Received May 31th, 2026; accepted June 12th, 2026. Date of publication June 30th, 2026
Digital Object Identifier: https://doi.org/10/25047/jtit.v13i1.500

LSTM Approached for Cassava Tapai Ripeness Identification

SHABRINA CHOIRUNNISA[1], MUHAMMAD IZZA ALFIANSYAH[1], KHAFIDURROHMAN AGUSTIANO[1], RIFDA HANIFAH AZZAHRA[2]

[1]Jurusan Teknologi Informasi Politeknik Negeri Jember, Jember, Indonesia
[2] Pendidikan Bahasa Inggris, Universitas Jember, Jember, Indonesia

CORESPONDING AUTHOR: Shabrina Choirunnisa (email:shabrinacnisa@polije.ac.id)

ABSTRACT Tapai singkong (cassava tapai) is a traditional Indonesian fermented food product whose quality is highly dependent on precise control of the fermentation process. Inconsistent fermentation outcomes arise from fluctuating environmental conditions including temperature, humidity, and fermentation gas levels making it difficult to reliably determine ripeness status without objective measurement tools. This study addresses the challenge of automated ripeness prediction by providing a controlled, head-to-head comparison of four machine learning approaches Logistic Regression, Support Vector Machine (SVM), Random Forest, and LSTM-based Recurrent Neural Network (RNN) on a single, uniformly preprocessed dataset of 600 time-series observations across three ripeness classes (unripe, ripe, overripe), collected from 10 fermentation trials spanning 60 hours each. All models were evaluated under identical preprocessing and hyperparameter settings using accuracy, precision, recall, F1-score, and confusion matrices to reveal per-class behavior. LSTM yielded the best test performance (96.46% accuracy; macro F1 = 0.93), Random Forest followed closely (93.70% accuracy; macro F1 = 0.94), while SVM and Logistic Regression obtained 91.28% and 90.31% accuracy respectively. This paper discusses the trade-off between predictive performance, temporal modeling capability, and interpretability, and recommends LSTM for high-accuracy quality control deployments where temporal dependencies are critical, and Random Forest as a strong, interpretable alternative for resource-constrained environments. Per-class metrics and experimental artifacts are provided to support reproducibility and practical adoption in traditional food production monitoring.

KEYWORDS: Tapai Singkong Fermentation, Ripeness Classification, Machine Learning, LSTM Time-Series, Fermentation Quality Control

I. INTRODUCTION

Cassava is one of the leading local food commodities in Indonesia, but has a relatively short shelf life after harvest [1]. To increase its economic value and extend its shelf life, cassava is processed by fermentation to become tapai [2]. Tapai cassava (tapai singkong) is a traditional food that has long been embedded in the culinary heritage of the country and is made through a fermentation process with the help of yeast [3]. However, its production still relies heavily on the subjective judgment of artisans when determining ripeness. Unlike others, tapai requires fermentation using yeast containing the Kapang Amylomyces Rousi, Mucor sp, Rhizopus sp, Khamir Saccharomycopsis fibuligera, Candida Utilis, Pichia burtonii, Saccharomyces Cerevisiae, Saccharomycopsis Malanga, and the bacteria Pediococcus sp and Bacillus sp [4]. Cassava is one of the staple foods used as a source of carbohydrates after rice and corn [5]. Indonesians people rely on cassava tape not only as a source of income but also as a daily food ingredient [6]. This necessitates the production of high-quality cassava tapai to meet the community's food needs [6].

The problem that arises is that many members of the general public do not have knowledge about the level of maturity of cassava tapai fermentation. The fermentation process can last for ± 72 hours in semi-aerobic conditions [7]. The process of biochemical changes occurs due to the activity of microorganisms during the fermentation process [8]. But, the quality of tapai is significantly influenced by environmental conditions during fermentation including temperature, humidity, and fermentation gas levels all of which are inherently dynamic and difficult to control consistently [9]. The absence of a standardized measurement method presents a critical barrier to modernizing tapai production and achieving consistent quality at a larger scale.

Technological developments have brought major transformations in various aspects of life, including data management in the fields of artificial intelligence (Artificial Intelligence) and machine learning (Machine Learning) to innovations in data processing and decision making [10]. Increasingly tight business competition in the era of globalization requires companies, including the food industry, to be able to meet market needs quickly and precisely [11]. So far, traditional producers generally detect tapai maturity levels only based on instinct and a quick visual assessment, which often triggers subjective views and different quality standards between producers. To overcome this subjectivity, the majority of modern research has turned to using Convolutional Neural Networks (CNN) to extract spatial features of objects such as color and texture from static images. However, this CNN model is only able to analyze momentary conditions (present state) and ignores the dynamic nature of the fermentation process which is very time-dependent. This study is here to fill this gap by applying the Long Short-Term Memory (LSTM) approach which is able to model sequential maturity progression based on temporal data series. Through the transition from CNN-style static analysis to LSTM-based development analysis, this research succeeded in creating a monitoring system that is objective, accurate, and adaptive to the evolution of tapai quality over time.

This study aims to address the aforementioned challenge through an automated ripeness prediction approach based on machine learning, leveraging time-series data collected from environmental sensors throughout the fermentation process. Specifically, the study conducts a systematic and controlled comparison of four machine learning algorithms Logistic Regression, Support Vector Machine (SVM), Random Forest, and LSTM-based Recurrent Neural Network using a unified dataset of 600 time-series observations drawn from 10 fermentation trials spanning 60 hours each. All models are evaluated under identical preprocessing pipelines and hyperparameter configurations to ensure a fair and reproducible comparison. Beyond identifying the best-performing model, this study also examines the trade-off between predictive accuracy, temporal modeling capability, and interpretability, with the goal of providing practical recommendations for quality control system deployment in traditional food production environments with varying resource constraints.

The hope of this research is that it can overcome the limitations of conventional methods in the tapai industry. From an economic perspective, this data-based standardization can play a vital role in reducing the risk of financial loss due to commodities that are damaged before they can be marketed. From a quality assurance perspective, this automation can consistently determine the ideal alcohol content, acidity and texture for each production batch. So practically, the transition to a digital system offers significant operational efficiency by reducing dependence on skilled labor and minimizing human error.

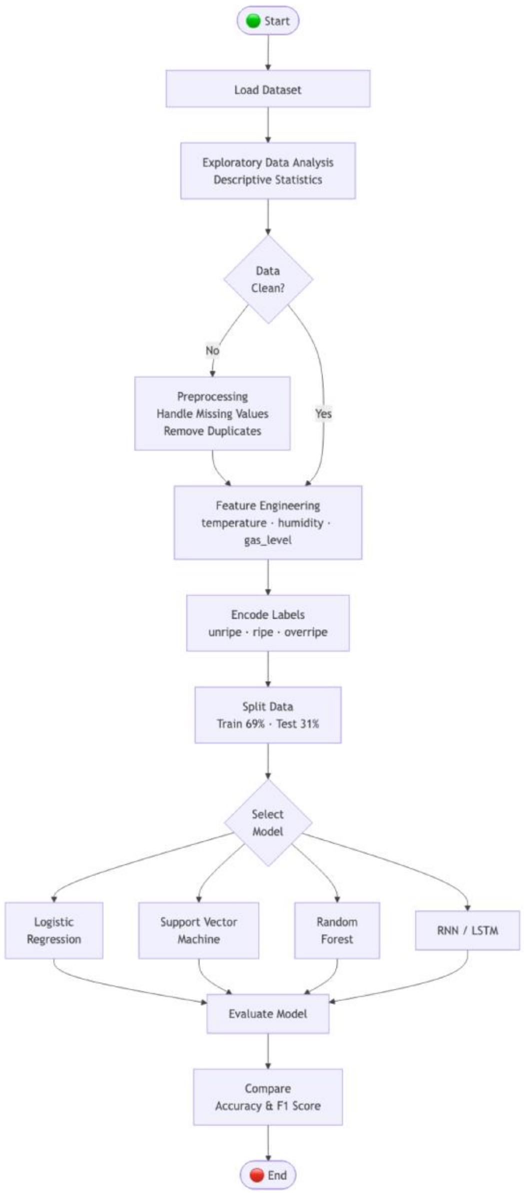
II. METHOD

The flow of this research is carried out as illustrated in the flowchart below. There are 4 models that will be used in this research, namely Logistic Regression, Support Vector Machine (SVM), Random Forest, and LSTM (Long Short-Term Memory). These four models were selected because they represent different architectural characteristics in terms of linearity, kernel-based transformation, ensemble learning, and temporal sequence modeling. At the same time, the selection allows a comprehensive comparison between classical machine learning approaches and deep learning. After sharing the dataset, all models will be subjected to the same preprocessing and trained with various parameters. Then all models will be tested and evaluated, so that the model with the best accuracy can be determined.

To ensure a fair comparison, all models will use the same dataset and go through identical pre-processing stages, including data cleaning, missing value handling, feature normalization or standardization, and dividing the data into training and testing data. Next, each model will be developed by adjusting hyperparameters to obtain the optimal configuration according to the characteristics of each algorithm. This process aims to minimize bias that can arise due to differences in data treatment during the training stage.

Model performance will be evaluated using several relevant measurement metrics, such as accuracy, precision, recall, F1-score, and confusion matrix. The evaluation results are then compared to analyze the ability of each model to carry out classification and identify its strengths and limitations. Through this approach, the research not only determines the best performing models, but also provides a deeper understanding of the effectiveness of various machine learning and deep learning approaches in solving the problem under study. The complete flow of the research stages carried out can be seen in Figure 1.

FIGURE 1. Flowchart Process



A. Dataset

This research utilized a dataset from monitoring the cassava tapai fermentation process which included 600 observations. The data was collected from 10 independent fermentation cycles, where each cycle was monitored continuously for 60 hours with data collection intervals once every hour. The data analyzed are presented in Table 1.

TABLE 1. Dataset Structure

Cycle	**Hours**	**Tempe rature**	**Humi dity**	**Gas Levels**	**Status**
-	-	-	-	-	unripe
-	-	-	-	-	ripe
-	-	-	-	-	overrip e

The analyzed data shows a distribution pattern based on maturity status (such as unripe, ripe and overripe) which is visualized in the following image to facilitate trend interpretation.

FIGURE 2. Raw Data Visualization

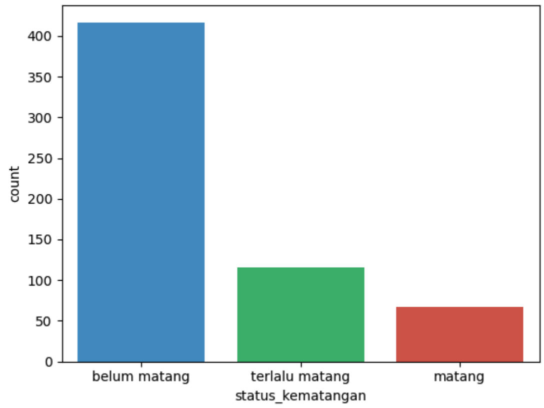


Figure 2 shows the distribution of maturity status which consists of three categories, namely not yet mature, too mature, and mature. The immature category has the largest number, namely around 415 data or more than two-thirds of the total data observed. Meanwhile, the overcooked category has around 115 data, and the overcooked category is the smallest with around 67 data.

B. Preprocessing

One of the most important stages in machine learning development is dataset preprocessing [12]. Data preprocessing is an essential part of developing a classification model [12]. Before training, the dataset must be preprocessed first to ensure compatibility with the models used. It is important to note that the cleaned data produced by this preprocessing pipeline is used exclusively for classical machine learning models (Logistic Regression, SVM, and Random Forest). The RNN model, on the other hand, uses the original raw dataset, as it requires data with temporal sequence interpretation specifically, the time-series ordering of sensor readings across each fermentation trial.

The target class distribution of raw dataset is heavily imbalanced, as shown in Table 2.

TABLE 2. Dataset Classes Distributions

Class	Count
Unripe	417
Ripe	67
Overripe	116

Imbalanced data is one of the crucial problems in machine learning and data mining which may provide low accuracy in minority classes and makes the classification method not fully optimized [13]. Due to the significant imbalance in class distribution, this research will apply Synthetic Minority Over-Sampling Technique (SMOTE) to overcome class imbalance [14]. SMOTE generates synthetic data points for the minority classes by interpolating between existing samples in the feature space, rather than simply duplicating existing records. This balancing step aims to ensure that the developed classical models are able to learn features from all classes equally and improve generalization on previously unrecognized data patterns.

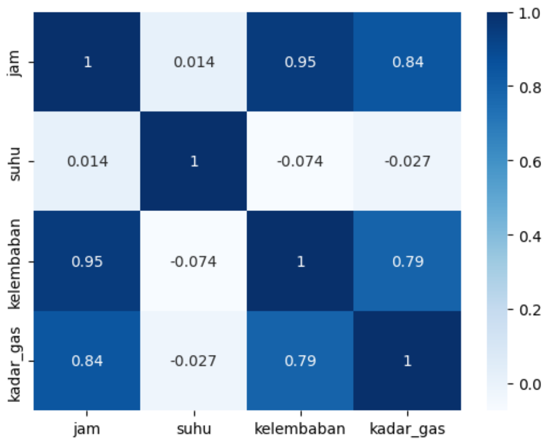
After applying SMOTE, each class contains 417 records, resulting in a balanced dataset of 1,251 records in total, as shown in Table 3 below.

TABLE 3. Dataset Comparison Before and After Preprocessing

Class	Before	After
Unripe	417	417
Ripe	67	417
Overripe	116	417

In the dataset used, not all features are relevant for applying conventional classical models, so feature selection is done selectively to optimize model performance. Therefore, the classical model applied only utilizes a few main columns, namely jam (hours), kelembaban (humidity), and kadar_gas (gas levels). The correlation between these features can be observed in Figure 3, which shows the linear relationship and potential multicollinearity between variables.

FIGURE 3. Dataset Features Correlation



The dataset is then divided into two parts, namely training data and testing data. As much as 70% of the total data was used as training data to build the model, while the remaining 30% was used as testing data to evaluate the model's ability to make predictions on data that was not involved during the training process. Data distribution is carried out randomly with a consistent scheme

C. Classification Process

a. Logistic Regression

The first classification approach employs Logistic Regression, a linear discriminant model that estimates class membership probabilities through a softmax activation function applied to a linear combination of input features. Logistic Regression is a suitable choice for developing a prediction system based on numerical data [15]. Prior to model fitting, the input features are standardized to zero mean and unit variance, and the target labels are numerically encoded to represent each ripeness category. The dataset is partitioned into training and testing subsets using a stratified random split, retaining 31% of observations for evaluation. Each hourly observation is treated as an independent sample; no temporal dependencies between consecutive time steps are modeled. The model learns a separate weight vector for each class, and the decision boundary is defined as a linear hyperplane in the three-dimensional feature space.

TABLE 4. Result of Logistic Regression Training

Class	Precision	Recall	F1-Score	Support
Unripe	0.96	0.92	0.94	136
Ripe	0.82	0.88	0.85	130
Overripe	0.93	0.90	0.92	147
Accuracy			**0.90**	**413**
Macro avg	0.90	0.90	0.90	413
Weighted avg	0.91	0.90	0.90	413

From the Table 4, the model achieves an overall accuracy of 0.9031 and a macro-averaged F1-score of 0.90. The relatively lower precision for the **ripe** class (0.82) suggests that the linear decision boundary struggles to cleanly separate this intermediate class from its adjacent categories.

b. Support Vector Machine

The second approach applies a Support Vector Machine (SVM) classifier using the Radial Basis Function (RBF) kernel. One of the supervised learning algorithms that is widely used in image classification is Support Vector Machine [16]. Unlike Logistic Regression, the RBF kernel implicitly maps the input features into a higher-dimensional space, enabling the model to learn non-linear decision boundaries. The SVM optimization objective seeks the hyperplane that maximizes the margin between the nearest support vectors of each class pair. The same data preparation procedure used in the section before feature standardization and label encoding is applied prior to training. Each hourly observation continues to be treated as an independent data point, with no sequential context incorporated.

TABLE 5. Result of Support Vector Machine Training

Class	Precision	Recall	F1-Score	Support
Unripe	0.95	0.95	0.95	136
Ripe	0.85	0.88	0.86	130
Overripe	0.94	0.90	0.92	147
Accuracy			**0.91**	**413**
Macro avg	0.91	0.91	0.91	413
Weighted avg	0.91	0.91	0.91	413

As shown in Table 5, the SVM achieves an accuracy of 0.9128 and a macro-averaged F1-score of 0.91, representing a modest improvement over Logistic Regression. The RBF kernel's capacity to capture non-linear relationships in the feature space contributes to better discrimination of the *ripe* class, as evidenced by the increase in precision from 0.82 to 0.85 relative to the linear baseline.

c. Random Forest

The third method employs a Random Forest classifier, an ensemble learning algorithm that constructs a large number of decision trees in parallel through bootstrap aggregating and random feature subsampling. Random Forest can be an effective, objective, and non-destructive method for determining maturity levels [17]. In this implementation, 200 individual decision trees are trained on randomly drawn subsets of the training data, each considering a random subset of features at each split node. The final class prediction for each sample is determined by majority vote across all trees. Unlike the previous models, Random Forest does not require explicit feature scaling, as decision tree splits are invariant to monotonic transformations of the input features. The same stratified train-test split is retained, and the target labels are numerically encoded prior to training.

TABLE 6. Result of Random Forest Classification Training

Class	Precision	Recall	F1-Score	Support
Unripe	0.98	0.95	0.96	136
Ripe	0.87	0.94	0.90	130
Overripe	0.96	0.93	0.94	147
Accuracy			**0.94**	**413**
Macro avg	0.94	0.94	0.94	413
Weighted avg	0.94	0.94	0.94	413

Table 6 represent the Random Forest model achieves the highest accuracy among the three non-temporal models at 0.9370, with a macro-averaged F1-score of 0.94. The ensemble mechanism substantially reduces variance compared to a single decision tree, and the model's inherent non-linearity allows it to capture interaction effects between humidity, gas concentration, and observation hour. Additionally, feature importance scores are available as a by-product of training, providing interpretable insight into predictor contributions.

d. Long Short-Term Memory (LSTM) Recurrent Neural Network

The fourth approach treats the fermentation data as a genuine time series and applies a Bidirectional Long Short-Term Memory (LSTM) network, a specialized variant of Recurrent Neural Networks (RNN). RNN are neural network architectures with hidden state and which use feedback loops to process a sequence of data that ultimately informs the final output [18]. Rather than treating each hourly observation as an independent sample, the dataset is restructured into overlapping sliding windows of 10 consecutive time steps, so that each training sample represents a contiguous sequence of 10 hours of sensor readings across three measured variables. The Bidirectional LSTM architecture processes each sequence in both forward and backward temporal directions, allowing the model to exploit contextual information from both past and future observations within each window. The network is trained by minimizing a multi-class classification loss function through an iterative gradient-based optimization procedure, and a stratified split is applied at the trial level to prevent data leakage across fermentation experiments.

The network architecture consists of the following sequential layers:

- **First LSTM Layer (64 units)**: Processes the input sequence and learns temporal patterns across the 3-hour observation window, passing its intermediate outputs forward to the next layer.
- **First Dropout Layer (rate = 0.2)**: Randomly deactivates 20% of neurons during training to reduce overfitting and improve generalization.
- **Second LSTM Layer (32 units)**: Further distills the temporal representations learned by the first layer into a compact feature vector summarizing the entire sequence.
- **Second Dropout Layer (rate = 0.2)**: Applies the same regularization mechanism as the first dropout layer to the condensed representations.
- **Output Layer (3 units)**: A fully connected layer that maps the final learned representation to probability scores for each of the three ripeness classes, using a softmax function to ensure the probabilities sum to one.

The architecture was configured to provide sufficient representational capacity while maintaining computational efficiency and minimizing overfitting. These hyperparameter values were determined based on empirical deep-learning practices and preliminary experimentation to achieve a balance between classification performance, model stability, and computational cost.

The model training process is carried out with an optimal hyperparameter configuration, namely 50 epochs with a batch size of 32 samples per iteration, which allows stable and efficient convergence to the complexity of the dataset. This setting was chosen to balance model accuracy and computing time, where each epoch represents one full iteration through the entire training data. The complete training process, including loss and accuracy curves per epoch, can be seen in Figure 4 and Figure 5.

FIGURE 4. LSTM Training Loss Progression

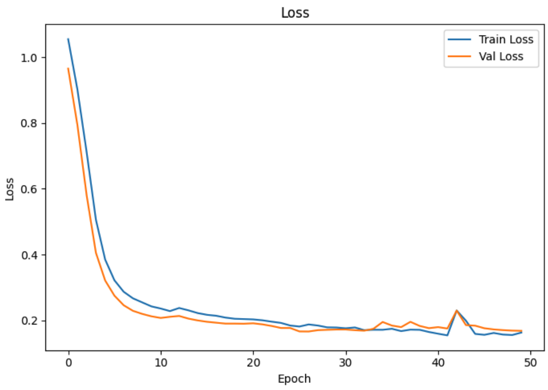
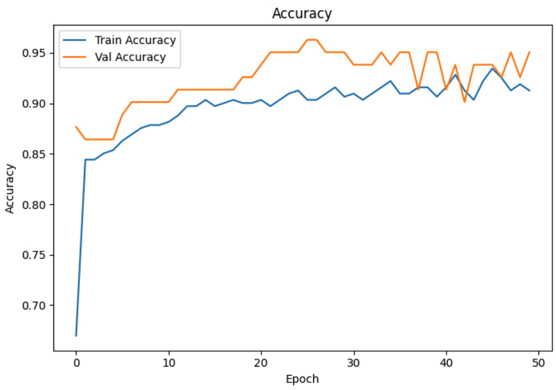


FIGURE 5. LSTM Training Accuracy Progression



The LSTM model achieves the highest overall accuracy of 0.9646 across all four methods. The near-perfect precision of 1.00 for the *unripe* class reflects the model's ability to reliably identify the early fermentation phase by leveraging the cumulative temporal pattern of rising gas concentration and humidity. The macro-averaged F1-score of 0.92 — computed on a smaller test set with class imbalance due to the sliding window construction procedure — reflects the model's strong generalization to unseen fermentation trials. The explicit modeling of temporal dependencies renders this approach the most scientifically appropriate for the fermentation domain, where the ripeness status at any given hour is inherently conditioned on the physiochemical trajectory of preceding hours. The detailed statistical can be found in Table 7.

TABLE 7. Result of Long Short-Term Memory Classification Training

Class	Precision	Recall	F1-Score	Support
Unripe	1.00	0.99	0.99	138
Ripe	0.83	0.86	0.84	22
Overripe	0.92	0.95	0.94	38
Accuracy			**0.96**	**198**
Macro avg	0.92	0.93	0.92	198
Weighted avg	0.97	0.96	0.97	198

D. Evaluation

The evaluation framework is grounded in standard information retrieval and binary classification theory, extended to the multi-class setting through macro and weighted averaging strategies. Standardization of evaluation for news recommendation systems remains minimal, despite the importance of these systems in addressing information overload in the digital era [19]. For each model, four primary metrics are computed: Precision, Recall, F1-Score, and Accuracy. These metrics are derived from the confusion matrix, which summarizes the agreement between predicted and true class labels on the held-out test set.

The foundation of all classification metrics used in this study is the **confusion matrix**, a square matrix of dimension $K \, x \, K$, where K denotes the number of target classes. In this study, K=3, corresponding to the three ripeness categories: *unripe* (belum matang), *ripe* (matang), and *overripe* (terlalu matang). Each entry C_{ij} of the confusion matrix represents the number of observations belonging to true class ii that the model predicted as class j.

For a given class, four fundamental counts are derived from the confusion matrix:

- **True Positives (TP)**: The number of instances correctly predicted as class identified.
- **False Positives (FP)**: The number of instances from other classes incorrectly predicted as class identified.
- **False Negatives (FN)**: The number of instances of class incorrectly predicted as another class identified.
- **True Negatives (TN)**: The number of instances correctly identified as not belonging to class identified.

These are the metrics used for evaluating the model trained:

$$Precission = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1\ Score = 2\ \frac{Precission\ .\ Recall}{Precision + Recall} \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (4)$$

III. RESULT AND DISCUSSION

The comparison of all four classification models based on the macro-averaged Precision, Recall, F1-Score, and overall Accuracy computed are represented in Table 8.

TABLE 8. Models Result Comparisson

Model	Precission	Recall	F1	Accuracy

Logistic Regression	0.90	0.90	0.90	0.9031
Support Vector Machine	0.91	0.91	0.91	0.9128
Random Forest	0.94	0.94	0.94	0.9370
LSTM	0.92	0.93	0.92	0.9646

Among the four models, the LSTM achieves the highest overall Accuracy (0.9646), reflecting the advantage of explicitly modeling temporal dependencies in the fermentation time series. Random Forest achieves the highest macro-averaged F1-Score (0.94) among the non-temporal models, demonstrating that ensemble-based non-linear classifiers are substantially more effective than linear approaches for this classification task. The SVM improves upon Logistic Regression across all metrics by leveraging the RBF kernel to capture non-linear decision boundaries. Logistic Regression, as the simplest baseline, still achieves competitive performance with an Accuracy of 0.9031, confirming that the three sensor features carry strong discriminative information even when temporal context is disregarded.

The lower macro-averaged F1-Score of the LSTM (0.92) relative to its high Accuracy (0.9646) reflects the influence of class imbalance in the LSTM test set — specifically, the disproportionately small support for the *ripe* class (22 instances) after sliding-window construction and trial-level partitioning. The macro average penalizes this underperformance equally regardless of class frequency, whereas the weighted-average F1-Score of 0.97 is more representative of the LSTM's practical utility across the observed class distribution.

The proposed Long Short-Term Memory Model (LSTM) shows a stronger capability in capturing the temporal dynamics of the fermentation process. In contrast to classical models, which generally handle observations independently, and CNN-based approaches, which are primarily designed to extract spatial patterns, LSTM architectures are specifically designed to learn sequential dependencies from time series data. By exploiting fermentation-related features, including temperature, humidity, fermentation duration (hours), and gas concentration, the proposed method can effectively model fermentation progression over time and provide a more accurate maturity classification. These results demonstrate that the LSTM-based approach is competitive and offers a meaningful contribution to the field, especially for applications involving continuous monitoring and intelligent assessment of fermented maturity of cassava tapai.

IV. CONCLUSION

This study addressed the objective of comparing four machine learning classification models Logistic Regression, Support Vector Machine (SVM), Random Forest, and Long Short-Term Memory (LSTM) for classifying the ripeness of tapai singkong into three categories (belum matang, matang, and terlalu matang) based on fermentation time-series sensor data comprising temperature, humidity, and gas concentration, evaluated using macro-averaged Precision, Recall, F1-Score, and Accuracy. The experimental results confirm that all four models achieved competitive classification performance, with Logistic Regression attaining an Accuracy of 90.31%, SVM 91.28% via an RBF kernel that captured non-linear decision boundaries, and Random Forest achieving the highest macro-averaged F1-Score of 0.94 at 93.70% accuracy among non-temporal models, demonstrating the effectiveness of ensemble-based learning for this task. The LSTM model, which explicitly models temporal dependencies through sliding-window sequence construction, achieved the highest overall Accuracy of 96.46%, validating the hypothesis that temporal modeling enhances classification performance for sequential fermentation sensor data; its comparatively lower macro-averaged F1-Score of 0.92 relative to Random Forest is attributable to class imbalance in the LSTM test partition arising from trial-level partitioning, as reflected by its weighted-average F1-Score of 0.97. These findings affirm that LSTM-based architectures are best suited for high-accuracy tapai fermentation monitoring where temporal dynamics are critical, while Random Forest remains a strong and interpretable alternative for resource-constrained deployments. Future research may extend this work by incorporating attention-based or Transformer architectures to improve sequence modeling, applying data augmentation and oversampling strategies to address class imbalance, expanding the dataset across more fermentation trials and environmental conditions, and exploring real-time embedded deployment of the proposed models for practical quality control in traditional food production.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the faculty members and academic advisors for their intellectual guidance in machine learning, time-series analysis, and traditional food technology; to the laboratory and technical staff who facilitated the systematic collection of temperature, humidity, and fermentation gas sensor data across ten fermentation trials; to the open-source communities behind scikit-learn, TensorFlow, Keras, NumPy, and pandas, whose freely available tools enabled the implementation and evaluation of the Logistic Regression, Support Vector Machine, Random Forest, and LSTM-based models presented herein; and to the reviewers and editors for their constructive evaluations. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCE

[1] I. Mualimah, E. Brata Tyarini, D. Ayu Aryati, and U. Palangka Raya, “Profil Karakteristik Organoleptik Tapai Singkong (Manihot esculenta): Analisis Rasa, Tekstur dan Aroma pada Berbagai Variasi Konsentrasi Ragi,” *JURNAL BIOSHELL*, vol. 15, no. 1, pp. 157–165, Mar. 2026, doi: 10.56013/BIO.V15I1.5502.

[2] C. Cpz *et al.*, “Pengoptimalan Pengolahan Singkong Menjadi Produk Pangan Dalam Meningkatkan Pendapatan Masyarakat Desa,” *Welfare : Jurnal Pengabdian Masyarakat*, vol. 1, no. 3, pp. 449–454, Sep. 2023, doi: 10.30762/WELFARE.V1I3.642.

[3] H. Syahrumsyah, H. Al Hafidz, M. Jurusan Teknologi Hasil Pertanian, F. Pertanian, and U. Mulawarman Jl Pasir Balengkong Kampus Gunung Kelua, “Karakteristik organoleptik dan kimia tape singkong (Manihot esculenta) varietas mentega dengan pra-perlakuan perendaman dalam sari buah nangka (Artocarpus heterophyllus),” *Journal of Tropical AgriFood*, vol. 2, no. 2, pp. 90–96, Jul. 2021, doi: 10.35941/JTAF.2.2.2020.4121.90-96.

[4] A. Subayu and Supatman, “Deteksi Tingkat Kematangan Fermentasi Singkong (Tape Singkong) Menggunakan Convolutional Neural Network (CNN),” *Journal Of Information System And Artificial Intelligence*, vol. 2, no. 2, pp. 136–141, May 2022, doi: 10.26486/JISAI.V2I2.68.

[5] S. A. Pangestu, “Pengaruh Penambahan Madu Terhadap Hedonik, Hardness, Total Padatan Terlarut, dan Nilai pH Tape Singkong,” *Jurnal Teknologi Pangan*, vol. 8, no. 1, pp. 22–26, Jun. 2024, doi: 10.14710/JTP.2024.38492.

[6] S. Laili, S. Sudarti, and T. Prihandono, “Analisis Perubahan Suhu Pada Proses Fermentasi Tape Singkong (Manihot esculenta),” *Wahana Fisika*, vol. 7, no. 1, pp. 19–30, Jun. 2022, doi: 10.17509/wafi.v7i1.40198.

[7] D. S. Rejeki, E. N. Alfiraza, and A. A. Kusuma, “PENENTUAN KADAR ALKOHOL DARI TAPAI UBI UNGU (Ipomoea batatas (L.) Lam) DAN TAPAI SINGKONG (Manihot esculenta Crantz) DENGAN VARIASI WAKTU FERMENTASI,” *Bhamada: Jurnal Ilmu dan Teknologi Kesehatan (E-Journal)*, vol. 16, no. 1, pp. 22–30, Apr. 2025, doi: 10.36308/JIK.V16I1.775.

[8] S. Susilawati *et al.*, “ANALISIS KUALITATIF DAN KUANTITATIF PADA TAPAI SINGKONG DAN BERAS KETAN,” *Journal of Food Security and Agroindustry*, vol. 2, no. 1, pp. 17–26, Feb. 2024, doi: 10.58184/JFSA.V2I1.281.

[9] E. Sijabat, N. L. Gaol, A. Wahyuningtyas, and I. Marvie, “Pengaruh Konsentrasi Ragi, Waktu Pemasakan dan Penyimpanan Terhadap Perubahan Fisikokimia dan Sensori Tapai Talas Kimpul (Xanthosoma sagittifolium),” *Jurnal Teknologi Pangan*, vol. 18, no. 2, pp. 75–87, 2024, doi: 10.33005/JTP.V18I2.4766.

[10] G. F. Wijaya and D. Yuniarto, “Tinjauan Penerapan Machine Learning pada Sistem Rekomendasi Menggunakan Model Klasifikasi,” *Populer: Jurnal Penelitian Mahasiswa*, vol. 3, no. 4, pp. 144–153, Dec. 2024, doi: 10.58192/POPULER.V3I4.2798.

[11] Y. A. Saputra and D. Mahdiana, “Analisis Penerapan Machine Learning, Deep Learning, dan Data Mining dalam Prediksi Penjualan di Industri Otomotif,” *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 6, pp. 1743–1755, Jun. 2025, doi: 10.52436/1.JPTI.826.

[12] S. S. Berutu, H. Budiati, J. Jatmika, and F. Gulo, “Data preprocessing approach for machine learning-based sentiment classification,” *JURNAL INFOTEL*, vol. 15, no. 4, pp. 317–325, Nov. 2023, doi: 10.20895/INFOTEL.V15I4.1030.

[13] F. R. Lumbanraja *et al.*, “Implementasi SMOTE dan Support Vector Machine Pada Klasifikasi Data Tidak Seimbang Metilasi Arginin,” *Jurnal Pepadun*, vol. 5, no. 1, pp. 27–37, Apr. 2024, doi: 10.23960/PEPADUN.V5I1.209.

[14] A. A. Dhani, T. A. Y. Siswa, and W. J. Pranoto, “Perbaikan Akurasi Random Forest Dengan ANOVA Dan SMOTE Pada Klasifikasi Data Stunting,” *Teknika*, vol. 13, no. 2, pp. 264–272, Jul. 2024, doi: 10.34148/TEKNIKA.V13I2.875.

[15] N. Aulia, I. Iqbal, and H. Asrawi, “Analisis Prediktif Tingkat Kematangan Alpukat Menggunakan Algoritma Logistic Regression,” *Jurnal Ilmu Komputer Aceh*, vol. 3, no. 1, pp. 152–163, Mar. 2026, doi: 10.51179/ILKA.V3I1.46.

[16] B. Laoli, I. Muslem, and H. Asrawi, “Klasifikasi Tingkat Kematangan Buah Pisang Menggunakan Algoritma Support Vector Machine,” *Jurnal Ilmu Komputer Aceh*, vol. 3, no. 1, pp. 142–151, Mar. 2026, doi: 10.51179/ILKA.V3I1.45.

[17] A. Penelitian *et al.*, “Klasifikasi Kematangan Buah Nanas (Ananas comosus) Berdasarkan Data Hidung Elektronik Multisensor Menggunakan Algoritma Random Forest,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 12, no. 1, pp. 35–44, Apr. 2026, doi: 10.25077/TEKNOSI.V12I1.2026.35-44.

[18] S. Das, A. Tariq, T. Santos, S. S. Kantareddy, and I. Banerjee, “Recurrent Neural Networks (RNNs): Architectures, Training Tricks, and Introduction to Influential Research,” *Neuromethods*, vol. 197, pp. 117–138, 2023, doi: 10.1007/978-1-0716-3195-9_4/FIGURES/9.

[19] I. K. Phan and Y. Yuricha, “Perancangan Skema Evaluasi untuk Sistem Rekomendasi Berita Menggunakan Metrik Precision, Recall, dan F1-Score,” *TIN: Terapan Informatika Nusantara*, vol. 6, no. 8, pp. 1351–1360, Jan. 2026, doi: 10.47065/TIN.V6I8.9106.

MUHAMMAD IZZA ALFIANSYAH, Muhammad Izza Alfiansyah was born in Jember, East Java, in 2004. He is a D3 Information Technology graduate from Politeknik Negeri Jember with a GPA of 3.93/4.00. Currently, he works as a Full Stack Developer at ERA Indonesia, where he develops scalable web and mobile applications while integrating artificial intelligence solutions into enterprise systems. He can reached via email at iansyah724@gmail.com.