Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.4, August 2026, pp. 4133~4143
e-ISSN: 2548-6861

Comparative Analysis of Random Forest and Long Short-Term Memory for Predicting Optical Power Degradation in FTTH Networks

Hermansyah [1]*, Taufiq [2]*, Defry Hamdhana [3]*, Munirul Ula [4]*, Muhammad Ikhwanus [5]*
*Master of Information Technology Study Program, Malikussaleh University, Lhokseumawe, Indonesia,
waxherman@gmail.com[1], taufiq.te@unimal.ac.id[2], defryhamdhana@unimal.ac.id[3], munirulula@unimal.ac.id[4], ikhwanus@unimal.ac.id[5]

Article Info

Article history:

Received 2026-06-26
Revised 2026-08-10
Accepted 2026-08-13

Keyword:

FTTH,
Long Short-Term Memory,
Optical Power Degradation,
Predictive Maintenance,
Random Forest.

ABSTRACT

Fiber-to-the-Home (FTTH) networks are widely used to provide high-speed broadband services, but optical power degradation can reduce network performance and service quality. This study compares Random Forest (RF) and Long Short-Term Memory (LSTM) for predicting FTTH network conditions classified as Normal, Warning, and Critical. The study used 63,145 historical records collected from 58 Optical Network Terminals (ONTs) between March and May 2026. To provide a fair comparison, RF was trained using engineered tabular features, including lag and rolling-window statistics, while LSTM used six-step sequential data representing approximately the previous six hours. Model performance was evaluated using accuracy, precision, recall, F1-score, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), training time, and inference time. The results show that RF substantially outperformed LSTM, achieving 98.41% accuracy, precision, recall, and F1-score, with an MAE of 0.0173 and RMSE of 0.1420. RF also required only 2.667 seconds for training and 0.0102 ms for inference, compared with 189.98 seconds and 0.1856 ms for LSTM. Per-class evaluation confirmed that RF performed well across all network conditions, with precision and recall above 99% for Normal, above 94% for Warning, and above 90% for Critical. A strict chronological train-test split further confirmed the robustness of RF, which achieved 98.59% accuracy. Feature importance analysis showed that historical optical power, particularly lag-based features, was the most influential predictor of network degradation. These findings indicate that FTTH optical power degradation can be effectively modeled using engineered tabular features rather than a purely sequential approach. Finally, the RF model was integrated into a web-based monitoring dashboard with WhatsApp-based early warnings to support proactive FTTH network maintenance.

I. INTRODUCTION

The rapid advancement of information and communication technology over the past decade has been accompanied by a significant increase in demand for digital services, including video streaming, cloud computing, the Internet of Things (IoT), and other data-intensive applications [1]. These growing demands require network infrastructures that provide high bandwidth, low latency, and high reliability [2], driving the transition from copper-based access networks to optical fiber technology. Among the available optical access solutions, Fiber-to-the-Home (FTTH) has become one of the most widely adopted technologies [3], delivering optical signals directly from the service provider to end users while offering high bandwidth capacity, superior transmission stability, and strong immunity to electromagnetic interference [4], [5].

Despite these advantages, FTTH networks remain susceptible to optical signal attenuation, which is reflected by a reduction in the optical power received at the Optical Network Terminal (ONT) [6]. Optical attenuation is influenced by multiple factors, including fiber length, the number of splices, connector quality, installation quality, equipment aging, and environmental conditions such as temperature and humidity. If left undetected, these factors may gradually degrade network performance and eventually lead to service disruptions [7]–[9]. Conventional maintenance practices are predominantly reactive, with corrective actions taken only after service failures occur or customer complaints are received. Such an approach increases network downtime and operational costs, highlighting the need for predictive maintenance strategies capable of identifying potential failures before they affect service quality.

Recent advances in machine learning and deep learning have accelerated the development of predictive models for network performance analysis. Random Forest (RF) [10], [11], an ensemble learning algorithm based on multiple decision trees, has demonstrated strong capability in handling high-dimensional data while providing feature importance analysis to identify the most influential variables [12], [13]. Beyond network engineering, Random Forest has been successfully applied across diverse domains, including healthcare cost prediction [25], privacy-preserving distributed evaluation [26], mental health prediction [27], and general decision-tree-based responsible-AI applications [28], while related supervised approaches such as k-Nearest Neighbor have also been used for comparable prediction tasks [35], underscoring the algorithm's versatility for structured, tabular prediction problems such as the one addressed in this study. Rashed et al. (2020) reported prediction accuracies ranging from 92% to 96% for optical network performance prediction using Random Forest, identifying transmission distance and the number of splices as the dominant factors affecting signal degradation [14]. Meanwhile, Long Short-Term Memory (LSTM), an advanced variant of Recurrent Neural Networks (RNNs) designed to overcome the vanishing gradient problem, has been widely employed for time-series forecasting of network performance. Zhang et al. (2018) and Kim and Cho (2019) achieved prediction accuracies exceeding 90% with RMSE values below 5%, demonstrating the effectiveness of LSTM in modeling temporal network behavior [15], [16]. LSTM-based sequential modeling has also been applied directly to optical power prediction in FTTH networks [8] and, more broadly, to other time-series domains such as time-series forecasting in general [17], stock price prediction [31], and streamflow prediction in hydrology [32], illustrating the wide applicability of LSTM for sequential pattern learning across disciplines.

Recent studies further reinforce these findings. Putra et al. (2025) applied Random Forest Regression to operational FTTH data from PT Telkom Akses Indonesia and achieved an R^2 of 0.9832, an RMSE of 0.2363, and an MAE of 0.1009 [39]. Soothar et al. (2024) proposed a hybrid CNN–LSTM model for optical signal degradation detection, achieving an accuracy of up to 99%, outperforming standalone Random Forest models, which achieved accuracies between 96% and 97% [40]. Zhang et al. (2024) employed an attention-based Bi-LSTM model to predict optical fiber cable degradation, reporting an accuracy of 97.5% [41]. Similarly, Kumar et al. (2023) reported coefficients of determination (R^2) greater than 0.90 for LSTM-based optical transmission quality prediction [42]. More recently, Li et al. (2025) integrated Random Forest and LSTM for predictive maintenance in Dense Wavelength Division Multiplexing (DWDM) networks, achieving an overall prediction accuracy of 98%, surpassing the performance of individual models [43].

Although these studies have demonstrated promising results, most previous research has focused either on a single predictive approach or on generic hybrid architectures. Furthermore, few studies have directly compared the performance of Random Forest and LSTM for predicting optical power degradation in FTTH networks using real operational and physical variables simultaneously, such as fiber distance, the number of splices, the number of connectors, and equipment age. Consequently, existing studies generally address either current network conditions or future prediction accuracy, without jointly analyzing the causal factors of degradation and their temporal evolution within a consistent and fair evaluation framework.

To address this research gap, this study proposes a predictive framework for analyzing optical power degradation in FTTH networks by comparing Random Forest and Long Short-Term Memory (LSTM) using real historical operational data collected from PT. Jaringanku Nusantara. Random Forest is employed to predict network conditions and identify the dominant factors contributing to optical power degradation, whereas LSTM is utilized to capture the temporal patterns of optical power measurements over time. The novelty of this study lies in three main contributions: (1) the utilization of real operational data incorporating measurable physical and environmental variables; (2) a dedicated preprocessing strategy that aligns temporal information between the tabular feature-engineering pipeline and the sequential data pipeline, enabling a fair and unbiased comparison between Random Forest and LSTM; and (3) the integration of prediction results into a web-based dashboard equipped with an early warning system for network monitoring. Therefore, the objective of this study is to analyze and compare the performance of Random Forest and LSTM in predicting optical power degradation in FTTH networks using evaluation metrics including Accuracy, Precision, Recall, F1-score, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE), while investigating the applicability of the proposed framework as an early warning system to support preventive maintenance and improve FTTH network reliability.

II. METHODS

A. Research Type and Study Area

This study adopts a quantitative research design using an experimental and predictive approach. The experimental approach involves developing, training, and evaluating Random Forest and Long Short-Term Memory (LSTM) models to assess their predictive performance. Meanwhile, the predictive approach is employed to forecast optical power degradation in Fiber-to-the-Home (FTTH) networks based on patterns extracted from historical operational data. The study was conducted at PT. Jaringanku Nusantara, using data obtained from the company's FTTH network monitoring system and optical power measurement records maintained as part of its routine network operations.

B. Variables and Data Collection

The research variables were categorized into informational, dynamic, and static variables, as presented in Table I. The optical_power variable serves as the prediction target, while the remaining variables represent factors that influence the quality of the optical signal.

TABLE I
RESEARCH VARIABLES

Category	Variable	Unit	Description
Informational	timestamp, ont_id	–	Measurement time and ONT identifier, used as the basis for lag and rolling feature generation
Dynamic (Target)	optical_power	dBm	Received optical signal power, serving as the primary prediction target
Static	fiber_distance	km	Fiber cable length between the Optical Line Terminal (OLT) and Optical Network Terminal (ONT)
Static	number_of_splices	points	Number of optical fiber splices along the transmission path
Static	number_of_connectors	units	Number of optical connectors installed along the fiber link
Dynamic	device_age, temperature, humidity	years / °C / %	Equipment age and environmental conditions affecting network performance

Data were collected using two complementary techniques. First, documentation was employed to obtain historical records of optical power measurements and FTTH network parameters from the company's network monitoring system and technical reports. Second, semi-structured interviews were conducted with network engineers and technical staff to gather supporting information regarding operational conditions, the causes of optical signal degradation, and existing network maintenance practices. The combination of these data collection methods ensured the completeness and reliability of the dataset for developing and evaluating the proposed predictive models.

C. Research Procedure

The research was conducted through a series of systematic stages. It began with a comprehensive literature review, followed by problem identification and formulation, as well as the determination of the research objectives and expected contributions. Subsequently, historical FTTH operational data were collected and subjected to data preprocessing, including data cleaning, feature engineering, normalization, and dataset preparation. The processed data were then used to develop predictive models based on Random Forest and Long Short-Term Memory (LSTM). The performance of both models was evaluated using appropriate prediction accuracy and error metrics, followed by a comparative analysis and discussion of the experimental results. Finally, conclusions were drawn based on the findings of the study. The overall research procedure is illustrated in Figure 1.

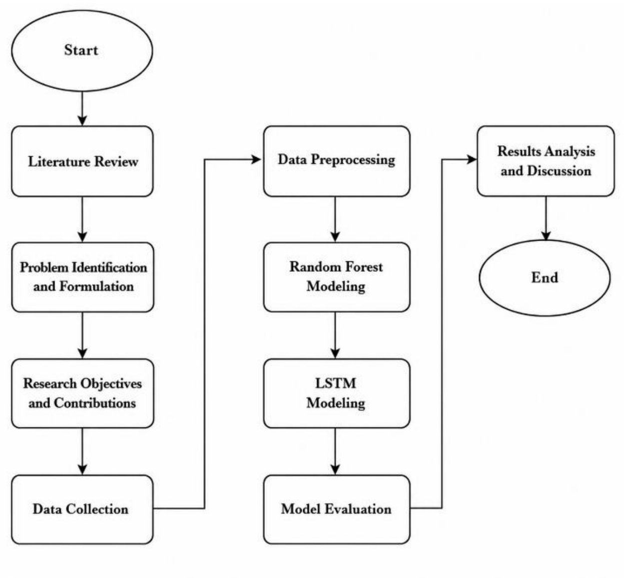

Figure 1. Research Stage Flow

D. Model Implementation and Comparative Evaluation Scenario

The Random Forest and Long Short-Term Memory (LSTM) models were implemented using the same historical dataset, encompassing data preprocessing, model training, and performance evaluation, as illustrated in Figure 2. The workflow began with data acquisition from the Optical Line Terminal (OLT) logs and the Network Management System (NMS) over an observation period of 3–6 months. The collected data then underwent an initial preprocessing stage consisting of data cleaning, missing value handling, and Min–Max normalization.

Since LSTM is specifically designed to process sequential time-series data, whereas Random Forest operates on independent tabular data, a dedicated preprocessing strategy was introduced to ensure that both models received equivalent temporal information, thereby enabling a fair and unbiased (apple-to-apple) comparison. This preprocessing strategy allowed the comparative evaluation to focus on the intrinsic predictive capabilities of the two algorithms rather than differences arising from data representation.

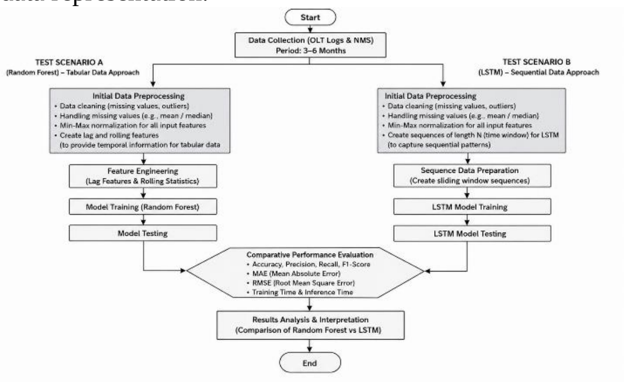

Figure 2. Implementation and Testing Scenarios of Random Forest and LSTM

1) Scenario A – Random Forest with Feature Engineering:
Since Random Forest does not possess an internal memory mechanism, temporal information was incorporated through feature engineering. This process involved generating lag features, including optical power and environmental variables at time steps $t-1$, $t-2$, $t-3$, representing observations from the three preceding measurement points for each ONT, as well as rolling window statistics, namely the moving average and standard deviation of optical power and temperature computed over a sliding window of three previous observations. To prevent temporal data leakage, all lag and rolling-window features were computed strictly from past observations only, using a one-step backward shift prior to window aggregation, ensuring that no information from the prediction time point or later was used during feature construction. The engineered dataset was subsequently divided into training and testing sets using an 80:20 stratified random split (train–test split approach, stratified by class label to preserve the class proportions in both subsets). The Random Forest model was trained using 200 decision trees (*n_estimators* = 200), with *max_features* set to sqrt and *class_weight* set to balanced to mitigate the effects of class imbalance.

2) Scenario B – LSTM with Sequential Data: For the Long Short-Term Memory (LSTM) model, the dataset was transformed into a three-dimensional tensor consisting of *(samples, time steps, features)*, with a sequence length of six time steps (TIME_STEPS = 6) per ONT, corresponding to approximately the preceding six hours given the near-hourly measurement interval. Unlike Random Forest, LSTM directly receives the raw sequential observations and utilizes its internal gating mechanism to automatically learn both short-term and long-term temporal dependencies without requiring manual feature engineering. The dataset was partitioned using a time-series split, preserving the chronological order of observations without random shuffling. The LSTM architecture consisted of one LSTM layer with 64 hidden units, followed by two dropout layers, one dense layer with 32 neurons, and an output layer containing three neurons with Softmax activation, resulting in a total of 19,843 trainable parameters. Both LSTM and dense layers were regularized using an L2 penalty ($\lambda = 0.001$) on the kernel and recurrent weights to reduce overfitting, and were followed by dropout layers with rates of 0.3 and 0.2, respectively. The model was compiled with the Adam optimizer [18] using a learning rate of 0.0005 and categorical cross-entropy loss, and was trained with a batch size of 64 for a maximum of 100 epochs. Training employed early stopping (monitoring validation loss, patience of 25 epochs, restoring the best weights) and a learning-rate reduction schedule (ReduceLROnPlateau, factor of 0.5, patience of 8 epochs, minimum learning rate of 1×10^{-6}) to stabilize convergence, consistent with established loss-function and optimization practices for deep learning models [18], [19].

The LSTM architecture is governed by three primary gating mechanisms, namely the forget gate, input gate, and output gate, which regulate the information to be discarded, retained, and propagated as predictions, respectively [17], [29], [30]. The sigmoid and hyperbolic tangent functions used in these gates are further discussed by Zhou and Chen [21], while the optimization of network weights during training follows backpropagation-based gradient descent, as outlined by Rahman and Hossain [18]. These operations are mathematically defined as follows:

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right) \tag{1}$$

$$\begin{aligned} i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_c[h_{t-1}, x_t] + b_c) \end{aligned} \tag{2}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{3}$$

$$\begin{aligned} o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\ h_t &= o_t \odot \tanh(C_t) \end{aligned} \tag{4}$$

Where x_t denotes the input vector at time step t, h_{t-1} is the hidden state from the previous time step, h_t is the hidden state at time step t, and $C_t$represents the cell state that stores long-term information. The forget gate (f_t) determines which information from the previous cell state should be discarded, the input gate (i_t) controls which new information should be stored, and $\tilde{C}_t$denotes the candidate cell state generated from the current input and previous hidden state. The output gate (o_t) determines the information propagated to the next hidden state. The function $\sigma(\cdot)$denotes the sigmoid activation function, $\tanh(\cdot)$represents the hyperbolic tangent activation function, $\odot$denotes element-wise multiplication, while $W$and $b$are the weight matrices and bias vectors associated with each gate, respectively.

III. RESULT AND DISCUSSION

A. Dataset Description and Distribution

The dataset used in this study consists of historical Fiber-to-the-Home (FTTH) network data containing optical power measurements collected from Optical Network Terminals (ONTs). The dataset includes the variables timestamp, ont_id, optical_power, fiber_distance,number_of_splices, number_of_connectors, device_age, temperature, humidity, and the corresponding network condition label.

A total of 63,145 records were collected from 58 ONTs during the observation period from March 1 to May 31, 2026. Based on the class distribution, the Normal condition accounted for 50,670 records (80.24%), followed by the Warning condition with 8,873 records (14.05%), and the Critical condition with 3,602 records (5.70%), as illustrated in Figure 3. Measurements were recorded at a near-hourly interval per ONT, with a median inter-measurement interval of approximately 60–68 minutes, although the interval varied slightly across devices due to differences in the monitoring system's polling schedule. The number of records per ONT ranged from 300 to 1,704 (mean ≈ 1,089), reflecting differences in the operational duration and connectivity uptime of individual ONTs during the observation period. The Normal, Warning, and Critical labels were assigned based on optical power thresholds consistent with the receiver sensitivity budget defined in the ITU-T G.984.2 GPON standard and internal operational guidelines used by the network operator [22], [23], [24]: optical power readings above approximately −24 dBm were categorized as Normal, readings between approximately −24 dBm and −27 dBm as Warning, and readings below approximately −27 dBm as Critical, reflecting progressively higher risk of service-affecting signal loss.

The class distribution indicates a substantial class imbalance, with the Normal class representing the majority of observations. Nevertheless, the Warning and Critical classes remain crucial, as they represent the early and severe stages of optical signal degradation. These minority classes provide essential information for developing reliable predictive models capable of supporting proactive network maintenance and early fault detection.

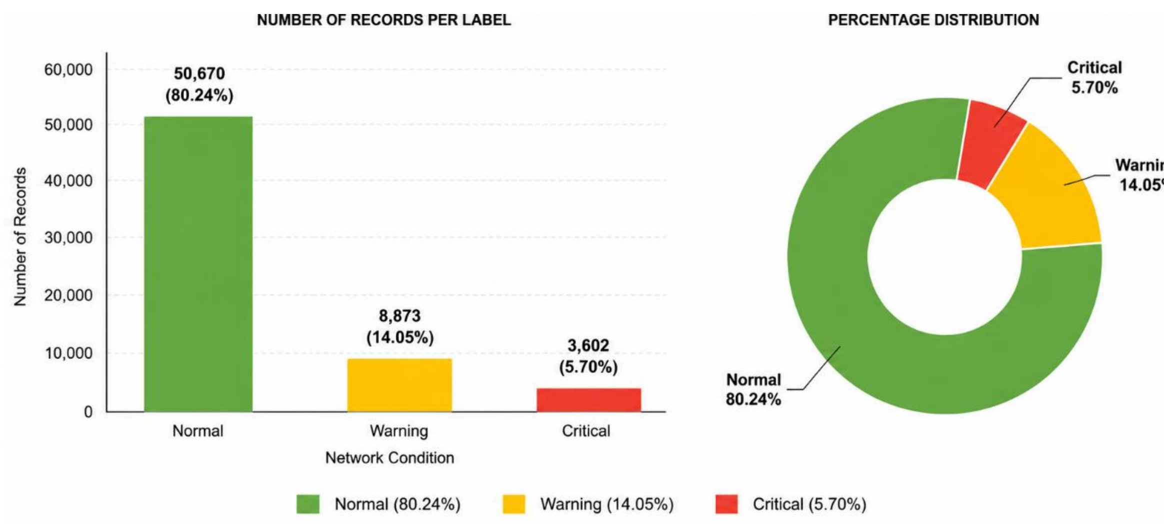

Figure 3. FTTH Network Condition Label Distribution

C. Model Training Results

The Random Forest model was trained using 50,376 training data and tested on 12,595 data, with 200 decision trees (n_estimators = 200), max_features = sqrt, and class_weight = balanced. The LSTM model was trained using 50,214 training sequences and tested on 12,583 sequences, with class weights of 0.415 (Normal), 2.384 (Alert), and 5.873 (Critical) to address label imbalance, as summarized in Table II.

TABLE II
IMPLEMENTATION CONFIGURATION OF RANDOM FOREST AND LSTM MODELS

Component	**Random Forest**	**LSTM**
Input Features	11 engineered tabular features	4 dynamic features, input shape (6, 4)
Training / Testing Data	50,376 / 12,595 records	50,214 / 12,583 sequences
Main Hyperparameters	*n_estimators* = 200, *max_features* = sqrt, *class_weight* = balanced	64 LSTM units, 32-neuron dense layer, 19,843 trainable parameters
Training Time	2.667 s	189.98 s (best model at epoch 36)
Inference Time	0.0102 ms per record	0.1856 ms per record

The feature importance analysis in Figure 4 shows that historical optical_power features, especially op_lag_1, have the most dominant influence on degradation predictions (31.07%), followed by op_roll_mean (22.93%), op_lag_2 (18.88%), and op_lag_3 (13.09%), while physical features like cable_length, number_of_connectors, number_of_splices, and device_age contribute less. This suggests that the condition of optical power in previous periods plays a big role in determining the current state of the network.

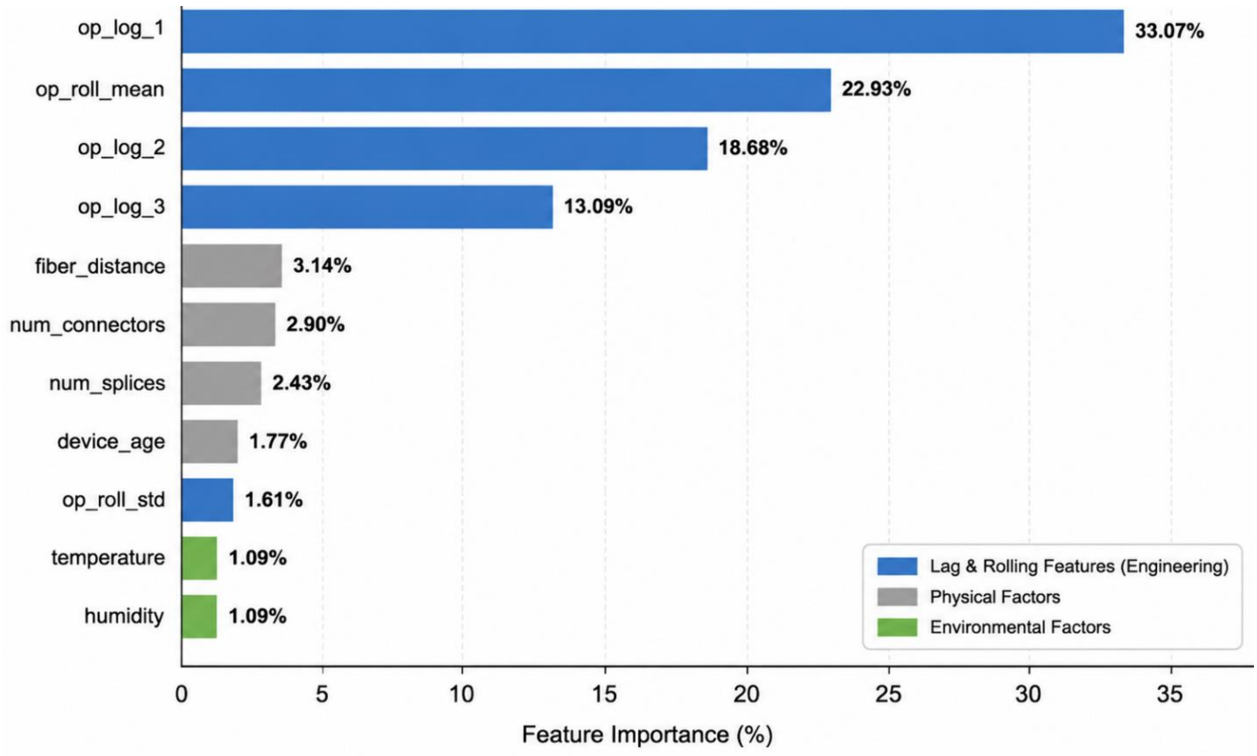

Figure 4. Feature Importance Model Random Forest

The LSTM learning curves shown in Figure 5 indicate that the training loss decreases steadily throughout the training process, whereas the validation loss remains consistently higher and exhibits noticeable fluctuations. Similarly, the training accuracy reaches a relatively high level, while the validation accuracy remains considerably lower. This performance gap suggests that the model learns the patterns in the training data effectively but has limited ability to generalize to unseen validation data, indicating a tendency toward overfitting. It is worth noting that the internal validation accuracy shown in Figure 5 (approximately 35–37%) is notably lower than the accuracy obtained on the independent test set (70.08%). This discrepancy is attributable to the internal validation split being drawn as a trailing 15% slice of the per-ONT-concatenated training sequences, which does not sample uniformly across all ONTs; consequently, the internal validation subset likely over-represents a small number of ONTs with atypical label distributions, whereas the final test set, obtained through stratified sampling across all ONTs, provides a more representative and less biased performance estimate. This finding suggests that the modest reported test performance of LSTM reflects a combination of limited feature availability and this internal validation artifact rather than solely an inherent limitation of the sequential modeling approach, and that adopting a stratified or per-ONT-balanced validation split is recommended for future refinement of the LSTM training procedure.

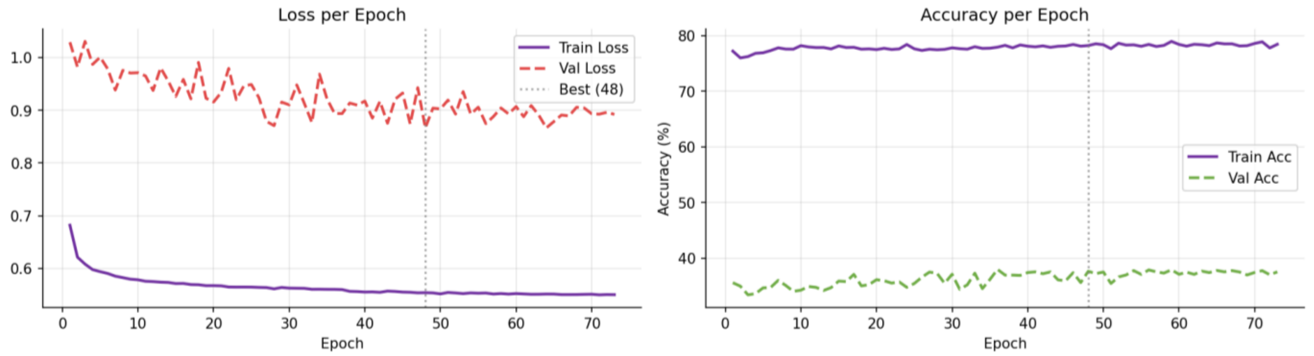

Figure 5. Loss and Accuracy Curve per Epoch in LSTM Model Training

D. Model Performance Evaluation and Comparison

The performance comparison between the Random Forest and Long Short-Term Memory (LSTM) models, presented in Table III and Figure 6, demonstrates that Random Forest consistently outperformed LSTM across all evaluation metrics. Random Forest achieved an accuracy, precision, recall, and F1-score of 98.41%, with a Mean Absolute Error (MAE) of 0.0173 and a Root Mean Square Error (RMSE) of 0.1420.

Although this study addresses a categorical prediction task, MAE and RMSE were additionally reported because the three condition classes (Normal, Warning, Critical) were encoded as an ordinal sequence (0, 1, 2) representing increasing severity of degradation. In this context, MAE and RMSE were computed between the ordinal-encoded true and predicted labels, providing a severity-aware error measure that penalizes a prediction error between distant classes (e.g., Normal predicted as Critical) more heavily than a prediction error between adjacent classes (e.g., Normal predicted as Warning), a distinction that standard accuracy or F1-score cannot capture [36], [37], [38]. In contrast, the LSTM model achieved an accuracy of 70.08%, precision of 66.33%, recall of 70.08%, and an F1-score of 67.46%, accompanied by a MAE of 0.4565 and an RMSE of 0.8781. Furthermore, LSTM required substantially longer training and inference times than Random Forest, indicating a significantly higher computational cost while delivering lower predictive performance on the FTTH optical power degradation dataset.

TABLE III
COMPARISON OF EVALUATION RESULTS OF RANDOM FOREST AND LSTM MODELS

Metrik Evaluasi	Random Forest	LSTM
Accuracy (%)	98,41	70,08
Precision (%)	98,41	66,33
Recall (%)	98,41	70,08
F1-Score (%)	98,41	67,46
MAE	0,0173	0,4565
RMSE	0,1420	0,8781
Train time (s)	2,667	189,98
Inference (ms)	0,0102	0,1856

It should be noted that the identical value (98.41%) reported for Random Forest's accuracy, precision, recall, and F1-score in Table III reflects weighted-average metrics; by definition, the weighted-average recall is mathematically equivalent to overall accuracy, while the weighted-average precision and F1-score happened to round to the same value given the strong dominance of the Normal class. To verify that this figure does not conceal poor performance on the minority classes, per-class metrics were additionally computed: Random Forest achieved a precision and recall above 99% for the Normal class, 94.19% precision and 95.42% recall for the Warning class, and 93.67% precision and 90.79% recall for the Critical class, corresponding to a macro-averaged precision, recall, and F1-score of 95.77%, 95.22%, and 95.49%, respectively. This confirms that Random Forest retained strong discriminative capability even for the minority Critical class, despite the overall class imbalance. Furthermore, because the models were trained using a stratified random 80:20 split, an additional evaluation was performed using a strict chronological (time-based) split, in which the most recent 20% of observations, ordered by timestamp, were held out for testing rather than being randomly sampled. Under this stricter, leakage-resistant validation scheme, Random Forest achieved an accuracy of 98.59%, comparable to or slightly higher than under the random split, indicating that its performance is driven by genuine predictive signal in the lag and rolling-window features rather than by information leakage across the train-test boundary. The superior performance of Random Forest can be attributed to its ability to exploit both the original technical variables and the engineered temporal features, including op_lag_1, op_lag_2, op_lag_3, op_roll_mean, and op_roll_std, which explicitly capture the historical behavior of optical power. These engineered features enable the model to incorporate temporal information while maintaining a tabular data representation, thereby improving its predictive capability.

LSTM relies exclusively on dynamic features represented as sequential data, making its performance highly dependent on the consistency of temporal patterns and the availability of sufficiently large sequential datasets for effective learning. Furthermore, Random Forest exhibited substantially shorter training and inference times because its tree-based ensemble architecture is computationally less complex and does not require iterative gradient-based optimization, as is the case with deep neural networks such as LSTM.

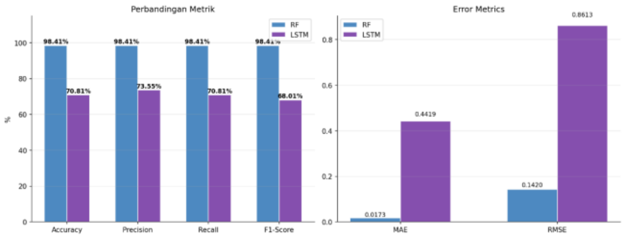

Figure 6. Comparison of Evaluation Metrics for Random Forest and LSTM

The confusion matrix presented in Figure 7 further supports these findings, consistent with confusion-matrix-based evaluation practices commonly used for multi-class prediction problems [33], [34]. Random Forest correctly predicted 10,055 Normal, 1,690 Warning, and 650 Critical instances, with relatively few prediction errors across all classes, including the minority classes. These results demonstrate the model's ability to maintain balanced predictive performance despite the presence of class imbalance.

The LSTM model correctly predicted only 6 Warning instances, while 1,665 Normal instances were incorrectly predicted as Critical. This substantial prediction error indicates that the sequential model struggled to distinguish the transitional Warning class from the Normal and Critical classes. Consequently, the LSTM model exhibited limited discriminative capability for intermediate degradation states, which contributed to its lower overall prediction performance compared with Random Forest.

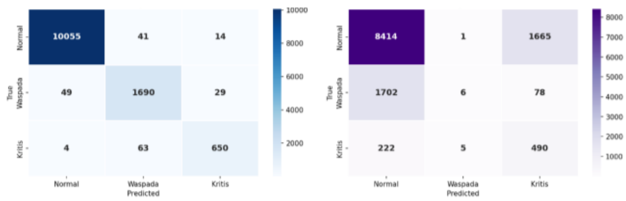

Figure 7. Comparison of Random Forest and LSTM Confusion Matrices

E. Discussion

The findings of this study are consistent with those reported by Rashed et al. [14] and Putra et al. [39], who demonstrated the effectiveness of Random Forest in modeling nonlinear relationships within tabular optical network data, achieving excellent predictive performance in terms of R^2, RMSE, and MAE. In the present study, the characteristics of the FTTH dataset are influenced not only by temporal variations in optical power but also by static technical factors, such as fiber distance, number of splices, number of connectors, and device age. Consequently, these data are more appropriately represented as tabular data enriched with engineered temporal features rather than as purely sequential data. This explains why Random Forest, augmented with lag features and rolling statistical features, was able to capture optical power degradation patterns more effectively than the sequential LSTM model.

The relatively lower performance of LSTM observed in this study differs from the findings reported by Soothar et al. [40], Zhang et al. [41], and Kumar et al. [42], who achieved prediction accuracies ranging from 97% to 99% using LSTM-based and hybrid CNN–LSTM models. Several factors may account for this discrepancy. First, the present study employed a more limited feature set, consisting of only four dynamic variables without incorporating static physical features. Second, the historical dataset covered a relatively short observation period of only three months, which may have limited the model's ability to learn long-term temporal dependencies. Third, the selected six-step (approximately six-hour) time window may not have been optimal for capturing the transitional characteristics of the Warning class. As suggested by Li et al. [43], hybrid approaches combining Random Forest and LSTM have the potential to integrate the strengths of tabular feature learning and sequential pattern modeling, making them a promising direction for future research.

From a practical perspective, the superior prediction performance and computational efficiency of Random Forest make it a more suitable candidate for deployment in a real-time early warning system. The model required only 2.667 seconds for training and 0.0102 ms per record for inference, substantially outperforming LSTM in computational efficiency. The prediction results enable network administrators and maintenance engineers to prioritize maintenance activities according to the predicted network condition. ONTs predicted as Warning can be subjected to closer monitoring, while those predicted as Critical can be immediately scheduled for inspection and corrective maintenance. Furthermore, by integrating the prediction model with a WhatsApp-based alert feature, the dashboard automatically flags ONTs predicted as Warning or Critical and prepares the corresponding notification message; the current implementation requires the network administrator to manually confirm and send this message through WhatsApp Web, rather than dispatching it automatically from a dedicated admin number. This semi-automated design still substantially reduces the manual effort of identifying at-risk ONTs, enabling maintenance personnel to take preventive action before service degradation or network failures occur, while leaving message dispatch under administrator supervision.

Despite these promising results, this study has several limitations. First, the dataset was limited to a single operator (PT. Jaringanku Nusantara), 58 ONTs, and a three-month observation period (March–May 2026); the monthly accuracy trend of Random Forest declined slightly over this period (98.69% in March, 98.20% in April, and 98.37% in May), while LSTM declined more markedly (75.46%, 71.61%, and 64.96%, respectively), suggesting the possible onset of concept drift and indicating that the current dataset may not fully capture longer-term or seasonal variations in network degradation across a wider range of operating conditions. Second, this study did not incorporate additional operational information such as network fault ticket histories, detailed physical cable conditions, or technician maintenance records. Third, the current WhatsApp-based alert feature is semi-automated: the dashboard automatically identifies ONTs predicted as Warning or Critical and drafts the corresponding alert message, but dispatch still requires the network administrator to manually confirm and send the message through WhatsApp Web, rather than fully automated delivery from a dedicated admin number. Because message dispatch depends on manual confirmation, no automated delivery log was available, and a systematic quantitative evaluation of notification delivery success rate and response latency was therefore not conducted within the scope of this study; future work should implement fully automated dispatch (e.g., via an official WhatsApp Business API) to enable such an evaluation. Future research should consider extending the observation period, validating the framework across multiple ONT populations and network operators, integrating maintenance and customer fault data, investigating hybrid predictive models such as Random Forest–LSTM or CNN–LSTM, exploring adaptive threshold-tuning strategies inspired by reinforcement learning [20] for dynamically adjusting Warning and Critical alert boundaries, and implementing the proposed prediction framework within a real-time operational network monitoring system.

IV. CONCLUSION

This study compared the performance of Random Forest and Long Short-Term Memory (LSTM) for predicting optical power degradation in Fiber-to-the-Home (FTTH) networks using 63,145 historical records collected from 58 Optical Network Terminals (ONTs). The experimental results demonstrate that Random Forest consistently outperformed LSTM across all evaluation metrics, achieving an accuracy, precision, recall, and F1-score of 98.41%, with a Mean Absolute Error (MAE) of 0.0173 and a Root Mean Square Error (RMSE) of 0.1420. In contrast, LSTM achieved an accuracy of 70.08%, with an MAE of 0.4565 and an RMSE of 0.8781, while also requiring significantly longer training and inference times. These findings indicate that Random Forest is more suitable for predicting optical power degradation in FTTH networks, primarily because its combination of engineered temporal features and technical variables effectively captures the characteristics of operational FTTH data.

This study contributes by proposing a fair comparative framework between tabular and sequential learning models through equivalent temporal preprocessing, identifying historical optical power features as the most influential predictors of network degradation, and integrating the Random Forest model into a web-based dashboard with a WhatsApp-based early warning system to support preventive maintenance. Nevertheless, the study is limited by its single-operator, 58-ONT, three-month dataset, the absence of detailed maintenance and fault history data, and the lack of a quantitative evaluation of WhatsApp notification delivery performance. Robustness checks using a strict chronological train-test split and per-class evaluation confirmed that Random Forest's performance was not an artifact of random data splitting or class imbalance, though longer-term, multi-operator validation remains necessary. Future work should utilize longer-term, multi-site historical datasets, investigate hybrid approaches such as Random Forest–LSTM or CNN–LSTM, quantitatively evaluate the early warning notification system in operational deployment, and integrate the proposed framework with real-time network monitoring systems to enable automated early fault detection and more responsive maintenance operations.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to PT. Jaringanku Nusantara for providing access to the historical FTTH network dataset and the technical information that supported this research.

REFERENCES

[1] H. Hermansyah, A. Khaidar, N. Nurdin, and S. Kurnia, “Implementation of Static Routing and Quality of Service for Optimization of Network Traffic Management on Cisco Routers,” J. Artif. Intell. Softw. Eng., vol. 5, no. 3, pp. 1222–1230, 2025.

[2] Y. Chen and M. Xu, “Predictive analytics for optical network degradation using ensemble learning,” IEEE Trans. Netw. Manag., vol. 33, no. 1, pp. 1–10, 2025.

[3] D. Prasetyo and A. Wibowo, “Prediksi kualitas sinyal optik menggunakan LSTM pada jaringan FTTH,” J. Telekomun. dan Komput., vol. 14, no. 2, pp. 101–109, 2024.

[4] N. Ahmad and A. Rahman, “Hybrid Random Forest and LSTM model for network performance prediction,” Int. J. Intell. Eng. Syst., vol. 17, no. 1, pp. 210–220, 2024.

[5] S. Liu and K. Zhang, “FTTH network challenges in supporting IoT and cloud services,” Comput. Networks, vol. 243, pp. 110–118, 2024.

[6] R. Singh and P. Kumar, “Deep learning approaches for predictive maintenance in optical networks,” IEEE Access, vol. 11, pp. 78900–78912, 2023.

[7] A. Putra and D. Nugroho, “Penerapan machine learning untuk prediksi gangguan jaringan optik,” J. RESTI, vol. 6, no. 3, pp. 450–458, 2022.

[8] L. Zhou and H. Chen, “Optical power prediction in FTTH networks using LSTM,” Opt. Switch. Netw., vol. 48, pp. 100–108, 2023.

[9] R. Hakim and F. Siregar, “Pendekatan pemeliharaan preventif jaringan FTTH berbasis data historis,” J. Inform., vol. 17, no. 1, pp. 55–63, 2023.

[10] X. Li and Y. Zhao, "Random Forest-based classification for network performance degradation," Appl. Soft Comput., vol. 120, pp. 108–115, 2022.

[11] P. Almeida and L. Costa, "FTTH network monitoring and fault detection using machine learning," IEEE Commun. Mag., vol. 59, no. 2, pp. 40–46, 2021.

[12] H. Wang and Q. Liu, "Analysis of optical signal attenuation in FTTH networks," Optik (Stuttg)., vol. 242, pp. 167–174, 2021.

[13] B. Santoso and R. Pratama, "Analisis redaman jaringan FTTH berbasis serat optik," J. Teknol. Inf. dan Komun., vol. 11, no. 2, pp. 85–92, 2022.

[14] M. Rashed, S. Ahmed, and M. Hossain, "Optical network performance analysis using Random Forest algorithm," Opt. Fiber Technol., vol. 56, pp. 102–110, 2020.

[15] Y. Zhang and X. Wang, "LSTM-based network traffic prediction for optical networks," IEEE Access, vol. 6, pp. 35612–35621, 2018.

[16] J. Kim and S. Cho, "Time-series performance prediction in communication networks using LSTM," J. Netw. Comput. Appl., vol. 129, pp. 12–21, 2019.

[17] D. Prasetyo and E. Santoso, "Implementasi LSTM untuk prediksi data deret waktu," J. Ilmu Komput., vol. 19, no. 1, pp. 44–52, 2024.

[18] A. Rahman and M. Hossain, "Backpropagation optimization techniques in deep learning," Appl. Soft Comput., vol. 118, pp. 108–116, 2022.

[19] P. Singh and R. Kumar, "Loss functions in deep learning for regression problems," Expert Syst. Appl., vol. 210, pp. 118–125, 2023.

[20] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction and recent developments," IEEE Trans. Neural Networks, vol. 31, no. 11, pp. 4485–4496, 2020.

[21] Y. Zhou and K. Chen, "Activation functions in deep neural networks: A review," Neural Networks, vol. 158, pp. 95–110, 2023.

[22] A. Pratama and Y. Nugroho, "Analisis kualitas layanan jaringan FTTH berbasis Passive Optical Network," J. Teknol. Telekomun., vol. 9, no. 1, pp. 15–24, 2024.

[23] M. Rahman and S. Hossain, "Performance evaluation of FTTH networks using GPON architecture," Int. J. Electr. Comput. Eng., vol. 13, no. 4, pp. 3890–3898, 2023.

[24] D. Tsonev and H. Haas, "Fiber-to-the-Home networks for high-speed broadband access," IEEE Commun. Mag., vol. 59, no. 3, pp. 62–68, 2021.

[25] J. Gao and Y. Liu, "Prediction and the influencing factor study of colorectal cancer hospitalization costs in China based on machine learning-random forest and support vector regression: a retrospective study," Front Public Heal., vol. 12, no. 1211220, 2024, doi: 10.3389/fpubh.2024.1211220.

[26] Y. Zhou, H. Shen, and M. Zhang, "A Distributed and Privacy-Preserving Random Forest Evaluation Scheme with Fine Grained Access Control," Symmetry (Basel)., vol. 14, no. 2, p. 415, 2022, doi: 10.3390/sym14020415.

[27] E. R. B. Sebayang, Y. H. Chrisnanto, and M. Melina, "Klasifikasi Data Kesehatan Mental di Industri Teknologi Menggunakan Algoritma Random Forest," IJESPG (International J. Eng. Econ. Soc. Polit. Gov., vol. 1, no. 3, pp. 237–253, 2023.

[28] H. Blockeel, "Decision trees: from efficient prediction to responsible AI," Front. Artif. Intell., 2023.

[29] R. A. Sembiring and Y. Sary, "Analisa dan Implementasi Long Short-Term Memory (LSTM) dalam Kebutuhan Persediaan Barang di PT. Gunung Sari Indonesia," J. Ilm. Tek. Mesin, Elektro dan Komput., vol. 5, no. 3, pp. 97–105, 2025, doi: 10.51903/juritek.v5i3.5578.

[30] M. A. Saputra and T. Sugihartono, "Evaluasi Kinerja Model LSTM untuk Prediksi Risiko Penyakit Jantung Menggunakan Dataset," J. Pendidik. dan Teknol. Indones., vol. 5, no. 7, pp. 1823–1833, 2025, doi: 10.52436/1.jpti.821.

[31] M. Tshamaroh, N. S. Nasution, N. Nadhirah, R. A. Alfira, and Z. Xintong, "Amazon Stock Price Prediction Using Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)," Public Res. J. Eng. Data Technol. Comput. Sci., vol. 3, no. 1, pp. 1–9, 2025, doi: 10.57152/predatecs.v3i1.1656.

[32] Q. Yu, B. A. Tolson, H. Shen, M. Han, J. Mai, and J. Lin, "Enhancing Long Short-Term Memory (LSTM)-based Streamflow Prediction with a Spatially Distributed Approach," Hydrol. Earth Syst. Sci., vol. 28, pp. 2107–2122, 2024, doi: 10.5194/hess-28-2107-2024.

[33] M. R. Humaidi and A. Maulani, "Klasifikasi Naïve Bayes dan Confusion Matrix pada Pengguna Aplikasi E-Commerce di Play Store," J. Ilm. Inform., vol. 8, no. 2, pp. 132–139, 2023.

[34] N. Hidayah and D. Dodiman, "Implementasi algoritma multinomial naïve bayes, TF-IDF dan confusion matrix dalam pengklasifikasian saran monitoring dan evaluasi mahasiswa terhadap dosen teknik informatika Universitas Dayanu Ikhsanuddin," in Jurnal Akademik Pendidikan Matematika, 2024, pp. 8–15.

[35] N. Nurdin, "Analisa Data Mining Dalam Memprediksi Masyarakat Kurang Mampu Menggunakan Metode K-Nearest Neighbor," J. Inform. Dan Tek. Elektro Terap., vol. 12, no. 2, 2025.

[36] A. Botchkarev, "Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology," Interdiscip. J. Information, Knowledge, Manag., vol. 16, pp. 1–18, 2021.

[37] T. O. Hodson, "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Which Metric," Geosci. Model Dev., vol. 15, no. 14, pp. 5481–5493, 2022.

[38] C. C. Aggarwal, "Evaluation Metrics for Regression Models in Machine Learning," ACM Comput. Surv., vol. 55, no. 6, pp. 1–36, 2023.

[39] A. Putra et al., "Prediction of Optical Fiber Network Attenuation to Assess Network Performance Using Random Forest Regression," Int. J. Comput. Sci., vol. 20, no. 3, pp. 455–463, 2025.

[40] M. Soothar et al., "Hybrid CNN-LSTM Model for Fault Detection and Performance Analysis in Optical Fiber Networks," Opt. Fiber Technol., vol. 78, pp. 103–112, 2024.

[41] Y. Zhang et al., "Prediction of Optical Fiber Cable Degradation Based on Bi-LSTM and Attention Mechanism," Sensors, vol. 24, no. 12, p. 4012, 2024.

[42] R. Kumar et al., "LSTM-Based Model for Reliable Optical Transmission in Flexible Optical Networks," Opt. Switch. Netw., vol. 47, pp. 100–109, 2023.

[43] X. Li et al., "AI-Driven Predictive Maintenance for DWDM Optical Fiber Networks Using Random Forest and LSTM," Opt. Commun., vol. 540, pp. 129–138, 2025.