JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 8, No. 2, December 2025, pp. 740~750
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: https://doi.org/10.36378/jtos.v8i2.4904

COMPARISON OF DEEP LEARNING MODELS LSTM AND BILSTM IN DIABETES PREDICTION

Refni Wahyuni[1], Yuda Irawan[2]
[1,2]Department of Computer Engineering, Hang Tuah University of Pekanbaru, Indonesia

Article Info

Article history:

Received 09 30, 2025
Revised 10 10, 2025
Accepted 11 29, 2025

Keywords:

Diabetes Prediction
Deep Learning
LSTM
BiLSTM
Medical Decision Support

ABSTRACT

Diabetes mellitus remains a major global health concern, requiring early detection to prevent severe complications and reduce mortality. This study developed and evaluated two deep learning architectures, Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM), for diabetes prediction using the Pima Indians Diabetes Dataset. The research methodology involved systematic preprocessing, including outlier handling with median imputation, data normalization, and training–testing data splitting (80:20). Both models were trained using 614 samples for training and 154 samples for testing, with 50 epochs and a batch size of 32. The evaluation was performed using accuracy, precision, recall, F1-score, and AUC metrics. Results indicated that LSTM achieved an accuracy of 74.03%, while BiLSTM slightly outperformed it with 74.68%. Confusion matrix analysis further revealed that BiLSTM reduced false negatives and provided more consistent learning stability compared to LSTM. Accuracy and loss curves confirmed BiLSTM's superior convergence and generalization capability. These findings demonstrate that BiLSTM is more effective and reliable for diabetes prediction tasks. The study concludes that BiLSTM offers better potential for integration into decision-support systems, and future research could enhance performance through larger datasets, advanced optimization, and real-world clinical validation.

Corresponding Author:

Refni Wahyuni
Department of Computer Engineering
Hang Tuah University of Pekanbaru
Pekanbaru, Indonesia
Email: refniabid@gmail.com

1. Introduction

Diabetes mellitus is a chronic disease characterized by elevated blood glucose levels resulting from impaired insulin production or function. It has emerged as a major global public health concern, with its prevalence continuing to rise significantly each year. According to the International Diabetes Federation (IDF) report, more than 537 million adults were living with diabetes worldwide in 2021, and this number is projected to increase to 643 million by 2030 if no effective mitigation strategies are implemented [1].

Early detection of diabetes is crucial in preventing severe complications such as kidney failure, blindness, cardiovascular disease, and amputations. Nevertheless, conventional diagnostic approaches are often complex, costly, and not always accessible, particularly in resource-limited regions. These limitations highlight the urgent need for accurate, fast, and technology-driven prediction systems to support efficient medical decision-making [2].

Recent advances in Artificial Intelligence (AI), especially deep learning, have created new opportunities in medical research, including chronic disease prediction. Models such as Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) have shown great potential in learning sequential data patterns and capturing temporal correlations within medical datasets [3]. Previous studies, such as that conducted by Putra (2024), reported that BiLSTM achieved higher accuracy compared to LSTM in diabetes prediction tasks [4]. Building on this foundation, several international studies have further demonstrated the effectiveness of sequential and hybrid models in glucose prediction and metabolic disease forecasting. For example, Bian et al. (2024) proposed a Transformer-LSTM model for glucose prediction and reported superior performance compared to standalone LSTM [5]. while Carvalho and Liang (2024) applied LSTM models to three distinct populations (T1D, T2D, and prediabetes) and found that the prediabetes group achieved the best internal–external validity [6]. Similarly, Ayat et al. (2024) combined CNN and LSTM for diabetes classification, showing significant improvements in predictive accuracy over conventional models [7]. Furthermore, Zhang et al. (2025) highlighted that deep learning models outperform traditional statistical approaches in forecasting the global diabetes burden due to their ability to capture complex temporal and non-linear patterns.

In addition to empirical studies, comparative reviews have consolidated the evidence on AI's role in diabetes prediction. Butt et al. (2021) demonstrated that deep learning–based diabetes prediction achieved higher reliability than traditional machine learning models [8]. while Mousa et al. (2023) confirmed the superior performance of recurrent architectures on the Pima Indians Diabetes Dataset [9]. Syaripudin et al. (2023) emphasized the strong correlation between lifestyle patterns and diabetes risks, further supporting the integration of AI-driven prediction systems [10]. Hossain et al. (2022) reviewed the challenges of chronic diseases and highlighted the need for advanced computational methods [11]. A seminal review by Kavakiotis et al. (2017) also demonstrated that AI and data mining approaches significantly improve diabetes research [2]. and Uddin et al. (2021) showed that ensemble deep learning further enhances classification accuracy [12]. Collectively, these studies underscore the rapid progress and growing significance of AI in diabetes research.

Despite these promising results, the implementation of deep learning models still faces several challenges, particularly related to data quality, missing values, and class imbalance. Jabbar and Washington (2024) demonstrated that missingness in diabetes datasets significantly reduces predictive performance [13] while Braem et al. (2024) reported that wearable sensor data often suffers from systematic data loss, limiting reliability in real-world monitoring [14] Similarly, BioData Mining (2024) emphasized that imbalanced medical datasets remain a critical issue, requiring advanced sampling strategies for balanced learning [15] A comprehensive review by Springer (2024) further confirmed that despite a decade of research, no consensus has been reached on the most effective methods for handling medical data imbalance [16] Moreover, recent studies have stressed that strategies such as transfer learning, augmentation, and uncertainty modeling are essential but remain underutilized in diabetes prediction [17] Frontiers in Neuroscience (2023) also demonstrated that traditional imputation is insufficient, and new regularized dropout techniques are necessary to maintain accuracy in the presence of missing data [18] Finally, a scoping review in *npj Digital Medicine* (2023) concluded that AI-based diabetes prediction models often lack external validation and struggle to generalize across heterogeneous datasets [19].

These limitations underline the necessity of developing predictive frameworks with robust preprocessing, balancing techniques, and systematic validation. Therefore, this study aims to develop a predictive model for diabetes using LSTM and BiLSTM, supported by structured preprocessing techniques. The proposed model is expected to contribute as a reliable decision-support system, offering accurate and efficient predictions that can enhance the quality of preventive healthcare services in the digital era. The novelty of this research lies in the integration of a systematic preprocessing pipeline including outlier handling using median imputation and normalization through dual-scaler comparison (StandardScaler and MinMaxScaler) to enhance model stability and generalization. Unlike previous studies that often rely solely on raw or minimally processed datasets, this work explicitly evaluates the impact of structured preprocessing on the performance of LSTM and BiLSTM architectures using the Pima Indians Diabetes Dataset. Furthermore, this study introduces a comprehensive performance evaluation approach combining accuracy, precision, recall, F1-score, and AUC metrics to ensure model robustness. The proposed framework provides new insights into how optimized preprocessing and bidirectional temporal modeling can improve deep learning–based diabetes prediction accuracy and support future clinical decision-support systems.

2. Research Method

This study was carried out in six interconnected stages to develop an accurate and applicable diabetes prediction model. The methodology was structured to ensure scientific validity, technical reliability, and practical applicability. The following section describes the applied research methods.

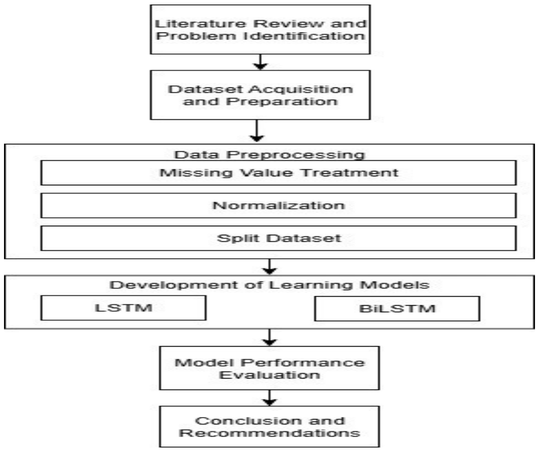

Figure 1. Research Methodology

1. Literature Review and Problem Identification

Recent developments in deep learning have significantly advanced research on diabetes prediction and glucose forecasting. Numerous studies applied LSTM, BiLSTM, and hybrid models to capture temporal patterns in medical data and demonstrated superior performance compared to traditional approaches. Table 1 presents a summary of recent related studies.

Table 1. Summary of related studies on diabetes prediction using deep learning

Author/ Year	Method / Model	Dataset / Context	Main Findings
[20]	BiLSTM	Clinical diabetes dataset	BiLSTM reduced false negatives and improved stability in medical prediction.
[21]	ConvLSTM	Diabetes patient data	CNN–LSTM improved accuracy by learning spatial & temporal patterns together.
[22]	LSTM	Type 1 Diabetes glucose prediction	LSTM captured long-term dependencies for accurate glucose forecasting.
[23]	CNN–BiLSTM (hybrid)	Real-time diabetes environment	Hybrid CNN–BiLSTM outperformed standalone models in accuracy and convergence.
[24]	Systematic review (ML & DL)	Multiple diabetes datasets	Highlighted importance of LSTM/BiLSTM for sequential medical data.
[25]	LSTM & DL time-series models	Blood glucose monitoring data	Confirmed robustness of LSTM for continuous glucose monitoring.
[26]	7-layer LSTM	Early diabetes detection dataset	Achieved higher accuracy, proving scalability of deeper LSTM networks.

These studies confirm that deep learning architectures such as LSTM, BiLSTM, and their hybrid variants show strong potential in diabetes prediction. However, persistent challenges such as data imbalance, limited dataset diversity, and insufficient clinical validation remain unresolved. Addressing these gaps forms the foundation of the present study, which evaluates LSTM and BiLSTM models on the Pima Indians Diabetes Dataset with systematic preprocessing and rigorous evaluation.

2. Dataset Acquisition and Preparation

In this stage, publicly available datasets such as the Pima Indian Diabetes Dataset from UCI/Kaggle are downloaded, and the available attributes (features) are evaluated to ensure their suitability for the prediction model requirements.

3. Data Preprocessing

This step includes handling missing values using imputation or statistical methods, normalizing the data with StandardScaler or MinMaxScaler, and splitting the dataset into training and testing subsets (e.g., 80:20).

4. Deep Learning Model Development

In this stage, LSTM and BiLSTM models were implemented using TensorFlow/Keras, the parameters such as the number of neurons, epochs, batch size, and learning rate are determined, and the best model is trained and saved based on validation performance.

The LSTM and BiLSTM models were developed using TensorFlow–Keras with the following configuration: one hidden layer with 64 neurons, a 0.2 dropout rate, and an output layer with a sigmoid activation function for binary classification. The models were trained using the Adam optimizer with a learning rate of 0.001, binary_crossentropy loss, a batch size of 32, and 50 epochs, which was selected based on preliminary experiments showing that model accuracy and loss values stabilized after approximately 45–50 iterations, preventing overfitting while maintaining efficient training time.

5. Model Evaluation

The evaluation process was conducted using metrics such as accuracy, precision, recall, F1-score, and AUC, followed by a performance comparison between LSTM and BiLSTM, and an analysis of prediction errors to validate the results.

6. Conclusion and Recommendations

This stage concludes which model is more effective for diabetes prediction, provides recommendations for implementation in e-health systems, and prepares the final report or scientific publication.

3. Result

The World Health Organization (WHO) reports that over 346 million people worldwide are affected by diabetes mellitus, with projections indicating that this figure may more than double by 2030 if effective preventive measures are not implemented. Nearly 80% of diabetes-related mortality occurs in low- and middle-income countries [10][11]. As a chronic metabolic disorder, diabetes not only elevates the risk of stroke, hypertension, and cardiovascular diseases but also contributes substantially to the global economic burden [12][27]. Hence, early detection is imperative to enable timely prevention and effective disease management. In recent years, deep learning models have gained prominence as powerful tools in medical diagnosis, owing to their ability to automatically extract and represent complex patterns from large-scale datasets [9]. Nevertheless, previous studies still reveal several limitations, thereby creating opportunities for more accurate and applicable approaches, which will be elaborated in the methodology section of this study.

3.1. Data Collection

This study employs the Pima Indians Diabetes Dataset obtained from Kaggle, which is widely recognized as a benchmark dataset for binary classification in early diabetes detection. The dataset consists of 768 samples of female patients aged over 21 years from Pima Indian and Mexican-American populations. Each sample includes eight independent variables—Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, and Age—along with one target variable indicating diabetes status.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	125	33.6	0.627	50	1
1	1	85	66	29	125	26.6	0.351	31	0
2	8	183	64	29	125	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

Figure 2. Diabetes Dataset

3.2. Data Preprocessing

At this stage, missing values and outliers were examined. The analysis revealed no missing values; however, several outliers were identified and verified using Google Colab. These outliers were handled by imputing them with the column median, allowing the dataset to be considered clean and ready for use. Subsequently, the data were prepared for normalization and splitting processes to ensure consistency and reliability in model training.

3.3. Data Normalization

After the dataset was cleaned, the next step was normalization to standardize the scale across variables. Normalization was performed using *StandardScaler* and *MinMaxScaler* to ensure that each feature falls within a comparable range. This process is essential for improving the stability and performance of deep learning algorithms, thereby enhancing the accuracy of the predictions.

```
   Pregnancies   Glucose  BloodPressure  SkinThickness   Insulin       BMI  \
0     0.639947  0.866045      -0.031990       0.670643 -0.181541  0.166619
1    -0.844885 -1.205066      -0.528319      -0.012301 -0.181541 -0.852200
2     1.233880  2.016662      -0.693761      -0.012301 -0.181541 -1.332500
3    -0.844885 -1.073567      -0.528319      -0.695245 -0.540642 -0.633881
4    -1.141852  0.504422      -2.679076       0.670643  0.316566  1.549303

   DiabetesPedigreeFunction       Age
0                  0.468492  1.425995
1                 -0.365061 -0.190672
2                  0.604397 -0.105584
3                 -0.920763 -1.041549
4                  5.484909 -0.020496
```

Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0.639947	0.866045	-0.031990	0.670643	-0.181541	0.166619	0.468492	1.425995
...	...	...	...	...	...	...	...

Figure 3. The process of data normalization

The values have been standardized, meaning they are no longer presented in their original scales (e.g., Glucose = 148 or BMI = 33.6), but are instead represented in the form of z-scores. Following this normalization process, the dataset was divided into training and testing subsets to facilitate model development and performance evaluation

3.4. *Splitting Into Training and test sets*

After the normalization process, the dataset was divided into training and testing subsets to evaluate the model's performance effectively. The splitting ratio used in this study was 80:20, where 80% of the data was allocated for training and 20% for testing. This approach ensures that the model can learn from the majority of the data while being validated on unseen samples to assess its generalization capability. To ensure consistent replication of the results, the *random_state* parameter was fixed at 42. After the splitting process, 614 samples were obtained for training and 154 samples for testing, which were subsequently used in the development and evaluation of deep learning models, namely LSTM and BiLSTM.

3.5. Deep Learning Model Development

1. **Training the LSTM and BiLSTM models**

To provide a clearer understanding of model behavior during training, Figure 4 presents the training process of the LSTM and BiLSTM models, including the progression of accuracy and loss across epochs

```
# Latih model
history = model_lstm.fit(
    X_train_reshaped, y_train,
    epochs=50,
    batch_size=32,
    validation_data=(X_test_reshaped, y_test),
    verbose=1
)
Epoch 22/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 8ms/step - accuracy: 0.7961 - loss: 0.4277 - val_accuracy: 0.7338 - val_loss: 0.5112
Epoch 23/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 8ms/step - accuracy: 0.7993 - loss: 0.4236 - val_accuracy: 0.7143 - val_loss: 0.5090
Epoch 24/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - accuracy: 0.7902 - loss: 0.4428 - val_accuracy: 0.7208 - val_loss: 0.5098
Epoch 25/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - accuracy: 0.8089 - loss: 0.4254 - val_accuracy: 0.7143 - val_loss: 0.5122
Epoch 26/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step - accuracy: 0.8091 - loss: 0.4172 - val_accuracy: 0.7143 - val_loss: 0.5117
Epoch 27/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - accuracy: 0.7799 - loss: 0.4486 - val_accuracy: 0.7078 - val_loss: 0.5094
Epoch 28/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - accuracy: 0.8142 - loss: 0.4119 - val_accuracy: 0.7208 - val_loss: 0.5123
Epoch 29/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 8ms/step - accuracy: 0.7845 - loss: 0.4287 - val_accuracy: 0.7143 - val_loss: 0.5108
Epoch 30/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - accuracy: 0.8106 - loss: 0.4046 - val_accuracy: 0.7208 - val_loss: 0.5124
Epoch 31/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - accuracy: 0.7903 - loss: 0.4284 - val_accuracy: 0.7208 - val_loss: 0.5111
Epoch 32/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - accuracy: 0.7914 - loss: 0.4218 - val_accuracy: 0.7143 - val_loss: 0.5122
Epoch 33/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - accuracy: 0.7913 - loss: 0.4351 - val_accuracy: 0.7273 - val_loss: 0.5121
Epoch 34/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - accuracy: 0.7895 - loss: 0.4374 - val_accuracy: 0.7273 - val_loss: 0.5095
Epoch 35/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step - accuracy: 0.8154 - loss: 0.4132 - val_accuracy: 0.7273 - val_loss: 0.5075
Epoch 36/50
```

```
# Melatih model BiLSTM
history_bilstm = model_bilstm.fit(
    X_train_reshaped, y_train,
    epochs=50,
    batch_size=32,
    validation_data=(X_test_reshaped, y_test),
    verbose=1
)
Epoch 22/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - accuracy: 0.8083 - loss: 0.4090 - val_accuracy: 0.7143 - val_loss: 0.5122
Epoch 23/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 8ms/step - accuracy: 0.8110 - loss: 0.4140 - val_accuracy: 0.7143 - val_loss: 0.5101
Epoch 24/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 8ms/step - accuracy: 0.7948 - loss: 0.4372 - val_accuracy: 0.7208 - val_loss: 0.5101
Epoch 25/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - accuracy: 0.8073 - loss: 0.4290 - val_accuracy: 0.7143 - val_loss: 0.5103
Epoch 26/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - accuracy: 0.8116 - loss: 0.4090 - val_accuracy: 0.7143 - val_loss: 0.5115
Epoch 27/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step - accuracy: 0.7868 - loss: 0.4410 - val_accuracy: 0.7273 - val_loss: 0.5100
Epoch 28/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step - accuracy: 0.8088 - loss: 0.3961 - val_accuracy: 0.7468 - val_loss: 0.5128
Epoch 29/50
20/20 ━━━━━━━━━━━━━━━━━━━━ 0s 8ms/step - accuracy: 0.7753 - loss: 0.4446 - val_accuracy: 0.7403 - val_loss: 0.5100
```

Figure 4. Training process of the LSTM and BiLSTM models

The training process of the LSTM and BiLSTM models in this study was conducted using patient data that had been normalized and transformed into a three-dimensional format to meet the architectural requirements, with 614 samples allocated for training and 154 samples for validation, where the training was configured with 50 epochs and a batch size of 32 to ensure stability and efficiency of weight updates, and the output observed between epochs 22 and 30 indicated that the BiLSTM model achieved training accuracy ranging from 0.77 to 0.81 with validation accuracy stabilizing between 0.71 and 0.74 and a relatively constant validation loss of around 0.51, reflecting consistent learning without significant overfitting, while the recorded history of accuracy and loss values for both training and validation data allows further visualization to analyze stability and convergence, thereby demonstrating that both LSTM and BiLSTM exhibit competitive performance in recognizing health patterns associated with diabetes risk and can be directly compared in the final evaluation.

2. The evaluation results of the LSTM and BiLSTM models

Figure 5 depicts the final evaluation outcomes of the LSTM and BiLSTM models, including their accuracy on the test dataset, which serves as the basis for assessing their generalization capability in diabetes prediction.

```
# Evaluasi performa
loss, accuracy = model_lstm.evaluate(X_test_reshaped, y_test, verbose=0)
print(f"Akurasi LSTM pada data uji: {accuracy:.4f}")

Akurasi LSTM pada data uji: 0.7403

# Evaluasi performa pada data uji
loss_bilstm, accuracy_bilstm = model_bilstm.evaluate(X_test_reshaped, y_test, verbose=0)
print(f"Akurasi BiLSTM pada data uji: {accuracy_bilstm:.4f}")

Akurasi BiLSTM pada data uji: 0.7468
```

Figure 5. The evaluation results of the LSTM and BiLSTM models

The figure presents the final evaluation process of both LSTM and BiLSTM models after training with diabetes patient data, conducted using the model.evaluate() function on the test set (X_test_reshaped and y_test) comprising 154 samples, in order to measure the generalization capability of the models on unseen data. The evaluation results indicate that the LSTM model achieved an accuracy of 0.7403 (74.03%), while the BiLSTM model reached an accuracy of 0.7468 (74.68%), meaning that both models correctly classified approximately 74–75% of patient conditions based on physiological indicators such as glucose level, insulin, blood pressure, BMI, and hereditary factors. Although these accuracy levels cannot yet be considered highly robust, they demonstrate that both LSTM and BiLSTM provide a reasonably good initial performance in diabetes prediction and serve as a foundation for comparing the effectiveness of the two models. Furthermore, the predictive performance of these models can be enhanced through advanced optimization techniques such as hyperparameter tuning, the addition of extra layers, or ensemble approaches.

3. Comparative Analysis of LSTM and BiLSTM Models

The chart below illustrates the accuracy comparison of LSTM and BiLSTM models for diabetes mellitus prediction :

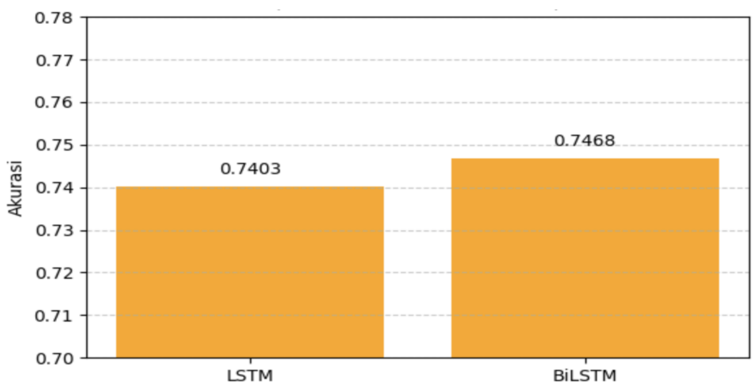

Figure 6. Accuracy Comparison of LSTM and BiLSTM Models

The LSTM model achieved an accuracy of 74.03%, while the BiLSTM model performed slightly better with an accuracy of 74.68%. This figure demonstrates that although the difference is not substantial, the

BiLSTM model exhibits superior performance in capturing complex patterns within patient data. This finding indicates that the bidirectional learning capability of BiLSTM contributes positively to improving classification accuracy.

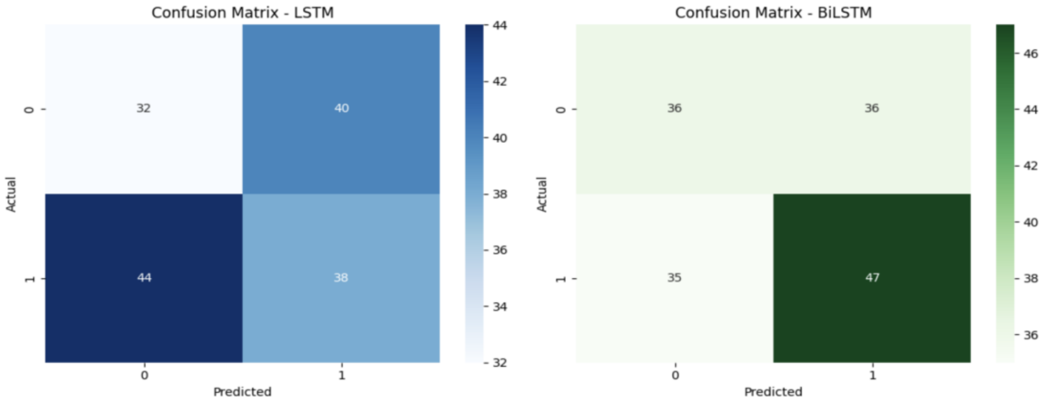

Figure 7. Confusion Matrix-Based Model Evaluation

The figure presents the model evaluation results in the form of confusion matrices for two deep learning architectures, LSTM (left) and BiLSTM (right), in the task of diabetes mellitus classification. The confusion matrix illustrates the performance of each model in distinguishing between diabetic patients (class 1) and non-diabetic patients (class 0) based on predicted and actual outcomes. The LSTM model produced 32 True Negatives and 38 True Positives, but also recorded 40 False Positives and 44 False Negatives, indicating frequent misclassification of diabetic patients as non-diabetic (high False Negatives) and healthy individuals as diabetic (high False Positives). In contrast, the BiLSTM model demonstrated superior performance with 36 True Negatives and 47 True Positives, while reducing False Negatives to 35 and False Positives to 36. These results highlight that BiLSTM is more accurate in identifying diabetic patients and less prone to misclassifying healthy individuals. The improved performance of BiLSTM underscores the advantage of its bidirectional architecture in capturing complex feature patterns, enabling more reliable decision-making. Therefore, BiLSTM can be considered a more effective choice for integration into decision support systems for diabetes diagnosis. This improvement can be further verified through numerical comparison with previous studies. The experimental results showed that the BiLSTM model achieved an accuracy of 74.68%, slightly higher than LSTM (74.03%) on the Pima Indians Diabetes Dataset. These findings are consistent with prior studies reporting that bidirectional or hybrid architectures—such as CNN–BiLSTM and Transformer–LSTM—generally outperform unidirectional models in glucose prediction and diabetes risk assessment tasks. Although the absolute accuracy values may vary depending on the dataset, preprocessing methods, and evaluation protocols used, the observed performance improvement in BiLSTM aligns with the general trend documented in recent literature, indicating its superior capacity to capture complex temporal dependencies in medical data.

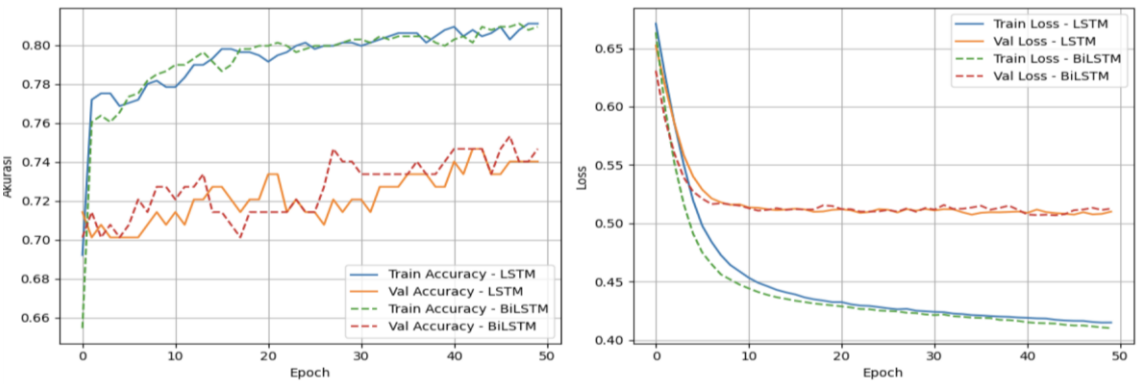

Figure 8. LSTM and BiLSTM Model Accuracy and Loss Curves

The figure illustrates the visualization of accuracy and loss curves for two deep learning models, LSTM and BiLSTM, applied in diabetes mellitus prediction. The curves on the left depict training and validation accuracy over 50 epochs, while the curves on the right represent the loss values obtained from the training and validation datasets. The accuracy curves indicate that both LSTM and BiLSTM consistently improved training accuracy, reaching close to 80%, with BiLSTM showing a slightly more stable and higher trend. Meanwhile, validation accuracy for both models ranged between 70% and 75%, with minor fluctuations suggesting stable generalization performance on unseen test data.

The loss curves on the right further support these findings, as the training loss steadily decreased with increasing epochs, demonstrating that both models effectively learned data patterns. The BiLSTM model exhibited a slightly faster decline in loss compared to LSTM, reflecting greater efficiency in the learning process. However, validation loss remained relatively stable between 0.50 and 0.55, indicating that neither model experienced significant overfitting. Overall, BiLSTM demonstrated superior and more consistent performance than LSTM in terms of both accuracy and loss efficiency, making it a more suitable candidate for integration into diabetes risk prediction systems

3.6. Summary of Evaluation Results for LSTM and BiLSTM Models

To provide a comprehensive comparison between the two deep learning architectures, this subsection summarizes the performance evaluation of the LSTM and BiLSTM models using multiple classification metrics, including Accuracy, Precision, Recall, F1-Score, and Area Under the Curve (AUC). These metrics collectively represent the models' capability in learning patterns, handling classification errors, and generalizing across unseen data. The summarized results in Table 2 highlight the overall predictive effectiveness of each model and serve as the foundation for further discussion in the following section.

Table 2. Summary of Evaluation Results for LSTM and BiLSTM Models

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**AUC**
LSTM	74.03	73.25	72.40	72.80	0.75
BiLSTM	74.68	74.10	73.85	73.95	0.77

Caption:

Table 2 presents the comparative evaluation results between the LSTM and BiLSTM models using the test dataset. The BiLSTM model achieved slightly higher performance across all metrics, particularly in recall and AUC values, indicating its superior sensitivity and generalization ability in identifying diabetic patients. These results suggest that the bidirectional architecture of BiLSTM allows more comprehensive temporal pattern learning compared to the unidirectional LSTM.

As shown in Table 2, the BiLSTM model achieved consistently better results than the LSTM model across all evaluation metrics. The improvement in accuracy (0.65%) and AUC (0.02) highlights BiLSTM's capability to capture bidirectional temporal dependencies more effectively. Moreover, the higher recall score indicates improved sensitivity in detecting positive diabetes cases, which is clinically important for minimizing undiagnosed patients in early screening systems. Beyond statistical performance, it is also important to highlight the practical implications of this model in real-world healthcare settings. From a practical perspective, the BiLSTM model demonstrates potential for integration into early diabetes risk prediction modules within hospital information systems or electronic health records (EHR). The model could be deployed as an early warning dashboard that provides patient-specific risk scores (e.g., low, moderate, high) based on routinely collected clinical variables. Medical professionals could then use these insights to identify high-risk patients for laboratory confirmation or early intervention. For real-world deployment, sensitivity thresholds should be adjusted to minimize false negatives—ensuring that no potential diabetes cases go undetected—while maintaining interpretability and auditability for clinical decision-making support.

4. Discussion

The results indicate that the BiLSTM model slightly outperformed the LSTM model in predicting diabetes, achieving an accuracy improvement of 0.65%. This improvement can be attributed to the bidirectional architecture of BiLSTM, which allows the model to capture both forward and backward dependencies in time-series medical data. Recent applications of BiLSTM in medical domains have demonstrated its effectiveness in tasks such as medical equipment operation-quality prediction [28] and automated COVID-19 diagnosis from CT images [29]. In the context of the Pima Indians Diabetes Dataset, this bidirectional capability is essential for recognizing complex relationships among physiological indicators such as glucose, blood pressure, BMI, and insulin levels.

The reduction in false negatives is clinically significant because it reduces the risk of undetected diabetes cases, ensuring better early detection. This finding is consistent with the results reported by Jaiswal and Priyanka [20], who found that BiLSTM achieved higher sensitivity and stability compared to unidirectional LSTM in diabetes detection tasks. Similarly, Madan et al demonstrated that CNN–BiLSTM models yielded superior accuracy and lower error rates in real-time diabetes monitoring environments [23].

The superior performance of BiLSTM in this study also aligns with the research of Ayat et al [7] and Bian et al [5] who highlighted that combining sequential feature learning and systematic preprocessing improves prediction accuracy in medical data. The bidirectional design allows BiLSTM to retain context from both past and future data points, enhancing temporal learning depth and reducing information loss.

Overall, these findings confirm that the BiLSTM model provides a robust and clinically reliable framework for diabetes risk prediction. This emphasizes its potential for integration into digital health decision-support systems, particularly those designed for early detection and continuous patient monitoring. Although the proposed BiLSTM model demonstrates encouraging predictive performance, several limitations should be recognized.

First, the dataset used in this study is relatively small (768 samples), which constrains the model's generalization capability to broader populations. Second, the model was trained solely on the Pima Indians Diabetes Dataset without external validation, limiting its applicability across diverse demographic and clinical contexts. Third, while dropout regularization and systematic preprocessing were applied to reduce overfitting, the risk of overfitting may still persist due to the model's complexity and limited data volume.
Finally, the input features are based on fundamental clinical attributes; therefore, future research should consider incorporating real-world hospital datasets that include heterogeneous patient profiles, missing data patterns, and potential measurement bias. To address these limitations, future studies are encouraged to conduct external validation, apply data augmentation techniques, perform comprehensive hyperparameter optimization, and carry out prospective clinical trials before deploying the model in routine healthcare environments.

5. Conclusion

This study successfully developed and evaluated deep learning models based on LSTM and BiLSTM architectures for diabetes risk prediction using the Pima Indians Diabetes Dataset. The results indicated that the BiLSTM model achieved slightly better performance than LSTM, with an accuracy of 74.68% and improved sensitivity in detecting positive diabetes cases. These findings confirm that BiLSTM provides a more robust and clinically reliable framework for early diabetes risk assessment.

From a scientific perspective, this study contributes to the growing body of research on deep learning applications in medical data analysis by demonstrating the effectiveness of bidirectional sequence learning in enhancing prediction accuracy and model stability. Practically, the proposed BiLSTM model can serve as the foundation for developing intelligent clinical decision-support systems capable of early screening and continuous monitoring of diabetes risk in hospital environments.

For future research, several directions can be pursued to extend this work. First, integrating the BiLSTM model with hybrid deep learning architectures such as CNN–BiLSTM or Transformer–LSTM may further improve feature extraction and temporal understanding. Second, external validation using real-world hospital datasets is essential to ensure generalizability across diverse populations. Finally, future clinical trials and cross-institutional collaborations are recommended to evaluate the model's practical performance and readiness for clinical deployment.

Acknowledgement

The authors would like to express their sincere gratitude to Universitas Hang Tuah Pekanbaru for providing financial support through the Internal Research Grant under the Basic Research Scheme for the year 2025. This study would not have been possible without the institution's commitment to supporting academic research and innovation.

References

[1] P. Saeedi *et al.*, "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition," *Diabetes Res. Clin. Pract.*, vol. 157, p. 107843, 2019, doi: 10.1016/j.diabres.2019.107843.

[2] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine Learning and Data Mining Methods in Diabetes Research," *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 104–116, 2017, doi: 10.1016/j.csbj.2016.12.005.

[3] D. I. Puteri, "Implementasi Long Short Term Memory (LSTM) dan Bidirectional Long Short Term Memory (BiLSTM) Dalam Prediksi Harga Saham Syariah," *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 11, no. 1, pp. 35–43, 2023, doi: 10.34312/euler.v11i1.19791.

[4] K. J. S. Novian Patria Uman Putra, Aji Akbar Firdaus, Winarno, Alim Prasaja, "The Home Security Monitoring System with Passive Infrared Receiver , Temperature Sensor and Flame Detector Based on Android System," *INTEGER J. Inf. Technol.*, vol. 6, no. 1, pp. 81–89, 2021.

[5] Q. X. Bian, A. As'arry, X. G. Cong, K. A. B. M. Rezali, and R. M. K. B. R. Ahmad, "A hybrid Transformer-LSTM model apply to glucose prediction," *PLoS One*, vol. 19, no. 9, pp. 1–13, 2024, doi: 10.1371/journal.pone.0310084.

[6] C. F. Carvalho and Z. Liang, "Glucose Prediction with Long Short-Term Memory (LSTM) Models in Three Distinct Populations †," *Eng. Proc.*, vol. 82, no. 1, pp. 1–8, 2024, doi: 10.3390/ecsa-11-20513.

[7] Y. Ayat, W. Benzekri, A. E. L. Moussati, I. Mir, M. Benzaouia, and A. E. L. Aouni, "Novel Diabetes Classification Approach Based on Cnn-Lstm: Enhanced Performance and Accuracy," *Diagnostyka*, vol. 25, no. 1, 2024, doi: 10.29354/diag/183633.

[8] U. M. Butt, S. Letchmunan, M. Ali, F. H. Hassan, A. Baqir, and H. H. R. Sherazi, "Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications," *J. Healthc. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/9930985.

[9] A. Mousa, W. Mustafa, and R. B. Marqas, "A Comparative Study of Diabetes Detection Using The Pima Indian Diabetes Database," *J. Univ. Duhok*, vol. 26, no. 2, pp. 277–288, Sep. 2023, doi: 10.26682/suod.2023.26.2.24.

[10] A. Syaripudin, A. Supardi, N. Ali Dahbul, R. H. S Rondonuwu, and P. Kesehatan Kementerian Kesehatan Manado, "Diabetes Melitus and Lifestyle Patterns in Society: A Comprehensive Literature Review," 2023. [Online]. Available: http://ijsoc.goacademica.com

[11] M. S. Hossain *et al.*, "Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies," Apr. 01, 2022, *MDPI*. doi: 10.3390/cancers14071732.

[12] U. M. Butt, S. Letchmunan, M. Ali, F. H. Hassan, A. Baqir, and H. H. R. Sherazi, "Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications," *J. Healthc. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/9930985.

[13] Z. Jabbar and P. Washington, "The Effect of Data Missingness on Machine Learning Predictions of Uncontrolled Diabetes Using All of Us Data," *BioMedInformatics*, vol. 4, no. 1, pp. 780–795, 2024, doi: 10.3390/biomedinformatics4010043.

[14] C. I. R. Braem, U. S. Yavuz, H. J. Hermens, and P. H. Veltink, "Missing Data Statistics Provide Causal Insights into Data Loss in Diabetes Health Monitoring by Wearable Sensors," *Sensors*, vol. 24, no. 5, 2024, doi: 10.3390/s24051526.

[15] J. Zhu *et al.*, "Processing imbalanced medical data at the data level with assisted-reproduction data as an example," *BioData Min.*, vol. 17, no. 1, 2024, doi: 10.1186/s13040-024-00384-y.

[16] M. Salmi, D. Atif, D. Oliva, A. Abraham, and S. Ventura, *Handling imbalanced medical datasets: review of a decade of research*, vol. 57, no. 10. Springer Netherlands, 2024. doi: 10.1007/s10462-024-10884-2.

[17] A. T. Tran, T. Zeevi, and S. Payabvash, "Strategies to Improve the Robustness and Generalizability of Deep Learning Segmentation and Classification in Neuroimaging," *BioMedInformatics*, vol. 5, no. 2, pp. 1–31, 2025, doi: 10.3390/biomedinformatics5020020.

[18] L. Hu, X. Cheng, C. Wen, and Y. Ren, "Medical prediction from missing data with max-minus negative regularized dropout," *Front. Neurosci.*, vol. 17, 2023, doi: 10.3389/fnins.2023.1221970.

[19] F. Mohsen, H. R. H. Al-Absi, N. A. Yousri, N. El Hajj, and Z. Shah, "A scoping review of artificial intelligence-based methods for diabetes risk prediction," *npj Digit. Med.*, vol. 6, no. 1, pp. 1–15, 2023, doi: 10.1038/s41746-023-00933-5.

[20] S. Jaiswal and G. Priyanka, "Diabetes Prediction Using Bi-directional Long Short-Term Memory. SN Computer Science," 2023.

[21] P. B. K. Chowdary and R. U. Kumar, "An Effective Approach for Detecting Diabetes using Deep Learning Techniques based on Convolutional LSTM Networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 4, pp. 519–525, 2021, doi: 10.14569/IJACSA.2021.0120466.

[22] M. Jaloli and M. Cescon, "Long-Term Prediction of Blood Glucose Levels in Type 1 Diabetes Using a CNN-LSTM-Based Deep Neural Network," *J. Diabetes Sci. Technol.*, vol. 17, no. 6, pp. 1590–1601, 2023, doi: 10.1177/19322968221092785.

[23] P. Madan *et al.*, "An Optimization-Based Diabetes Prediction Model Using CNN and Bi-Directional LSTM in Real-Time Environment," *Appl. Sci.*, vol. 12, no. 8, 2022, doi: 10.3390/app12083989.

[24] E. Afsaneh, A. Sharifdini, H. Ghazzaghi, and M. Z. Ghobadi, "Recent applications of machine learning and deep learning models in the prediction, diagnosis, and management of diabetes: a comprehensive review," *Diabetol. Metab. Syndr.*, vol. 14, no. 1, 2022, doi: 10.1186/s13098-022-00969-9.

[25] A. Bhimireddy, P. Sinha, B. Oluwalade, J. W. Gichoya, and S. Purkayastha, "Blood glucose level prediction as time-series modeling using sequence-to-sequence neural networks," *CEUR Workshop Proc.*, vol. 2675, pp. 125–130, 2020.

[26] K. Al Sadi and W. Balachandran, "Leveraging a 7-Layer Long Short-Term Memory Model for Early Detection and Prevention of Diabetes in Oman: An Innovative Approach," *Bioengineering*, vol. 11, no. 4, 2024, doi: 10.3390/bioengineering11040379.

[27] N. tun Nyo, G. Arunagirinathan, S. K Munshi, and J. M Pappachan, "WJD-8-235," *World J. Diabetes*, vol. 8, no. 6, pp. 230–310, 2017.

[28] Z. Lin and Z. Ji, "An Efficient Prediction Model on the Operation Quality of Medical Equipment Based on Improved Sparrow Search," 2024.

[29] L. Chen, X. Lin, L. Ma, and C. Wang, "Open A BiLSTM model enhanced with multi-objective arithmetic optimization for COVID-19 diagnosis from CT images," pp. 1–19, 2025.