International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22S, 2025
https://www.theaspd.com/ijes.php

Genetic Algorithm-Based Hyperparameter Optimization For Diabetes Type Prediction Using SVM And LSTM

Mustafa M. Abd Zaid[1]*, Ali Kareem[2], Ahmed Kateb Jumaah Nussairi[3], Putra Sumari[4], Lamyaa Sabeeh Ashour[5], Chasith Somsak[4], Deenesha Murugun[4] and Gan Zhong Li[4]
[1]College of Technical Engineering, Islamic University, Najaf, Iraq
[2]Faculty of Electrical, Computer, IT, & Biomedical Engineering, Islamic Azad University, Qazvin, Iran
[3]Al-Manara College for Medical Sciences, Amarah, Iraq
[4]School of Computer Science, University Sains Malaysia, 11800 Peneng, Malaysia
[5]Department of Mathematics, College of Basic Education, University of Misan

Abstract: *Diabetes is a chronic illness that needs the most accurate and efficient predictive algorithms for early detection and management. This paper utilizes Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) networks to predict diabetes from the Pima Indians Diabetes dataset. The model takes an enhancement through genetic algorithms that optimize hyperparameters using HPO. The SVM model was tested with four kernel functions, which included linear, polynomial, radial basis function (RBF), and sigmoid, while LSTM model training occurred with Adam, SGD, and RMSprop optimizers. The data preparation process began by handling missing value cases, applying feature scaling, and adjusting the unbalanced data classes after dividing the dataset for training 70% and testing 30%. The experimental outcomes validate the HPO method as a practical improvement for prediction accuracy. Both optimized SVM models using RBF kernel and LSTM model with Adam optimizer produced high accuracy rates, but the SVM model delivered 94%, and the LSTM model reached 90%. This study's findings highlight improved machine learning by optimization methods for enhancing diabetes prediction precision, which supports the development of trustworthy diagnostic healthcare instruments.*
Keywords: *Diabetes Prediction, SVM, LSTM, Genetic Algorithm, Hyperparameter Optimization*

1. INTRODUCTION

Diabetes is a worldwide health emergency because it currently affects about 537 million adults aged 20 to 79 years, and experts forecast this number will rise to 643 million by 2030 and continue to increase to 783 million by 2045 [1]. The world faces a global health emergency due to diabetes because the condition stands among the top disease causes of death, resulting in 6.7 million deaths in 2021. The current diagnosis methods essentially use clinical tests together with physician examinations for early-stage diabetes detection, but they are slow and expensive, with sometimes irregular accuracy levels. This issue has prompted a growing interest in machine learning (ML) and deep learning (DL) algorithms for diabetes detection, offering higher accuracy, scalability, and automation in diagnosis [2].

The two diabetes types are type 1, which shows insulin deficiency, and type 2, which demonstrates insulin resistance. A diabetes condition without proper control results in multiple severe complications that include cardiovascular disease and neuropathy, nephropathy, and retinopathy. The number of individuals affected by diabetes climbed dramatically between 1980, when WHO reported 108 million diabetics, and 2014, when WHO recorded 422 million patients, according to their reports [3]. The continuous development of diabetes and its multiple risk elements results in traditional statistical methods becoming unable to track data because of non-linear patterns. This has led to the detection of artificial intelligence-driven models for more accurate illness prediction and risk stratification.

ML and DL trends top in set of fields such as medical fields [4], [5], agriculture [6], [7], geographical [8], [9], [10], and others. SVMs, and LSTM networks, have become top-performing techniques for healthcare solutions. The SVM system efficiently analyzes high-dimensional datasets with non-linear relationships while allowing LSTM to extract long-term dependencies from sequential medical records suitable for departmental use for patient record analysis. However, the performance of these models is highly dependent on hyperparameter selection, which, if not optimized, can lead to suboptimal generalization and lower predictive accuracy.

This paper employs a GA-based hyperparameter optimization system to increase the accuracy of SVM and LSTM models as addressing solutions for this challenge. Related studies used traditional methods of grid search and random search to find hyperparameters, but both methods suffer from high computation time and often converge near local optimal solutions. GA benefits from natural selection and evolutionary principles to establish an effective search procedure that locates the highest-performing hyperparameter configurations that help achieve better model robustness and generalization.

This study uses the PIDD benchmark as its dataset for diabetes prediction analysis. It compares linear and polynomial SVM kernels, RBF kernels, and sigmoid kernels to three LSTM optimizers (Adam, RMSprop, and SGD) under both non-optimized and optimized hyperparameter conditions. This paper aims to use SVM and LSTM for diabetes predictions through Genetic Algorithm-based hyperparameter optimization to create an accurate and robust predictive model.

2. RELATED WORK

Rajendra & Latif. (2021) suggested a predictive model incorporating diagnostics obtained from the PIMA and Vanderbilt datasets. This paper used logistic regression as its core model, feature selection was done in two ways, and ensemble techniques were used to improve predictions beyond single-model performance. The experiments achieved a 78% accuracy for Dataset 1 through the Max Voting ensemble and 93% accuracy for Dataset 2 through the Max Voting and Stacking ensemble. The algorithm's accuracy was significantly enhanced compared to a single model [11].

The authors Driouich et al. (2024) proposed an FSVM called SMOTE Density Based Fuzzy Support Vector Machine (SMOTE-DB-FSVM) as a classification model based on FSVM to improve diabetes detection. This methodology includes five fundamental procedures, starting with data cleaning followed by density-based filtering, then feature selection for identifying key attributes, assessing confidence scores of minority class points, and concluding with using SMOTE to balance the data. A comparison of various kernel function versions in the SVM model was made to maximize classification results through metaheuristic estimation of kernel parameters. The accuracy rates reached 91.31% by SMOTE-DB-FSVM for detecting diabetes in its early stages [12].

The PIMA Indian feature dataset received training from Shams et al.(2025) through the utilization of Random Forest (RF), Logistic Regression (LR), K-Nearest neighbor (KNN), Naïve Bayes (NB), Histogram Gradient Boost (HGB) and Gated Recurrent Unit (GRU) machine learning models. The researchers employed RFE-GRU as a novel method to identify PIDD. RFE is important for feature selection in the training dataset, as features contribute to successfully predicting the target. At the same time, the GRU handles the challenge of vanishing and inflating the gradient of the features resulting from RFE. The RFE-GRU model achieved high precision, recall, F1-score, accuracy, and Area Under the Curve (AUC), achieving 90.50%, 90.70%, 90.50%, 90.70%, and 92.78%, respectively [13].

In 2025, researchers introduced a deep learning method for early diabetes detection using a PIMA dataset containing numerical values. The method converts numerical data into images based on the robust representation of CNN models. Three classification strategies are applied to the resulting diabetes image data: feeding diabetes images into ResNet18 and ResNet50 CNN models, fusing deep features of ResNet models with support vector machines (SVM), and selecting fusion features classified by SVM. The classification accuracy using the SVM/cubic model with 500 selected features was 92.19% [14].

Gill et al. (2022) aimed to develop a classifier and compare various data mining techniques based on their accuracy in detecting diabetes based on various symptoms and features. Machine learning methods were used on the diabetes dataset, available through the Biostatistics program at Vanderbilt. Using a genetic algorithm as a feature selection method and a Random Forest for classification yielded the highest accuracy of 93.95% [15].

According to Naseem et al. (2022), six Machine learning techniques like Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Support Vector Machine (SVM), Logistic Regression, Artificial Neural Network (ANN), and Long Short-Term Memory (LSTM) were adopted to detect deadly illnesses. They capture the model's performance using accuracy, precision, recall, and F score. Among algorithms, RNN performs best compared to other algorithms, with an F score of 65%, accuracy of 81%, and precision of 75%. However, recall is higher with the ANN model, which is 56%. With this patient

health monitoring system's implementation, physicians can diagnose the disease early [16].
Srinivasu et al. (2022) created a novel technique to forecast type 2 diabetes using portions of genomic DNA. They implemented automated matching and feature selection processes to correlate specific gene profiles with the training set. The model was validated with tabular information and tested with an artificial intelligence model with recurrent neural network (RNN), long short-term memory (LSTM), and gated recurrent units (GRU) components. The model produced moderate performance, in the accuracy RNN and LSTM combined with weight optimization performed the best Accuracy: 81.0%, Sensitivity: 81.5%, Specificity: 79.3%, F1-Score: 0.856 and MCC: 0.568 [17].
Khademi et al. (2022) suggested various approaches to diagnose it, so classification is one of the primary methods. An ensemble classifier was used to apply support vector machine (SVM), k-nearest neighbour (KNN), and whale optimization algorithm (WOA). WOA creates weights for every classifier to increase the accuracy of diabetes classifications. According to the implementation findings, the intended ensemble classifier attained an accuracy rate of 83% [18].
Table 1 show summery results of related works.

Table 1: Results of Related Works

Ref	Dataset	model	Accuracy
[11]	PIMA and Vanderbilt	Logistic Regression, ensemble techniques- Max Voting, and Stacking	ensemble techniques= 78% for Dataset 1, achieved with Dataset2 = 93%
[12]	PIMA	SOMTE-DB-FSVM	SOMTE-DB-FSVM= 91.31
[13]	PIMA	Random Forest (RF), Logistic Regression (LR), K-Nearest neighbor (KNN), Naïve Bayes (NB), Histogram Gradient Boost (HGB), and Gated Recurrent Unit (GRU), Recursive Feature Elimination GRU (RFE-GRU)	RFE-GRU= 90.70%
[14]	PIMA	ResNet18 , ResNet50 CNN, SVM	SVM/cubic model with 500 selected features = 92.19%
[15]	PIMA	genetic algorithm with random forest	GA+RF= 93.95%
[16]	PIMA	Support Vector Machine (SVM), Logistic Regression, Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM)	RNN=81%
[17]	PIMA	Recurrent Neural Network (RNN) components, Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU)	RNN+LSTM(WO)= 81%
[18]	Private	ensemble classifier to apply support vector machine (SVM), k-nearest neighbor (KNN), and whale optimization algorithm (WOA)	ensemble classifier+ WOE=83%

3. METHODOLOGY

Our methodology consists of a set of phases: dataset collection, processing, models for prediction, enhancement of models using genetic algorithms for hyperparameters optimization, evaluation of models, and comparison between them, as shown in Figure 1.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22S, 2025
https://www.theaspd.com/ijes.php

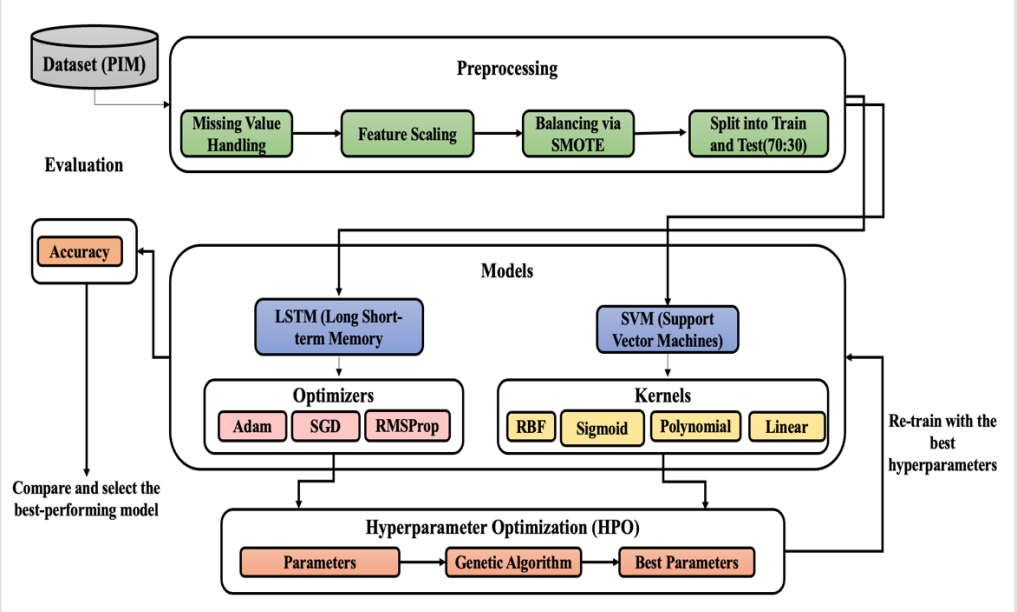

Figure 1: Methodology of Study

3.1 Phase1: Dataset Description

The dataset used in this research is PIMA. The PIMA Diabetes Database is a commonly used dataset for predicting diabetes [19]. This dataset consists of 768 rows and 9 columns. The features in the columns are glucose, pregnancy, skin thickness, blood pressure, BMI, insulin, age, and results. The dataset label distribution appears in Figure 2. The outcome variable predicts the patient's diabetes status since it determines whether someone has diabetes or not. The information can be viewed and used in files that employ the CSV file format. The PIMA dataset functions as the most comprehensive dataset for diabetes prediction. A wide range of characteristics from more than 8,000 people enables this study to deliver accurate, current health and lifestyle information. Researchers can depend on this dataset because scientists have extensively tested and verified it to predict diabetes.

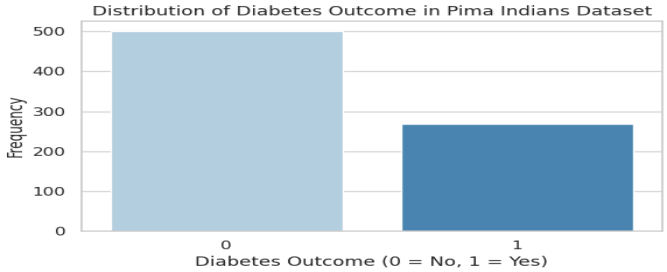

Figure 2: Number of Each Class in PIMA

The predictive glucose values, pregnancy, skin thickness, blood pressure, BMI, insulin, and age values appear in Figure 3.

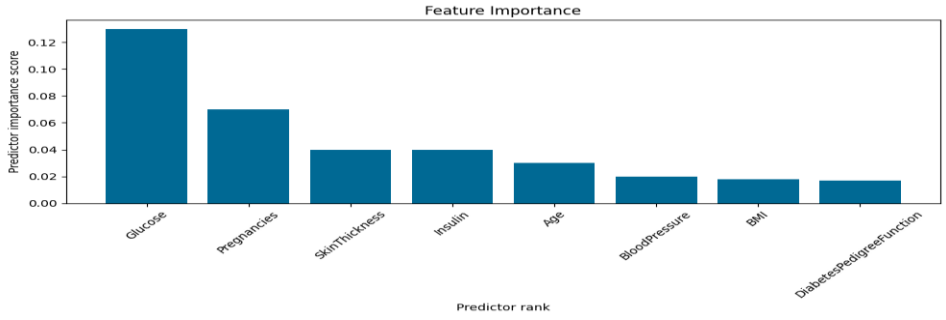

Figure 3: The Predictive Values of Features

The correlation matrix between diabetes diagnosis parameters is as shown Figure 4.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22S, 2025
https://www.theaspd.com/ijes.php

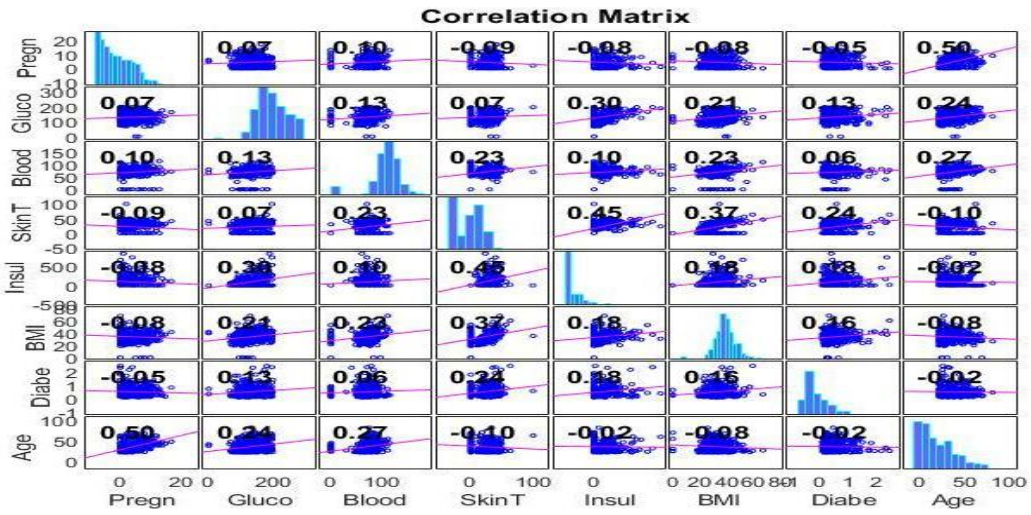

Figure 4: Correlation Matrix for PIMA

3.2 Phase 2: Preprocessing

Standard data preparation methods that apply to the Pima Indians Diabetes dataset clean up the datasets before entering the model. A result with a binary value (0= no diabetes, 1= diabetes) and numerical values make up the dataset. The usual procedures for preprocessing are as follows:

3.2.1 Handling Missing Values

Many things may need to be fixed with data input, such as broken machinery or unfinished surveys, all of which can lead to missing values in a dataset. Machine learning relies heavily on handling missing variables properly, as they have the potential to impact model performance [20]. Although there are no obvious missing values in the Pima dataset, characteristics such as Glucose, Blood Pressure, Skin Thickness, Insulin, and BMI include columns with invalid zero values. Putting NaN in lieu of zeros in these columns will make them imply nothing.

3.2.2 Feature Scaling

The dataset includes variables with varying scales, such as glucose levels, insulin levels, and body mass index (BMI). Algorithms rely on features that have been standardized, which is a statistical technique that rescales features in a dataset to have specific properties, usually a mean of 0 and a standard deviation of 1 [21]; its formula is:

$$z = (x - \mu) / \sigma \quad (1)$$

where:

z is the standardized value,

x is the original value,

μ is the mean of the feature,

σ is the standard deviation of the feature.

3.2.3 Dealing with Imbalanced Classes

As shown in Figure 2, PIMA is an unbalanced dataset with a mean and an uneven class distribution. The machine learning task requires an imbalance solution using SMOTE, which stands for Synthetic Minority Over-sampling Technique [22]. SMOTE makes the model more effective and robust by adding new instances to the minority class. Users can quickly implement SMOTE with their datasets through multiple machine-learning packages, including imbalanced-learn in Python.

3.2.4 Splitting the Dataset

The division of datasets in model construction ensures that models successfully predict new examples. Partitioning available data into multiple subgroups remains an essential procedure for training and testing objectives [23]. The dataset will be divided into 70% dedicated to training purposes and 30% dedicated to testing.

3.3 Phase 3: Models to Predict Diabetes

The predictive analysis of patient diabetes types utilizes patient information through Support Vector

Machine (SVM) and Long Short-Term Memory (LSTM) networks during this phase. This assignment justifies using these models because they each bring specific characteristics that make them ideal for this work. The following part explains both models alongside their prediction capabilities.

3.3.1 SupportVector Machine (SVM)

The first analytical model in this study involves a Support Vector Machine (SVM). The supervised machine-learning technique operates as a regression and classification solution while capable of supervised learning applications [24]. A linear model named this algorithm applies border and cloud boundaries to divide the diabetes dataset into separate categories. SVM determines the best-separating plane through maximum margins by finding the distance between the closest points of different classes in a specific dataset. The SVM algorithm depends on a decision boundary system for defining decision boundaries [25]. SVM demonstrates its leading capability for diabetes prediction by detecting complex nonlinear relationships, which ordinary methods would overlook [26]. With its advantage, SVM can detect patterns present in the data, leading to improved prediction accuracy. The ability of SVM to process high-dimensional data makes it efficient when predicting diabetes since analysis relies on several variables [27].

3.3.2 Long Short-Term Memory (LSTM)

The second machine learning model used for diabetes diagnosis within this study is LSTM deep networks. RNNs modified as Long Short-Term Memories (LSTMs) process data sequences, including text strings, audio patterns, and time-based datasets [28]. LSTMs use several levels of memory cells to track data dependencies precisely, according to [29]. LSTMs utilize this mechanism to detect information patterns at different time scales and solve problems like diabetes diagnosis using continuous patient records. The important step starts with creating a model by determining model layers and linking them with optimization and a loss function before fitting training data. The training process utilizes training datasets while executing the model several times through parameter modifications to enhance prediction accuracy. The accuracy evaluation of the model occurs through its execution against experimental diabetes datasets to determine its performance level.

3.4 Phase 4: Enhance Models using Genetic Algorithm (GA)

A Genetic Algorithm (GA) is used to optimize the hyperparameters of both the SVM and LSTM models. The GA finds the best configurations by acting out a natural selection process, whereby individuals (i.e., sets of hyperparameters) change during generations [30]. The models' performance is used to determine the fitness function. Relevant measures, including recall, accuracy, and F1 score, are used. Iteratively generating new populations via selection, crossover, mutation, and encoding hyperparameters as chromosomes are all part of GA [31]. The process continues until convergence is reached or a preset stopping threshold is satisfied. Automatically adjusting the hyperparameters enables the models to attain optimum predicted performance.

3.4.1 Hyperparameter Optimization for SVM

After creating the initial structure of SVM, we tune the hyperparameters to further improve the model's accuracy. The hyperparameters for the Support Vector Machine (SVM) include C. This first parameter assigns a weight to each data point based on its error value. A more considerable C value means the SVM tries to classify all points correctly, while a smaller C value means the SVM tries to minimize the error as much as possible [32]. The kernel's second parameter transforms the input data into a higher feature space. Kernel types include linear, polynomial, Radial Basis Function (RBF) and sigmoid [33]. The following relationships show the kernels used in the problem of diabetes diagnosis.

The Mathematical Formula of linear Kernel [34]

$$K(x, y) = x.y \quad (2)$$

The Mathematical Formula of polynomial Kernel [35]

$$K(x, y) = (ax.y + b)^d \quad (3)$$

The Mathematical Formula of RBF kernel [36]

$$K(x, y) = e^{-\|x-y\|^2/\sigma^2} \quad (4)$$

The Mathematical Formula of sigmoid kernel [37]

$$K(x, y) = ae^{\frac{\gamma^2}{\|x-y\|^2/\sigma^2}} - 1 \quad (5)$$

The gamma third parameter is used in the Radial Basis Function (RBF) kernel and controls data spread in the feature space [38]. As gamma increases, the SVM model arranges data with less spread, meaning that the influence of each data point becomes more localized in the feature space [39]. The degree parameter is specific to the polynomial kernel and indicates the degree of the polynomial used to transform the data into a higher-dimensional space. The higher degree of models enables them to detect sophisticated data patterns. The coef0 fourth parameter functions with both polynomial and sigmoid kernels by controlling the dimensionality transformation of data [40]. The decision boundary shape and behaviour depend enormously on this parameter during the use of low-degree kernels [41]. Training applications of the shrinking heuristic depend on the value of the shrinking parameter. Enabling SVM implements an algorithm to shrink the problem size, which shortens training periods, mainly when processing extensive data. Enabling the probability mode adds a fifth parameter in SVM that enhances the model by delivering probability ratings for each class valuable prediction for confidence assessment [42].

3.4.2 Hyperparameter Optimization for LSTM

When applying Long Short-Term Memory (LSTM) networks for diabetes prediction, proper hyperparameter adjustment must be made to reach maximum performance levels. The LSTM requires adjusting multiple parameters named hyperparameters, which involve layer counts, unit distribution, activation type, sequence type selection, learning rate adjustment, specific loss calculation, early stopping threshold, maximum epochs, and batch quantity and dropout mechanism. The number of layers defines the depth of LSTM layers within the network, and the number of units indicates the number of LSTM units (or neurons) inside each layer [43]. Each LSTM unit output transformation uses an activation function between sigmoid and tanh [44]. The return sequence type controls network output selection between single and sequential values, determining the BPTT or truncated BPTT usage [45]. The learning rate controls how quickly the model updates its weights during training, and the loss function [46], such as mean squared error (MSE) or cross-entropy loss, helps the model optimize its weights by minimizing the error between predicted and actual outputs. Early stopping accuracy prevents overfitting by halting training once accuracy plateaus [47]. The number of epochs refers to how many complete passes the model makes over the dataset, while the batch size dictates how many data samples are processed before the model's weights are updated [48]. The dropout rate defines the proportion of units randomly deactivated during training to prevent overfitting [49].

3.5 Phase 5: Evaluate the Models

Phase 5 involves evaluating the models to see how well they forecast, with accuracy being the leading indicator. With its measurement of the percentage of examples identified adequately out of the total number of instances in the dataset, accuracy is a commonly used assessment criterion in classification issues [50]. The following is the formula for determining accuracy in this study:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

3.6 Phase 6: Comparison between Models

This paper analyzes and evaluates how Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) perform in type-diabetes prediction following Genetic Algorithms (GA) training. At this point, automatic evaluation of the prediction capacities takes place through accuracy analysis. We conduct this assessment to determine the prediction accuracy levels for both models on the testing data and analyze their strengths and weaknesses. The success of support vector machines (SVMs) depends on evaluating hyperparameters C and gamma and multiple kernel choices consisting of linear, polynomial, RBF and sigmoid. To evaluate an LSTM model, one needs to assess parameters consisting of both LSTM layer numbers and unit quantities, learning rate modifications, and dropout rate modifications for evaluating temporal relationship capabilities. The paper evaluates both models' training efficiency, computational capability, and generalization ability. The provided data set requires us to evaluate the final results from each predictive model for optimal diabetes diagnosis performance. The evaluation outcomes will lead to selecting the optimal operational model with high predictive accuracy and efficient computational performance.

4. RESULT AND DISCUSSION

This section investigates Support Vector Machines (SVM), Long Short-Term Memory (LSTM), and their variants enhanced by genetic algorithms for diabetes-type prediction. The next section focuses on evaluating the performance results obtained from different models.

4.1 Result

We present results that refer to SVM and LSTM analysis execution results.

4.1.1 Performance of SVM Without Hyperparameter Optimization (HPO)

The SVM model evaluation process started with the default parameter to determine its baseline performance capabilities. Four kernel functions' classification abilities were evaluated by combining linear, polynomial, radial basis function (RBF), and sigmoid. Figure 5 reveals the total accuracies obtained from each kernel, while Figure 6 displays confusion matrices to reveal model classification outcomes.

The linear kernel reached 89% accuracy in testing. The models produced these accuracy statistics according to its confusion matrix: TP (102) and TN (103) against FP (13) and FN (12). The linear kernel accomplishes good classification results in most cases. However, it demonstrates failures in correctly identifying many diabetic and non-diabetic cases because it struggles to detect intricate patterns in the dataset.

The accuracy of the polynomial kernel reached 89.13% while producing results of TP = 101, TN = 104, FP = 14, and FN = 11. The polynomial kernel reduces false positive outcomes compared with linear kernel performance, thus improving the representation quality of decision boundaries. The number of incorrectly identified diabetic cases remains at the same level because such challenges persist with correct diagnosis.

The RBF kernel improved, resulting in 92.10% accuracy and statistical values of TP = 105, TN = 107, FP = 10, and FN = 8. This kernel shows improved capability to detect the non-linear relationships in the dataset through its reduced occurrence of false positives and false negatives, which leads to better classification performance.

The accuracy measure for the sigmoid kernel was 92.10%, while TP = 105, TN = 107, FP = 10, and FN = 8. The simulated data reveal that the sigmoid kernel offers marginal improvement in accuracy because it creates effective complex decision boundaries compared to the RBF kernel.

These obtained results make the significance of selecting appropriate kernels for SVM-based diabetes classification apparent. The confusion matrices show that linear and polynomial kernels achieve reasonable accuracy but face difficulties dealing with data points that are not separable using linear classification. RBF and sigmoid kernels achieve higher classification accuracy because they create better models of complex decision borders. Additional improvements must occur through hyperparameter optimization because it enables maximum predictive performance and minimum classification errors.

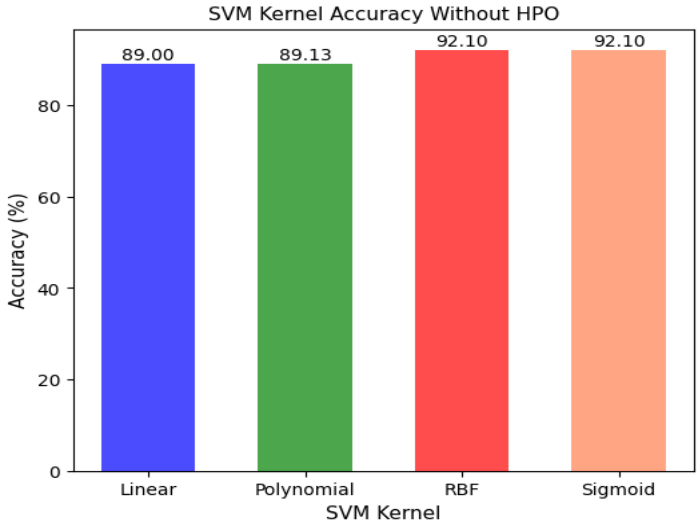

Figure 5: Performance of SVM without HPO

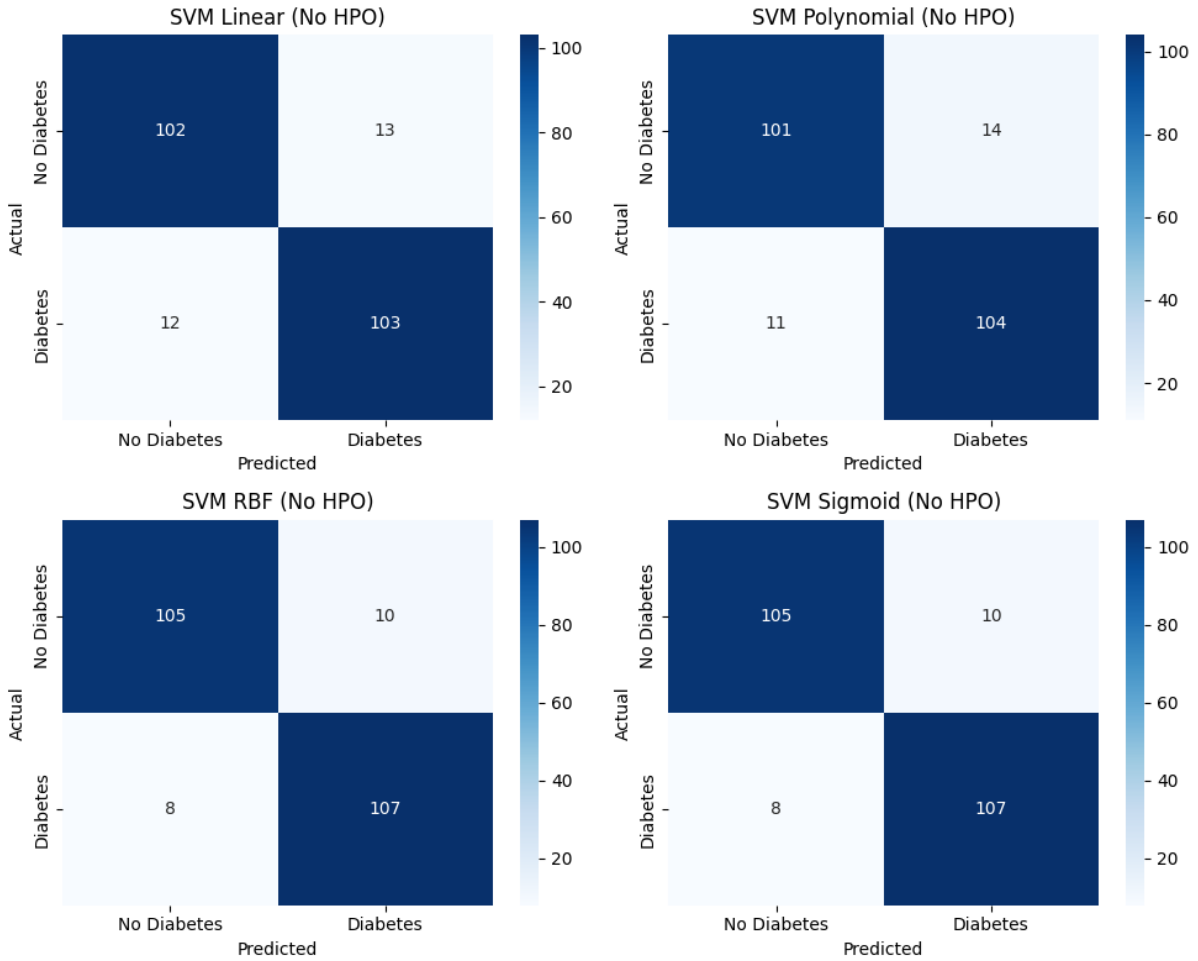

Figure 6: Confusion Matrics for SVM without HPO

4.1.2 Performance of SVM with Hyperparameter Optimization (HPO)

The second evaluation of the SVM model concentrated on implementing hyperparameter optimization (HPO) to enhance performance levels. After HPO implementation, we evaluated four kernel functions, including linear, polynomial, and radial basis functions (RBF) and sigmoid, for their classification ability. The accuracy information regarding each model can be found in Figure 7 alongside the confusion matrices, which display classification results as shown in Figure 8.

The linear kernel, after hyperparameter enhancement, achieved an accuracy of 90%. The model produced results in its confusion matrix comprising TP = 103 together with TN = 104 while counting FP = 12 and FN = 11. The performance showed a minimal improvement over the baseline test without HPO, though it still resulted in several cases of false diagnosis between diabetic and non-diabetic patients. Hyperparameter adjustment reduced false positive and false negative outcomes while better refining the linear decision limits within the dataset.

The polynomial kernel outperformed initial results by reaching a 93% accuracy rate compared to the 89.13% accuracy rate before HPO implementation. In its confusion matrix, TP = 106, TN = 108, FP = 9, and FN = 7. This enhancement in accuracy and a decrease in false positives and negatives refers to the effectiveness of the polynomial kernel after optimization. By optimizing the polynomial kernel, boundaries became more defined, thus producing fewer misclassification errors.

The RBF kernel brought the most pronounced results which enabled a 94% accuracy rate. Results from the confusion matrix showed TP = 107 while TN = 109 alongside FP = 8 and FN = 6. The modified RBF kernel achieved better relations between dataset variables through hyperparameter optimization thus decreasing both false positive and false negative rates. The kernel achieved maximum accuracy in classification which made it suitable for this application.

After HPO optimization, the sigmoid kernel reached a 92.20% accuracy rate yet demonstrated TP = 106, TN = 106, FP = 9, and FN = 9. The sigmoid kernel demonstrated better performance than its baseline counterpart by achieving fewer misclassifications, even though the performance boost was less pronounced than the RBF kernel's. The evaluation matrix demonstrates that the sigmoid kernel functions properly identified relevant decision boundaries without achieving better performance than the RBF kernel in this evaluation.

The obtained results demonstrate that hyperparameter optimization techniques effectively enhance the performance capabilities of SVM models. Following optimization, the RBF kernel became the most efficient by reaching the highest accuracy with the lowest misclassification rates. HPO generated confusion matrices that proved how the method successfully cut errors and enhanced classification results. Additional optimization can be implemented to reach superior performance levels for diabetes diagnosis.

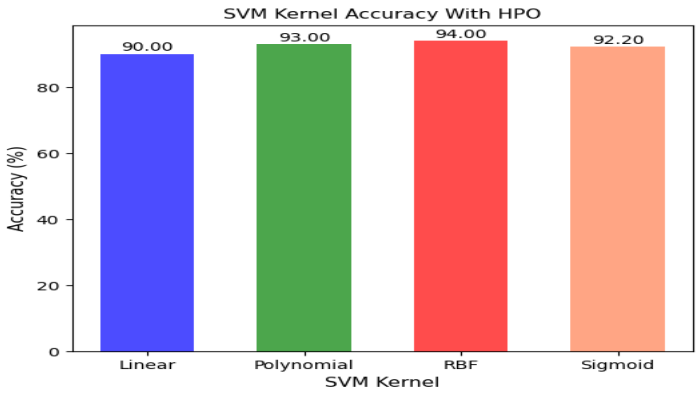

Figure 7: Performance of SVM with HPO

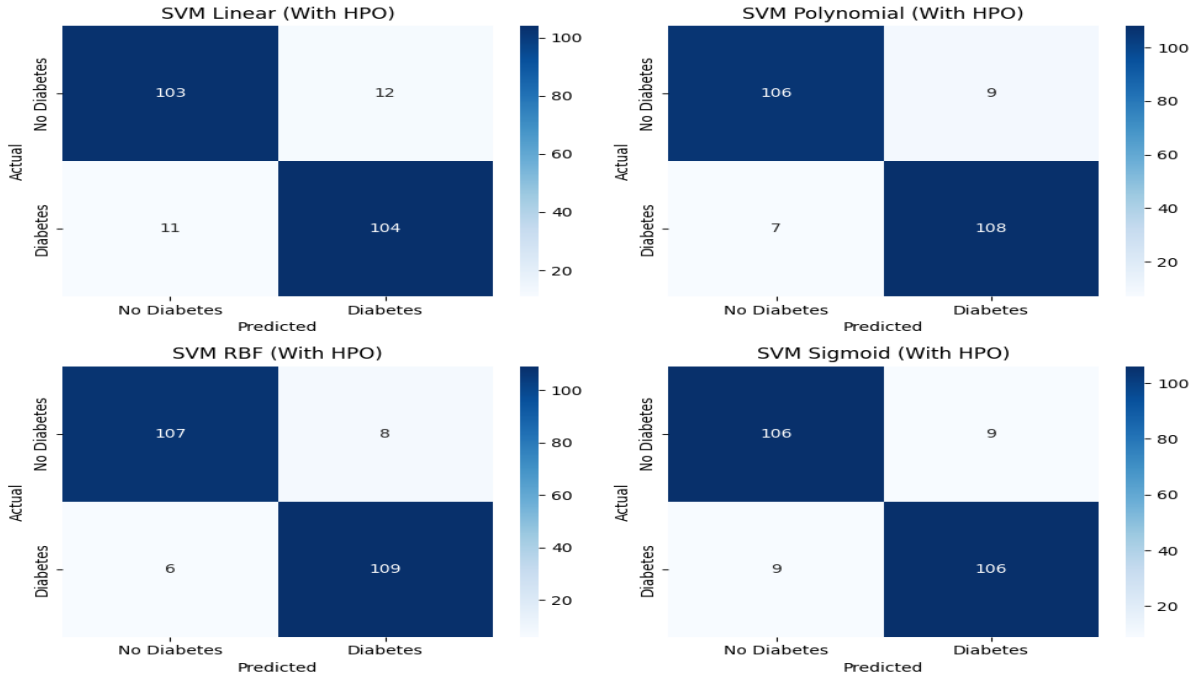

Figure 8: Confusion Matrics for SVM with HPO

4.1.3 Performance of LSTM Without Hyperparameter Optimization (HPO)

To establish performance standards, the LSTM model was first evaluated without modifying its hyperparameters. The classification performance evaluation utilized three optimization methods: Adam, RMSprop, and SGD. Figure 9 shows the accuracy result of all applied algorithms, and Figure 10 illustrates the confusion of matrices that reveal more detailed information about the model's classification performance.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22S, 2025
https://www.theaspd.com/ijes.php

The Adam optimizer reached 85% accuracy during its execution.
The confusion matrix analysis revealed 95 instances of correct positive predictions while also reporting 101 correct negative classifications, 20 incorrect positive predictions, and 14 incorrect negative ones. Adam delivered accurate results, but its confusion matrix analysis indicates many misidentified instances, especially incorrect positive and negative predictions. Adam proves generally effective but demonstrates limited accuracy when distinguishing between diverse patterns of diabetic and non-diabetic medical cases, which results in incorrect classifications.
When applied to this task, the RMSprop optimizer reached 84% accuracy, although it performed slightly worse than Adam. The RMSprop confusion matrix demonstrated that TP totalled 92, with TN counting 101, while FP equalled 23 and FN equalled 14. The values of false positives and false negatives in RMSprop evaluation reveal that the algorithm has greater difficulty than Adam when detecting proper diabetic versus non-diabetic category divisions. The effectiveness of the RMSprop optimizer is restricted when applied to complex datasets because it leads to increased classification error rates.
The SGD optimizer produced an 80% accuracy, the worst outcome of the three evaluated algorithms. The SGD confusion matrix displayed a TP of 90 and TN of 94, while FP reached 25 and FN attained 21. Incorrect positive and negative diagnoses from SGD demonstrate poor discrimination capability between classification groups, resulting in major classification errors. The wide application of SGD fails to achieve optimal results for this specific task when used without proper parameter adjustments.
The experiment results demonstrate the performance variations and implementation challenges between optimization algorithms in LSTM models. Adam generated the optimum accuracy, but its higher error rates for false positive and false negative classifications show that the model has problems with precise data classification. The performance of RMSprop and SGD remained reasonable, but misclassification proved more challenging. Exposure to parameter optimization methods will allow these optimizers to improve their performance levels, leading to diminished misclassification errors and stronger classification accuracy results.

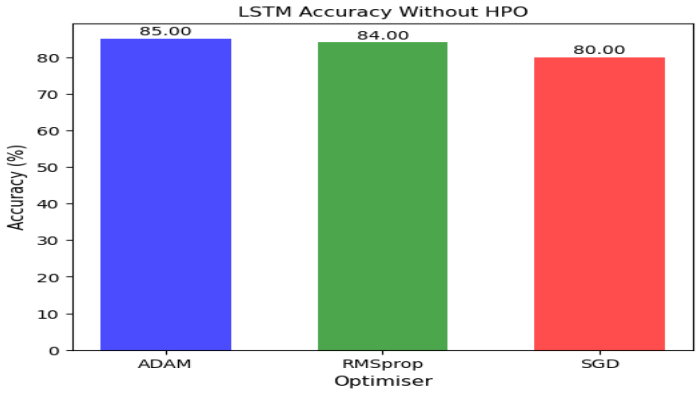

Figure 9: Performance of LSTM without HPO

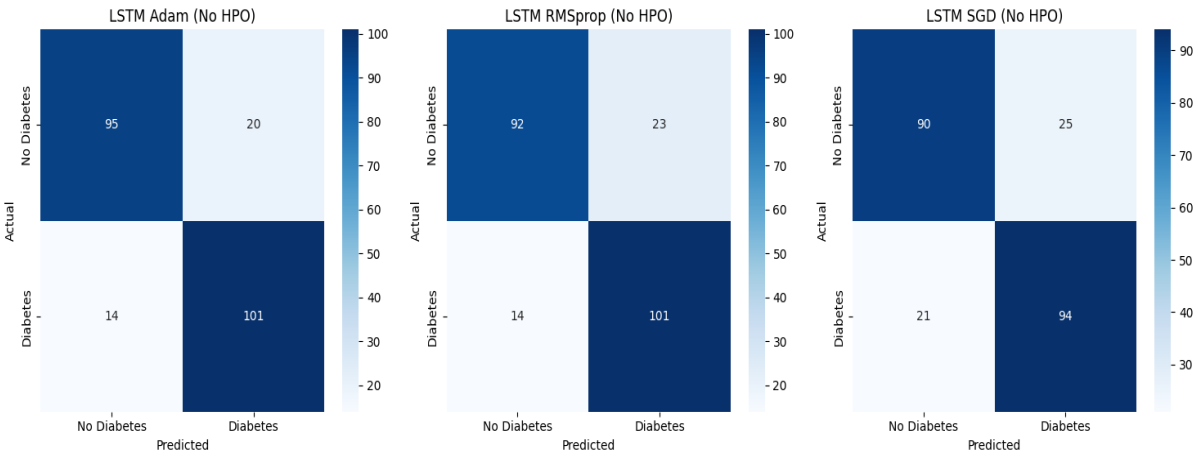

Figure 10: Confusion Matrics for LSTM without HPO

4.1.4 Performance of LSTM with Hyperparameter Optimization (HPO)

The evaluation of the LSTM model with hyperparameter optimization (HPO) was conducted to assess enhancements in classification performance. The model used hyperparameter tuning of Adam, RMSprop and SGD optimization algorithms to test their efficiency and accuracy improvements. The accuracy analysis of each algorithm appears in Figure 11, together with confusion matrix visualizations for further understanding in Figure 12.

After optimizing its hyperparameters, the Adam optimizer delivered an accuracy level of 90%. The confusion matrix yielded TP = 345, TN = 340, FP = 50 and FN = 40 results. The Adam optimizer combined with HPO achieved higher accuracy performance against the baseline of 85% while exhibiting a reduced rate of misclassification errors. Some false positive and false negative results persist despite improving overall classification through hyperparameter optimization because the model encounters limitations in diabetic and non-diabetic case differentiation.

Through HPO, the RMSprop optimizer obtained 88% accuracy. RMSprop produced a confusion matrix with TP set to 325, TN equaled 335, FP totaled 55, and FN amounted to 50. The accuracy surpassed the baseline level (84%), but misclassification rates stayed significant, with higher false positive and false negative results. This suggests that while RMSprop benefits from hyperparameter optimization, it still struggles with properly classifying all instances, particularly when dealing with complex patterns in the data.

Although enhanced with HPO, the SGD optimizer delivered an accuracy rate of 80%, which ranks it as the least effective of the three optimization approaches. The confusion matrix for SGD contained TP values of 310, TN values of 300, FP values of 80, and FN values of 75. The performance improved from baseline according to HPO, yet the results indicated major classification errors with numerous false positive and false negative errors. Additional strategies or modifications may be necessary to upgrade the effectiveness of SGD when performing this competition.

The analysis of LSTM with optimized hyperparameters demonstrated that the Adam optimizer produced better classification accuracy than the other tested functionalities, RMSprop and SGD. The enhanced results from hyperparameter optimization do not eliminate complete misclassification errors since optimizers produce false positive and false negative outcomes. Additional model enhancement through advanced hyperparameter optimization alongside alternative algorithm assessment remains crucial for increasing the accuracy and reducing the errors during diabetes case classification.

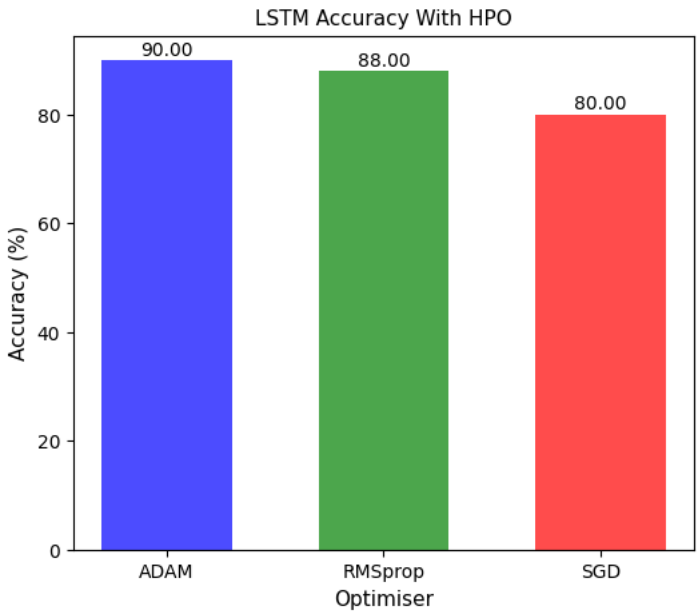

Figure 11: Performance of LSTM with HPO

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22S, 2025
https://www.theaspd.com/ijes.php

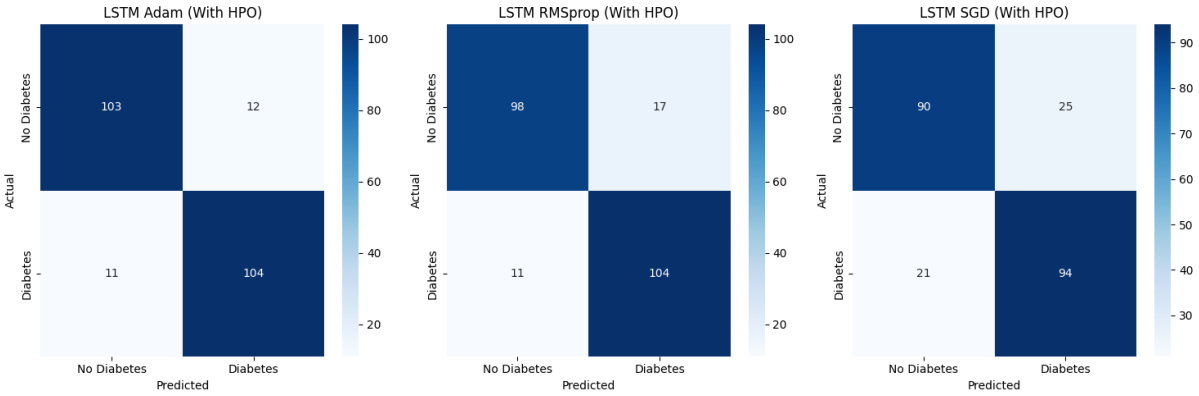

Figure 12: Confusion Matrics for LSTM with HPO

4.2 Discussion and Comparison

Various fundamental points arise from the comparison between SVM and LSTM models. The SVM models, particularly those using the RBF and sigmoid kernels, consistently performed better than the LSTM models, especially after hyperparameter optimization. After optimization, the RBF kernel produced an accuracy level of 94% while surpassing all other models, including LSTM, with 90% accuracy under Adam optimizer conditions. This suggests that SVM models, particularly with non-linear kernels, are more effective at classifying diabetes cases in this dataset, especially when fine-tuned through hyperparameter optimization.

In the given dataset, LSTM models demonstrated promising results but could not achieve better performance than the SVM models. The LSTM model failed to achieve maximum value from the data structure despite its sequential pattern detection capacity compared to SVM models. The enhanced outcome from HPO indicates that LSTM models maintain potential for additional development when utilizing the Adam optimizer for improved execution. The potential of LSTM to tackle text summarization depends on additional neural architecture optimizations for complete performance optimization.

The performance improvement of SVM and LSTM models relied heavily on applying effective hyperparameter optimization. After adjusting parameters, SVM achieved its best results by combining RBF and sigmoid kernels, but LSTM benefited the most from the Adam optimizer during optimization. Despite the better performance of the SVM models, the results highlight the importance of model selection and hyperparameter tuning in maximizing the predictive capabilities of both traditional machine learning models and deep learning architectures. Future studies on enhanced optimizers and hybrid methods might result in better performance and classification accuracy achievements. All information about accuracy measurements is presented in Table 2.

Table 2: Overall Results

	Model	Without HPO	With HPO
SVM	Linear	89%	90%
	Polynomial	89.13%	93%
	RBF	92.10%	94%
	Sigmoid	92.10%	92.20%
LSTM	Adam	85%	90%
	RMSprop	84%	88%
	SGD	80%	80%

5. CONCLUSION AND FUTURE WORK

In this paper, an examination has been conducted to evaluate how Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) networks perform in predicting diabetes by using Pima Indian Diabetes dataset data. Without HPO in place, the RBF kernel model achieved an accuracy level of 92.1%, according to study findings. After applying Genetic Algorithm genetic algorithm-based tuning, the RBF kernel improved significantly, reaching 94%, showcasing its strength in modelling non-linear relationships. Combined with the polynomial kernel, the sigmoid kernel achieved better accuracy through optimization, resulting in 91.19% and 91.12% final performance levels. Adam optimizer produced the

upper levels of accuracy at 85% for LSTM models without requiring any adjustments to their hyperparameters. The implementation of GA resulted in 90% accuracy for Adam-optimized LSTM and 88% accuracy for RMSProp, while the performance of SGD remained at 80%. The experimental results validate adaptive optimizers such as Adam as performance boosters for GA-driven optimization strategies. The study proves that hyperparameter optimization techniques lead to higher accuracy levels in result predictions for SVM and LSTM models which detect diabetes cases. The results demonstrate why applying genetic algorithms leads to successful optimization of machine learning models for healthcare systems. In future work the authors demands the evaluation of alternate optimization approaches together with ensemble and Bayesian Optimization and the development of healthcare data sets that work towards making models more understandable to attain easy clinical deployment.

REFERENCES

[1] G. M. Barlow and R. Mathur, "Type 2 Diabetes and the Microbiome," Journal of the Endocrine Society, vol. 7, no. 2, Dec. 2022, doi: 10.1210/jendso/bvac184.
[2] S. Asif et al., "Advancements and Prospects of Machine Learning in Medical Diagnostics: Unveiling the Future of Diagnostic Precision," Archives of Computational Methods in Engineering, Jun. 2024, doi: 10.1007/s11831-024-10148-w.
[3] A. Barnwal et al., "Sugar and stops in drivers with insulin-dependent type 1 diabetes," Accident Analysis & Prevention, vol. 173, p. 106692, Aug. 2022, doi: 10.1016/j.aap.2022.106692.
[4] A. Abed Mohammed, P. Sumari, and kassem Attabi, "Hybrid K-means and Principal Component Analysis (PCA) for Diabetes Prediction," International Journal of Computing and Digital Systems, vol. 15, no. 1, pp. 1719–1728, Jun. 2024, doi: 10.12785/ijcds/1501121.
[5] L. S. Ashour et al., "Non-overlapping Patch-Based Pre-trained CNN for Breast Cancer Classification," Iraqi Journal for Computer Science and Mathematics, vol. 6, no. 2, Jun. 2025, doi: 10.52866/2788-7421.1271.
[6] B. H. Jawad, O. A. Alwesabi, N. Abdullah, and A. A. Mohammed, "Hybrid Artificial Neural Network Activation Function to Reduce Water Wastage in Agricultural Irrigation," IEEE Access, vol. 13, pp. 93302–93322, 2025, doi: 10.1109/ACCESS.2025.3573678.
[7] A. A. Mohammed, H. Alqaseer, P. Sumari, and M. M. A. Zaid, "Durian Fruit Diseases Detection Using Convolutional Neural Networks (CNN) and Pre-trained Models," 2025, pp. 97–116. doi: 10.1007/978-981-96-5220-4_9.
[8] M. M. Abd Zaid, A. A. Mohammed, and P. Sumari, "Classification of Geographical Land Structure Using Convolution Neural Network and Transfer Learning," Journal of Computer Science, vol. 20, no. 12, pp. 1580–1592, Dec. 2024, doi: 10.3844/jcssp.2024.1580.1592.
[9] M. M. Abd Zaid, A. A. Mohammed, and P. Sumari, "Remote Sensing Image Classification Using Convolutional Neural Network (CNN) and Transfer Learning Techniques," Journal of Computer Science, vol. 21, no. 3, pp. 635–645, Mar. 2025, doi: 10.3844/jcssp.2025.635.645.
[10] M. M. Abd Zaid, A. Abed Mohammed, and P. Sumari, "Classification of Road Features Using Convolutional Neural Network (CNN) and Transfer Learning," International Journal of Computing and Digital Systems, vol. 17, no. 1, pp. 1–12, Jan. 2025, doi: 10.12785/ijcds/1571031764..
[11] P. Rajendra and S. Latifi, "Prediction of diabetes using logistic regression and ensemble techniques," Comput. Methods Prog. Biomed. Update, vol. 1, p. 100032, 2021. doi: 10.1016/j.cmpbup.2021.100032.
[12] A. Driouich, A. El Ouissari, K. El Moutaouakil, and I. Akharraz, "A metaheuristic for fuzzy density based SVM and confidence SMOTE for early prediction of diabetes," Stat., Optim. Inf. Comput., 2024. doi: 10.19139/soic-2310-5070-1348.
[13] M. Y. Shams, Z. Tarek, and A. M. Elshewey, "A novel RFE-GRU model for diabetes classification using PIMA Indian dataset," Sci. Rep., vol. 15, no. 1, p. 982, 2025. doi: 10.1038/s41598-024-82420-9.
[14] M. F. Aslan and K. Sabanci, "A novel proposal for deep learning-based diabetes prediction: converting clinical data to image data," Diagnostics, vol. 13, no. 4, p. 796, 2023. doi: 10.3390/diagnostics13040796.
[15] S. Gill and P. Pathwar, "Prediction of diabetes using various feature selection and machine learning paradigms," in Modern Approaches in Machine Learning & Cognitive Science: A Walkthrough, Cham: Springer, 2022, pp. 133-146.
[16] A. Naseem, R. Habib, T. Naz, M. Atif, M. Arif, and S. Allaoua Chelloug, "Novel Internet of Things based approach toward diabetes prediction using deep learning models," Front. Public Health, vol. 10, p. 914106, 2022. doi: 10.3389/fpubh.2022.914106.
[17] P. N. Srinivasu, J. Shafi, T. B. Krishna, C. N. Sujatha, S. P. Praveen, and M. F. Ijaz, "Using recurrent neural networks for predicting type-2 diabetes from genomic and tabular data," Diagnostics, vol. 12, no. 12, p. 3067, 2022. doi: 10.3390/diagnostics12123067.
[18] F. Khademi, M. Rabbani, H. Motameni, and E. Akbari, "A weighted ensemble classifier based on WOA for classification of diabetes," Neural Comput. Appl., pp. 1-9, 2022.
[19] H. Naz and S. Ahuja, "Deep learning approach for diabetes prediction using PIMA Indian dataset," Journal of Diabetes & Metabolic Disorders, vol. 19, no. 1, pp. 391–403, Jun. 2020, doi: 10.1007/s40200-020-00520-5.
[20] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona, "A survey on missing data in machine learning," Journal of Big Data, vol. 8, no. 1, p. 140, Oct. 2021, doi: 10.1186/s40537-021-00516-9.
[21] S. Mohapatra and N. Chaudhary, "Statistical Analysis and Evaluation of Feature Selection Techniques and implementing Machine Learning Algorithms to predict the Crop Yield using Accuracy Metrics," Engineered Science, 2022, doi: 10.30919/es8d787.

[22] D. Elreedy and A. F. Atiya, "A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance," Information Sciences, vol. 505, pp. 32–64, Dec. 2019, doi: 10.1016/j.ins.2019.07.070.
[23] S. Albahra et al., "Artificial intelligence and machine learning overview in pathology & laboratory medicine: A general review of data preprocessing and basic supervised concepts," Seminars in Diagnostic Pathology, vol. 40, no. 2, pp. 71–87, Mar. 2023, doi: 10.1053/j.semdp.2023.02.002.
[24] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised Classification Algorithms in Machine Learning: A Survey and Review," 2020, pp. 99–111. doi: 10.1007/978-981-13-7403-6_11.
[25] M. Aslani and S. Seipel, "Efficient and decision boundary aware instance selection for support vector machines," Information Sciences, vol. 577, pp. 579–598, Oct. 2021, doi: 10.1016/j.ins.2021.07.015.
[26] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review," Information, vol. 15, no. 4, p. 235, Apr. 2024, doi: 10.3390/info15040235.
[27] T. Alghamdi, "Prediction of Diabetes Complications Using Computational Intelligence Techniques," Applied Sciences, vol. 13, no. 5, p. 3030, Feb. 2023, doi: 10.3390/app13053030.
[28] J. Oruh, S. Viriri, and A. Adegun, "Long Short-Term Memory Recurrent Neural Network for Automatic Speech Recognition," IEEE Access, vol. 10, pp. 30069–30079, 2022, doi: 10.1109/ACCESS.2022.3159339.
[29] G. van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," Artificial Intelligence Review, vol. 53, no. 8, pp. 5929–5955, Dec. 2020, doi: 10.1007/s10462-020-09838-1.
[30] F. I. Alarsan and M. Younes, "Best Selection of Generative Adversarial Networks Hyper-Parameters Using Genetic Algorithm," SN Computer Science, vol. 2, no. 4, p. 283, Jul. 2021, doi: 10.1007/s42979-021-00689-3.
[31] L. Tani, D. Rand, C. Veelken, and M. Kadastik, "Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics," The European Physical Journal C, vol. 81, no. 2, p. 170, Feb. 2021, doi: 10.1140/epjc/s10052-021-08950-y.
[32] T. Imam, K. M. Ting, and J. Kamruzzaman, "z-SVM: An SVM for Improved Classification of Imbalanced Data," 2006, pp. 264–273. doi: 10.1007/11941439_30.

[33] Y. Li, X. Li, M. Guo, C. Chen, P. Ni, and Z. Huang, "Regression analysis and its application to oil and gas exploration: A case study of hydrocarbon loss recovery and porosity prediction, China," Energy Geoscience, vol. 5, no. 4, p. 100333, Oct. 2024, doi: 10.1016/j.engeos.2024.100333.
[34] Ghojogh, Benyamin, et al. "Reproducing Kernel Hilbert Space, Mercer's Theorem, Eigenfunctions, Nystr\" om Method, and Use of Kernels in Machine Learning: Tutorial and Survey." arXiv preprint arXiv:2106.08443 (2021).
[35] D.-X. Zhou and K. Jetter, "Approximation with polynomial kernels and SVM classifiers," Advances in Computational Mathematics, vol. 25, no. 1–3, pp. 323–344, Jul. 2006, doi: 10.1007/s10444-004-7206-2.
[36] Q. Liu, C. Chen, Y. Zhang, and Z. Hu, "Feature selection for support vector machines with RBF kernel," Artificial Intelligence Review, vol. 36, no. 2, pp. 99–115, Aug. 2011, doi: 10.1007/s10462-011-9205-2.
[37] A. Oleynik, A. Ponosov, V. Kostrykin, and A. v. Sobolev, "Spatially localized solutions of the Hammerstein equation with sigmoid type of nonlinearity," Journal of Differential Equations, vol. 261, no. 10, pp. 5844–5874, Nov. 2016, doi: 10.1016/j.jde.2016.08.026.
[38] Y. Zhao, J. Pei, and H. Chen, "Multi-layer radial basis function neural network based on multi-scale kernel learning," Applied Soft Computing, vol. 82, p. 105541, Sep. 2019, doi: 10.1016/j.asoc.2019.105541.
[39] A. Chaudhuri and K. De, "Fuzzy Support Vector Machine for bankruptcy prediction," Applied Soft Computing, vol. 11, no. 2, pp. 2472–2486, Mar. 2011, doi: 10.1016/j.asoc.2010.10.003.
[40] Arqué Martínez, Arnau. Kernel methods with mixed data types and their applications. MS thesis. Universitat Politècnica de Catalunya, 2023.
[41] A. Ruiz and P. E. Lopez-de-Teruel, "Nonlinear kernel-based statistical pattern analysis," IEEE Transactions on Neural Networks, vol. 12, no. 1, pp. 16–32, 2001, doi: 10.1109/72.896793.
[42] J. Kruppa et al., "Probability estimation with machine learning methods for dichotomous and multicategory outcome: Theory," Biometrical Journal, vol. 56, no. 4, pp. 534–563, Jul. 2014, doi: 10.1002/bimj.201300068.
[43] M. Cai and J. Liu, "Maxout neurons for deep convolutional and LSTM neural networks in speech recognition," Speech Communication, vol. 77, pp. 53–64, Mar. 2016, doi: 10.1016/j.specom.2015.12.003.
[44] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," Neurocomputing, vol. 503, pp. 92–108, Sep. 2022, doi: 10.1016/j.neucom.2022.06.111.
[45] G. Bellec, F. Scherr, E. Hajek, D. Salaj, R. Legenstein, and W. Maass, "Biologically inspired alternatives to backpropagation through time for learning in recurrent neural nets," Jan. 2019, [Online]. Available: http://arxiv.org/abs/1901.09049
[46] L. N. Smith and N. Topin, "Super-convergence: very fast training of neural networks using large learning rates," in Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications, T. Pham, Ed., SPIE, May 2019, p. 36. doi: 10.1117/12.2520589.
[47] E. Agliari, F. Alemanno, M. Aquaro, and A. Fachechi, "Regularization, early-stopping and dreaming: A Hopfield-like setup to address generalization and overfitting," Neural Networks, vol. 177, p. 106389, Sep. 2024, doi: 10.1016/j.neunet.2024.106389.
[48] Hoffer, Elad, Itay Hubara, and Daniel Soudry. "Train longer, generalize better: closing the generalization gap in large batch training of neural networks." Advances in neural information processing systems 30 (2017).
[49] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, May 2013, pp.

8609–8613. doi: 10.1109/ICASSP.2013.6639346.

[50] G. M. Foody, "Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification," Remote Sensing of Environment, vol. 239, p. 111630, Mar. 2020, doi: 10.1016/j.rse.2019.111630.