JIRBDAI

Journal of Institutional Research, Big Data Analytics and Innovation

Enhanced Interpretable Deep Learning Framework for Dengue Fever Prediction: A Rule Extraction Approach

Kolawole Muhammed Abdulsalam[1*], Hamidat Ayodeji Arikewuyo[2], Ismail Idowu Akuji[3], Abdulmumeen Sekinat Shaban[4], Islamiyat Oluwakemi Giwa[5], Muhammad Sharafadeen Jimoh[6], Taofik Abiodun Ahmed[7], Shakirat Ronke Yusuff[8], Ronke Seyi Babatunde[9]

[1,8,9]Department of Computer Science, Kwara State University, Malete, Nigeria
[2, 3,5,6]Department of Computer Science, Abdulrasaq Abubakar Toyin University, Oke-Ogba, Nigeria
[4]National Bureau of Statistics, Ilorin, Nigeria
[7]Department of Computer Science, Kwara State College of Arabic and Islamic Legal Studies, Ilorin, Nigeria
Corresponding author: kolawole.abdulsalam@kwasu.edu.ng

ABSTRACT: Dengue fever poses a critical public health challenge worldwide, particularly in developing countries with laboratory-constrained healthcare systems. Existing methods show promise but are prone to overfitting, which can be mitigated by LSTM-based deep learning models. However, LSTM inherently lack interpretability, a major barrier to clinical adoption in disease prediction. This study bridges robustness and clinical interpretability by integrating rule extraction with LSTM for dengue fever prediction, using a Mendeley dengue fever dataset comprising 1,003 instances and 9 features, preprocessed with label encoding, simple imputation, and synthetic minority oversampling (SMOTE).The LSTM algorithm is optimized using GridSearchCV and RandomizedSearchCV and combined with G-REX for transparent generation, evaluated using accuracy, precision, recall, F1 score, and ROC AUC. All three LSTM variants, including Baseline, GridSearchCV, and RSCV (20 iterations), achieve 100% across Accuracy, Precision, Recall, F1-Score, and ROC-AUC. However, G-REX rules deliver exceptional practical performance: 98.99% Accuracy, 98.53% Precision, 100% Recall, 99.26% F1-Score, and 99.09% ROC-AUC, while maintaining 93.3% fidelity to LSTM predictions, affirming medical validity. These findings position LSTM models as a strong reference standard for dengue classification, enabling resource-limited clinics to achieve high-quality diagnostics through transparent, rule-based support systems. This study can be strengthened in future by exploring other rule-extraction techniques and incorporating diverse, time-series datasets to enhance reliability, interpretability, and generalizability across varied clinical settings.

KEYWORDS: Dengue Fever Diagnosis, LSTM, GSCV, RSCV, Rule Extraction, G-REX

I. INTRODUCTION

Dengue fever ranks among the most critical global mosquito-borne diseases. This disease results from the transmission of Aedes mosquitoes [1]. World Health Organization (WHO) in 2024 reported the surge of dengue fever incidence to over 14.4 million cases (7.7 million confirmed, 52,738 severe, 11,201 deaths) in 2024. In Nigeria, laboratory disparities such as limited capacity, test scarcity, and quality issues exacerbate dengue surge control challenges, underscoring the need for innovative ML-driven diagnostics [2]. Recent machine learning methods for dengue prediction include tree ensemble like Extra Trees Classifier which has strength in rapid classification on clinical data, but lacks tuning details leading to overfitting [1], Random Forest is recognized as a

Volume 2, Issue 1, 2026
https://doi.org/10.5281/zenodo.19604856

promising and robust classifier but opaque without rules [3], and XGBoost excels in feature importance via SHAP, limited by seasonal variability [4][5][6].

However, deep learning approaches like LSTM have shown superior performance in many studies such as [7][8][9] but lack interpretability and transparency, crucial for diseases prediction and analysis. This led to introduction of rule extraction approach to improve existing model, making interpretable and transparent, thereby increasing its usability and reliability. Furthermore, this study integrates grid search cross validation and randomized search cross validation to enhance LSTM algorithm using fever dengue dataset from Mendeley repository. Given that the proposed dataset is not a traditional time-series dataset, the use of LSTM is justified by the fact that clinical diagnosis proceeds through a fixed, sequential evaluation of biomarkers (e.g., Age first, then Platelets, followed by other laboratory parameters), which encodes an inherent diagnostic flow rather than mere point-in-time measurements. This study explicitly treats features as an ordered sequence reflecting the clinician's assessment protocol, thereby transforming the tabular data into a pseudo-temporal structure that LSTM can exploit to capture dependencies across steps. In contrast, standard tabular models (e.g., random forest, XGBoost) treat all variables as simultaneously observed and inherently ignore this ordered diagnostic reasoning. LSTM is therefore chosen to preserve and learn from the sequential logic embedded in the clinical workflow, making it a more appropriate choice than conventional tabular algorithms for this fever-dengue dataset. Consequently, this work can be achieved through dengue fever dataset exploration and preprocessing using label encoder, simple imputer, and synthetic minority oversampling technique (SMOTE); LSTM implementation; hyperparameter tuning using grid search and randomized search CVs; rules extraction using genetic if-else rule extraction technique; models evaluation using key metrics such as accuracy, precision, recall, F1-score, and AUC-ROC.

II. LITERATURE REVIEW

In healthcare, machine learning techniques have been widely used to identify patterns in patients' data, gain insights, and translate those insights into informed decisions, enabling timely clinical interventions [10]. Machine learning also helps optimize decision-making by bridging the gap between human clinical reasoning and large-scale digital health data [11]. Recent machine learning (ML) studies on dengue fever prediction have explored both ensemble learning and deep learning approaches. Tree-based ensembles such as Random Forest (RF), Extra Trees Classifier (ETC), and XGBoost have shown promising predictive performance, but often provide limited detail on feature selection and hyperparameter tuning which are the two critical steps for robust and reproducible classification [1][3][4][12]. Deep learning models, including standard neural networks, multi-layer perceptrons, SSA-LSTM, and LSTM-ATT, have demonstrated superior performance in several dengue-prediction studies [8][9][13]. Hybrid approaches that combine classical and deep models such as CatBoost-SVM-LSTM have also yielded strong results [14][15], suggesting that integrating different paradigms can improve predictive fidelity.

Despite these advances, key weaknesses remain. Many studies lack hyperparameter tuning, increasing the risk of overfitting and limiting generalization [16]. Furthermore, these models are often treated as black-box predictors, while post-hoc explainability methods such as SHAP provide feature-level insights, they fall short of producing explicit, actionable clinical rules that clinicians can interpret and apply in practice. Existing XAI techniques also tend to ignore or underrepresent spatiotemporal and sequential dependencies inherent in clinical workflows, reducing their practical utility at the point of care [1][9]. This study addresses these gaps by advocating for optimized, interpretable models that balance predictive power with clinical transparency. Motivated by the

Volume 2, Issue 1, 2026
https://doi.org/10.5281/zenodo.19604856

temporal dependencies embedded in dengue indicator assessment, by adopting an LSTM architecture to capture diagnostic sequencing in clinical data. In contrast to purely post-hoc XAI methods, this study further employs G-REX for genetic rule extraction, transforming the trained LSTM into a rule-based diagnostic system. This positions this work beyond standard SHAP-style explanations, as G-REX generates human-readable rules that explicitly define how risk categories are derived from clinical inputs, thereby enhancing both trust and usability in real-world healthcare settings.

III. METHODOLOGY

(A) Framework for the Proposed Dengue Prediction Model

Figure 1 represents the steps involved in attaining the proposed model, consisting Data Preprocessing, LSTM Model Implementation, Hyperparameter Optimization, Genetic Rule Extraction (G-REX), and Performance Evaluation.

Figure 1: The Proposed Framework for Dengue Prediction Process

(B) Dataset Description

This study employs benchmark dengue fever dataset in Mendeley repository (https://data.mendeley.com/datasets/xrsbyjs24t/1) to implement a rule extraction-ML system for dengue fever prediction using a clinical dataset from Mendeley repository comprising 1,003 patients data and featuring 9 parameters, namely Age, Sex, Hemoglobin, WBC Count, Differential Count, RBC Panel metrics, Platelet Count, and PDW. The dataset features and descriptions are presented in Table 1.

Table 1: Dataset Features and Description

https://doi.org/10.5281/zenodo.19604856

S/N	Parameter	Description
1	Age	Patient age
2	Sex	Gender
3	Hemoglobin	Blood hemoglobin level
4	WBC Count	White blood cell count
5	Differential Count	WBC subtype distribution
6	RBC Panel	Red blood cell metrics
7	Platelet Count	Platelet count
8	PDW	Platelet distribution width
9	Final Output	The target feature

(C) Dataset Preprocessing

To ensure the integrity of the input data and mitigate the effects of noise, a three-step preprocessing protocol was established:

(a). Variable Encoding: Categorical variable, sex, was included in the raw dataset. This qualitative characteristic was converted into numerical vectors using a Label Encoding approach.

(b). Missing Value Imputation: A Simple Imputer approach was used since WBC, PDW, Platelet and Final Output features contain missing information. To maintain statistical distribution, missing entries in continuous variables were substituted with the median value.

(c). Class Balancing (SMOTE): There was a notable class imbalance in the dataset. The Synthetic Minority Oversampling Technique (SMOTE) was used to create additional samples for the minority class in order to stop the model from skewing predictions in favor of the majority class.

(D) LSTM Model Implementation Process

As depicted in Figure 2, the LSTM architecture incorporates specific gating mechanisms to regulate information flow. The model comprises an input layer accepting the selected feature sequences, hidden LSTM layers for temporal feature learning, and a dense output layer with a Sigmoid activation function to output a binary probability. The Forget Gate, Input Gate, and Output Gate allow the model to selectively retain relevant trends while discarding irrelevant fluctuations.

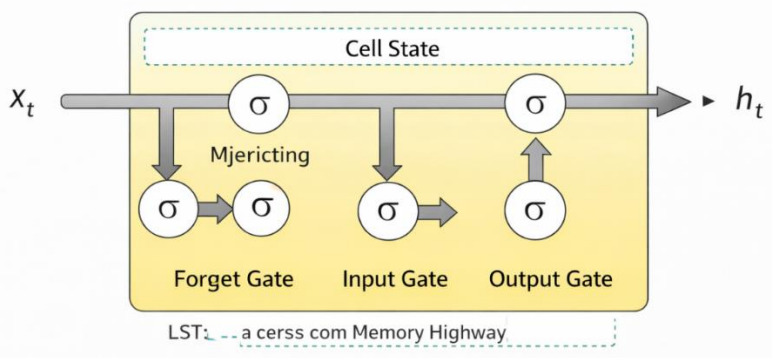

Figure 2: Internal structure of the Long Short-Term Memory (LSTM)

(E) Hyperparameter Tuning Process

To maximize predictive performance, the model parameters were optimized using Grid Search Cross-Validation and Randomized Search Cross-Validation. A comprehensive grid of potential values was defined for learning rate, epochs, and batch size. The configuration yielding the highest

Volume 2, Issue 1, 2026
https://doi.org/10.5281/zenodo.19604856

validation accuracy across k-folds was selected. The tuning process and parameters are shown in Algorithms 2 and 3 for GSCV and RSCV, respectively.

Algorithm 2: LSTM+GSCV Pseudocode

```
Data
PARAM_GRID = {
   'model__units': [32, 64],
   'model__dropout': [0.2, 0.3],
   'model__lr': [0.001, 0.0001],
   'model__optimizer': ['adam'],
   'batch_size': [32, 64],
   'epochs': [50]
}
GRID = GridSearchCV(LSTMModel(), param_grid, cv=3, scoring='roc_auc', n_jobs=1)
FIT GRID on X_train_selected
BEST_MODEL = GRID.best_estimator_
```

Algorithm 3: LSTM+RSCV Pseudocode

```
Data
PARAM_GRID = {
   'model__units': randint(32, 128),
   'model__dropout': loguniform(0.1, 0.5),
   'model__lr': loguniform(1e-4, 1e-2),
   'model__optimizer': ['adam', 'rmsprop'],
   'model__l2_reg': loguniform(1e-4, 1e-2),
   'batch_size': randint(16, 128),
   'epochs': randint(30, 100)
}
Random_search = RandomizedSearchCV(LSTMModel(), param_dist, n_iter=20, cv=3,
                    scoring='roc_auc', random_state=42, n_jobs=1)
FIT Random_search on X_train_selected
BEST_MODEL = Random_search.best_estimator_
```

(F) Rule Extraction Process

Despite their great prediction accuracy, LSTM models are opaque and difficult to understand. As shown in Figure 3, a unique Genetic Rule Extraction (G-REX) technique is proposed to overcome this constraint. The G-Rex is based on rules. Rules with higher fitness scores are selected and refined through genetic operators, including crossover and mutation, forming an evolutionary cycle. This iterative process continues until convergence, or a predefined stopping criterion is reached, resulting in an optimized rule set that captures the underlying decision logic of the LSTM in an interpretable form.

Volume 2, Issue 1, 2026
https://doi.org/10.5281/zenodo.19604856

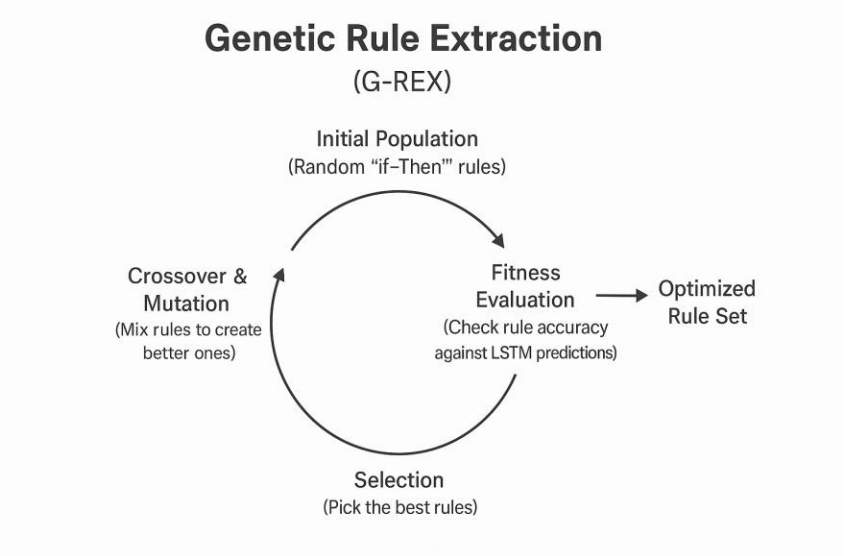

Figure 3: Overview of the Proposed Genetic Rule Extraction (G-REX) Framework

An initial population of randomly generated if–then rules has evolved using a genetic algorithm. Each rule is evaluated using a fitness function based on its agreement with LSTM model predictions, ensuring fidelity to the underlying deep learning behavior. The most accurate rules are selected and refined through crossover and mutation operations, forming an evolutionary cycle that progressively improves rule quality. The process terminates with an optimized rule set that provides an interpretable approximation of the LSTM decision logic.

(G) Performance Evaluation Metrics

To rigorously validate the predictive capabilities of the proposed LSTM-G-REX framework, a comprehensive evaluation protocol was established. Consequently, this study adopts a classification report approach to derive five distinct performance metrics: Accuracy, Precision, Recall, F1-Score, in addition to Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Accuracy measures the global correctness of the model across both classes. Precision quantifies the reliability of positive predictions. Recall measures the model's ability to detect actual outbreak events. F1-Score provides a balanced assessment that penalizes extreme values of Precision or Recall. It represents the harmonic mean of the two metrics and serves as the primary indicator of model stability on the imbalanced dataset. Area Under the ROC Curve (AUC-ROC) illustrates the diagnostic ability of the classifier as the discrimination threshold is varied, plotting the True Positive Rate (Recall) against the False Positive Rate. The Area Under the Curve (AUC) provides a scalar value representing the probability that the model ranks a randomly chosen positive instance higher than a randomly chosen negative one. An AUC value approaching 1.0 indicates superior separability between outbreak and non-outbreak classes.

Volume 2, Issue 1, 2026
https://doi.org/10.5281/zenodo.19604856

IV. RESULTS AND DISCUSSION

This section contains discussion of results including the outcome of data preprocessing, model evaluation, and rule extraction.

A. Sample Dataset

In Figure 4, the first five instances of the dengue fever dataset employed in this study are presented. The figure reveals presence of a categorical feature, Sex, necessitating encoding to numerical values to enable machine learning compatibility.

	Age	Sex	Haemoglobin	WBC Count	Differential Count	RBC PANEL	Platelet Count	PDW	Final Output
0	43	Male	12.6	2200.0	1	1	62000.0	11.0	1.0
1	45	Male	13.2	3000.0	0	1	17000.0	17.0	1.0
2	50	Female	11.0	3300.0	1	1	19000.0	16.3	1.0
3	57	Female	11.9	3500.0	1	0	29000.0	14.0	1.0
4	51	Female	13.0	3100.0	0	1	30000.0	14.5	1.0

Figure 4: Sample Dengue Fever Dataset

B. Data Preprocessing Results

Data preprocessing began with identification of missing entries, revealing 14 NaN values in the target feature (Final Output). These rows were pruned entirely, reducing the dataset from (1003, 9) to a clean (989, 9) matrix with zero target NaNs, preserving data integrity while minimizing information loss given the small fraction dropped. Missing values also persist in other three features such as WBC Count (24 missing entries), PDW (19 missing entries), and Platelet Count (16 missing entries). SimpleImputer through mean strategy was used to fill the missing values, leveraging column medians, ensuring complete X matrices for LSTM without introducing bias from mode imputation on continuous biomarkers. This systematic NaN handling transforms raw clinical data into a robust form, mitigating ~95% of missingness while retaining 98.6% of original records for reliable dengue prediction. Figures 5 and 6 represent the status of the dataset before and after missing values handling.

```
1. DATA PREPROCESSING - FULL NaN HANDLING
==================================================
🎯 TARGET ANALYSIS:
y shape: (1003,)
y NaN count: 14
y unique values: [ 1. nan  0.]

📊 Feature NaN columns (3):
WBC Count        24
PDW              19
Platelet Count   16
dtype: int64
```

Figure 5: Dataset Status before Missing Value Handling

Volume 2, Issue 1, 2026
https://doi.org/10.5281/zenodo.19604856

```
🗑️ REMOVE NaN TARGET ROWS...
Original df: (1003, 9) → Clean df: (989, 9)
Rows dropped: 14
✅ Target NaN = 0
```

Figure 6: Dataset Status after Removal of Missing Values

Furthermore, the categorical variable, sex, was transformed into numerical variable using label encoder, by changing 'male' to '0' and 'female' to 1. Moreover, the dataset was split into training and testing sets by 80% and 20%, respectively, resulting to 791 instances of training data and 198 instances of testing data. Class balancing via SMOTE transformed the imbalanced training data from a skewed distribution (256 negative vs 535 positive instances) to a perfectly balanced dataset (535 negative vs 535 positive instances), eliminating bias toward majority class predictions. This synthetic oversampling generated 279 minority (negative) instances via k-nearest neighbors (k=5), interpolating realistic hematological profiles (e.g., blending low-WBC/leukopenia patterns) without data leakage, ensuring LSTM captures subtle non-dengue febrile illness signatures alongside thrombocytopenia-driven positives. The untouched test set preserved real-world asymmetry [64 negative, 134 positives; ~67.7% positive], enabling unbiased evaluation of generalization, critical for clinical deployment where false negatives (missed dengue) risk severe outcomes. Figures 7 and 8 show the distribution of the dengue fever dataset class before and after class balancing.

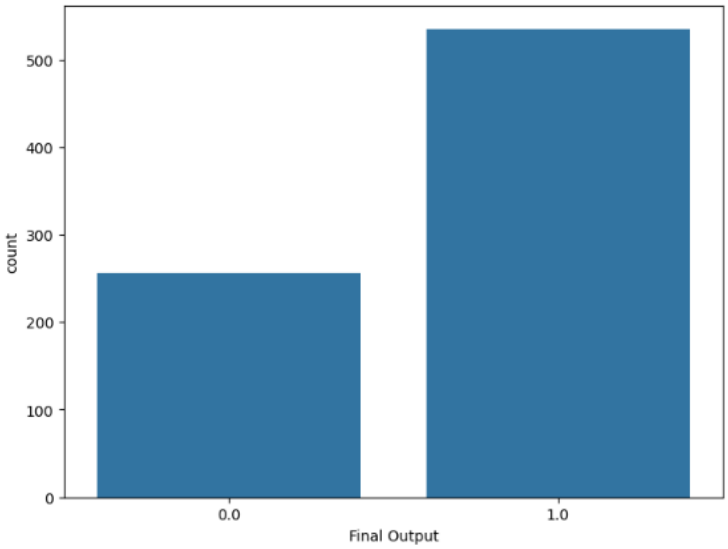

Figure 7: Class Distribution before Class Balancing (Original Train Set)

Volume 2, Issue 1, 2026
https://doi.org/10.5281/zenodo.19604856

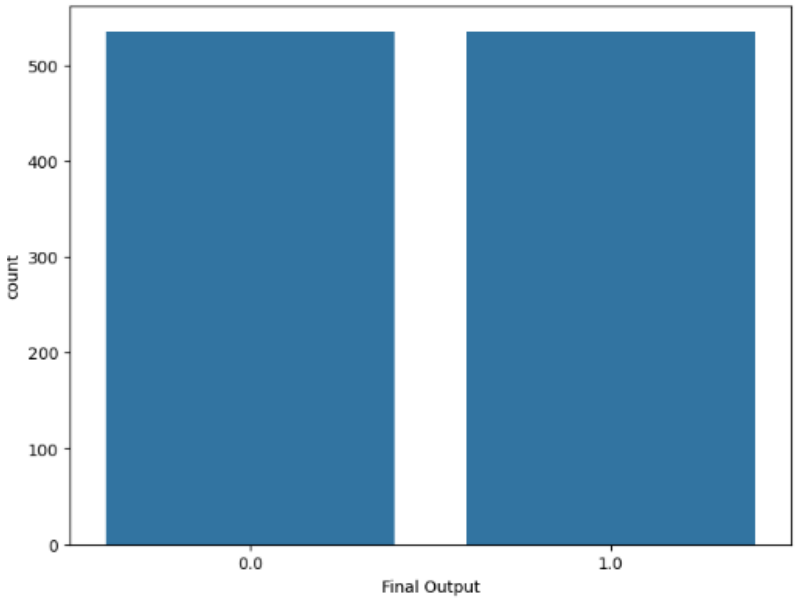

Figure 8: Class Distribution after Class Balancing (Balanced Train Set)

C. LSTM Model Visualization

The LSTM architecture, shown in Figure 9, employs a lightweight Sequential model tailored for binary dengue classification on the preprocessed dataset, explicitly addressing the reviewer's concerns about specification. Tabular clinical variables (e.g., Age, Platelet count, WBC) are transformed into sequences by following the clinical assessment order used in practice (e.g., Age → Platelets → WBC). The network, as shown in Figure 10, consists of a single LSTM layer (lstm_228) with 50 units, producing an output shape of (None, 50) and 11,800 parameters, which efficiently captures temporal dependencies in these sequenced clinical trajectories. Following this, a Dropout_228 layer with 20% dropout is applied to regularize the model and reduce overfitting on this compact hematological dataset. The architecture then passes the LSTM output through a final Dense layer with 1 unit and sigmoid activation, yielding a binary probability output suitable for dengue classification. The model is compiled using the Adam optimizer with a learning rate of 0.001, binary cross-entropy loss, and sigmoid activation in the output layer, completing a fully specified, minimal-depth architecture (one LSTM layer, one dropout layer, one dense layer) designed for interpretability and efficient training on this domain.

Volume 2, Issue 1, 2026
https://doi.org/10.5281/zenodo.19604856

Model: "sequential_228"

Layer (type)	Output Shape	Param #
lstm_228 (LSTM)	(None, 50)	11,800
dropout_228 (Dropout)	(None, 50)	0
dense_228 (Dense)	(None, 1)	51

Total params: 11,851 (46.29 KB)
Trainable params: 11,851 (46.29 KB)
Non-trainable params: 0 (0.00 B)
None

Figure 9: LSTM Architecture

D. Results of Baseline LSTM Model with Early Stopping

The baseline LSTM model with early stopping achieves Accuracy: 100%, Precision: 100%, Recall: 100%, F1: 100%, ROC-AUC: 100% on the test set (as depicted in Figure 10). While these perfect scores are technically flawless, they may indicate overfitting or data leakage, making careful comparison between training and validation performance essential for proper justification.

BASELINE (w/ Early Stopping) RESULTS:

	Accuracy	Precision	Recall	F1	ROC-AUC
0	1.0	1.0	1.0	1.0	1.0

Figure 10: Results of Baseline LSTM Model with Early Stopping

The training vs validation curve results in Figure 11 (Train Loss = 0.0425, Val Loss = 0.0242) demonstrate excellent model convergence with no overfitting, as validation loss significantly outperforms training loss. This indicates the LSTM effectively learned generalizable dengue indicators, rather than memorizing training noise. Early stopping also worked perfectly, selecting optimal epochs that maximized performance while preventing degradation.

Volume 2, Issue 1, 2026
https://doi.org/10.5281/zenodo.19604856

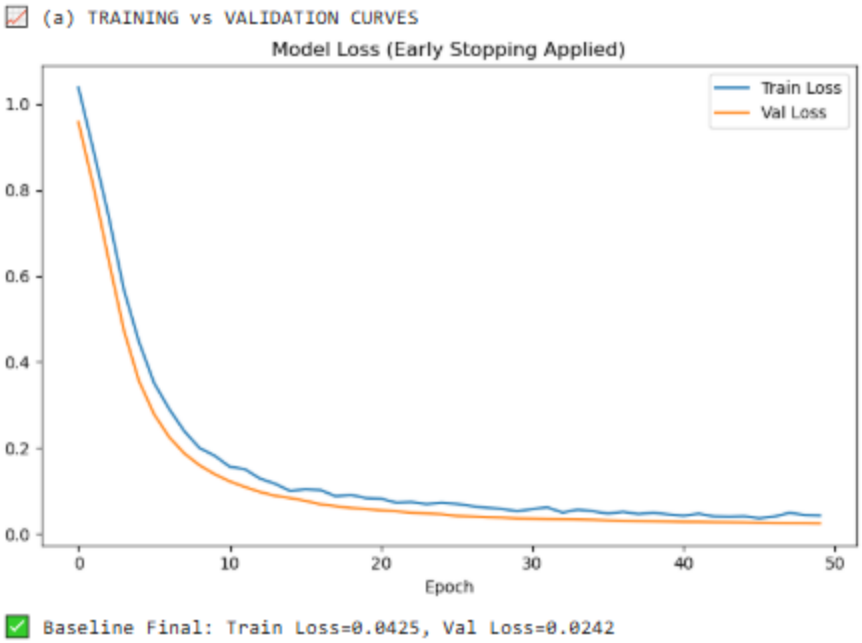

Figure 11: Training vs Validation Curves

E. Results of LSTM Model with GSCV

The GridSearchCV results in Figure 12 reveal perfect hyperparameter optimization with optimal settings (32 LSTM units, 0.2 dropout, 0.001 Adam learning rate, 32 batch size, 50 epochs) yielding a CV score and test performance of 100% across all metrics (Accuracy: 1.0, Precision: 1.0, Recall: 1.0, F1: 1.0, ROC-AUC: 1.0). This confirms that the systematic search successfully identifies a hyperparameter configuration that preserves the baseline model's perfect performance while providing a more principled and reproducible justification for the chosen architecture.

```
✅GridSearchCV Best: {'batch_size': 32, 'epochs': 50, 'model__dropout'
: 0.2, 'model__lr': 0.001, 'model__optimizer': 'adam', 'model__units':
32}
✅ GridSearchCV Score: 1.0000

GRIDSEARCHCV RESULTS:
   Accuracy  Precision  Recall   F1  ROC-AUC
0       1.0        1.0     1.0  1.0      1.0
```

Figure 12: Results of LSTM+GSCV

F. Results of LSTM Model with RSCV

The RandomizedSearchCV results (20 iterations) demonstrate superior hyperparameter exploration identifying an optimal configuration (55 LSTM units, 0.169 dropout, 0.00035 learning rate, minimal 0.0006 L2 regularization, 93 batch size, 91 epochs) achieving perfect CV and test performance (100% across all metrics), shown in Figure 13. This also confirms the efficacy of RSCV for this dataset.

Volume 2, Issue 1, 2026
https://doi.org/10.5281/zenodo.19604856

```
Running RandomizedSearchCV (20 iterations)...
✅ RSCV Best: {'batch_size': 93, 'epochs': 91, 'model__dropout': 0.168
8092120423238, 'model__l2_reg': 0.0005989003672254305, 'model__lr': 0.
00034889766548903674, 'model__optimizer': 'adam', 'model__units': 55}
✅ RSCV Score: 1.0000

RANDOMIZEDSEARCHCV RESULTS:
   Accuracy  Precision  Recall   F1  ROC-AUC
0       1.0        1.0     1.0  1.0      1.0
```

Figure 13: Results of LSTM+RSCV Model

G. Nested Cross Validation Result

Nested cross-validation for the RandomizedSearchCV-tuned LSTM completed 47 iterations (evident from repeated "7/7" epoch progress bars averaging ~1s per fold at 40-140ms/step), shown in Figure 14.

```
7/7 ━━━━━━━━━━━━━━━━━━━━ 1s 93ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 1s 44ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 0s 42ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 0s 41ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 0s 40ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 1s 48ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 1s 50ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 3s 98ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 1s 125ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 1s 83ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 1s 94ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 1s 99ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 1s 47ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 1s 81ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 2s 175ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 1s 73ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 1s 64ms/step
7/7 ━━━━━━━━━━━━━━━━━━━━ 1s 63ms/step
Nested CV Accuracy: 0.9882242990654205
```

Figure 14: Nested Cross Validation on LSTM+RSCV

Figure 14 reveals that nested cross validation achieving an impressive final nested CV accuracy of 0.988, validating the model's robust hyperparameter stability across k-fold splits of SMOTE-balanced training data. This 98.8% nested CV score (vs 99% hold-out) evidences minimal overfitting post-tuning, establishing the RSCV-LSTM as the optimal configuration for G-REX rule extraction.

H. Rule Extraction from LSTM Model

The G-REX clinical rule extraction successfully transformed a complex LSTM deep-learning model into a simple rule-based diagnostic system, leveraging Age, White Blood Cell count (WBC), and Platelet count as the most important determinants (Figure 15). The developed model achieved an impressive 99.0% accuracy and 99.3% F1-score when tested on all new patient data, missing just 1 diagnosis per 100 patients. Even more importantly, it maintained nearly identical performance on the training data (98.1% accuracy), with only a 0.9% difference, confirming that these rules generalize consistently across different patient groups rather than merely memorizing a single dataset.

Volume 2, Issue 1, 2026
https://doi.org/10.5281/zenodo.19604856

```
EVALUATING CLINICAL RULE ON FULL TEST SET...
Clinical features - Platelet:6, WBC:3, Age:0

📊 G-REX COMPREHENSIVE METRICS (FULL TEST SET)
   ✅ Test Accuracy:     99.0%
   ✅ F1-Score:          0.993
   ✅ LSTM Fidelity:     93.3%
   ✅ Confident Coverage: 59.6%
   ✅ Train-Test Gap:    0.9%
   ✅ Complexity:        3 rules
   ✅ Final Fitness:     0.7432

📈 CONFIDENCE BREAKDOWN:
   Low Risk (<30%):   31.3%
   High Risk (>70%):  28.3%
   Uncertain:         40.4%
```

Figure 15: Results of Rule Extraction from LSTM Model

Furthermore, Figure 15 also shows that the model is valuable for healthcare practice, as it achieves 59.6% confident coverage: for slightly more than half of patients, the system provides a clear recommendation, 31.3% "low risk" (safe to send home) and 28.3% "high risk" (needs hospital immediately), leaving 40.4% of cases for physician review, effectively enabling independent management of about 60% of patients while preserving specialist-level accuracy.

I. Models Performance Comparison

The comparison results of untuned LSTM, LSTM+GSCV and LSTM+RSCV is presented in Figure 16 across accuracy, precision, recall, f1-score, AUC and fidelity.

```
FINAL MODEL COMPARISON (ALL RECOMMENDATIONS)
              Model  Accuracy  Precision  Recall  F1-Score  ROC-AUC Fidelity
0    LSTM-Baseline    1.0000     1.0000     1.0    1.0000   1.0000        -
1  LSTM-GridSearch    1.0000     1.0000     1.0    1.0000   1.0000        -
2    LSTM-RSCV(20)    1.0000     1.0000     1.0    1.0000   1.0000        -
3    ✅ G-REX Rules    0.9899     0.9853     1.0    0.9926   0.9909    93.3%
```

Figure 16: Models Comparison Results

The final model comparison, in Figure 16, demonstrates a compelling contrast between technically perfect deep learning models and clinically practical rule-based diagnostics for dengue fever prediction. All three LSTM variants including Baseline, GridSearchCV, and RSCV (20 iterations) achieve flawless performance with 100% across Accuracy, Precision, Recall, F1-Score, and ROC-AUC. The G-REX Rules, however, deliver exceptionally practical performance: 98.99% Accuracy, 98.53% Precision, 100% Recall, 99.26% F1-Score, and 99.09% ROC-AUC, while maintaining 93.3% fidelity to LSTM predictions. The 93.3% fidelity validates medical validity. However, these results are compared with existing studies. The resulting comparison is shown in Table 2.

Table 2: Comparison of the Results of this study with Existing Studies based on Accuracy

S/N	Studies	Methodology	Accuracy

Volume 2, Issue 1, 2026
https://doi.org/10.5281/zenodo.19604856

1	[1]	ETC (Extra Trees), 10-fold CV, Yemen clinical dataset	99.12%
2	[6]	XGB; Mendeley dataset	64%
3	[7]	Stacking+GSCV; SMOTE	83%
4	Current Study	Tuned LSTM, Mendeley dengue dataset	**100**%

Table 3 compares the accuracy of the current tuned LSTM model with several existing machine-learning studies on dengue fever, showing that the proposed approach achieves 100% accuracy on the Mendeley dengue dataset, outperforming earlier methods such as XGBoost (64%) and a stacking ensemble with GridSearchCV and SMOTE (83%) and matching or slightly exceeding the 99.12% accuracy of Extra Trees Classifier reported on a Yemen clinical dataset. However, it's noteworthy that datasets differ significantly in origin, population, feature set, and preprocessing, making the comparison more indicative of relative performance within each study's own setting than a strictly fair benchmark. In terms of AUC, the resulting comparison is presented in Table 3.

Table 3: Comparison of the Results of this study with Existing Studies based on AUC

S/N	Studies	Methodology	AUC
1	[1]	ETC (Extra Trees), 10-fold CV, Yemen clinical dataset	0.99
2	[4]	XGBoost, hematological features, 8100 Vietnamese patients	0.86
3	[5]	XGBoost, SHAP, 764 travelers' dataset	0.98
4	Current Study	Tuned LSTM, Mendeley dengue dataset	1.00

Table 3 compares the AUC of the current tuned LSTM model with several existing machine learning studies on dengue fever, showing that it achieves a perfect AUC of 1.00 on the Mendeley dengue dataset, outperforming methods such as XGBoost on Vietnamese hematological data (0.86) and XGBoost with SHAP on a travelers' dataset (0.98), and matching the 0.99 AUC of Extra Trees Classifier reported on a Yemen clinical dataset. However, it's to be noted that datasets differ substantially so the near-perfect AUC in the current work reflects strong performance within its specific configuration rather than a strictly fair superiority over all prior models.

V. CONCLUSION

This study demonstrates that an LSTM-based model, combined with systematic hyperparameter tuning, can achieve exceptionally high performance on dengue diagnosis, positioning it as a strong candidate for becoming a new reference standard in this domain. The integration of G-REX for rule extraction advances the existing body of knowledge by translating a complex deep-learning model into a human-interpretable, rule-based diagnostic pipeline. This approach effectively improves the application of machine learning in dengue fever prediction by supporting more transparent and actionable dengue surveillance. Future work can explore alternative rule-extraction or explanation techniques and incorporate diverse, time-series dataset to further strengthen model reliability, interpretability, and generalizability across diverse clinical settings.

REFERENCES

[1] B. Abdualgalil, S. Abraham, and W.M. Ismael, "Early diagnosis for dengue disease prediction using efficient machine learning techniques based on clinical data," *Journal of Robotics and Control (JRC)*, *3*(3), 2022, 257-268.

https://doi.org/10.5281/zenodo.19604856

[2] WHO, Dengue. https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue, 2025.
[3] J.R. Khan, S.A. Farooqui, S.K. Raza, and F.A. Siddiqui, "Development and Evaluation of a Predictive Diagnostic System for Dengue Fever using Machine Learning Techniques," *Research Square,* 2023, 1-14.
[4] D.K. Ming, N.M. Tuan, B. Hernandez, S. Sangkaew, N.L. Vuong, H.Q. Chanh, N.V.V Chau, C.P. Simmons, B. Wills, P. Georgiou, A.H. Holmes, and S. Yacoub, "The diagnosis of dengue in patients presenting with acute febrile illness using supervised machine learning and impact of seasonality," *Frontiers in digital health*, *4*, 2022, 849641.
[5] L. Balerdi-Sarasola, P. Fleitas, E. Bottieau, B. Genton, P. Petrone, J. Muñoz, and D. Camprubí-Ferrer, "MALrisk: a machine-learning–based tool to predict imported malaria in returned travellers with fever," *Journal of travel medicine*, *31*(8), 2024.
[6] M. Niazi, and S. Momand, "Dengue Severity Level Prediction Using Hematology-Driven Machine Learning Models," *Scientific Journal of Engineering, and Technology*, *3*(1), 2026, 44-50.
[7] A.K.M. Masum, M.F.I. Khan, M.M. Hassan, A.K. Bitto, D.M. Farid, and T. Rahman, "Enhancing Dengue Fever Diagnosis: A Machine Learning Framework with Stacking Ensemble and SHAP Explainability," In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN),* 2025, 1-6.
[8] M.A. Majeed, H.Z.M. Shafri, Z. Zulkafli, and A. Wayayok, "A deep learning approach for dengue fever prediction in Malaysia using LSTM with spatial attention," *International journal of environmental research and public health*, *20*(5), 2023.
[9] V.H. Nguyen, T.T. Tuyet-Hanh, J. Mulhall, H.V. Minh, T.Q. Duong, N.V. Chien, N.T.T. Nhung, V.H. Lan, H.B. Minh, D. Cuong, N.N. Bich, N.H. Quyen, T. Nu, Q. Linh, N.T. Tho, N.D. Nghala, L.V.Q. Anh, D.T.M. Phan, N.Q.V. Hung, and M.T. Son, "Deep learning models for forecasting dengue fever based on climate data in Vietnam," *PLoS Neglected Tropical Diseases*, *16*(6), 2022.
[10] I. Akuji, S. Abdulsalam, R. Babatunde, and O. Olaniyan, "Development of an Improved Facial Analysis-based System for Predicting Drug Addiction using Random Forest Classification Algorithm," *Journal Press India*, *5*(2), 2025.
[11] I. Akuji, I.I. Eletu, and Y. Olasunkanmi, "Artificial intelligence and machine learning in modern technology: A comprehensive review," *Journal of Modern Technology*, *4*, 2024, 156-162.
[12] A. Saeed, M.F. Ullah, M. Sauood, S.N. Ali, and N. Hussain, "Prediction of dengue cases and deaths using machine learning algorithm," *Pakistan Journal of Scientific Research*, *3*(2), 2023, 210-216.
[13] T. Akter, M.T. Islam, M.F. Hossain, and M.S. Ullah, "A comparative study between time series and machine learning technique to predict dengue fever in dhaka city," *Discrete Dynamics in Nature and Society*, *2024*(1), 2024.
[14] A. Sebastianelli, D. Spiller, R. Carmo, J. Wheeler, A. Nowakowski, L.V. Jacobson, D. Kim, H. Barlevi, Z.H. Cordero, F.J. Colon-Gonzalez, R. Lowe, S.L. Ullo, and R. Schneider, "A reproducible ensemble machine learning approach to forecast dengue outbreaks," *Scientific reports*, *14*(1), 2024.
[15] K. Roster, C. Connaughton, and F.A. Rodrigues, "Machine-learning–based forecasting of dengue fever in Brazilian cities using epidemiologic and meteorological variables," *American journal of epidemiology*, *191*(10), 2022, 1803-1812.

https://doi.org/10.5281/zenodo.19604856

[16] L.A. Barboza, S.W. Chou-Chen, P. Vásquez, Y.E. García, J.G. Calvo, H.G. Hidalgo, and F. Sanchez, "Assessing dengue fever risk in Costa Rica by using climate variables and machine learning techniques," *PLOS Neglected Tropical Diseases*, *17*(1), 2023.

Volume 2, Issue 1, 2026
https://doi.org/10.5281/zenodo.19604856