INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

# Classification of Cocoa Beans Based on the Grade Level of Fermentation using KNN Algorithm

**Emmanuel D Santos**[1]*, **Diane B Remot**[2]

**1** Information Technology Department, University of Eastern Philippines, University Town, Northern Samar, Philippines
**2** Information Technology Department, Eastern Visayas State University, Tacloban City, Leyte, Philippines

OPEN ACCESS

**Received:** 22-08-2023
**Accepted:** 25-03-2024
**Published:** 21-05-2024

**Citation:** Santos ED, Remot DB (2024) Classification of Cocoa Beans Based on the Grade Level of Fermentation using KNN Algorithm. Indian Journal of Science and Technology 17(21): 2166-2176. https://doi.org/10.17485/IJST/v17i21.1997

* **Corresponding author**.

emmanuel.santos@evsu.edu.ph

**Funding:** None

**Competing Interests:** None



Published By Indian Society for Education and Environment (iSee)

**ISSN**
Print: 0974-6846
Electronic: 0974-5645

## Abstract

**Objectives:** This study aims to develop a cacao bean fermentation prototype, specifically this study covers: (1) Development of cacao bean fermentation device using Arduino microcontroller, (2) Creating a model that will be used in evaluating the fermented cocoa bean using KNN Algorithm, and (3) Evaluating the developed system using the ISO 9126 software quality model in terms of functionality, reliability, usability, efficiency, maintainability, and portability. **Methodology:** The researcher collected, presented, and analyzed the necessary data and information to address the research objectives. Reasons and justifications for the Developmental Model, Gantt chart, Data Gathering Procedure and Technique, Context Diagram, Data and Process Model, and Schematic Diagram are given. **Findings:** This study aimed to address these issues by automating the cacao bean fermentation process. A novel device was created and assessed using ISO 9126 and the K-Nearest Neighbor algorithm. The results demonstrated excellent performance in functionality (4.77), reliability (4.56), usability (4.58), efficiency (4.6), maintainability (4.6), and portability (4.56). The device effectively mixed cacao beans during fermentation and achieved an 85.3% accuracy rate in image classification. **Novelty:** In comparison to existing methods, the device surpassed expectations in terms of fermentation control and software reliability. Overall, the developed Cacao Bean Fermentation Device significantly enhances functionality and control, ensuring consistent quality in chocolate production.

**Keywords:** Arduino Microcontroller; KNN algorithm; Image Analysis; Cacao Bean Fermentation

## 1 Introduction

Cacao Lavezares Chocolates and Confiserie, a small-scale business farm based in Brgy. Urdaneta, Lavezares, Northern Samar, produces and develops cacao beans. The conventional method of cacao fermentation in Lavezares involves harvesting, sorting, opening pods, fermentation, mixing, drying, and sorting and grading. The cacao beans

are placed in large, shallow fermentation boxes or fermenting containers lined with banana leaves. The beans are covered with sweet pulp and left to ferment for several days, typically between 3-7 days. Cocoa yields and quality are diminished by prolonged dry seasons, rainfall, extreme temperatures, and the emergence of novel pests and diseases.

Anggraini et al.[1] study discovered a model that uses artificial neural network models to quantify the standard and accurate measurement of the degree of fermentation of cocoa beans. This model can then use color feature extraction—that is, the average value of RGB and L*a*b—to segment, calculate, and grade classification. It has been discovered that the Multilayer Perceptron (MLP) Artificial Neural Network (ANN) is better than previous models, achieving 94% training and validation accuracy.

Alhasani[2] elevated the accuracy of breast cancer classification through the application of information gain feature selection and machine learning techniques on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. The presented model attains a remarkable maximum classification accuracy of 100%, accompanied by a weighted average precision and recall of 100%, using a C4.5 decision tree. In comparison, an SVM achieves an accuracy of 98.42%, with corresponding weighted averages for precision and recall at 98.17% and 98.58%, respectively, utilizing a C4.5 decision tree. Additionally, the NB algorithm achieves an accuracy of 96%, with weighted average precision and recall values of 18.57% and 50%, respectively. A thorough comparison of the proposed model's results with similar studies reveals substantial advancements, signifying new possibilities in the realm of breast cancer detection.

The constructed irrigation control device was assessed using the ISO 9126 - Software Product Quality Model in a study by Villaluz[3]. The findings indicated that the Arduino-based prototype device had the ability to save water for irrigation of farmlands and rice fields.

As a result, it has been found that there are several reasons why automating the cacao bean fermentation procedure may be necessary or beneficial. This paper involves two methods: the development of the prototype device using Arduino Microcontroller, and the image processing phase and the grading classification phase. In the image processing phase, the researcher utilized orange software to initiate data computation for an algorithm by simplifying data processing using the training and forecasting functions. In a study conducted by Brosas et al.[4] the researcher tested the trained images in the ANFIS algorithm to determine the accuracy in predicting the result using Receiver Operating Characteristic Analysis, Confusion Matrix, Test and Score, and Predictions. In this study, this technique was adopted by the researcher. The algorithm consists of five (5) stages. First was importing the image dataset into orange software and identifying image classifications. The second was preprocessing the image data in Orange using a built-in image processor. Third was selecting the KNN algorithm in testing and comparing. Fourth was testing images to compare accuracy rates. The Fifth was to evaluate the cacao beans fermented in the developed device.

## 2 Methodology

In the study of Iguazu[4], researchers utilized image processing and machine learning techniques to develop an algorithm specifically aimed at classifying oranges by size. They conducted testing using 204 orange images categorized into three size classes: small, medium, and large. Through the application of supervised learning techniques, the algorithm demonstrated promising results. It achieved an impressive overall accuracy rate of 82% and a coefficient of 0.66, showcasing its efficacy in comparison to traditional methods.

This study focused on the development of an automated cacao bean processing system that can be used to monitor the fermenting, drying, and roasting stages of the cacao bean. These automated devices are equipped with a variety of sensors in order to identify a high-quality cacao bean. The microcontroller used in the device is Arduino ATMega with a DC Motor that rotates 180 degrees to continuously perform the mixing process. The fermented cacao bean will be evaluated using K-Nearest Neighbors Algorithm (KNN). The fermented cacao will be graded to determine the samples submitted for categorization according to their grade level value (Good Quality, Slightly Good Quality and Bad Quality) using the Image Analytics using Orange software. The grading of the cacao bean were adapted from the traditional grading system of Cacao Lavezares Chocolates and Confiserie. The images were trained in the classification of cacao bean quality taken from the fermented cacao bean of Cacao Lavezares Chocolates and Confiserie. Four hundred data sets were loaded in the orange software for Image Analytics 50% were trained as Good Quality and 50% were trained as Bad Quality cacao beans.

1. **Process:**

- Phase 1: Development of cacao bean fermentation device.

To ferment and dry cacao beans, Figure 1 shows the process model of phase 1: fermentation and drying process of the study. In the fermentation phase, the researcher developed an Arduino microcontroller-based device built with Ph. level, temperature,

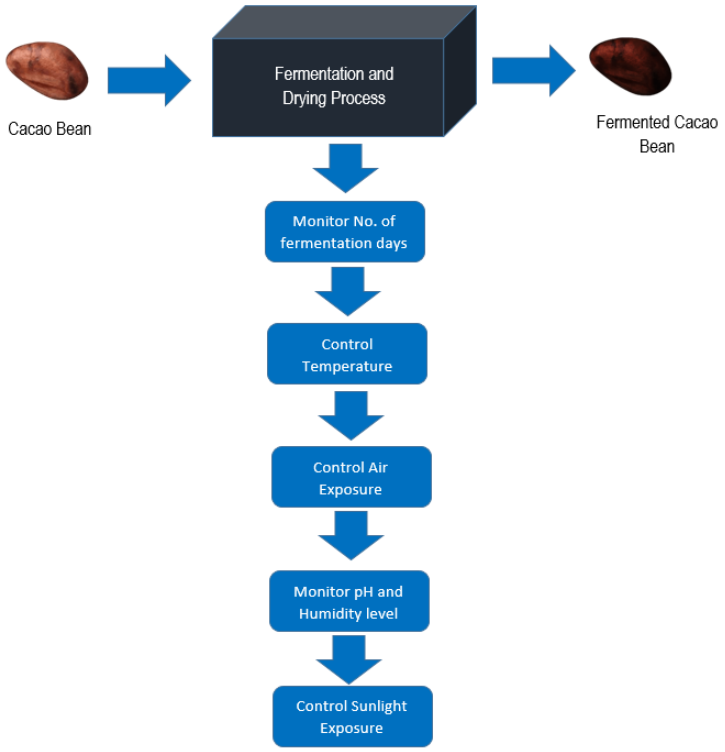



**Fig 1. Fermentation and Drying Process**

and humidity sensors. The fermentation process monitors the number of days of the cacao bean in a modified greenhouse box. Monitoring of pH and humidity levels and controlling temperature, air, and sunlight exposure are included in this procedure.

- Grading of the fermented cocoa bean grading quality process using KNN Algorithm.

Figure 2 depicts the adapted cacao bean grading classification process using the KNN technique. The fermented and dried cacao bean quality was tested using this technique.

The researchers collected categorized samples to form the system's dataset. These samples consist of beans manually classified by the CEO of Lavezares Chocolate and Confeserie, a scientist and an expert in cacao classification. The dataset comprises 400 sample beans, with 200 beans classified as good quality fermented beans and another 200 classified as bad quality fermented beans. After pre-processing procedures, image classification is the next step, as mentioned in the methodology of the study.

## 3 Results and Discussion

Through experimentation and data analysis, this phase presents the results of applying the K-Nearest Neighbors (KNN) algorithm to categorize cocoa beans based on their fermentation grades. The subsequent discussion provides valuable insights and interpretations of the effectiveness and implications of the data model classification approach.

Cacao bean fermentation is a crucial step in producing high-quality cocoa beans. In Cacao Lavezares Chocolates and Confiserie Cacao, the fermentation process has been conducted using manual labor, which according to them, can lead to inconsistent results and increased labor costs. The researcher developed a new cacao bean fermentation device to address these issues.

Figure 3 shows that the developed device, based on Madarang et al.[5], underwent modifications to enhance its functionality. The cover and base sections were constructed using metal materials, ensuring durability and stability. The internal mechanism features a roller screen with a size equivalent to a $\frac{1}{2}$ pillow block, facilitating efficient bean processing. A No. 12 screen with a circular diameter is utilized within the box for optimal performance. A stainless-steel roller is employed to support the drying

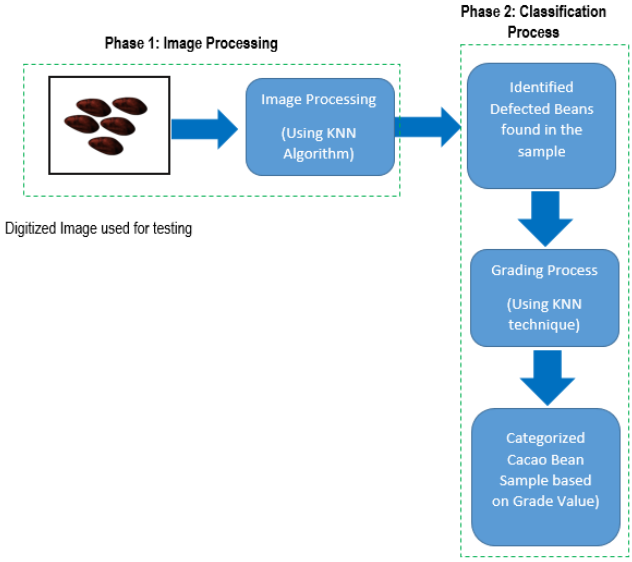



**Fig 2. Image Processing**

**Fig 3. Cocoa bean fermenting device, 360 degrees rotating container**

process, securely holding the beans in place.

To test the quality of the cacao bean fermented in the device, the researcher collected 400 images of the fermented cacao bean in the traditional fermentation procedure of Lavezares. These images are categorized by Lavezares according to their grade level value (Good Quality and Bad Quality). The researcher utilized the KNN algorithm. The model evaluation resulted in 85.3% classification accuracy. The model was evaluated using the test and score widget in Orange data mining software through a five-fold cross-validation.

As presented in Figure 4, the evaluation results indicate that the KNN model has an accuracy rate of 85.3% in classification. This percentage indicates the proportion of data instances that were correctly classified.

Figure 5 shows that in 200 images classified as bad quality, 91.0% of the images are accurately classified. In the good quality category 79.5% are classified accurately.

To better illustrate how accurate the model is, an ROC curve was utilized, which compares the true positive rate and false positive rate of the model. The results of the ROC analysis are shown in Figures 6 and 7. These figures indicate that both the ”Good Quality” and ”Bad Quality” true positive rates, which are represented by a green line, are above the 0.5 threshold indicated

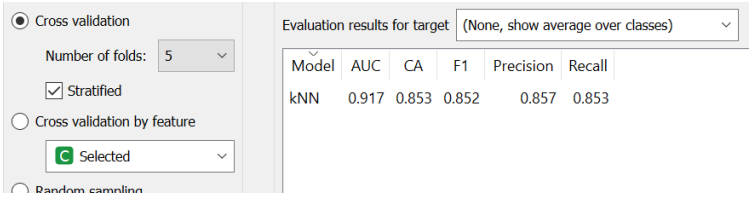



**Fig 4. Evaluation of Accuracy Result**

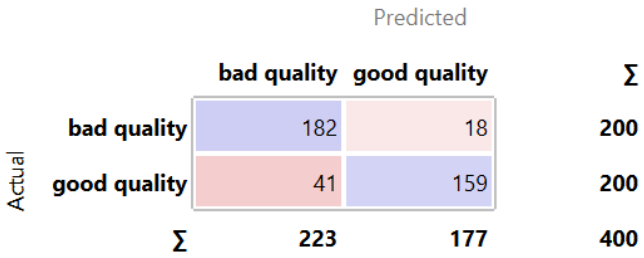



**Fig 5. Classification of Accuracy Result**

by a red dotted line. This suggests that the model has more true positive results than false positives and therefore performed well.

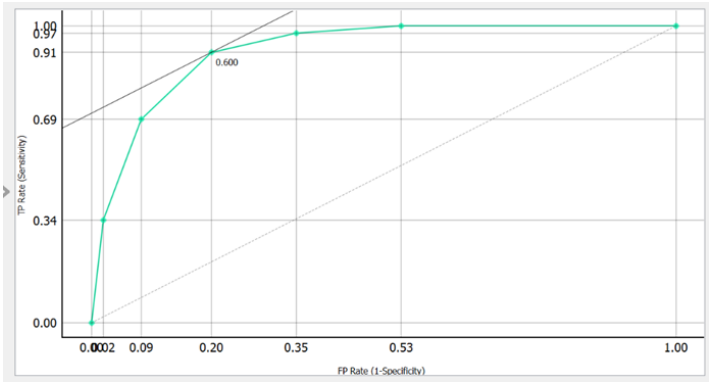



**Fig 6. ROC curve for "Bad Quality"**

Figure 8 shows that the cacao bean fermentation device can maintain a temperature between 45 and 50 degrees. According to the International Cocoa Organization (ICCO), the temperature range for fermenting cacao beans is generally between 45°C to 50°C (113°F to 122°F). They state that this temperature range is optimal for the growth of microorganisms that are responsible for the fermentation process, which is necessary for developing the desired chocolate flavor.

Images of fermented cacao beans were captured after 6 days of fermentation. Twenty (20) fermented cacao beans were tested using the developed KNN model in Orange software. Figure 9 shows that 100% of the fermented cacao beans through the developed device are predicted as "Good Quality".

The employees of Cacao Lavezares Chocolates and Confiserie served as the evaluators in the conduct of this study. The researcher used ISO 9126 questionnaire for evaluation of the existing fermentation procedure as well as in the developed device. It was both evaluated in terms of its functionality, reliability, usability, efficiency, maintainability and portability. The evaluation result is depicted in Figure 10.

Figure 11 depicts that the Cacao Bean Fermentation Device outperforms the existing fermentation procedure in all evaluated areas related to the fermentation process. The Cacao Bean Fermentation Device receives a score of 5 out of 5 in consistently mixing cacao beans during fermentation, monitoring the temperature level of the fermentation device, and controlling the

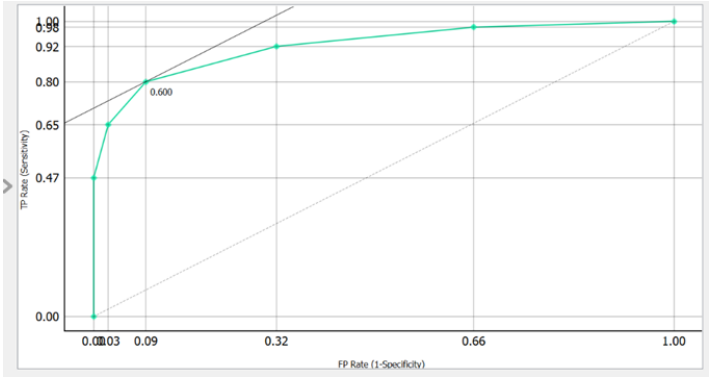



**Fig 7. ROC Curve for "Good Quality"**

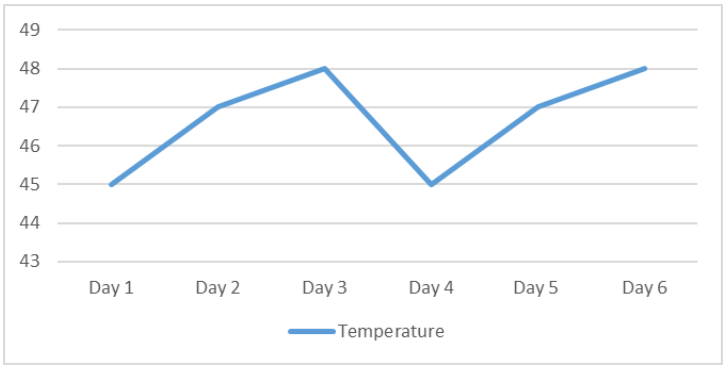
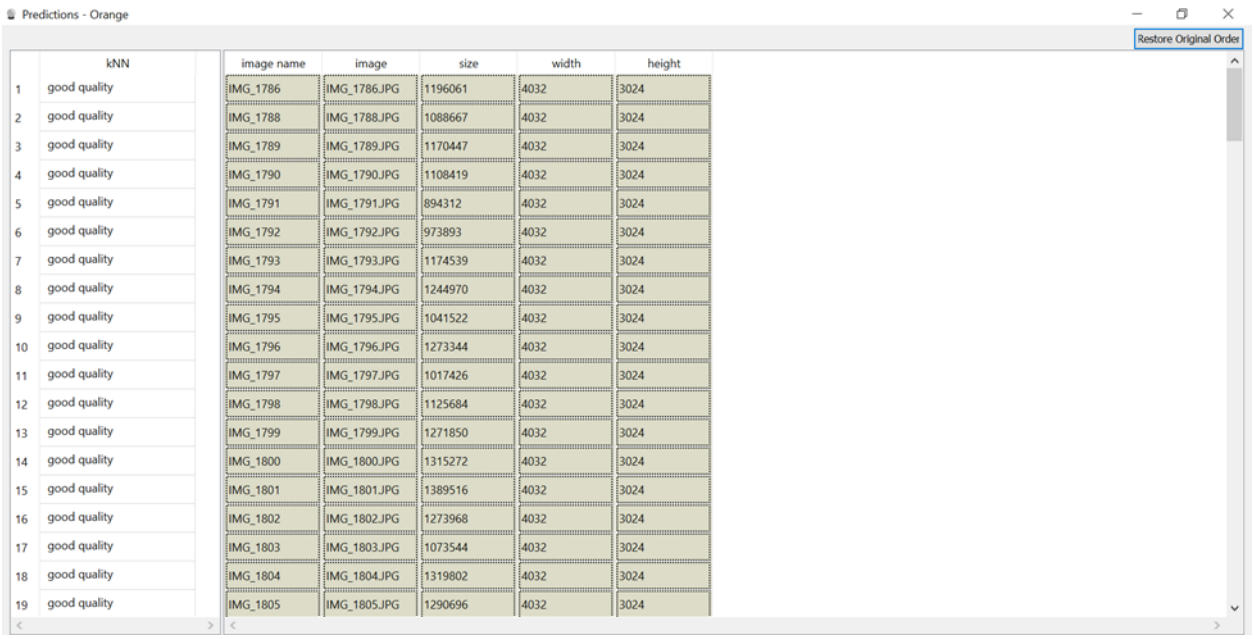



**Fig 8. Graph for the cacao fermentation device temperature and cacao bean container rotation speed**




| | kNN | image name | image | size | width | height |
|---|---|---|---|---|---|---|
| 1 | good quality | IMG_1786 | IMG_1786.JPG | 1196061 | 4032 | 3024 |
| 2 | good quality | IMG_1788 | IMG_1788.JPG | 1088667 | 4032 | 3024 |
| 3 | good quality | IMG_1789 | IMG_1789.JPG | 1170447 | 4032 | 3024 |
| 4 | good quality | IMG_1790 | IMG_1790.JPG | 1108419 | 4032 | 3024 |
| 5 | good quality | IMG_1791 | IMG_1791.JPG | 894312 | 4032 | 3024 |
| 6 | good quality | IMG_1792 | IMG_1792.JPG | 973893 | 4032 | 3024 |
| 7 | good quality | IMG_1793 | IMG_1793.JPG | 1174539 | 4032 | 3024 |
| 8 | good quality | IMG_1794 | IMG_1794.JPG | 1244970 | 4032 | 3024 |
| 9 | good quality | IMG_1795 | IMG_1795.JPG | 1041522 | 4032 | 3024 |
| 10 | good quality | IMG_1796 | IMG_1796.JPG | 1273344 | 4032 | 3024 |
| 11 | good quality | IMG_1797 | IMG_1797.JPG | 1017426 | 4032 | 3024 |
| 12 | good quality | IMG_1798 | IMG_1798.JPG | 1125684 | 4032 | 3024 |
| 13 | good quality | IMG_1799 | IMG_1799.JPG | 1271850 | 4032 | 3024 |
| 14 | good quality | IMG_1800 | IMG_1800.JPG | 1315272 | 4032 | 3024 |
| 15 | good quality | IMG_1801 | IMG_1801.JPG | 1389516 | 4032 | 3024 |
| 16 | good quality | IMG_1802 | IMG_1802.JPG | 1273968 | 4032 | 3024 |
| 17 | good quality | IMG_1803 | IMG_1803.JPG | 1073544 | 4032 | 3024 |
| 18 | good quality | IMG_1804 | IMG_1804.JPG | 1319802 | 4032 | 3024 |
| 19 | good quality | IMG_1805 | IMG_1805.JPG | 1290696 | 4032 | 3024 |

**Fig 9. Prediction Result**

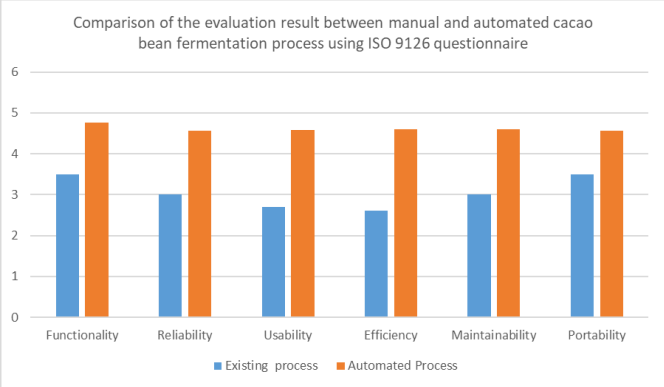



**Fig 10. Comparison of the evaluation results during testing**

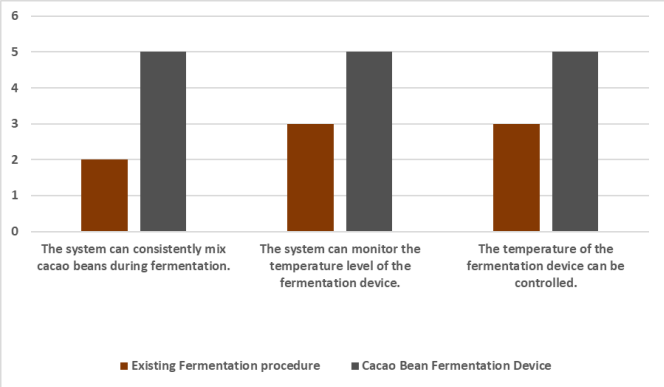



**Fig 11. Comparison of the existing fermentation procedure and cacao bean fermentation device in terms of functionality**

temperature of the fermentation device.

In contrast, the existing fermentation procedure receives a lower score of 2 out of 5 in consistently mixing cacao beans during fermentation, indicating that the mixing process may not be uniform and consistent. Additionally, the existing fermentation procedure receives a score of 3 out of 5 in monitoring the temperature level of the fermentation device and controlling the temperature of the fermentation device. This suggests that the existing fermentation procedure may not have an effective temperature monitoring and control system in place to ensure that the temperature remains within the desired range during the fermentation process.

Overall, the data suggests that the Cacao Bean Fermentation Device is a more functional compared to the existing fermentation procedure when it comes to the mixing, monitoring, and control of the fermentation process.

Based on the result shown in Figure 12, it appears that the developed fermentation device outperforms the existing fermentation process in all areas related to software reliability. The developed device has a higher score in avoiding failure as a result of faults in the software (4.8 vs. 3), maintaining a specified level of performance in cases of software faults or interface infringement (4.44 vs. 3.1), re-establishing a specified level of performance and recovering data directly affected in the case of a failure (4.7 vs. 3.33), adhering to standards, conventions or regulations relating to reliability (5 vs. 2.8), and enabling the user to understand whether the software is suitable, and how it can be used for particular tasks and conditions of use (4.52 vs. 2.7).

These differences suggest that the developed device is more reliable, robust, and user-friendly compared to the existing fermentation procedure.

Figure 13 shows comparison of the survey result in terms of usability, the result shows that the user can recognize how to use the system easily (Existing Fermentation procedure = 2.3, Cacao Bean Fermentation Device = 4.9). The Cacao Bean Fermentation Device performs better in this criterion than the Existing Fermentation procedure, indicating that users can recognize how to use it more easily.

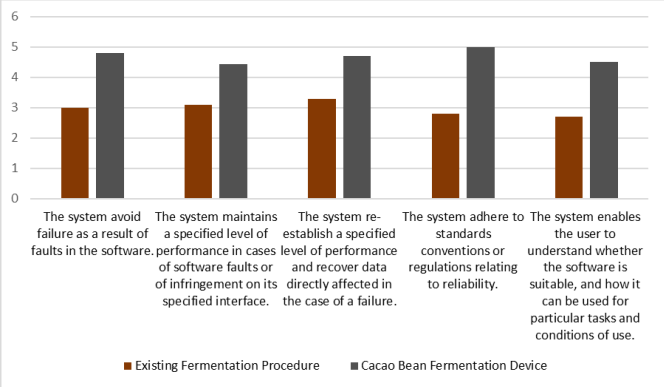



**Fig 12. Comparison of the existing fermentation procedure and cacao bean fermentation device in terms of reliability**

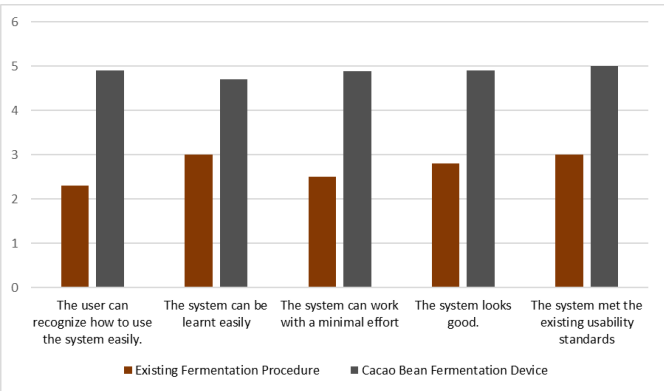



**Fig 13. Comparison of the existing fermentation procedure and cacao bean fermentation device in terms of usability**

The system can be learned easily (Existing Fermentation procedure = 3, Cacao Bean Fermentation Device = 4.7). The Cacao Bean Fermentation Device performs slightly worse than the Existing Fermentation procedure in this criterion, indicating that users may need slightly more time to learn how to use it.

The system can work with minimal effort (Existing Fermentation procedure = 2.5, Cacao Bean Fermentation Device = 4.88). The Cacao Bean Fermentation Device again performs better in this criterion than the Existing Fermentation procedure, indicating that users can use it with less effort.

The system looks good (Existing Fermentation procedure = 2.8, Cacao Bean Fermentation Device = 4.9). Both systems perform well in this criterion, with the Cacao Bean Fermentation Device performing slightly better than the Existing Fermentation procedure.

The system meets the existing usability standards (Existing Fermentation procedure = 3, Cacao Bean Fermentation Device = 5). Both systems meet the existing usability standards, with the Cacao Bean Fermentation Device performing better than the Existing Fermentation procedure.

Overall, the data suggests that the Cacao Bean Fermentation Device has better usability than the Existing Fermentation procedure in most evaluated criteria, except for the criterion of learnability. The Cacao Bean Fermentation Device performs better in the criteria of user recognition, minimal effort, good appearance, and meeting existing usability standards.

Based on the data depicted in Figure 14, it can be concluded that the Cacao Bean Fermentation Device performs better than the Existing Fermentation Procedure in terms of providing appropriate response and processing times, using appropriate amounts and types of resources, and adhering to standards or conventions relating to efficiency.

The Cacao Bean Fermentation Device scored higher in all three categories compared to the Existing Fermentation Procedure. This suggests that the Cacao Bean Fermentation Device is more efficient and effective in performing its function, providing a faster and more reliable response with optimal resource usage.

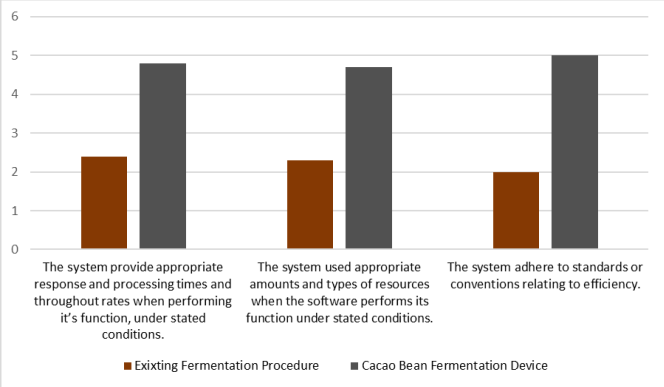



**Fig 14. Comparison of the existing fermentation procedure and cacao bean fermentation device in terms of efficiency**

It is important to note that the data only considers the performance and efficiency of the system and does not take into account other factors such as cost, maintenance, and ease of use. Therefore, a more comprehensive evaluation of the two systems is necessary to determine the best option for a specific use case or scenario.

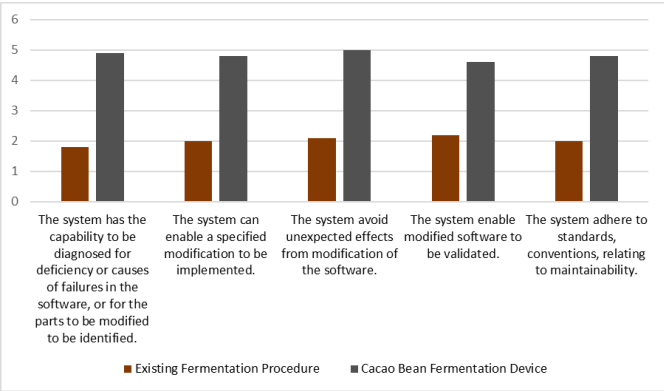



**Fig 15. Comparison of the existing fermentation procedure and cacao bean fermentation device in terms of maintainability**

Based on the ratings shown in Figure 15, it appears that the Cacao Bean Fermentation Device performs better in terms of diagnosing and modifying the software, as well as avoiding unexpected effects from modifications, and adhering to maintainability standards. However, the Existing Fermentation Procedure seems to be slightly better in terms of enabling modifications to be implemented, while both systems have similar ratings for enabling modified software to be validated.

The Figure 16 represents the comparison of the existing fermentation procedure and the proposed cacao bean fermentation device in terms of the system's capability to be adopted and installed in different specified environments, co-exist with other independent software, and adhere to standards and conventions relating to portability.

The system can be adopted for different specified environments without applying actions or means other than those provided for this purpose for the software considered, with a score of 2 for the existing fermentation procedure and 4.56 for the cacao bean fermentation device, indicating a significant improvement in the device's capability to adapt to different environments.

The system can be installed in a specified environment, with a score of 2.3 for the existing fermentation procedure and 4.78 for the cacao bean fermentation device, indicating a significant improvement in the device's ease of installation.

The system can co-exist with other independent software in a common resource, with a score of 1.7 for the existing fermentation procedure and 5 for the cacao bean fermentation device, indicating a significant improvement in the device's capability to work with other software applications.

The system can be used in place of another specified software product for the same purpose in the same environment, with a score of 1.6 for the existing fermentation procedure and 4.9 for the cacao bean fermentation device, indicating a significant improvement in the device's capability to replace other software products for the same purpose.

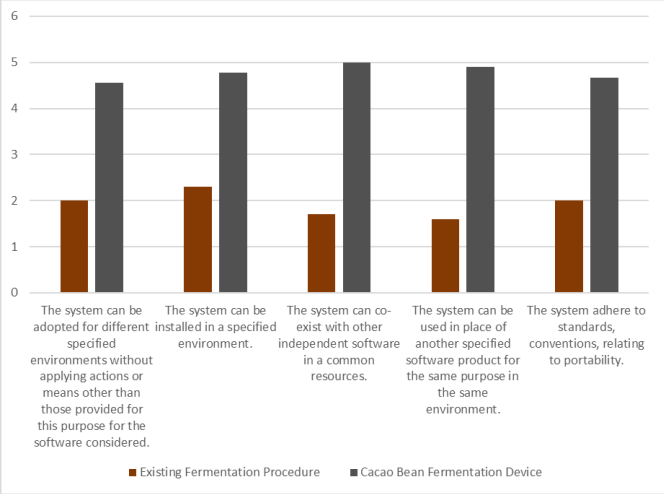



**Fig 16. Comparison of the existing fermentation procedure and cacao bean fermentation device in terms of portability**

Lastly, the system adheres to standards, conventions relating to portability, the cacao bean fermentation device met the requirements for portability standards.

## 4 Conclusion

The researcher designed a prototype with durability and stability, featuring a metal construction for the cover and base sections. Its internal mechanism includes a roller screen of appropriate size for efficient bean processing and a No. 12 circular screen for optimal performance. A stainless-steel roller ensures secure bean placement during the drying process, and essential electrical components such as a coil heater and a 12-volt dynamo are incorporated for effective heating and smooth operation. The device also employs galvanized iron sheets on the inner walls to enhance heat absorption, ensuring optimal performance. With dimensions of approximately 4 feet in height and 3 feet in width, the device offers a spacious container capable of accommodating up to 5 kilograms of cacao beans for efficient and sizable batches during fermentation. Results showed that the developed device could maintain the optimal temperature range for fermenting cacao beans and consistently mix cacao beans during fermentation.

The developed KNN model accurately classified images of fermented cacao beans as ”Good Quality” or ”Bad Quality” with an 85.3% classification accuracy rate. The study captured pictures of fermented cacao beans after 6 days of fermentation and tested 20 beans using a KNN model in orange software. The results showed that all the beans tested were predicted to be of good quality.

In summary, the developed system meets the ISO 9126 software quality model’s functionality, reliability, usability, efficiency, maintainability, and portability standards. This indicates that the system is effective, dependable, user-friendly, efficient, maintainable, and easily deployable, making it suitable for its intended purposes.

### Recommendations

1. Develop a bigger device that can be used in fermenting a bigger number of cacao beans.
2. Develop a standalone user-friendly application to evaluate the fermented cacao beans without using the orange software.
3. Evaluate the impact of the device on the quality of cacao beans and the overall profitability of cacao farming and processing. This could involve conducting surveys or interviews with farmers and processors to assess the device’s effectiveness and identify improvement areas.

## References

1) Anggraini CD, Putranto AW, Iqbal Z, Firmanto H, Riza DFA. Preliminary study on development of cocoa beans fermentation level measurement based on computer vision and artificial intelligence. In: 2021 International Conference on Green Agro-industry and Bioeconomy;vol. 924 of IOP Conference Series: Earth and Environmental Science. IOP Publishing. 2021;p. 1–8. Available from: https://dx.doi.org/10.1088/1755-1315/924/1/012019.

2) Alhasani AT, Alkattan H, Subhi AA, El-Kenawy ESM, Eid MM. A Comparative Analysis of Methods for Detecting and Diagnosing Breast Cancer Based on Data Mining. *Journal of Artificial Intelligence and Metaheuristics*. 2023;4(2):08–17. Available from: https://doi.org/10.54216/JAIM.040201.
3) and LRV. Project BIS-IG: Prototyping of Wireless Irrigation Control System Using Arduino Microcontroller with SMS Notification. *Indian Journal of Science and Technology*. 2022;15(21):1012–1021. Available from: https://dx.doi.org/10.17485/ijst/v15i21.19.
4) Santiago WE, Lopes GD, Faceto AD, Júnior ARL, Caldeira RF. Classification of oranges using digital image processing. *Acta Iguazu*. 2019;8(2):49–58. Available from: https://e-revista.unioeste.br/index.php/actaiguazu/article/view/19227.
5) Madarang BP, G C Magulod J, Ramos HJB. Design and Development of Solar-Powered Mechanical Dryer for Small-Scale Cacao Processing. *Asia Pacific Journal of Multidisciplinary Research*. 2019;7(2):17–23. Available from: http://www.apjmr.com/wp-content/uploads/2019/05/APJMR-2019-7.2.003.pdf.