app9194195

Ginanjar, Y., Apiatno, A., & Amanda, H. (2020). Kinerja Rantai Pasokan Kopi Arabika Java Preanger Di Kabupaten Sumedang Jawa Barat. *E-Jurnal Manajemen Universitas Udayana,* 9(11), 3511-3532. https://doi.org/10.24843/EJMUNUD.2020.v09.i11.p05

Ginting, C.P., & Kartiasih, F. (2019). Analisis Ekspor Kopi Indonesia Ke Negara-Negara Asean. *Jurnal Ilmiah Ekonomi Dan Bisnis,* 16(2), 143-157. https://doi.org/10.31849/jieb.v16i2.2922

Ibnu, M. (2019). Sustainability Standards and Certification Dan Pathway Menuju Produksi Kopi Berkelanjutan Di Indonesia (Sustainability Standards and Certification and Pathway Toward Sustainable Coffe Production in Indonesia). *Journal of Extension and Development,* 1, 97-108. https://doi.org/10.23960/jsp.Vol1.No2.2019.21

ICO - International Coffee Organization (2020). *Total exports by all exporting countries.* Available at: https://www.ico.org/historical/1990 onwards/PDF/1e-exports.pdf

Jamil, A.S. (2019). Daya Saing Perdagangan Kopi Indonesia di Pasar Global. *Agriekonomika,* 8(1), 26-35. https://doi.org/10.21107/agriekonomika.v8i1.4924

Jian, M.S., & Pan, C.J. (2022). Blockchained Industry Information Handoff Based on Internet of Things Devices with Intelligent Customized Object Recognition. *Sensors,* 22(6), 1-15. https://doi.org/10.3390/s22062312

Kirli, D., Couraud, B., Robu, V., Salgado-Bravo, M., Norbu, S., Andoni, M. et al. (2022). Smart contracts in energy systems: A systematic review of fundamental approaches and implementations. *Renewable and Sustainable Energy Reviews,* 158, 1-28. https://doi.org/10.1016/j.rser.2021.112013

Kudus, A., Widayat, H.P., & Abubakar, Y. (2019). Kriteria Mutu Kopi Arabika Gayo Pada Beberapa Koperasi dan Eksportir di Aceh Tengah. *Jurnal Ilmiah Mahasiswa Pertanian,* 4(2), 274-279. https://doi.org/10.17969/jimfp.v4i2.11018

Landínez, S.P.C., Rodríguez, P.E.C., & Gómez, D.F.S. (2019). Design and implementation of a software for the traceability of coffee processing. *Ciencia Tecnologia Agropecuaria,* 20(3), 523-536. https://doi.org/10.21930/rcta.vol20_num3_art1588

Ligar, B.W., Madenda, S., Mardjan, S.S., & Kusuma, T.M. (2023). Functional Requirements and Traceability System Information Modeling on Java Preanger Coffee Supply Chain. *2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)* (99-104). https://doi.org/10.1109/ICIMCIS56303.2022.10017857

Nasution, T.H., Rumansa, M., & Lukman-Adlin, H. (2019). Designing the quality of coffee bean detection application using Hue Saturation Intensity. *IOP Conference Series: Materials Science and Engineering,* 648(1), 1-7. https://doi.org/10.1088/1757-899X/648/1/012036

Nepal, U., & Eslamiat, H. (2022). Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs. *Sensors,* 22(2), 1-15. https://doi.org/10.3390/s22020464

Pradana, I.G.M.T., Djatna, T., & Hermadi, I. (2020). Blockchain Modeling for Traceability Information System in Supply Chain of Coffee Agroindustry. *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (217-224). Depok, Indonesia. https://doi.org/10.1109/ICACSIS51025.2020.9263214

Randy, P., & Marianti (2019). Analisis Rantai Nilai Kopi Untuk Meningkatkan Nilai Tambah Pada Petani Di Indonesia. *Jurnal Administrasi Bisnis,* 5(2), 57-69.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (779-788). https://doi.org/10.1109/CVPR.2016.91

Saberi, S., Kouhizadeh, M., Sarkis, J., & Shen, L. (2019). Blockchain technology and its relationships to sustainable supply chain management. *International Journal of Production Research,* 57(7), 2117-2135. https://doi.org/10.1080/00207543.2018.1533261

Siswanto. (2019). *Rancang Bangun Sistem Informasi Traceability Jagung Pakan.* Institut Pertanian Bogor.

Thiruchelvam, V., Mughisha, A.S., Shahpasand, M., & Bamiah, M. (2018). Blockchain-based technology in the coffee supply chain trade: Case of Burundi coffee. *Journal of Telecommunication, Electronic and Computer Engineering,* 10(3-2), 121-125.

Thuan, D. (2021). *Evolution of Yolo Algorithm and Yolov5: the State-of-the-Art Object Detection Algorithm.* Oulu University of Applied Sciences.

Wang, J., Yue, H., & Zhou, Z. (2017). An improved traceability system for food quality assurance and evaluation based on fuzzy classification and neural network. *Food Control,* 79, 363-370. https://doi.org/10.1016/j.foodcont.2017.04.013

Wang, P., Tseng, H.W., Chen, T.C., & Hsia, C.H. (2021). Deep convolutional neural network for coffee bean inspection. *Sensors and Materials,* 33(7), 2299-2310. https://doi.org/10.18494/SAM.2021.3277

Yang, S., Li, C., Mei, Y., Liu, W., Liu, R., Chen, W. et al. (2021). Determination of the Geographical Origin of Coffee Beans Using Terahertz Spectroscopy Combined with Machine Learning Methods. *Frontiers in Nutrition,* 8, 1-10. https://doi.org/10.3389/fnut.2021.680627

Yustisar, M. (2018). Model Identifikasi dan Analisis Risiko Rantai Pasok Agroindustri Kopi Gayo dengan Pendekatan Fuzzy Analytichal Hierarchy Process. *Journal of Informatics and Telecommunication Engineering,* 1(2), 51-57. https://doi.org/10.31289/jite.v1i2.1425