Journal of Applied Data Sciences
Vol. 6, No. 4, December 2025, pp. 2651-2665

A Hybrid LSTM–Stacking–SMOTE Model for Weather-Aware Palm Oil Price Prediction Addressing Data Imbalance and Forecast Accuracy

Kusmanto[1,*], S. Subagio[2], Erni Manja[3]

[1,2]*Faculty of Computer Science, Department of Informatics Engineering, Universitas Al Washliyah Labuhanbatu, Rantauprapat, Indonesia*

[3]*Faculty of Economics, Department of Management, Universitas Al Washliyah Labuhanbatu, Rantauprapat, Indonesia*

(Received: February 18, 2025; Revised: May 15, 2025; Accepted: August 14, 2025; Available online: September 26, 2025)

Abstract

Accurate forecasting of palm oil prices is crucial for agribusiness decision-making due to high market volatility influenced by dynamic weather conditions. This study proposes a novel hybrid deep learning model combining Long Short-Term Memory (LSTM), Stacking Ensemble, and Synthetic Minority Over-sampling Technique (SMOTE) to improve predictive accuracy and handle class imbalance in price trend classification. The model was trained using a multivariate time-series dataset sourced from Kaggle, consisting of daily records of temperature, humidity, rainfall, and palm oil prices. A binary classification scheme was applied by labeling instances as either price increase (class 1) or price stable/decrease (class 0), based on a 0% price change threshold. Four experimental configurations were evaluated: standard LSTM, LSTM + SMOTE, LSTM + Stacking, and the proposed LSTM + SMOTE + Stacking. The proposed model outperformed all baselines, achieving the highest accuracy of 83.12%, an F1-score of 0.8466, MAE of 0.1688, RMSE of 0.4109, and a perfect recall of 1.0000, indicating excellent sensitivity to minority class trends. In contrast, the standard LSTM achieved only 77.32% accuracy and an F1-score of 0.7224, showing limited ability in handling imbalanced data. Visualization of loss curves and confusion matrices confirmed the model's learning stability and classification effectiveness. This study contributes a novel integration of ensemble learning and oversampling in time-series commodity forecasting and demonstrates the effectiveness of this approach in capturing weather-driven price patterns, offering a robust framework for predictive analytics in agriculture.

Keywords: Palm Oil Price Prediction, LSTM, Stacking Ensemble, SMOTE, Deep Learning, Weather Data

1. Introduction

Palm oil is one of Indonesia's most strategic agricultural commodities, contributing significantly to national economic output [1], [2]. As the world's largest producer of Crude Palm Oil (CPO), Indonesia maintains a highly dynamic market in terms of production, distribution, and pricing. However, CPO prices are notoriously volatile and subject to a wide range of influencing factors, both internal and external [3], [4]. Among internal factors, weather conditions—such as rainfall, temperature, and humidity—have a direct impact on fresh fruit bunch (FFB) yields [5], [6], [7], which in turn affect supply levels and market prices [8], [9], [10]. These complexities underscore the need for advanced computational approaches, particularly AI-based algorithms, to achieve reliable price forecasting [11], [12]. Although Convolutional Neural Networks (CNNs) were originally developed for image analysis, they have also been applied to time-series prediction tasks by transforming sequential data into visual representations [13], [14]—using methods such as Gramian Angular Fields and Recurrence Plots [15], [16].

Additionally, 1D-CNNs can be applied directly to time-series data by extracting local features through convolutional filters [17], [18], [19]. CNNs offer notable advantages, including efficient feature extraction, faster training times, and robustness to noise due to their parameter sharing mechanisms [20], [21]. However, they lack internal memory, which limits their ability to model long-term dependencies—a critical requirement for price forecasting in time-series contexts [22], [23]. To overcome this limitation, CNNs are often used as preliminary feature extractors, followed by sequential models like LSTM or GRU for temporal learning [24], [25], [26]. Previous research has explored hybrid approaches combining econometric methods and machine learning. For instance, [3] demonstrated the strength of integrating the

***Corresponding author: Kusmanto (kusnabara03@gmail.com)**
DOI: https://doi.org/10.47738/jads.v6i4.922

Auto-Regressive Distributed Lag (ARDL) model with LSTM to capture both short- and long-term price dynamics of CPO. Their study also utilized LASSO and SHAP for variable selection and interpretability, producing high predictive accuracy using a 17-year dataset. Despite these strengths, the model relied heavily on macroeconomic variables and lacked generalizability, having been validated only on Malaysian data.

Additionally, the complexity of LSTM may hinder practical adoption in industry settings with limited technical capacity [3]. A related study [5] focused on daily weather prediction using a BiLSTM model enhanced with SMOTE and ADASYN to address class imbalance. The model employed five years of real synoptic data and used multi-category contingency tables for evaluation, enhancing both accuracy and applicability for flight safety. Despite the robust implementation, the study was geographically limited to a single airport, and the heavy use of oversampling raised concerns of overfitting. Furthermore, generalizing the model to other regions would require retraining due to differences in local meteorological conditions [5].

While LSTM has shown great promise in time-series forecasting, it still suffers from several limitations when used as a standalone model. These include susceptibility to overfitting, especially in the presence of limited or imbalanced datasets [27], [28], and sensitivity to hyperparameter tuning [29], [30]. To overcome these challenges, ensemble learning—particularly stacking—has emerged as a viable approach to improve model accuracy and stability [31], [32], [33]. In stacking, multiple base learners (e.g., individual LSTM models) are combined using a meta-learner, thereby increasing predictive performance through aggregated learning [34]. Another critical issue in real-world datasets is class imbalance, where the model is biased toward majority trends due to skewed data distributions [21], [35]. In the context of palm oil price prediction, this is particularly problematic as extreme price shifts are underrepresented in historical records. To address this, the SMOTE is used to generate synthetic minority samples, improving model balance and performance in regression or classification tasks involving imbalanced data [35], [36], [37]. Given this background, this study aims to develop an optimized palm oil price prediction model that integrates LSTM, stacking ensemble, and SMOTE techniques. By addressing both temporal dependency and data imbalance, the proposed model is expected to enhance predictive accuracy and generalization, thereby offering a valuable decision-support tool for stakeholders in the palm oil industry—including farmers, plantation companies, and policy makers [38], [39].

2. Literature Review

The forecasting of agricultural commodity prices, particularly for palm oil, has received considerable attention in both economic and technological domains due to the high price volatility that significantly impacts stakeholders ranging from smallholder farmers to large-scale industry actors [40], [41]. Palm oil prices are influenced by various factors, including global demand and supply, currency exchange rates, trade policies, and—most fundamentally—production conditions [3]. Among these, weather factors play a pivotal role in determining productivity. Variations in rainfall, temperature, humidity, and sunlight intensity can directly affect crop yields and thus market supply and price movement. Therefore, incorporating weather variables into forecasting models is essential to improve both the accuracy and contextual relevance of predictions [4].

In recent years, advances in artificial intelligence—especially machine learning and deep learning—have significantly improved the modeling of complex phenomena such as price forecasting [42], [43]. LSTM, a specialized form of Recurrent Neural Network (RNN), has emerged as a powerful architecture for sequential data tasks [44], [45]. Designed to overcome the limitations of traditional RNNs in capturing long-term dependencies, LSTM incorporates memory cells and gated mechanisms (input, forget, and output gates) that allow it to retain and propagate relevant information over time. This makes LSTM particularly well-suited for time-series tasks like commodity price forecasting. However, LSTM models are not without drawbacks. They tend to be computationally intensive, sensitive to overfitting, and require careful data handling and hyperparameter tuning [46], [47].

To overcome the limitations of single-model architectures, ensemble learning has gained prominence. One particularly effective approach is stacking, which combines multiple base learners through a meta-learner to improve prediction accuracy and model robustness. In the context of price forecasting, stacking allows the integration of various LSTM configurations or the combination of LSTM with other architectures such as GRU, CNN, or even non-neural models like XGBoost. The primary strength of stacking lies in its ability to reduce prediction error by leveraging the strengths of diverse models. However, this approach also introduces challenges, such as increased computational complexity and the need for larger datasets to adequately train multiple models [48], [49].

Beyond model architecture, data distribution remains a critical challenge. Historical palm oil price data often exhibits strong class imbalance, with the majority of values clustered within a narrow range, while extreme price changes occur infrequently. This imbalance can lead to biased models that favor dominant patterns and overlook rare but significant fluctuations. To mitigate this, the SMOTE is employed [35], [50]. SMOTE synthesizes new data points for the minority class using interpolation between nearest neighbors, thereby enriching the dataset and enabling better generalization of the model in regression contexts. As the relationship between weather and crop yields becomes better understood, numerous studies have integrated weather data into price prediction models. These studies consistently show that incorporating weather variables enhances forecast accuracy compared to models relying solely on historical price data. LSTM models have also demonstrated superior performance over traditional statistical approaches such as ARIMA in capturing nonlinear and temporal patterns [51], [52], [53]. However, the combined application of stacking and SMOTE in commodity price forecasting remains underexplored, presenting an opportunity to contribute new insights and techniques to the field. By integrating the temporal strength of LSTM, the ensemble advantages of stacking, and the balancing capabilities of SMOTE, this study aims to develop a more reliable and accurate forecasting model for palm oil prices—particularly under the influence of fluctuating weather conditions. The resulting model is expected to support more informed decision-making in the palm oil industry and contribute to the advancement of data-driven agribusiness forecasting solutions.

3. Methodology

To achieve the objective of developing a reliable and accurate palm oil price forecasting model based on weather-related variables, a systematic and structured methodology is essential. This section outlines the research framework in detail, encompassing data collection and preprocessing procedures, the architectural design of LSTM model, the application of stacking as an ensemble learning strategy, and the use of the SMOTE for handling data imbalance. In addition, the evaluation metrics and computational tools used in model development are described. These methodological steps are designed to ensure that the resulting model not only achieves high predictive performance but is also robust to data variability and representative of real-world conditions.

3.1. Dataset Description

This study employs a time-series dataset sourced from the Kaggle platform, selected for its accessibility, completeness, and well-structured format. The dataset consists of five variables: Date, Temperature (°C), Humidity (%), Precipitation (%), and Palm Oil Price (IDR). The date variable functions as the chronological index required for sequence-based models such as LSTM. The weather variables serve as input features, while the palm oil price is the target output to be forecasted. All data points were cleaned to remove missing values and ensure temporal alignment across features. The values were then normalized using MinMaxScaler to standardize input ranges, an essential step for optimizing convergence during deep learning model training. To see the summary statistics presentation of the main numeric variables used, please see table 1.

Table 1. Presents the Summary Statistics of the Key Numerical Variables Used

Variable	Mean	Std Dev	Min	25%	50%	75%	Max
Temperature (°C)	18.99	16.86	14.00	5.00	21.00	30.00	39.00
Humidity (%)	68.52	20.48	20.00	57.00	70.00	83.00	109.00
Precipitation (%)	53.94	32.20	0.00	20.00	58.00	83.00	109.00
Price (IDR)	1609.69	37.07	1557.21	1572.23	1617.98	1648.03	1648.03

As shown table 1, the temperature variable has a wide range from 14°C to 39°C, indicating potential outliers or measurement variability. Humidity and precipitation show higher values and broader dispersion, consistent with tropical climate data. Palm oil price data, on the other hand, appears relatively stable with a narrow range of variation, highlighting the importance of using advanced models to capture subtle but critical changes. To evaluate the relationships between the input variables and the target variable (palm oil price), a correlation heatmap was generated (figure 1). The results indicate that temperature and humidity are positively correlated, suggesting that increases in ambient temperature tend to coincide with higher humidity levels in the observed regions. Precipitation exhibits a moderate correlation with palm oil prices, implying that rainfall levels may influence crop yield, thereby affecting market prices. However, the overall correlation coefficients between individual weather variables and the price are relatively low. This underscores the necessity of employing non-linear predictive models such as LSTM, which are capable of capturing complex multivariate dependencies and temporal patterns that linear models may fail to detect.

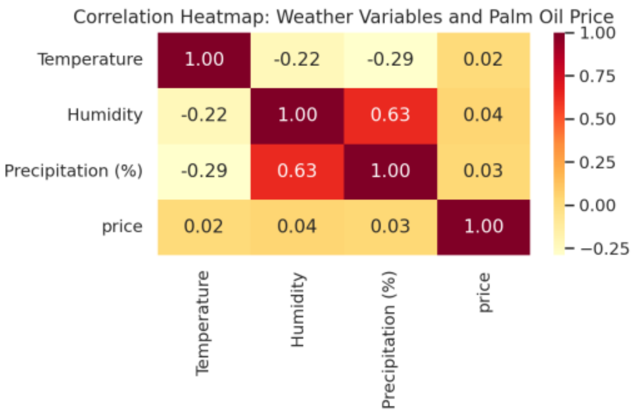


Figure 1. Correlation Heatmap: Weather Variables and Palm Oil Price

3.1. Research Design

This study adopts a quantitative research approach through computational experiments grounded in machine learning techniques. The primary objective is to develop and optimize a time-series-based forecasting model for palm oil prices using the LSTM architecture. To enhance predictive performance, the model integrates a stacking ensemble strategy and employs the SMOTE to address data imbalance issues. The research process is structured into several stages: data collection and preprocessing, model development, training and evaluation, and result analysis. A detailed overview of the research workflow is presented in figure 2.

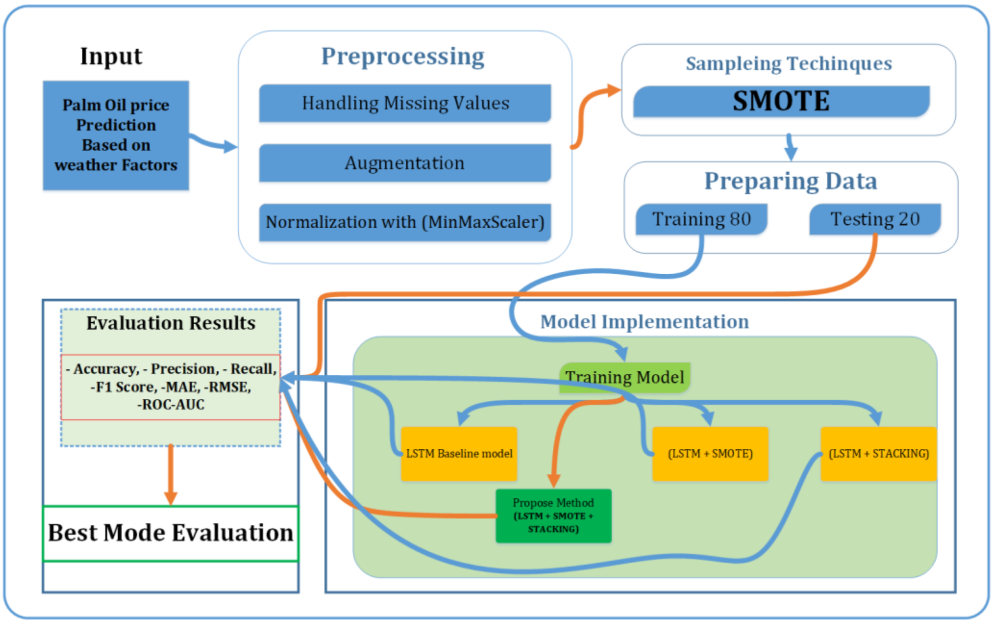


Figure 2. Research Framework

Figure 2 illustrates the overall research framework for palm oil price prediction based on weather-related factors, employing deep learning techniques and advanced data preprocessing. The process begins with the input of palm oil price data and weather variables, followed by preprocessing steps that include handling missing values, data augmentation (if needed), and normalization using MinMaxScaler to ensure consistent feature scaling. In the data preprocessing stage, a class labeling process is performed to convert the price regression data into a binary classification

ISSN 2723-6471

format, namely class 1 (price increase) and class 0 (stable/decrease). This labeling is performed by comparing the palm oil price on day t with the price on the previous day (t-1). If the price difference exceeds the 0% threshold (positive), the data is labeled 1 (increase); whereas if the price difference is zero or negative, it is labeled 0 (decrease/stable). This strategy aims to simplify the prediction task and adapt the model to the classification approach used in the study. After preprocessing, SMOTE is applied to address class imbalance and reduce model bias toward majority patterns. The balanced dataset is then split into 80% training and 20% testing data. Model implementation involves training four different configurations: baseline LSTM, LSTM with SMOTE, LSTM with stacking, and the proposed hybrid model combining LSTM + SMOTE + Stacking. Each model is evaluated using multiple performance metrics, including Accuracy, Precision, Recall, F1-score, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and ROC-AUC. These evaluations are analyzed to determine the most effective model for predicting palm oil prices under weather-influenced conditions.

This study employs a structured and comprehensive application of SMOTE and stacking ensemble to enhance model performance in commodity price prediction tasks. SMOTE is an oversampling method used to address class imbalance by generating synthetic samples for the minority class through interpolation between existing nearest neighbors. The synthetic data point is generated as follows:

$$x_{synthetic} = x_i + \lambda.\,(x_i{}^{(NN)} - x_i) \tag{1}$$

$\lambda \epsilon [0,1]$ is a random number. This process is repeated until a balanced class distribution is achieved. Stacking, on the other hand, is an ensemble learning technique that combines multiple base learners to generate individual predictions, which are then used as input for a meta-learner to produce the final output. The stacking model aims to minimize the aggregated prediction error by learning non-linear combinations of the base models' outputs, expressed mathematically as:

$$Min_H \sum_{i=1}^{n} \mathcal{L}(y_i, H\big(h_1(x_i), \ldots\ldots, h_k(x_i)\big)) \tag{2}$$

$\mathcal{L}$ is a loss function, such as Mean Squared Error (MSE) for regression or log-loss for classification tasks. The dataset is partitioned into two subsets: 80% for training and 20% for testing, to ensure model generalizability and avoid overfitting. The model implementation stage compares two main approaches: baseline models and the optimized stacking-based models. The stacking approach combines these same models into an ensemble framework to improve predictive accuracy. Upon training completion, all models are evaluated using a set of performance metrics, including Accuracy, Precision, Recall, F1 Score, MAE, RMSE, and ROC-AUC.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \tag{6}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - y_i \tag{7}$$

$$\eta = total\ amount\ of\ data\ (sampel), yi = actual\ value\ \left(\frac{real}{observation}\right) ke - i, \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y_i)^2} \tag{9}$$

$$AUC = \int_0^1 TPR(x)dx \tag{10}$$

The evaluation results serve as the basis for identifying the most accurate and reliable model for palm oil price prediction based on weather-related factors. This approach is designed to ensure that the final model is not only statistically robust but also adaptable to the complexities of real-world data.

3.2. Proposed Model

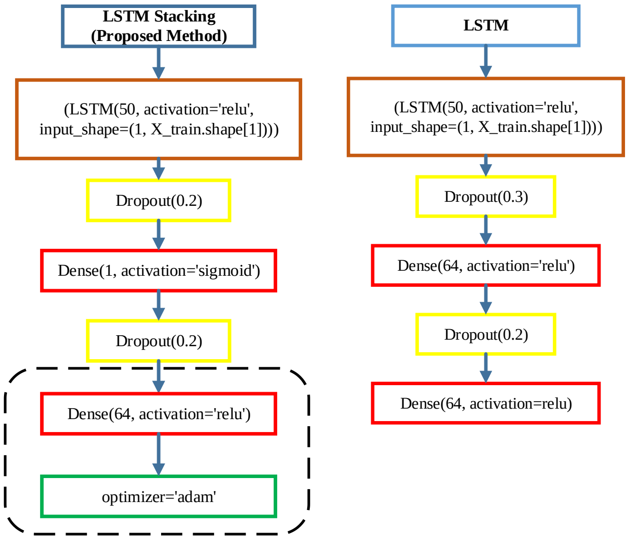
This study proposes a new forecasting model designed to improve the accuracy and stability of palm oil price predictions by incorporating weather factors as input variables. The model combines three main approaches: LSTM for time-series processing, SMOTE for handling data imbalance, and stacking ensemble to enhance prediction performance by combining multiple models. This integrated approach addresses the limitations found in individual models and aims to provide better results, especially in capturing price fluctuations influenced by weather patterns and historical trends. Four models were tested in this study to evaluate the impact of stacking and SMOTE on prediction performance, particularly in the context of imbalanced data. The comparison is presented in table 2.

Table 2. Comparative Summary of Model Characteristics

Specification	**Standard LSTM**	**LSTM + Stacking**	**LSTM + SMOTE**	**LSTM + Stacking + SMOTE (proposed)**
Model Complexity	Low	High	Medium	Very High
Imbalanced Data Handling	No	No	Yes	Yes
Ensemble Techniques	No	Yes	No	Yes
Prediction Accuracy	Sufficient	Good	Good	Very Good
Generalization Ability	Limited	Good	Fair	Very Good
Noise Robustness	Low	Medium	Good	High
Training Time	Fast	Longer	Medium	Longest
Advantages	Easy implementation	Strong model combination	Fair data representation	Combined benefits of stacking and SMOTE
Disadvantages	Overfitting, bias	Biased to majority class	Weak in ensemble	High complexity, computational cost

Table 2 summarizes the specifications and characteristics of the four forecasting models evaluated in this study. The standard LSTM model is the simplest in terms of implementation and training time but struggles with data imbalance and generalization, making it prone to overfitting. The LSTM + Stacking model improves prediction accuracy and generalization through ensemble techniques, yet remains vulnerable to minority class bias. Meanwhile, the LSTM + SMOTE model addresses data imbalance effectively, offering better noise robustness, though it lacks the benefits of model combination. The proposed model (LSTM + Stacking + SMOTE) integrates both enhancements—data balancing and ensemble learning—resulting in the highest accuracy, generalization capability, and robustness. However, it also introduces greater complexity and computational demands. This comparison enables the evaluation of how stacking and hyperparameter tuning contribute to improving commodity price prediction performance. The structure of the stacking model is illustrated in figure 3 below.

Figure 3 illustrates the architectural differences between (a) the proposed stacking model and (b) the baseline stacking model, both built on LSTM. In the proposed method, the model begins with an LSTM layer containing 50 units and ReLU activation, followed by a Dropout layer (0.2) to prevent overfitting. The intermediate output is then passed through a Dense layer with a sigmoid activation (Dense (1, activation='sigmoid')). Instead of treating this as the final output, the architecture applies another Dropout (0.2) followed by a Dense layer with 64 neurons and ReLU activation, representing a two-stage stacking approach. The Adam optimizer is used for training.



a) Stacking Proposed Method b) Baseline

Figure 3. Comparison of Stacking Architectures

In contrast, the baseline model adopts a more linear architecture: an LSTM layer followed by a Dropout (0.3), then two Dense layers (64 units, ReLU activation) separated by another Dropout (0.2), without any intermediate prediction output. The proposed design is expected to enhance performance by leveraging the intermediate prediction as an additional feature representation, ultimately improving final prediction accuracy. The component-wise comparison is shown in table 3.

Table 3. Component-wise Comparison Between the Proposed Method and the Baseline Model

Aspect	Proposed Method	Baseline Model
Initial Layer Output	Dense (1, activation='sigmoid')	No initial output
Final Dense Addition	Yes (Dense (64, activation='relu'))	Yes (Dense (64, activation='relu'))
Initial Dropout	0.2	0.3
Optimizer	Adam (explicitly defined)	Not specified (default optimizer assumed)
Focus	Combination of intermediate prediction and post-processing	Direct feature mapping through dense layers
Objective	Enhanced generalization and ability to handle non-linear stacking	Basic prediction without intermediate representation or tuning

source: processed data

Table 3 provides a component-level comparison between the proposed stacking method and the baseline model. The key distinction lies in the inclusion of an intermediate prediction layer (Dense (1, activation='sigmoid')) in the proposed model, which serves as a preliminary output and is further processed to enhance feature representation. Both models include a final Dense layer with ReLU activation, but the proposed method employs a lower dropout rate (0.2 vs. 0.3), potentially balancing regularization and learning capacity more effectively.

Additionally, the proposed model explicitly defines the Adam optimizer, offering better control over optimization behavior, whereas the baseline uses the default setting. In terms of architectural focus, the proposed model emphasizes a combination of intermediate prediction and post-processing, enabling stronger generalization and improved handling of nonlinear stacking interactions. The baseline, by contrast, follows a direct path to dense layers without intermediate refinement, limiting its ability to adapt to complex input dynamics.

4. Results and Discussion

This section presents the results of implementing and testing palm oil price prediction models based on weather-related factors using the LSTM architecture, the SMOTE, and stacking ensemble methods. Each model was evaluated using preprocessed data, which was partitioned into training and testing subsets. The analysis compares the performance of

ISSN 2723-6471

each configuration based on several evaluation metrics, with emphasis on prediction accuracy, generalization capability, and effectiveness in handling data imbalance. This section also includes a detailed explanation of the dataset, training procedures, and model evaluation outcomes.

4.1. Dataset Description

The dataset used in this study comprises five key variables: date, temperature, humidity, rainfall, and palm oil price. The data were sourced from the Kaggle platform: https://www.kaggle.com/code/theeyeschico/crop-analysis-and-prediction/input, which provides structured time-series information on historical weather conditions and commodity prices. In this study, the weather-related variables serve as input features, while the palm oil price acts as the target variable to be predicted. Figure 4 illustrates the class distribution of the target variable before and after the application of SMOTE. Prior to oversampling (left), the dataset exhibited a significant imbalance, with Class 0 (e.g., stable or decreasing prices) having substantially more instances than Class 1 (e.g., increasing prices). This imbalance could lead to biased model predictions favoring the majority class. After applying the SMOTE, the dataset was balanced, as shown on the right side of the figure. SMOTE synthetically generated new instances for the minority class (Class 1), resulting in equal representation of both classes. This balancing process is essential to improve model sensitivity, particularly in detecting rare but critical price fluctuation patterns.

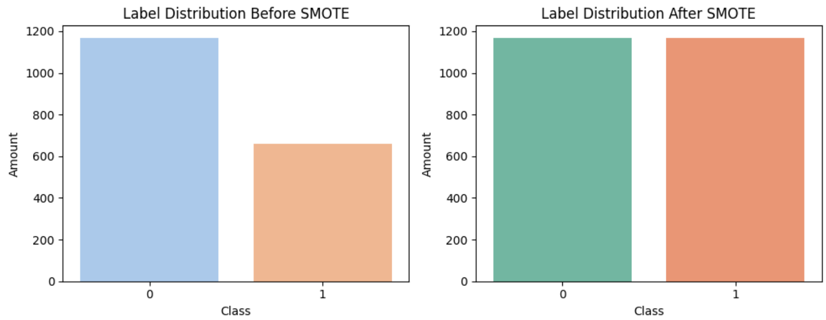
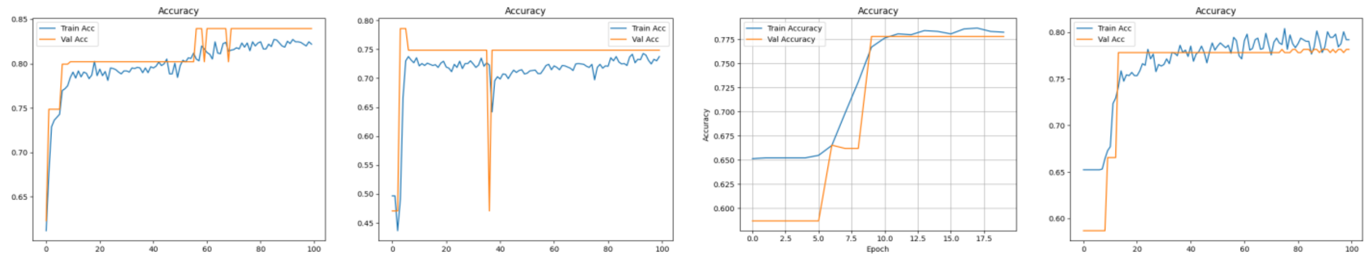


Figure 4. SMOTE Distribution Results

4.2. Model Training and Evaluation

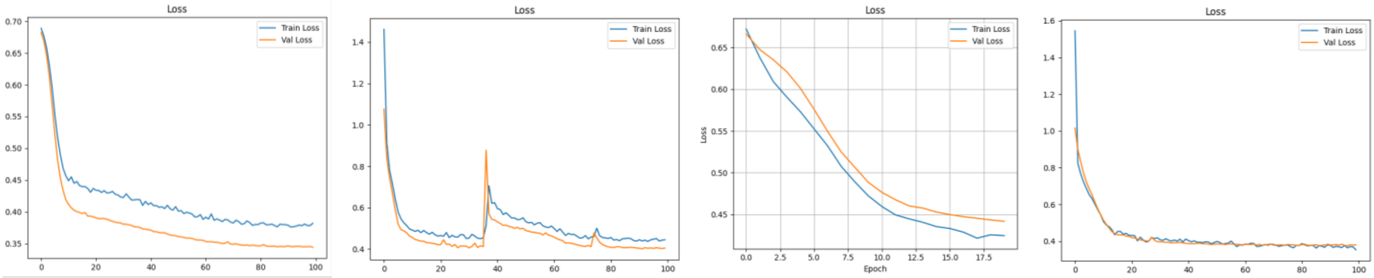
To assess the effectiveness of each model in learning and identifying data patterns, performance visualization during the training phase is essential. One of the key indicators analyzed is the accuracy curve across training epochs for both the training and validation sets. This metric provides insight into model stability, convergence behavior, and potential issues such as overfitting or underfitting. Figure 5 presents the training and validation accuracy curves for the four evaluated models: the baseline LSTM, LSTM with stacking, LSTM with SMOTE, and the proposed hybrid model combining LSTM, stacking ensemble, and SMOTE.



a) LSTM + Stacking + SMOTE (Proposed)
b) LSTM + SMOTE
c) LSTM + Stacking
d) Baseline LSTM

Figure 5. Compares the Training and Validation Accuracy Curves Across Four Different Model Configurations

The proposed model (figure 5a) demonstrates the most stable and consistently high accuracy throughout training, indicating strong learning convergence and minimal overfitting. Both training and validation curves steadily improve and remain close, suggesting good generalization. In contrast, the LSTM + SMOTE model (figure 5b) exhibits significant fluctuations in validation accuracy, implying instability and potential overfitting to minority class patterns generated by SMOTE. Similarly, the LSTM + Stacking model (figure 5c) shows a sharp rise in accuracy during early epochs but later plateaus, which may indicate early convergence but limited long-term learning. The baseline LSTM model (figure 5d) achieves acceptable accuracy but shows signs of noise and lower generalization performance compared to the optimized configuration. Overall, these results confirm that the proposed hybrid model not only improves accuracy but also delivers better learning stability and adaptability across epochs. Figure 6 presents the training and validation loss curves for the four model variants.



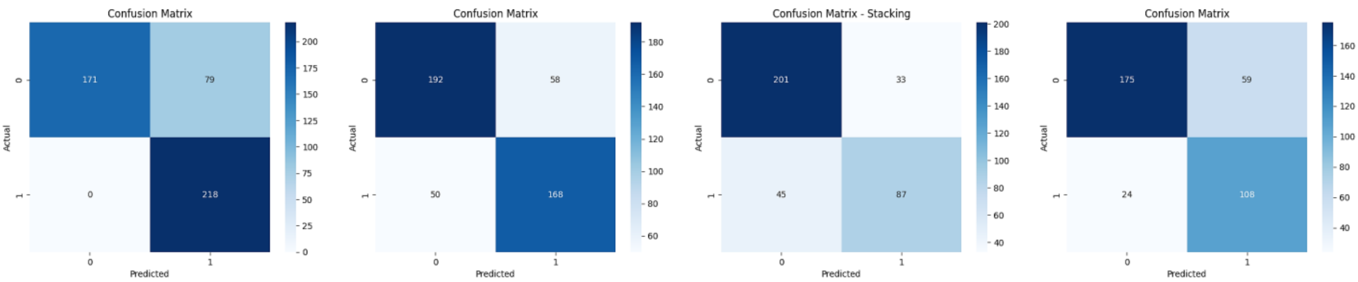
a) LSTM + Stacking + SMOTE (Proposed) b) LSTM + SMOTE c) LSTM + Stacking d) Baseline LSTM

Figure 6. Compares the Training and Validation Loss Curves Across Four Different Model Configurations

The proposed hybrid model (figure 6a) exhibits the most stable and consistently decreasing loss values across epochs for both training and validation sets. The convergence pattern suggests efficient learning with minimal overfitting, as indicated by the close alignment of training and validation loss curves. The LSTM + SMOTE model (figure 6b), while initially showing sharp loss reduction, displays sudden spikes in validation loss, indicating training instability and possible overfitting to synthetic minority data. The stacking-only model (figure 6c) shows gradual convergence but with a relatively larger gap between training and validation loss, hinting at limited generalization. The baseline LSTM model (figure 6d) achieves rapid convergence with minimal fluctuation but plateaus early and fails to achieve further improvement. This behavior reflects a lack of adaptive learning capacity in capturing more complex patterns. These results reinforce the effectiveness of the proposed model in reducing prediction error and improving model reliability through integrated optimization strategies.

4.3. Model Evaluation Results

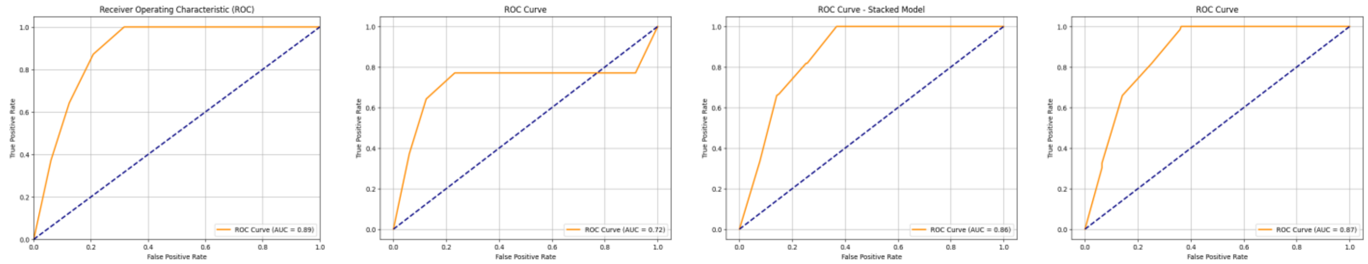
To assess the classification performance of the models in distinguishing palm oil price categories, a confusion matrix was employed as a visual evaluation tool. This matrix provides detailed insights into correct predictions (true positives and true negatives) as well as misclassifications (false positives and false negatives), which are essential for calculating precision, recall, and F1-score. In this study, the target variable was divided into two categories—such as “price decrease/stable” and “price increase”—to evaluate each model’s ability to detect price trend changes based on historical and weather-related data. Model performance was evaluated using both regression and classification metrics, including MAE, RMSE, and R-squared (R^2) for prediction accuracy, as well as Precision, Recall, and F1-score to assess pattern recognition capabilities. The results indicate that the proposed hybrid model (LSTM + SMOTE + Stacking) outperformed all other configurations, achieving the lowest error rates and highest predictive accuracy. It demonstrated consistent performance in capturing complex price patterns influenced by weather dynamics. The baseline LSTM model recorded the weakest performance, while the LSTM + SMOTE and LSTM + Stacking models showed notable improvements, though still inferior to the proposed approach. These findings confirm that combining data balancing techniques with ensemble learning substantially enhances the model’s accuracy and resilience in forecasting weather-driven commodity price fluctuations. Figure 7 shows the confusion matrix for the four model configurations.



a) Proposed b) LSTM + SMOTE c) LSTM + Stacking d) Baseline LSTM

Figure 7. Presents the Confusion Matrices for the Four Model Configurations

The proposed model (figure 7a) shows the most balanced and accurate classification, achieving perfect recall for the positive class (price increase), as indicated by 218 true positives and zero false negatives. Despite a moderate number of false positives (79), this model demonstrates a strong ability to detect minority class instances without sacrificing overall performance. The LSTM + SMOTE model (figure 7b) correctly identifies a high number of positive and negative cases but still misclassifies 58 negative and 50 positive instances, suggesting moderate improvement in balance handling yet limited precision. In contrast, the LSTM + Stacking model (figure 7c) achieves high true negative counts (201) but underperforms in positive class recognition, with only 87 true positives and 45 false negatives. This indicates a bias toward the majority class and reduced sensitivity to minority patterns. The baseline LSTM (figure 7d) has the lowest recall and precision, particularly evident in the misclassification of 59 negatives and 24 false negatives, reflecting its inability to adequately capture non-linear and imbalanced patterns in the dataset. These findings further confirm the superiority of the proposed hybrid model, particularly in contexts involving skewed distributions and complex weather-driven price fluctuations. To see a comparison of the Receiver Operating Characteristic (ROC) curves for the four model variants, see figure 8 below.



a) LSTM + Stacking + SMOTE (Proposed) b) LSTM + SMOTE c) LSTM + Stacking d) Baseline LSTM

Figure 8. Compares the Receiver Operating Characteristic (ROC) Curves For Four Model Variants

The proposed model (figure 8a) achieves the highest Area Under the Curve (AUC) score of 0.89, indicating superior classification performance across various threshold settings. The ROC curve demonstrates strong separability, with a high true positive rate (TPR) and low false positive rate (FPR), confirming the model's effectiveness in detecting positive class instances. In comparison, the LSTM + SMOTE model (figure 8b) records an AUC of 0.77, showing moderate improvement over the baseline but still limited in sensitivity. The LSTM + Stacking model (figure 8c) achieves an AUC of 0.85, reflecting better performance than SMOTE alone, but slightly below the proposed hybrid model. The baseline LSTM model (figure 8d) registers the lowest AUC of 0.67, indicating weak class separability and lower discriminatory capability. Its ROC curve closely follows the diagonal reference line, suggesting a near-random classification behavior. Overall, the ROC analysis reinforces the earlier findings that the proposed hybrid model consistently outperforms other configurations in terms of both precision and robustness, especially in handling imbalanced and nonlinear data.

4.4. Discussion

The experimental results confirm that applying SMOTE to balance the dataset has a positive impact on the model's ability to recognize rare price variations, such as sharp increases or decreases that are underrepresented in the historical data. When using the standard LSTM without additional preprocessing techniques, the model tends to overfit to dominant patterns (e.g., stable prices) and fails to capture significant changes. This limitation is reflected in the lower recall and F1-score for the minority class. Table 4 presents the quantitative evaluation results of the four tested models based on six key performance metrics: Accuracy, Precision, Recall, F1 Score, MAE, and RMSE. The proposed model (LSTM + Stacking + SMOTE) consistently outperforms all other configurations across these metrics, demonstrating superior predictive accuracy, improved error minimization, and stronger capability in identifying complex and rare price patterns.

Table 4. Evaluation Metrics

Model	Accuracy	Precision	Recall	F1 Score	MAE	RMSE
LSTM + Stacking + SMOTE (Proposed)	0.8312	0.7340	1.0000	0.8466	0.1688	0.4109
LSTM + SMOTE	0.7692	0.7434	0.7706	0.7568	0.2308	0.4804

LSTM + Stacking	0.7869	0.7250	0.6591	0.6905	0.2131	0.4616
Baseline LSTM	0.7732	0.6467	0.8182	0.7224	0.2268	0.4762

Table 4 summarizes the evaluation results of the four tested models across six performance metrics. The proposed model—LSTM + Stacking + SMOTE—achieved the highest scores in nearly all metrics. It recorded the best accuracy (0.8312), recall (1.0000), and F1 score (0.8466), indicating excellent capability in identifying both majority and minority classes. Notably, the perfect recall shows that the model was able to detect all instances of the minority class (i.e., significant price changes), without missing any relevant cases. Furthermore, the model also achieved the lowest error rates, with MAE of 0.1688 and RMSE of 0.4109, reinforcing its robustness and predictive reliability. In comparison, the LSTM + SMOTE model improved recall (0.7706) over the standard model (0.8182) but suffered from higher MAE and RMSE, suggesting that while it better handled data imbalance, it lacked fine-grained accuracy. The LSTM + Stacking model performed better in terms of overall accuracy (0.7869) but had the lowest recall (0.6591), reflecting limited sensitivity to minority class patterns. The baseline LSTM model showed the weakest performance in most metrics, particularly precision (0.6467) and F1 score (0.7224), underscoring its inability to generalize in complex, imbalanced data scenarios. These results highlight the advantage of combining data balancing and ensemble learning, affirming that the proposed hybrid model provides a more comprehensive and effective solution for weather-based palm oil price prediction.

Figure 9 displays a heatmap comparing the performance of all models across six evaluation metrics. Darker shades represent higher values for Accuracy, Precision, Recall, and F1 Score, while lighter shades in MAE and RMSE indicate better error performance (lower is better). The LSTM + SMOTE + Stacking (Proposed) model consistently shows the most favorable values, with the highest accuracy (0.8312), recall (1.0000), and F1-score (0.8466), along with the lowest MAE (0.1688) and RMSE (0.4109). In contrast, the baseline LSTM model shows weaker performance across most metrics, especially in Precision and RMSE. This heatmap offers a compact and interpretable visual comparison, reinforcing that the proposed hybrid approach provides a well-balanced model with superior classification power and minimal prediction error.

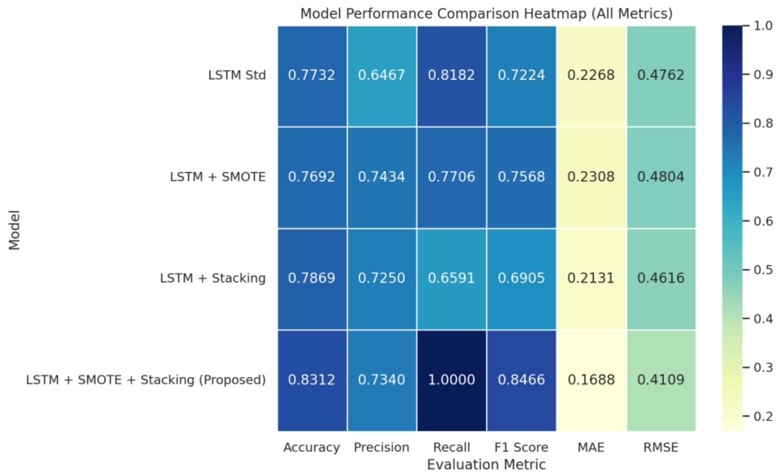


Figure 9. Comparing the Performance of All Models Across Six Evaluation Metrics

5. Conclusion

This study proposed a hybrid forecasting model for palm oil price prediction that integrates LSTM, stacking ensemble learning, and SMOTE. By incorporating weather-based variables as input features, the model aims to capture both linear and nonlinear dependencies, while addressing the challenge of imbalanced datasets commonly found in commodity pricing scenarios. Through comparative experiments involving four model configurations—baseline LSTM, LSTM + SMOTE, LSTM + Stacking, and the proposed hybrid approach—the results consistently demonstrate the superiority of the proposed model across multiple evaluation metrics. It achieved the highest accuracy (0.8312), perfect recall (1.0000), the highest F1 score (0.8466), and the lowest error values (MAE of 0.1688 and RMSE of 0.4109). These results validate the effectiveness of combining data balancing and ensemble learning in enhancing model generalization, particularly in forecasting complex, weather-driven price fluctuations. The findings contribute both theoretically and practically to the field of time-series prediction in agriculture and commodity economics. Future research may explore integrating additional external variables such as macroeconomic indicators or deploying real-time forecasting systems using online learning frameworks.

6. Declarations

6.1. Author Contributions

Conceptualization: K., S.S.; Methodology: K., E.M.; Software: K.; Validation: S.S., E.M.; Formal Analysis: K.; Investigation: K.; Resources: S.S., E.M.; Data Curation: K.; Writing – Original Draft Preparation: K.; Writing – Review and Editing: S.S., E.M.; Visualization: K.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

This work was funded by the Ministry of Higher Education, Science, and Technology, Directorate General of Research and Development, through the Regular Fundamental Research Program for the 2025 Fiscal Year.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Y. Asbur, "Improvement Of Soil Chemical Properties In Mature Oil Palm Plantations By Pruning And Immersing Of Weeds As Cover Crops," *Universal Journal Of Agricultural Research*, vol. 12, no. 1, pp. 188–194, 2024, Doi: 10.13189/Ujar.2024.120118.

[2] C. I. N. Indonesia, "Palm'journal," *Palmoilina.Asia*. vol. 7, no. 7, Pp.11-28, 2023, doi: 10.33541/japs.v7i1.5102,

[3] M. S. Bahar And A. A. Karia, "A Dual Methods Approach To Crude Palm Oil Price Forecasting In Malaysia : Insights From Ardl And Lstm," *Burgas Free University (BFU),* vol. 2024, no. 2008, pp. 106–124, 2024.

[4] M. Rashid, B. S. Bari, Y. Yusup, M. A. Kamaruddin, And N. Khan, "A Comprehensive Review Of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis On Palm Oil Yield Prediction," *Ieee Access*, vol. 9, no. 1, pp. 63406–63439, 2021, Doi: 10.1109/Access.2021.3075159.

[5] F. Danitasari, M. Ryan, D. Handoko, And I. Pramuwardani, "Improving Accuracy Of Daily Weather Forecast Model At Soekarno-Hatta Airport Using Bilstm With Smote And Adasyn," *Journal of Science Education Research*, vol. 10, no. 1, pp. 179–193, 2024, Doi: 10.29303/Jppipa.V10i1.5906.

[6] T. Benil, "Efficient Data Pruning Using Optimal Knn For Weather Forecasting In Cloud Computing," *International Journal Of Global Warming,* vol. 30, no. 2, pp. 137–151, 2023, Doi: 10.1504/Ijgw.2023.130984.

[7] I. Intan, S. Aminah, D. Ghani, And A. T. C. Koswara, "Machine Learning Performance Analysis Of Weather Forecasting Using Machine Learning Algorithms," *Research and Development Journal of Communication, Informatics, and Mass Media*, vol. 2021, no. 1 , pp. 1–8, 2021, Doi: 10.30818/Jpkm.2021.2060221.

[8] C. Mcknight, "Prices For A Second-Generation Biofuel Industry In Canada: Market Linkages Between Canadian Wheat And Us Energy And Agricultural Commodities," *Canadian Journal Of Agricultural Economics*, vol. 69, no. 3, pp. 337–351, 2021, Doi: 10.1111/Cjag.12295.

[9] M. R.L, “Market Efficiency And Price Risk Management Of Agricultural Commodity Prices In India,” *Journal Of Modelling In Management*, vol. 18, no. 1, pp. 190–211, 2023, Doi: 10.1108/Jm2-04-2021-0104.

[10] M. Garg, “Price Discovery Mechanism And Volatility Spillover Between National Agriculture Market And National Commodity And Derivatives Exchange: The Study Of The Indian Agricultural Commodity Market,” *Journal Of Risk And Financial Management*, vol. 16, no. 2, pp. 1-12, 2023, Doi: 10.3390/Jrfm16020062.

[11] A. P. Windarto, T. Herawan, And P. Alkhairi, “Prediction Of Kidney Disease Progression Using K-Means Algorithm Approach On Histopathology Data,” *In Artificial Intelligence, Data Science And Applications, Cham: Springer Nature Switzerland*, 2024, Pp. 492–497. Doi: 10.1007/978-3-031-48465-0_66.

[12] S. Defit, A. P. Windarto, And P. Alkhairi, “Comparative Analysis Of Classification Methods In Sentiment Analysis: The Impact Of Feature Selection And Ensemble Techniques Optimization,” *Telematika*, vol. 17, no. 1, pp. 52–67, 2024.

[13] J. D. Rosita P And W. S. Jacob, “Multi-Objective Genetic Algorithm And Cnn-Based Deep Learning Architectural Scheme For Effective Spam Detection,” *International Journal Of Intelligent Networks*, vol. 3, no. december 2021, pp. 9–15, 2022, Doi: 10.1016/J.Ijin.2022.01.001.

[14] N. N. Prakash, V. Rajesh, D. L. Namakhwa, S. Dwarkanath Pande, And S. H. Ahammad, “A Densenet Cnn-Based Liver Lesion Prediction And Classification For Future Medical Diagnosis,” *Scientific African*, vol. 20, no. 1, pp. 1-29, 2023, Doi: 10.1016/J.Sciaf.2023.E01629.

[15] J. Yu, “Recurrence Plot Image And Googlenet Based Historical Abuse Backtrace For Li-Ion Batteries,” *Journal Of Energy Storage*, vol. 74, no. 1, pp. 1-12, 2023, Doi: 10.1016/J.Est.2023.109378.

[16] G. Feng, “Research On Emotion Recognition In Electroencephalogram Based On Independent Component Analysis-Recurrence Plot And Improved Efficientnet,” *Journal Of Biomedical Engineering*, vol. 41, no. 6, pp. 1103–1109, 2024, Doi: 10.7507/1001-5515.202406029.

[17] X. Zhao, “Remote Sensing Image Change Detection Based On Improved Deeplabv3+ Siamese Network,” *Journal Of Geo-Information Science*, vol. 24, no. 8, pp. 1604–1616, 2022, Doi: 10.12082/Dqxxkx.2022.220029.

[18] H. A. Shah, F. Saeed, S. Yun, J. H. Park, A. Paul, And J. M. Kang, “A Robust Approach For Brain Tumor Detection In Magnetic Resonance Images Using Finetuned Efficientnet,” *Ieee Access*, vol. 10, no. 1, pp. 65426–65438, 2022, Doi: 10.1109/Access.2022.3184113.

[19] M. S. Puchaicela-Lozano, “Deep Learning For Glaucoma Detection: R-Cnn Resnet-50 And Image Segmentation,” *Journal Of Advances In Information Technology*, vol. 14, no. 6, pp. 1186–1197, 2023, Doi: 10.12720/Jait.14.6.1186-1197.

[20] S. Jameer, “A Dcnn-Lstm Based Human Activity Recognition By Mobile And Wearable Sensor Networks,” *Alexandria Engineering Journal*, vol. 80, no. 1, pp. 542–552, 2023, Doi: 10.1016/J.Aej.2023.09.013.

[21] J. A. Adisa, “The Effect Of Imbalanced Data And Parameter Selection Via Genetic Algorithm Long Short-Term Memory (Lstm) For Financial Distress Prediction,” I*aeng International Journal Of Applied Mathematics*, vol. 53, no. 3, pp. 1-12, 2023.

[22] S. K. Challa, “An Optimized-Lstm And Rgb-D Sensor-Based Human Gait Trajectory Generator For Bipedal Robot Walking,” *Ieee Sensors Journal*, vol. 22, no. 24, pp. 24352–24363, 2022, Doi: 10.1109/Jsen.2022.3222412.

[23] M. A. Khatun, “Deep Cnn-Lstm With Self-Attention Model For Human Activity Recognition Using Wearable Sensor,” Ieee Journal Of Translational Engineering In Health And Medicine, vol. 10, no. 1, pp. 1-12, 2022, Doi: 10.1109/Jtehm.2022.3177710.

[24] A. Sarkar, “An Effective And Novel Approach For Brain Tumor Classification Using Alexnet Cnn Feature Extractor And Multiple Eminent Machine Learning Classifiers In Mris,” *Journal Of Sensors*, vol. 2023, no. 1, pp. 1-12, 2023, Doi: 10.1155/2023/1224619.

[25] G. Kaur And A. Sharma, “A Deep Learning-Based Model Using Hybrid Feature Extraction Approach For Consumer Sentiment Analysis,” *Journal Of Big Data*, vol. 10, no. 1, pp. 1-12, 2023, Doi: 10.1186/S40537-022-00680-6.

[26] V. Sunanthini, “Comparison Of Cnn Algorithms For Feature Extraction On Fundus Images To Detect Glaucoma,” *Journal Of Healthcare Engineering*, vol. 2022, no. 1, pp. 1-12, 2022, Doi: 10.1155/2022/7873300.

[27] A. Gogineni, “Evaluating Machine Learning Algorithms For Predicting Compressive Strength Of Concrete With Mineral Admixture Using Long Short-Term Memory (Lstm) Technique,” *Asian Journal Of Civil Engineering*, vol. 25, no. 2, pp. 1921–1933, 2024, Doi: 10.1007/S42107-023-00885-X.

[28] A. Beaulieu, "Ultra-Wideband Data As Input Of A Combined Efficientnet And Lstm Architecture For Human Activity Recognition," *Journal Of Ambient Intelligence And Smart Environments*, vol. 14, no. 3, pp. 157–172, 2022, Doi: 10.3233/Ais-210462.

[29] K. Balaraman, "Hybrid Resnet And Bidirectional Lstm Based Deep Learning Model For Cardiovascular Disease Detection Using Ppg Signals," *Journal Of Machine And Computing*, vol. 3, no. 3, pp. 351–359, 2023, Doi: 10.53759/7669/Jmc202303030.

[30] I. Janbain, "Use Of Long Short-Term Memory Network (Lstm) In The Reconstruction Of Missing Water Level Data In The River Seine," *Hydrological Sciences Journal*, vol. 68, no. 10, pp. 1372–1390, 2023, Doi: 10.1080/02626667.2023.2221791.

[31] A. Rehman, T. Alam, M. Mujahid, F. S. Alamri, B. Al Ghofaily, And T. Saba, "Rdet Stacking Classifier: A Novel Machine Learning Based Approach For Stroke Prediction Using Imbalance Data," *Peerj Computer Science*, vol. 9, no. 1, pp. 1–28, 2023, Doi: 10.7717/Peerj-Cs.1684.

[32] V. Mohanavel, "Influence Of Stacking Sequence And Fiber Content On The Mechanical Properties Of Natural And Synthetic Fibers Reinforced Penta-Layered Hybrid Composites," *Journal Of Natural Fibers*, vol. 19, no. 13, pp. 5258–5270, 2022, Doi: 10.1080/15440478.2021.1875368.

[33] S. S. Rani, "An Automated Lion-Butterfly Optimization (Lbo) Based Stacking Ensemble Learning Classification (Selc) Model For Lung Cancer Detection," *Iraqi Journal For Computer Science And Mathematics*, vol. 4, no. 3, pp. 87–100, 2023, Doi: 10.52866/Ijcsm.2023.02.03.008.

[34] Y. S. Taspinar, "Classification By A Stacking Model Using Cnn Features For Covid-19 Infection Diagnosis," *Journal Of X Ray Science And Technology*, vol. 30, no. 1, pp. 73–88, 2022, Doi: 10.3233/Xst-211031.

[35] N. Mqadi, "A Smote Based Oversampling Data-Point Approach To Solving The Credit Card Data Imbalance Problem In Financial Fraud Detection," *International Journal Of Computing And Digital Systems*, vol. 10, no. 1, pp. 277–286, 2021, Doi: 10.12785/Ijcds/100128.

[36] A. Kishor, "Early And Accurate Prediction Of Diabetics Based On Fcbf Feature Selection And Smote," *International Journal Of System Assurance Engineering And Management*, vol. 15, no. 10, pp. 4649–4657, 2024, Doi: 10.1007/S13198-021-01174-Z.

[37] L. Azoulay-Younes, "Zinc Electrode Manufacturing Quality Control With Machine Learning: Using Smote & Image Augmentation To Prevent Overfitting," *Journal Of Engineering Research Kuwait*, vol. 2025, no. 1, pp. 1-12, 2025, Doi: 10.1016/J.Jer.2025.01.002.

[38] P. C. Y. Cheah, "Enhancing Financial Fraud Detection Through Addressing Class Imbalance Using Hybrid Smote-Gan Techniques," *International Journal Of Financial Studies*, vol. 11, no. 3, pp. 1-12, 2023, Doi: 10.3390/Ijfs11030110.

[39] N. A. A. Khleel, "A Novel Approach For Software Defect Prediction Using Cnn And Gru Based On Smote Tomek Method," *Journal Of Intelligent Information Systems*, vol. 60, no. 3, pp. 673–707, 2023, Doi: 10.1007/S10844-023-00793-1.

[40] S. R. Dani, S. Solikhun, And D. Priyanto, "The Performance Machine Learning Powel-Beale For Predicting Rubber Plant Production In Sumatera," *International Journal Of Engineering And Computer Science Applications (Ijecsa)*, vol. 2, no. 1, pp. 29–38, 2023, Doi: 10.30812/Ijecsa.V2i1.2420.

[41] I. Ranggadara, U. M. Buana, A. Ratnasari, And U. M. Buana, "Backpropagation Neural Network For Predict Sugarcane Stock Availability," *International Journal Of Advanced Trends In Computer Science And Engineering*, vol. 9, no. 5, pp. 8279–8284, 2020, Doi: 10.30534/Ijatcse/2020/197952020.

[42] L. Tong, "A Novel Deep Learning Bi-Gru-I Model For Real-Time Human Activity Recognition Using Inertial Sensors," *Ieee Sensors Journal*, vol. 22, no. 6, pp. 6164–6174, 2022, Doi: 10.1109/Jsen.2022.3148431.

[43] D. Malhotra, "Hybrid Deep Learning Model For Covid-19 Prediction Using Convolutional Neural Network (Cnn) And Bidirectional Long Short-Term Memory (Lstm) Network," *International Journal Of Computer Theory And Engineering*, vol. 15, no. 3, pp. 125–129, 2023, Doi: 10.7763/Ijcte.2023.V15.1341.

[44] S. Y. Xiong, "A Hybrid Cnn–Rnn Model For Rainfall–Runoff Modeling In The Potteruvagu Watershed Of India," *Clean – Soil, Air, Water*, vol. 10, no. 3, pp. 23-41, 2024, Doi: 10.1002/Clen.202300341.

[45] P. R. Shekar, A. Mathew, And K. V. Sharma, "A Hybrid Cnn–Rnn Model For Rainfall–Runoff Modeling In The Potteruvagu Watershed Of India," *Clean – Soil, Air, Water*, vol. n/a, no. n/a, pp. 23-41, 2024, Doi: 10.1002/Clen.202300341

[46] Enung, "Hourly Discharge Prediction Using Long Short-Term Memory Recurrent Neural Network (Lstm-Rnn) In The Upper Citarum River," *International Journal Of Geomate*, vol. 23, no. 98, pp. 147–154, 2022, Doi: 10.21660/2022.98.3462.

[47] A. Bâra And S. V. Oprea, "Machine Learning Algorithms For Power System Sign Classification And A Multivariate Stacked Lstm Model For Predicting The Electricity Imbalance Volume," *International Journal Of Computational Intelligence Systems*, vol. 17, no. 1, pp. 1-12, 2024, Doi: 10.1007/S44196-024-00464-1.

[48] S. Nizarudeen, "Multi-Layer Resnet-Densenet Architecture In Consort With The Xgboost Classifier For Intracranial Hemorrhage (Ich) Subtype Detection And Classification," *Journal Of Intelligent And Fuzzy Systems*, vol. 44, no. 2, pp. 2351–2366, 2023, Doi: 10.3233/Jifs-221177.

[49] R. E. Ako, "Pilot Study On Fibromyalgia Disorder Detection Via Xgboosted Stacked-Learning With Smote-Tomek Data Balancing Approach," *Nipes Journal Of Science And Technology Research*, vol. 7, no. 1, pp. 12–22, 2025, Doi: 10.37933/Nipes/7.1.2025.2.

[50] V. D. Gowda, "A Novel Rf-Smote Model To Enhance The Definite Apprehensions For Iot Security Attacks," *Journal Of Discrete Mathematical Sciences And Cryptography*, vol. 26, no. 3, pp. 861–873, 2023, Doi: 10.47974/Jdmsc-1766.

[51] M. Waqar, H. Dawood, H. Dawood, N. Majeed, A. Banjar, And R. Alharbey, "An Efficient Smote-Based Deep Learning Model For Heart Attack Prediction," *Scientific Programming*, vol. 2021, no. 1, pp. 1-12, 2021, Doi: 10.1155/2021/6621622.

[52] A. Arafa, "Rn-Smote: Reduced Noise Smote Based On Dbscan For Enhancing Imbalanced Data Classification," *Journal Of King Saud University Computer And Information Sciences*, vol. 34, no. 8, pp. 5059–5074, 2022, Doi: 10.1016/J.Jksuci.2022.06.005.

[53] J. Nanda, "Sshm: Smote-Stacked Hybrid Model For Improving Severity Classification Of Code Smell," *International Journal Of Information Technology Singapore*, vol. 14, no. 5, pp. 2701–2707, 2022, Doi: 10.1007/S41870-022-00943-8.