



(REVIEW ARTICLE)



A review of machine learning applications in Cacao Post-harvest management

Zhaira Marianne Ferrancol Ferran, Luisa Ann Tamayo Ongquit * and Edwin Romeroso Arboleda

Department of Computer, Electronics and Electrical Engineering, College of Engineering and Information Technology, Cavite State University, Philippines.

International Journal of Science and Research Archive, 2024, 11(01), 1540–1550

Publication history: Received on 23 December 2023; revised on 06 February 2024; accepted on 08 February 2024

Article DOI: <https://doi.org/10.30574/ijrsra.2024.11.1.0186>

Abstract

Cacao beans play a crucial role in chocolate production, making them a highly relevant crop. The post-harvesting stage of cacao beans, encompassing classification, quality assessment, and fermentation, holds significant importance. With the increasing demand for marketable cacao beans, the need for reliable, accurate, and fast technologies has emerged. This study provides a comprehensive review of machine learning techniques in the post-harvesting stage of cacao beans from 2016 to the present. Analyzing 36 studies, it focuses on classification, quality assessment, and fermentation. The proposed framework includes the application domain, learning algorithms, performance metrics, and reported impacts. Notably, it explores various machine learning applications like classification, quality assessment, and fermentation, highlighting commonly used algorithms like ANN, CNN, and SVM. In terms of performance metrics, GLCM achieved the highest accuracy (99.61%) in cacao classification, ANFIS excelled in quality assessment (99.715%), and k-NN emerged as the most accurate for fermentation. This review serves as a valuable resource for researchers in the cacao bean sector, offering insights into machine learning advancements.

Keywords: Cacao beans; Classification; Fermentation; Machine learning; Quality assessment; Post-harvesting stage

1. Introduction

Cacao (*Theobroma cacao*) is one of the most globally cultivated and renowned food crops. Commonly known as the main ingredient for chocolate, cacao yields various well-known products such as tableya, cocoa powder, cocoa liquor, cocoa mass, and other confectionery products. As time passed, cacao also found its place in the cosmetics and pharmaceutical industries, continuously having increasing demands in domestic and global markets.

Consistent cacao quality is crucial for production and farmer revenue. It is determined by aroma and flavor, which can vary based on post-harvest management practices. Sorting, fermentation, drying, classifying, and storing are important stages. Parameters like fermentation index, maturity level, and variety are used to evaluate and classify cacao beans.

Traditional cacao fermentation, using heap/box methods with banana leaves, and maturity assessment through visual cues and sound inspections are widely adopted but face limitations. The subjective nature of interpretation, reliance on manual labor, and environmental disturbances pose challenges. These methods, while valuable, may result in variations and compromised precision due to subjective judgments and external factors.

Acknowledging challenges in traditional post-harvest methods emphasizes the importance of training and exploring modern technologies for a sustainable cacao industry. Machine Learning (ML) in agriculture, a subset of Artificial Intelligence (AI), offers innovative solutions to overcome traditional farming limitations. Cacao production has seen a rise in ML applications, prompting a need for a comprehensive literature review on trends and developments in ML for cacao post-harvest management.

* Corresponding author: Luisa Ann Tamayo Ongquit

The paper offers a comprehensive examination of machine learning applications in cacao post-harvest cultivation, focusing on fermentation, quality assessment, and fruit classification. It explores research findings, trends, and the performance of various ML techniques and algorithms. This paper mainly aims to (1) provide a practical assessment of the current trends and development in the utilization of machine learning in cacao post-harvest management, (2) to investigate and evaluate the effectiveness of wide range of machine learning algorithms in this specific domain, and (3) to determine the most effective and efficient machine learning model algorithm for each cacao's post-harvest management field: fruit classification, quality assessment, and fermentation.

2. Methodology

This review systematically assesses the utilization of machine learning in cacao's post-harvest management. To ensure the validity of the findings, researchers explored various academic online databases and journals relevant to the chosen academe, such as IEEE Explore, Elsevier, PubMed, ResearchGate, Google Scholar, and other agricultural and technological journals. The use of related keywords such as "cacao post-harvest", "machine learning", "cacao bean quality assessment", and "cacao fermentation", together with operators, were capitalized to refine the search findings. Every search finding was thoroughly reviewed according to the inclusion and exclusion criteria to ensure that only necessary and relevant papers would be evaluated. The researchers selected peer-reviewed articles, conference papers, and reviews published between 2016 and the present, focusing on applying machine learning (ML) in cacao post-harvest management. The chosen studies were categorized into three post-harvest fields in cacao production: fruit classification, quality assessment, and fermentation. Upon having the final list of included papers, manual and systematic examination was conducted to extract the necessary data and information such as the title, author, methodology, utilized ML algorithms, and performance metric results. This review addresses thirty-six papers deemed pertinent and sourced from various databases.

3. Results

3.1. Individual Summary of selected publications by application area

This section provides a synopsis of the 36 selected literature under the 3 ML-application fields in cacao post-harvest management production. Table 1 summarizes the list of applications and the corresponding studies for each.

3.1.1. Classification of cacao pods

Accurate selection of cocoa beans is crucial for product quality [1]. However, manual classification during harvest can lead to pod contamination, contributing to a decline in cacao production [2]. Additionally, cross-pollination poses challenges in identifying cocoa varieties, impacting quality control and pricing [3]. The colour of pods serves as a ripening indicator linked to cocoa bean quality [4]. Adopting advanced technology can address these challenges and enhance the efficiency of cocoa production.

[1] proposed a CNN-based model for classifying healthy and diseased cacao pods and plants. The study used a labelled dataset of 312 images consisting of healthy, Monilia-damaged, and Phytophthora-damaged cacao. The YOLOv5m model was employed to classify healthy beans from unhealthy ones using phone images.

[2] employed an Artificial Neural Network (ANN) and Support Vector Machine (SVM) to distinguish healthy and unhealthy cacao beans without harvesting. The study utilized colour histograms and local binary pattern features, with the ANN classifier proving more accurate than SVM. The combination of CH and LBP, image processing techniques, and a large dataset enhanced the efficiency and reliability of bean segregation.

[5] propose a technique to determine the maturity of cacao pods through the sound they produce when tapped in different locations. Sounds were transformed into Mel spectrograms and classified as ripe or unripe using a trained CNN model. The CNN model outperformed other machine learning algorithms like SVM, KNN, and Decision trees in the study.

[6] focused on the UF-18 cacao variety, renowned for its prolonged lifespan and distinctive orange-yellow colour when ripe. They classified bean maturity based on pod physical characteristics such as colour, shape, and texture using Matlab and algorithms like OTSU, K-Means, and SVM, with SVM demonstrating excellent reliability. The study also incorporated sound analysis, tapping cacao beans four times to determine their maturity.

[7] present a method of detecting and segmenting cacao pod infection through visualization. The SVM model segregated healthy and infected fruits using four clusters and a quadratic kernel, while the K-means model clustered pixels based

on their color in the Lab color space. This method can improve the inspection process for farmers and agricultural technicians.

[3] developed a non-invasive cacao pod classifier using a Residual Network (ResNet) architecture based on Convolutional Neural Network (CNN). The study aimed to identify specific cacao varieties using pod images—Criollo, W-10, BR-25, and unknown. The ResNet model, trained using ImageAI, outperformed other CNN models, including VGG16, InceptionV3, and DenseNet.

[8] investigated cacao pod maturity classification using CocoaMFDB and various tools under different lighting conditions. Preprocessing techniques and color space analysis were applied, and feature extraction involved GLCM and MobileNet. The chi-square distance achieved the highest accuracy, emphasizing its impact on maturity determination.

[4] revealed ANN's usefulness in identifying the cacao pods' maturity. The model utilized image processing, and the algorithm was composed of RGB images of criollo cocoa pods with an amelonada shape, which could change its colour from green to yellow when the fruit was ripe. The process consisted of four steps: colour enhancement, segmentation, feature extraction and ANN classification. The suggested model yields much lower accuracy compared to similar previous studies.

[9] employ deep learning classifiers to swiftly identify healthy cacao pods and those affected by pests, enabling prompt responses to prevent losses. The study categorized cacao pod images into three conditions: healthy, pest attack, and black pod disease attack. Model training utilized k-fold cross-validation, and the proposed CNN model achieved the highest accuracy on the test set compared to MobileNetV3Small and NASNetMobile.

Another study by [10] utilizes AI and deep learning algorithms, specially the CNN model, to address the spread of infection to cacao pods. Severity of pods infection is classified into three categories using transfer learning from a pre-trained MobileNetV3Small architecture. The model's performance was compared to EfficientNetB0 and NASNetMobile, with the proposed CNN model outperforming the others.

3.1.2. Quality Assessment

The four main characteristics of cocoa beans are size, shape, colour, and texture. Cacao quality evaluation traditionally relies on manual inspection of those four characteristics. This approach is subjective, labor-intensive, and prone to human error. It is not suitable for large quantities of beans [11]. The constraints of manual inspection are being overcome by the increasing use of machine vision and automation technology, supported by artificial intelligence, to assess the quality of agricultural and food commodities.

[12] proposed MELS-SVM, an improved SVM model using the ensemble method. The model uses the least square approach to obtain hyperplanes and the One-Against-All method for multiclass procedures. The model extracts morphological parameters from cocoa bean images.

[13] developed an artificial neural network (ANN) to classify cocoa beans into good, fair, and bad based on attributes like slate, mould, free fatty acid, foreign matter, and bean count. The Multilayer Perceptron (MLP) model uses backpropagation for training and WEKA as a tool for implementing the algorithm on the cocoa bean data set, ensuring accurate classification of beans.

[14] propose a classifier system for the quality assessment of cocoa beans using image processing and a Backpropagation Neural Network Algorithm. It utilizes PCA for feature extraction and backpropagation for classification based on texture and color. The system aims to help farmers and traders reduce losses from manual selection.

[15] employed a supervised machine-learning method to classify cocoa beans into four categories: slate, purple, violet, and good. Supervised machine learning is a type of artificial intelligence that learns from labelled data and makes predictions based on the learned patterns.

[16] developed a method to classify cocoa beans into seven classes using GLCM, CNN feature extraction methods, and various classifiers. GLCM performed better and more reliable than CNN for classification. SVM and XGBoost classifiers were applied to the extracted features, showing higher accuracy than AdaBoost. GLCM features outperformed CNN features with both classifiers.

[17] implemented an automated method for cacao bean grading classification using image processing and an adaptive neuro-fuzzy inference system (ANFIS). The method involves defect identification and classification using CNN techniques, followed by grade level determination using ANFIS techniques that combine artificial neural networks and fuzzy logic systems.

[18] proposed a new method for classifying cocoa beans' pulp using computer vision techniques like k-means clustering, morphological operations, a bag of visual words, and a support vector machine classifier. This method aids in removing the pulp and efficiently estimates the quality of cocoa beans, dividing them into excellent and poor-quality classes.

[19] employ a decision tree-supervised machine learning approach to classify cocoa beans into four quality categories: good, slatty, mouldy, and weevilly. They capture digital images of cut cocoa beans, process and segment them, extract colour, texture, and statistical features, and store them in a feature database. They use the Classification and Regression Tree (CART) algorithm for classification.

[11] developed four machine-learning algorithms to categorize cocoa beans into four groups: large, medium, small, and rejected. Four machine learning algorithms (KNN, SVM, Decision Tree, Random Forest) were developed to categorize cocoa beans based on size. Image processing techniques were used to extract features, and redundant features were eliminated.

[20] suggested using CNN and PCA techniques to reduce the dimensionality and complexity of image data. CNN generated feature maps, while PCA transformed variables into uncorrelated components. SVM was employed as the classifier for cocoa bean categorization, with different kernels and parameters tested.

[21] proposed a cacao bean classifier employing a computer vision system that categorizes cacao beans based on their external physical quality: good, broken, clumped, flat, or mouldy. They use image processing and pre-trained CNN to extract and analyse the features from the datasets. The method consists of image acquisition, preprocessing, feature extraction, and classification.

[22] developed a computer vision system to classify cocoa beans into healthy and diseased ones using a dataset from Ecuador. They used YOLOv5 for image identification and preprocessing, focusing on the beans using interval analysis and mathematical morphology techniques.

[23] proposes two methods for implementing Machine Learning techniques in intelligent farming. The first assesses agricultural product quality using a feature classification framework, comparing methods like GLCM. The second method uses temporal data classification to monitor farmers' activities, employing a quaternion representation and HTM algorithm to predict and detect accidents or abnormal situations.

[24] developed a machine-learning model using enhanced image feature extraction and a regularized ANN model to classify cocoa beans into 14 grades based on extracted features. They used cut testing to visually inspect interior characteristics, such as colour, compartmentalization, and defects. They combined colour histograms and grey-level co-occurrence matrix (GLCM) features to represent the beans' colour and texture information.

[25] implemented convolutional neural networks (CNNs) to automate the cut-test process and predict cocoa bean quality scores based on colour and surface characteristics. The ResNet V2 152-layer model and data augmentation techniques were employed to enhance the CNN models' performance.

3.1.3. Fermentation

Cacao bean's fermentation is a crucial stage of its post-harvest management as it influences the development of the flavour [26], aroma, and processing properties [27] influencing its quality and price [28]. The standard method for classifying cacao beans' fermentation degree relies on manual estimation systems and demands specific expertise, which is prone to the inconsistency of prediction [29]. Also, some analytical equipment and methods were expensive for farmer groups. With this, the utilization of technology for swift and precise measurement of fermentation level motivated numerous scientists to develop a predictive model [28]

[30] used hyperspectral images and spectral classification techniques to study cocoa bean fermentation stages, identifying three categories: slightly, correctly, and highly fermented. A Support Vector Machine (SVM) accurately classified beans based on spectral signatures extracted using a superpixel segmentation algorithm. The study also analyzed bean uniformity and the impact of training sample numbers on classification performance.

[29] conducted an initial investigation into a computer vision and artificial intelligence-based system for measuring the fermentation level of cocoa beans. Using image analysis and color features, an Artificial Neural Network Multilayer Perceptron achieved high accuracy in classifying fermented and unfermented beans. This method provides a rapid and precise alternative to the traditional cut test for assessing cocoa bean quality.

[27] employed an electronic nose system with nine gas sensors to identify the fermentation level of cocoa beans. Six machine learning methods were employed to process sensor responses and classify fermentation time. The bootstrap forest algorithm achieved the highest accuracy with the lowest misclassification rate. Half of the algorithms successfully segregated cocoa beans based on fermentation level, while the other half were unsuccessful.

[28] proposed utilizing technology to measure the cacao bean fermentation rate precisely, focusing on forecasting pH and the fermentation index. Color features and a PLSR model demonstrated high accuracy in predicting both pH and the fermentation index, indicating their effectiveness as visual indicators for assessing cocoa bean fermentation.

[31] utilized machine vision and multiclass SVM classifiers for evaluating cocoa bean fermentation levels in quality control. Colour features were extracted from RGB, HSV, and YCbCr colour spaces to characterize fermentation. The proposed method achieved higher accuracy than the conventional cut-test method, demonstrating improved efficiency and reliable classification of cocoa beans based on fermentation level.

[32] used the k-Nearest Neighbors algorithm to classify cacao beans into three fermentation levels based on photograph colour features. Accuracy was assessed through function tests and statistical analysis, including a confusion matrix. The research highlights the potential of devices and machine learning to enhance cacao quality and value, reducing risks for farmers and traders. Further exploration is recommended, such as incorporating additional features and datasets, as well as exploring alternative classification algorithms for device improvement.

[33] used various computer vision methods to gauge the fermentation level of the cocoa beans. Four methods, including vector quantification, K-means, Fuzzy grouping, and average displacement, were compared and evaluated for training and classification. The dataset consisted of images taken during the fermentation process, and a reference pattern was developed using inputs from sensory profiles and cocoa testers.

[34] explores categorizing the degree of roasting of cocoa beans using smartphone camera images and an extreme learning machine (ELM) algorithm. Color features from RGB and CIELAB color spaces, as well as texture features from the grey-level co-occurrence matrix (GLCM), are used to generate a 22-dimensional feature vector for each region of interest (ROI) in the images.

[35] propose a novel method using unsupervised (K-means clustering) and supervised (SVM) machine learning to classify cocoa beans as well-fermented or over-fermented based on spectral signatures. This approach streamlines quality assessment, improves efficiency, and reduces costs without sample destruction.

Table 1 Overview of ML applications in cacao's various post-harvest fields

	Application	Topic	Number of Studies (n=36)	Sources / References
Post-Harvest	Classification	Healthiness	5	[1,2,7,9,10],
		Cacao's maturity and ripeness	4	[4-6,8]
		Classification of cacao variety	1	[3]
	Quality Assessment	Morphological features	15	[11-25]
	Fermentation	Fermentation level	8	[26,27,29-33,35]
		Fermentation index	1	[28]
		Cacao Bean's Roasting Degree	2	[34,36]

[36] demonstrated continuous monitoring of volatile compounds during cocoa refining using a gas sensor system. Trained Kernel Distribution Models (KDMs) characterized volatile profiles, identifying deviations in processing variables. An electronic nose (E-nose) tracked compounds validated by comparison with gas chromatography-mass spectrometry (GC-MS).

[26] achieved high classification rates for cocoa bean quality using a calorimetric sensor e-nose and near-infrared chemo-intermediary-dyes spectra technique. SVM and ELM algorithms yielded the highest classification rates, demonstrating the system's potential for nondestructive and in situ classification of cocoa bean fermentation levels.

3.2. Machine Learning Algorithms

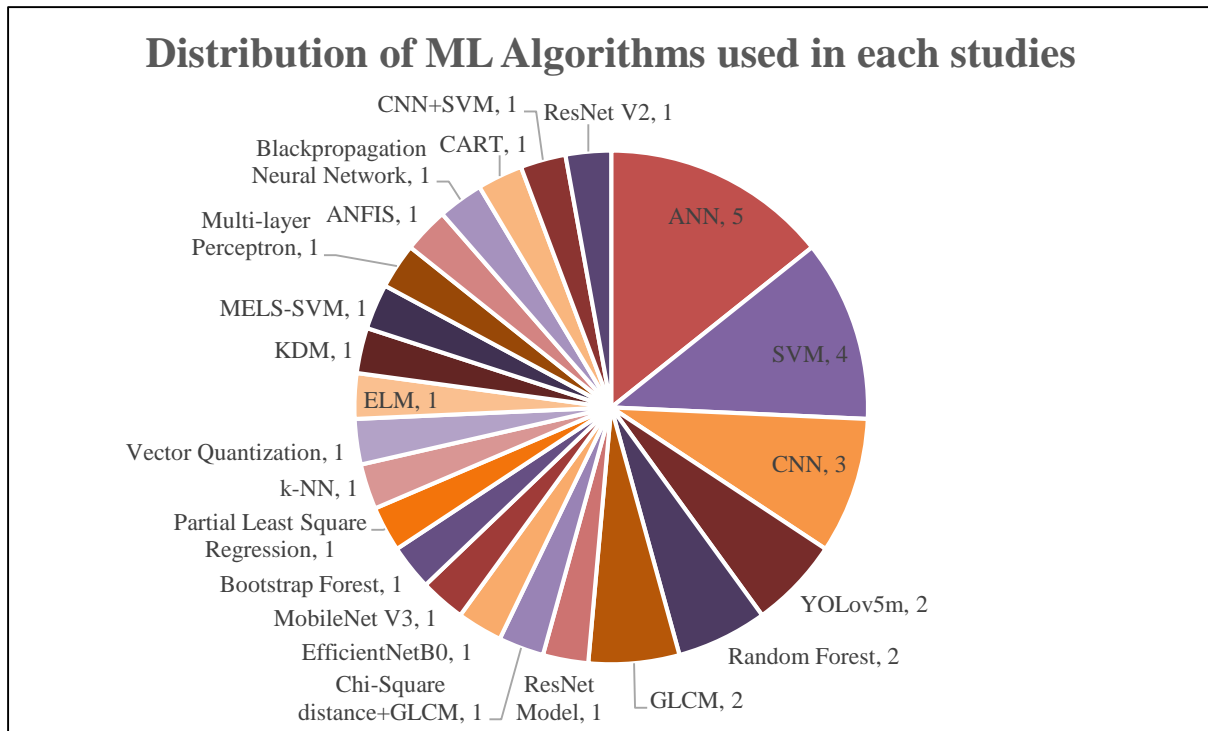


Figure 1 Distribution of ML Algorithms used in each study

Figure 1 illustrates the distribution of Machine Learning algorithms across the 36 analyzed studies. A total of 36 different algorithms were utilized, with several employed in only one study, showcasing the diverse approaches taken. These include CART, ResNet V2, Backpropagation Neural Network, ANFIS, Multi-Layer Perceptron, MELS-SVM, KDM, ELM, Vector Quantization, YOLOv2, Bootstrap Forest, MobileNet V3, EfficientNetB0, GLCM, Chi-Square distance+GLCM, and ResNet Model.

Contrastingly, some algorithms were more commonly used. Support Vector Machines (SVM) were applied in four studies, making it prominent, while Artificial Neural Network (ANN) featured in five studies, demonstrating versatility. Convolutional Neural Network (CNN) was utilized in three studies, mainly for image-related tasks. Random Forest appeared in two studies, and K-Nearest Neighbors (k-NN) in one study.

The accompanying pie chart visually depicts the algorithm distribution, emphasizing both the variety of approaches and the higher prevalence of certain algorithms like SVM, ANN, CNN, Random Forest, and k-NN. This suggests that while researchers explore a diverse set of ML algorithms, certain methodologies have gained popularity and are more frequently adopted.

Table 2 Top 3 Most-used ML algorithms

	Algorithm	Tasks	No. of studies
1	ANN	Classification	5
2	SVM	Classification, image processing	4
3	CNN	Classification	3

Table 2 presents the top three most frequently utilized Machine Learning algorithms among the remaining 33. These algorithms, widely applied across various domains, include Artificial Neural Network (ANN) for classification, Support Vector Machines (SVM) for classification and image processing, and Convolutional Neural Network (CNN) for classification.

Artificial Neural Network (ANN) stands out as the most commonly used algorithm, appearing in five studies. Its popularity is attributed to its capacity to model complex relationships and its versatility in handling diverse data types, making it effective for classification tasks.

Support Vector Machines (SVM) is the second most frequently employed algorithm, featured in four studies. Its ability to handle both linear and non-linear classification tasks, coupled with effectiveness in managing high-dimensional data, positions it as a preferred choice for classification and image processing tasks.

Convolutional Neural Network (CNN) is the third most commonly used algorithm, appearing in three studies. Known for its unique architecture designed for grid-like data like images, CNN has become the go-to choice for image classification tasks. Its capability to capture local spatial dependencies within images has contributed to its success in various computer vision applications.

The inclusion of these three algorithms in Table 2 underscores their prominence and widespread adoption in the research community. ANN, SVM, and CNN have consistently demonstrated their efficacy and versatility, making them go-to solutions for addressing diverse machine learning challenges.

3.3. Performance Metrics

Table 3 List of Machine Learning Algorithms utilized in Cacao Classification Post-Harvest Classification

	Application	ML Algorithms	Performance Metrics
Cacao Classification	Healthiness	YOLOv5m	Accuracy: 98%
		ANN	Accuracy: 98%
		SVM	Accuracy: 89%
		EfficientNetB0	Accuracy: 94%
		MobileNetV3	Accuracy: 91%
	Cacao's maturity and ripeness	CNN	Accuracy: 97%
		SVM	Accuracy: 98.33%
		Chi-square distance-GLCM	Accuracy: 99.61%
		ANN	Accuracy: 91%
	Cacao variety	ResNet Model	Accuracy: 95%

Table 4 List of Machine Learning Algorithms utilized in Cacao Quality Assessment Post-Harvest Classification

	Application	ML Algorithms	Performance Metrics
Quality Assessment	Morphological features	GLCM	Accuracy: 65.08%
		CNN	Accuracy: 90.67%
		CNN-SVM	Accuracy: 98.32%
		ANN	Accuracy: 85.36%

	Muticlass Ensemble Least-Squares Vector Machine (MELS-SVM)	Accuracy: 99.705%
	YOLOv5	Accuracy: 88.5%
	SVM	Accuracy: 97.57%
	Multi-Layer Perceptron (MLP)	Accuracy: 92.9714%
	ANFIS	Accuracy: 99.715%
	Backpropagation Neural Network (ANN)	Accuracy: 70%
	Classification and Regression Tree (CART)	Accuracy: 89.2%
	Random Forest	Accuracy: 75%
	ResNet V2	Accuracy: 92%

Table 5 List of Machine Learning Algorithms utilized in Cacao Fermentation Management

	Application	ML Algorithms	Performance Metrics
Fermentation	Fermentation level	SVM	Accuracy: 98.43% - 100%
		ANN	Accuracy: 94%
		Bootstrap Forest	Accuracy: 90.6%
		k-NN	Accuracy: 92.50%
		Random Forest	Accuracy: 97.7% (unbalanced dataset) Accuracy: 99.4% (balanced dataset)
	Fermentation index	Partial Least Square Regression	Accuracy: 99%
		Vector Quantization	Level of certainty: 75%
	Cacao Bean's Roasting Degree	ELM	Accuracy: 94.75%
		KDM	Error: ~1%

In the studies, machine learning algorithms are evaluated based on metrics like accuracy, error rates, and levels of certainty to assess classifier performance in doing specified tasks.

Table 3 outlines the learning algorithms employed for classifying cacao's healthiness, maturity, ripeness, and variety. Object detection via YOLOv5m and machine learning methods including ANN, SVM, CNNs (EfficientNetB0, MobileNetV3), and others like SVM, GLCM, and ResNet were used. Notably, the Chi-square distance metric, applied with GLCM as a feature extractor, achieved the highest accuracy in cacao classification. A study focusing on cacao pod maturity classification demonstrated that employing automation tools based on similarity measures, specifically utilizing the Chi-square distance metric with GLCM, significantly enhanced efficiency and accuracy in determining cacao pod maturity.

Moving to Table 4, which presents the performance of various machine learning algorithms for quality assessment of cacao beans based on morphological features, it can be observed that the ANFIS model achieved the highest accuracy of 99.715%, followed by the MCE model with 99.705%. The MCE model is a combination of multiple machine learning algorithms, which can improve the accuracy and robustness of the model. ANFIS, being a powerful machine learning algorithm that combines the capabilities of neural networks and fuzzy logic, likely demonstrated superior performance in classifying morphological features of cacao pods by effectively capturing the intricate patterns and relationships present in the data.

Examining the data in Table 5, the k-NN algorithm emerges as the most accurate for the fermentation application, achieving a remarkable accuracy score of 99.4% on a balanced dataset. This algorithm is a simple and effective instance-based learning algorithm that classifies new instances based on their similarity to the training data. The balanced

dataset contributed to the high accuracy of k-NN, where the number of instances in each class is approximately equal. This can help to avoid bias towards the majority class and improve the performance of the algorithm.

Using a variety of machine learning algorithms and techniques is significant for addressing the unique requirements of different cacao processing stages. The success of these models has the potential to enhance efficiency and accuracy in cacao cultivation and processing, benefiting the chocolate industry. However, the choice of algorithm should consider specific problem and data characteristics, along with other factors.

4. Conclusion

This study offers a comprehensive overview of the application of Machine Learning techniques in the post-harvest stage of cacao beans. The use of machine learning has greatly facilitated the identification and classification of cacao beans, providing efficient and accurate solutions. This review emphasizes the performance metrics associated with each of the machine learning algorithms identified in the collected studies.

Among 36 machine learning algorithms used in cacao pod classification, quality assessment, and fermentation, Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Convolutional Neural Networks (CNN) are the most commonly employed. The Gray Level Co-occurrence Matrix (GLCM) algorithm achieved the highest accuracy in cacao pod classification, while Adaptive Neuro-Fuzzy Inference System (ANFIS) achieved 99.715% accuracy in quality assessment. For fermentation, the k-Nearest Neighbors (k-NN) algorithm demonstrated 99.4% accuracy.

This comprehensive review provides insights into recent Machine Learning applications in the post-harvest stage of cacao beans. It highlights widely used algorithms, performance metrics, and the need for further advancements in the field. A valuable resource for researchers in the cacao bean industry seeking efficient and accurate solutions.

Compliance with ethical standards

Acknowledgments

The researchers extend their gratitude to all contributors who made this literature review possible, with special appreciation for Dr. Edwin R. Arboleda for his invaluable guidance. Sincere thanks are also extended to the authors of the examined literature for sharing their findings in this specific field.

Disclosure of conflict of interest

The authors declare no conflict of interest.

References

- [1] Ferraris S, Meo R, Pinaridi S, Salis M, Sartor G. Machine Learning as a Strategic Tool for Helping Cocoa Farmers in Côte D'Ivoire. *Sensors*. 2023 Sep 3;23(17):7632.
- [2] Baculio NG, Barbosa JB. An Objective Classification Approach of Cacao Pods using Local Binary Pattern Features and Artificial Neural Network Architecture (ANN). *IJST*. 2022 Mar 21;15(11):495–504.
- [3] Razon JJA, Alimuin RA, Arenas SU. Cacao Pod Variety Classification Using Res Net Modelled Convolutional Neural Network. *Journal of Green Engineering*. 2021;11(3):2132–46.
- [4] Veites-Campos SA, Ramírez-Betancour R, González-Pérez M. Identification of Cocoa Pods with Image Processing and Artificial Neural Networks. *IJAEMS*. 2018;4(7):510–8.
- [5] Bueno GE, Valenzuela KA, Arboleda ER. Maturity classification of cacao through spectrogram and convolutional neural network. *Jurnal Teknologi dan Sistem Komputer*. 2020 Jul 31;8(3):228–33.
- [6] Carbajosa PD, Lugay DC, Tan RKC. Maturity Classification of Cacao by Image Processing and Acoustic Sensing.
- [7] Tan DS, Leong RN, Laguna AF, Ngo CA, Amalin D, Alvindia D. A Method for Detecting and Segmenting Infected Part of Cacao Pods. 2016;4.
- [8] Ayikpa KJ, Mamadou D, Gouton P, Adou KJ. Classification of Cocoa Pod Maturity Using Similarity Tools on an Image Database: Comparison of Feature Extractors and Color Spaces. *Data*. 2023 May 30;8(6):99.

- [9] Godmalin RA, Aliac CJ, Feliscuzo L. Classification of Cacao Pod if Healthy or Attack by Pest or Black Pod Disease Using Deep Learning Algorithm. In: 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAET) [Internet]. Kota Kinabalu, Malaysia: IEEE; 2022 [cited 2024 Jan 12]. p. 1–5. Available from: <https://ieeexplore.ieee.org/document/9936817/>
- [10] Godmalin RA, Aliac CJ, Feliscuzo L. Cacao Pod Infection Level Classification Using Transfer Learning. In: 2023 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream) [Internet]. Vilnius, Lithuania: IEEE; 2023 [cited 2024 Jan 12]. p. 1–6. Available from: <https://ieeexplore.ieee.org/document/10135062/>
- [11] Akuli A, Das S, Bag AK, Parua S, Ghosh A. Morphological Image Processing for Cocoa Bean Classification. YMER. 2022 Jun 15;21(06):413–26.
- [12] Lawi A, Adhitya Y. Classifying Physical Morphology of Cocoa Beans Digital Images using Multiclass Ensemble Least-Squares Support Vector Machine. J Phys: Conf Ser. 2018 Mar;979:012029.
- [13] Folake A. A SCIENTIFIC APPROACH FOR CLASSIFICATION OF COCOA BEAN. EPRA International Journal of Multidisciplinary Research (IJMR). 2019;5(2).
- [14] Lestari U, Kumalasanti RA, Wulandari ES. Identifying the Quality System of Cocoa Beans to Increase Productivity Using Backpropagation Neural Network Algorithm: A Case Study at Sumber Rejeki Farmers Group, Patuk Gunung Kidul. J Phys: Conf Ser. 2019 Nov 1;1413(1):012033.
- [15] Antunes F, Canuto A, Bedregal B, Palmeira E, Silva I, Santos F. Supervised Methods Applied to the Construction of a Vision System for the Classification of Cocoa Beans in the Cut-Test. In: Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2019) [Internet]. Brasil: Sociedade Brasileira de Computação - SBC; 2019 [cited 2024 Jan 27]. p. 72–83. Available from: <https://sol.sbc.org.br/index.php/eniac/article/view/9273>
- [16] Adhitya Y, Prakosa SW, Köppen M, Leu JS. Feature Extraction for Cocoa Bean Digital Image Classification Prediction for Smart Farming Application. Agronomy. 2020 Oct 25;10(11):1642.
- [17] Obediencia DC, Villafuerte R, Brosas DG. Cacao Bean Quality Assessment Procedure: A Method for Classification Process. 2018;
- [18] Oña Oña AJ, Grijalva F, Proano K, Acuna B, Garcia M. Classification of Fresh Cocoa Beans with Pulp Based on Computer Vision. In: 2020 IEEE ANDESCON [Internet]. Quito, Ecuador: IEEE; 2020 [cited 2024 Jan 27]. p. 1–6. Available from: <https://ieeexplore.ieee.org/document/9272188/>
- [19] Tanimowo AO, Eludiora SI, Ajayi AO. A DECISION TREE CLASSIFICATION MODEL FOR COCOA BEANS QUALITY EVALUATION BASED ON DIGITAL IMAGING. 2020;
- [20] Ayikpa KJ, Mamadou D, Ballo AB, Gouton P, Adou KJ. Application based on Hybrid CNN-SVM and PCA-SVM Approaches for Classification of Cocoa Beans. IJACSA [Internet]. 2022 [cited 2024 Jan 27];13(9). Available from: <http://thesai.org/Publications/ViewPaper?Volume=13&Issue=9&Code=IJACSA&SerialNo=27>
- [21] Hortinela CC, Tupas KJR. Classification of Cacao Beans Based on their External Physical Features Using Convolutional Neural Network. In: 2022 IEEE Region 10 Symposium (TENSymp). IEEE; 2022.
- [22] Zhinin Vera LF, Zhiminaicela-Cabrera J, Suarez P, Pretel E, Chang O, Anton Castro F. Artificial Vision Techniques to Detect and Classify Cacao Beans. SSRN Journal [Internet]. 2022 [cited 2024 Jan 27]; Available from: <https://www.ssrn.com/abstract=4232560>
- [23] Adhitya Y. Machine Learning Based Classification Models in Smart Farming. 2023;
- [24] Eric O, Gyening RMOM, Appiah O, Takyi K, Appiahene P. Cocoa beans classification using enhanced image feature extraction techniques and a regularized Artificial Neural Network model. Engineering Applications of Artificial Intelligence. 2023 Oct;125:106736.
- [25] Sangsawad S, Jintawatsakoon S. Quality Assessment of Cocoa Bean Using Regression Convolutional Neural Network. In: 2023 27th International Computer Science and Engineering Conference (ICSEC) [Internet]. Samui Island, Thailand: IEEE; 2023 [cited 2024 Jan 27]. p. 1–6. Available from: <https://ieeexplore.ieee.org/document/10329764/>
- [26] Kutsanedzie FYH, Chen Q, Sun H, Cheng W. In situ cocoa beans quality grading by near-infrared-chemodyes systems. Anal Methods. 2017;9(37):5455–63.

- [27] Tan J, Balasubramanian B, Sukha D, Ramkissoon S, Umaharan P. Sensing fermentation degree of cocoa (*Theobroma cacao* L.) beans by machine learning classification models based electronic nose system. *J Food Process Engineering*. 2019 Oct;42(6):e13175.
- [28] Riza DFA, Putranto AW, Iqbal Z, Firmanto H, Anggraini CD. Prediction of Fermentation Index and pH of Cocoa (*Theobroma cacao* L.) Beans Based on Color Features (Cut Test) and Partial Least Square Regression Model. *fst*. 2023 Mar;11(1):54–62.
- [29] Anggraini CD, Putranto AW, Iqbal Z, Firmanto H, Al Riza DF. Preliminary study on development of cocoa beans fermentation level measurement based on computer vision and artificial intelligence. *IOP Conf Ser: Earth Environ Sci*. 2021 Nov 1;924(1):012019.
- [30] Sánchez K, Bacca J, Arévalo-Sánchez L, Arguello H, Castillo S. Classification of Cocoa Beans Based on their Level of Fermentation using Spectral Information. *TecnoL*. 2021 Jan 30;24(50):e1654.
- [31] Yro A, N'zi CE, Kpalma K. Cocoa Beans Fermentation Degree Assessment For Quality Control Using Machine Vision and Multiclass SVM Classifier. 2018;
- [32] Angelia RE, Linsangan NB. Fermentation Level Classification of Cross Cut Cacao Beans Using k-NN Algorithm. In: *In Proceedings of the 5th International Conference on Bioinformatics Research and Applications (ICBRA '18)*. 2018. p. 64–8.
- [33] Parra P, Negrete T, Llaguno J, Nino. Computer Vision Methods in the Process of Fermentation of the Cocoa Bean. In 2018.
- [34] Yang Y, Darwish AG, El-Sharkawy I, Zhu Q, Sun S, Tan J. Rapid determination of the roasting degree of cocoa beans by extreme learning machine (ELM)-based imaging analysis. *Journal of Agriculture and Food Research*. 2022 Dec;10:100437.
- [35] Gomez NA, Sanchez K, Arguello H. Non-Destructive Method for Classification of Cocoa Beans from Spectral Information. In: *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)* [Internet]. Bucaramanga, Colombia: IEEE; 2019 [cited 2024 Jan 12]. p. 1–5. Available from: <https://ieeexplore.ieee.org/document/8730257/>
- [36] Tan J, Kerr WL. Characterizing cocoa refining by electronic nose using a Kernel distribution model. *Food Science and Technology*. 2019 May;104:1–7.