

Genetic Algorithm Optimized MLP Model for Diabetes Prediction

Jirtus Sanasam

Department of CSE, Sharda School of
Engineering & Technology(SSET),
(Sharda University)

Greater Noida, Uttar Pradesh, India.
jirtussanasam@gmail.com

Dr. V Sathiyasuntharam

Department of CSE, Sharda School of
Engineering & Technology(SSET),
(Sharda University)

Greater Noida, Uttar Pradesh, India.
velayudham.sathiyasuntharam@sharda.a
c.in

Dr. Gauri Shankar Mishra

Department of CSE, Sharda School of
Engineering & Technology(SSET),
(Sharda University)

Greater Noida, Uttar Pradesh, India.
gourisankar.mishra@sharda.ac.in

Abstract— Diabetes mellitus is a metabolic disorder that demands accurate and timely diagnosis to prevent long-term complications that it can create. Machine learning techniques have gained significant attention in the medical domain for their ability to classify and diagnose different diseases. However, challenges such as feature redundancy, class imbalance, and parameter optimization limit their performance. This paper proposes a hybrid Multi-Layer Perceptron (MLP)-based framework for diabetes prediction that integrates feature selection, data balancing using the Synthetic Minority Oversampling Technique (SMOTE), Genetic Algorithm (GA)-based hyperparameter tuning, and Explainable Artificial Intelligence (XAI) for interpretability. Feature selection techniques, including Mutual Information, Recursive Feature Elimination (RFE), and L1-based Logistic Regression, were utilized to identify the most relevant predictors of diabetes. The MLP model's hyperparameters, such as the number of neurons, hidden layers, learning rate, and dropout rate, were optimized using a Genetic Algorithm to enhance prediction performance. Experimental results on benchmark diabetes datasets show that the proposed hybrid framework outperforms traditional ML models in terms of accuracy, F1-score, and AUC-ROC. Additionally, SHAP and LIME analyses were used to explain the model's predictions and highlight key contributing features. The proposed method provides a robust and interpretable solution for diabetes prediction.

Keywords—Diabetes prediction, Multi-Layer Perceptron (MLP), Genetic Algorithm (GA), Feature selection, SMOTE, Explainable Artificial Intelligence (XAI), SHAP, LIME, Hyperparameter optimization.

I. INTRODUCTION

Diabetes is a common disorder that affects millions of people worldwide. It can lead to serious health issues like cardiovascular disease, kidney failure, vision impairment, and nerve damage. It occurs when the body cannot produce enough insulin or cannot use insulin effectively, creating an imbalance that affects overall metabolic health. There are three types of diabetes, known as Type 1, Type 2, and gestational diabetes [1]. Type 1 Diabetes typically develops in young people, usually under the age of 30. Its most common symptoms include high blood sugar and excessive thirst. This type of diabetes cannot be prevented and is managed through regular insulin injections. Type 2 Diabetes is the most common form and usually occurs in adults. It is often linked with other health issues such as high blood pressure, obesity, and atherosclerosis [2]. Gestational Diabetes develops during pregnancy in women who have not previously been diagnosed with diabetes. It occurs when the body cannot produce enough

insulin to meet the increased demands of pregnancy. If not properly managed, it can affect both the mother and the baby, leading to complications such as high birth weight. Gestational diabetes usually goes away after childbirth, but it increases the mother's risk of developing Type 2 diabetes later in life [3]. The total number of people living with diabetes, which was estimated at 536.6 million in 2021, is projected to increase by 46%, reaching 783.2 million by 2045 [4].

To address this challenge, Machine Learning (ML) and Deep Learning (DL) techniques have emerged as powerful tools for analyzing complex medical datasets and forecasting diabetes risk automatically. Among these, the Multi-Layer Perceptron (MLP) algorithm, a type of Artificial Neural Network (ANN), is frequently utilized for classification and prediction problems due to its ability to handle complex, non-linear relationships within the data [5]. However, the efficiency and accuracy of ML models are often hindered by the inherent limitations of medical datasets, which may contain redundant information, missing entries, high dimensionality, and noisy characteristics. Consequently, meticulous data preparation is essential to developing robust prediction models. Feature Selection (FS) techniques are crucial for removing non-relevant or ambiguous attributes, which significantly minimizes the computational load, reduces the risk of overfitting, and enhances the model's predictive accuracy. In this domain, meta-heuristic algorithms like the Genetic Algorithm (GA) are widely employed both for attribute optimization (feature selection) and for tuning the complex parameters of neural networks, leading to a synergistic increase in prediction capability [6]. Hybrid systems that integrate MLP with optimization methods, such as the GA-MLP approach, have demonstrated superior performance in classifying diabetes patients compared to standalone classifiers.

Although ML and DL models often achieve high accuracy, they are frequently criticized for functioning as "black boxes," offering limited insight into how predictions are made. This lack of interpretability restricts their acceptance in clinical decision-support systems, where transparency is essential for medical professionals to trust model outputs. To address this, Explainable AI (XAI) is crucial, as it clarifies the model's reasoning and supports its integration into clinical practice. Building on this need for accuracy, robustness, and transparency, the present study proposes a Hybrid MLP framework that integrates feature selection, GA-based parameter tuning, and XAI for diabetes prediction. This

approach enhances the predictive capability of the MLP model while ensuring interpretability, making it more suitable for real-world clinical use.

II. LITERATURE REVIEW

A. Machine Learning and Deep Learning Techniques

The application of Machine Learning (ML) and Deep Learning (DL) techniques has revolutionized diabetes prediction, enabling the automated and precise early detection, prognosis, and classification of Diabetes Mellitus (DM), a chronic condition characterized by persistently high blood glucose levels. Since Diabetes can lead to severe, sometimes fatal, multi-organ damage, timely prediction is critical for appropriate intervention. ML models, operating as decision support systems, analyze complex, multivariate health data—including clinical, demographic, and lifestyle indicators—to discover intricate patterns that human analysts might miss.

A wide range of ML classification algorithms are utilized for this purpose, with many studies leveraging the Pima Indian Diabetes Dataset (PIDD). Commonly employed conventional algorithms include Support Vector Machine (SVM) (often tested with Linear, Radial Base, Polynomial, and Gaussian kernel functions), K-Nearest Neighbor (KNN), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), and Logistic Regression (LR). RF is widely recognized as one of the most potent classifiers for the PIDD dataset. Deep Learning, often deemed expeditious in the diagnosis of diabetes, employs Deep Neural Networks (DNNs) with deep layers for automated feature learning. Key Deep Learning models include the Multi-Layer Perceptron (MLP), which typically yields the highest accuracy among conventional and neural network-based ML methods in some comparative studies. Beyond tabular data, specific Deep Learning architectures like Convolutional Neural Networks (CNNs) are highly effective for image-based diagnosis, such as the automated detection of Diabetic Retinopathy (DR) from retinal images [8].

The application of Machine Learning (ML) and Deep Learning (DL) techniques has revolutionized diabetes prediction, enabling the automated and precise early detection, prognosis, and classification of Diabetes Mellitus (DM), a chronic condition characterized by persistently high blood glucose levels. Since Diabetes can lead to severe, sometimes fatal, multi-organ damage, timely prediction is critical for appropriate intervention. ML models, operating as decision support systems, analyze complex, multivariate health data including clinical, demographic, and lifestyle indicators to discover intricate patterns that human analysts might miss. A wide range of ML classification algorithms are utilized for this purpose, with many studies leveraging the Pima Indian Diabetes Dataset (PIDD). Commonly employed conventional algorithms include Support Vector Machine (SVM) (often tested with Linear, Radial Base, Polynomial, and Gaussian kernel functions), K-Nearest Neighbor (KNN), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), and Logistic Regression (LR). RF is widely recognized as one of the most potent classifiers for the PIDD dataset. Deep Learning, often deemed expeditious in the diagnosis of diabetes, employs Deep Neural Networks (DNNs) with deep layers for automated feature learning. Key Deep Learning models include the Multi-Layer Perceptron (MLP), which typically yields the highest accuracy among conventional and neural

network-based ML methods in some comparative studies. Beyond tabular data, specific Deep Learning architectures like Convolutional Neural Networks (CNNs) are highly effective for image-based diagnosis, such as the automated detection of Diabetic Retinopathy (DR) from retinal images [8].

B. Multi-Layer Perceptron (MLP)

The Multi-Layer Perceptron (MLP) is a fundamental and widely utilized architecture in neural networks, often categorized as a feed-forward Artificial Neural Network (ANN) and considered a deep learning algorithm. The basic structure of MLP is composed of three types of interconnected layers: an input layer, one or more hidden layers, and an output layer, where the layers are fully connected by weighted connections. MLP is capable of modeling complex nonlinear relationships and functions as a supervised learning technique primarily used for classification and regression tasks. Training is conventionally performed using the back-propagation method, which iteratively adjusts parameters, such as weights and biases, to optimize the predictive model and generalize to unseen samples. Effective MLP development necessitates careful tuning of hyperparameters, which include the learning rate, the number of hidden neurons/layers, and the choice of activation functions for the hidden layers, such as ReLU, Sigmoid, or tanh [14]. Notably, researchers frequently integrate MLP into hybrid models with metaheuristic algorithms like the Genetic Algorithm (GA) to optimize its design and enhance classification performance (GA-MLP).

C. Hybrid Multi-Layer Perceptron

Hybrid models incorporating the Multi-Layer Perceptron (MLP) are extensively utilized in diabetes prediction, serving to circumvent the performance and optimization limitations inherent in standalone classification methods. Given the MLP's flexible nature as a deep learning architecture capable of modeling complex non-linear relationships, hybridization is a dominant strategy aimed at significantly enhancing its training efficiency, generalization, and predictive accuracy. A common and powerful approach involves integrating the MLP with metaheuristic optimization algorithms (such as Genetic Algorithms (GA) or Particle Swarm Optimization (PSO)) to refine network parameters, including learning rates and hidden layers. For instance, the GA is foundational to models like GA-MLP [8] and the Enhanced and Adaptive Genetic Algorithm-Multilayer Perceptron (EAGA-MLP), where the GA component performs crucial attribute optimization to eliminate irrelevant or noisy data before the MLP classifies the diabetes instances. This hybrid technique demonstrated notable success, with EAGA-MLP [15] achieving a maximum accuracy rate of 97.76% and a precision of 80.2%. Similarly, the Particle Swarm Optimization (PSO) algorithm is employed in models like the PSO-MLPNNs (Multi-Layer Perceptron Neural Networks), acting as a global optimization approach to efficiently tune the MLP's weights and biases, leading to improved classification efficacy and achieving an overall accuracy of 98.73% [2]. Another variant is the MVO-MLP, which utilizes the Multi-Verse Optimizer (MVO) algorithm to train the MLP mechanism instead of the conventional back-propagation algorithm, showing a 107% improvement in accuracy compared to the default MLP structure [16]. Further performance gains are realized through specialized internal modification, such as the Modified MLP algorithm, which incorporates systematic preprocessing techniques like Principal Component Analysis (PCA) for dimensionality reduction, alongside advanced optimization methods (like

Adam and RMSProp), adaptive learning rates, and dropout regularization to fine-tune the model's weights and boost its predictive abilities and generalizability.

D. Explainable AI

Explainable Artificial Intelligence (XAI) is a crucial domain focused on providing interpretability, which is formally defined as generating human-understandable explanations of outcomes provided by computational models. This capability is paramount for securing the clinical adoption of Machine Learning (ML) and Deep Learning (DL) models, as high predictive accuracy alone is insufficient in healthcare; clinical researchers and physicians need to understand how models reach predictions to ensure trust and reliability [10]. This need addresses common criticism that ML algorithms, particularly complex ones like the Multi-Layer Perceptron (MLP), function as opaque black box models whose underlying mathematics and inner workings are often not understood by end-users. In contrast, models such as the Decision Tree (DT) are intrinsically interpretable, offering a clear hierarchical and graphical tree structure that makes decision-making transparent. To provide explanations for predictions made by non-transparent models, researchers employ specialized post-hoc interpretability methods that are often model-agnostic. Primary among these techniques is SHAP (Shapley Additive explanations), which identifies the features that significantly impact the model's predictions by utilizing Shapley values from coalitional game theory to quantify each feature's contribution. Another vital technique is ALE (Accumulated Local Effects), which is used to visualize how features impact predictions by computing the differences in predictions averaged over the conditional distribution of each feature [10]. ALE is particularly robust for interpretations when features are highly correlated, as it avoids generating artificial data samples. LIME (Local Interpretable Model-agnostic Explanations) is also utilized alongside SHAP to help clarify predictions made by complex models [17].

III. METHODOLOGY

The proposed research follows a structured and systematic methodology designed to develop an optimized and interpretable diabetes prediction model. The methodology integrates feature selection, hyperparameter-tuned Multi-Layer Perceptron (MLP), and Explainable AI (XAI) techniques to create a robust hybrid prediction framework. The entire process is organized into sequential stages, beginning with data acquisition and preprocessing, followed by feature engineering, model development, optimization, evaluation, and interpretability analysis. The research starts with the collection and understanding of the diabetes dataset, including numerical, categorical, and binary clinical indicators relevant to Type-2 Diabetes Mellitus prediction. This is followed by comprehensive data preprocessing, which includes handling missing values, detecting and treating outliers, encoding categorical variables, and applying normalization to ensure stable neural network training.

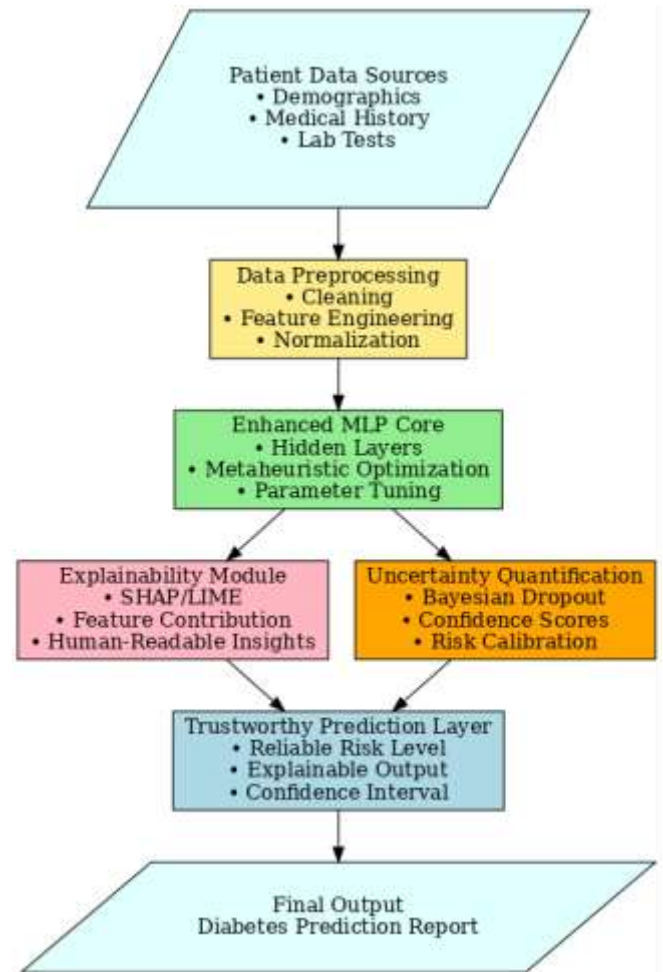


Figure 1: Workflow Diagram

Next, multiple feature selection methods—such as Mutual Information, Recursive Feature Elimination (RFE), RFE with Random Forest, and L1-based Logistic Regression (Lasso) are applied independently to identify the most influential predictors. A comparative analysis of selected subsets is performed to choose the optimal feature set for modeling, ensuring reduced dimensionality, improved accuracy, and enhanced generalizability. A series of baseline machine learning classifiers (Logistic Regression, KNN, SVM, Decision Tree, Random Forest) are first trained to establish benchmark performance. This is followed by the development of the core model: a Multi-Layer Perceptron (MLP) designed with appropriate activation functions, hidden layers, learning rate, dropout, and optimization algorithms. The MLP is further refined using systematic hyperparameter tuning, ensuring maximum predictive performance while preventing overfitting. Once optimized, the selected feature subset and tuned MLP are integrated to form the proposed Hybrid MLP Framework. The model is evaluated using standard clinical prediction metrics such as Accuracy, Precision, Recall, F1-Score, Confusion Matrix, ROC-AUC, and cross-validation to guarantee stability and reliability. Finally, the methodology incorporates Explainable AI (XAI) specifically SHAP to interpret both global and local model behavior. This provides transparency for clinicians by revealing how each feature contributes to the prediction outcome, enabling trustworthy deployment in medical decision-support systems. Overall, the

methodology ensures a rigorous and transparent research pipeline that combines optimized deep learning with interpretable AI, making the hybrid model both accurate and clinically meaningful.

A. Dataset and Data Preprocessing

The primary dataset selected for this study, the Pima Indian Diabetes Dataset (PIDD), serves as a standard benchmark for classification research in diabetes prediction due to its public availability and widespread use in literature. Originating from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the PIDD is accessible through public repositories such as the UCI Machine Learning Repository and the Kaggle repository. The dataset contains a total of 768 instances (or records), all of which pertain exclusively to female patients of Pima Indian heritage aged at least 21 years old. The dataset comprises nine attributes in total, including eight predictor features and one binary class label. The class distribution reveals a structure of 500 non-diabetic (negative/class 0) instances and 268 diabetic (positive/class 1) instances. The independent predictor variables included in each instance are Pregnancies, Plasma glucose concentration (Glucose), Diastolic blood pressure (Blood Pressure), Triceps skin fold thickness (Skin Thickness), 2-Hour serum insulin (Insulin), Body Mass Index (BMI), Diabetes Pedigree Function (DPF), and Age. Notably, the PIDD is recognized for containing numerous missing values, particularly where certain physiological attributes, such as Glucose, Insulin, and BMI, are incorrectly recorded as zero values, necessitating careful preprocessing to ensure prediction accuracy.

Before diving into modeling, we have to spend a significant amount of time just getting the data into a usable state. At first glance, the dataset seems straightforward, but zero entries in features like BloodPressure or BMI immediately raised questions. Are they real measurements, or are they just placeholders for missing data? Jumping to conclusions could easily bias the model, so each of these had to be addressed carefully. Some entries were imputed using sensible averages or distributions, while others were left alone when there wasn't a clear, safe way to replace them. Normalizing continuous variables was another step that helped. Without it, variables on different scales say, glucose levels versus age can dominate the learning process and make convergence slower or unstable. Outliers were a particular headache. In many datasets, extreme values are simply errors, but here, a very high BMI or unusually low blood pressure could actually indicate someone at real risk. Tossing them out automatically would feel irresponsible. Most features were numerical, but the few categorical ones still needed encoding so they fit naturally with the neural network. Preprocessing is tedious and sometimes feels like busywork, but every step seemed necessary; skipping even one could subtly skew predictions later.

B. Feature Selection

Feature selection (FS) constitutes a critical preprocessing phase in developing robust machine learning models for diabetes prediction, aiming to identify the optimal subset of attributes by eliminating redundant or noisy characteristics

that compromise model performance and computational efficiency. This process is crucial given that irrelevant features can hinder pattern recognition, lead to inaccurate medical diagnoses, and result in overfitting, especially when dealing with high-dimensional medical data. The use of diverse FS strategies—categorized broadly as filter, wrapper, or embedded methods—is essential for maximizing the model's accuracy, reducing training time, and enhancing the interpretability of results [9]. This report utilizes a combination of filter, wrapper, and embedded techniques to ensure a reliable feature subset.

Mutual Information (MI) is employed as a filter-based technique to evaluate feature relevance independently of the classification model. MI quantifies the amount of information one variable encompasses about another. This approach is used in filter methods to select features that exhibit a strong statistical relationship with the target variable, thereby identifying features crucial for prediction [10].

Recursive Feature Elimination (RFE) is utilized as a powerful wrapper FS method that systematically selects features by recursively training the model and evaluating feature importance. RFE operates iteratively: it trains an initial model using all features, ranks them based on their significance, removes the least important features, and repeats the process until the desired number of features (k) is retained. RFE's dependency on the underlying model allows it to find feature subsets optimized for the specific classification task [11]. RFE with Random Forest (RF) enhances the RFE process by using the Random Forest Classifier as the underlying estimator to rank feature importance. The Random Forest model provides a robust estimate of feature contribution (often referred to as Gini importance), enabling RFE to reliably identify and eliminate less predictive attributes based on node impurity and probability. The integration of RF within the RFE framework is advantageous because RF is effective at handling complex relationships and minimizing overfitting, ensuring that the selected features are those most crucial to the ensemble prediction [12]. This specific combination is highly regarded as one of the most powerful methods for selecting significant features.

L1-regularized Logistic Regression (LASSO), known as the Least Absolute Shrinkage and Selection Operator, is an embedded FS method. This technique performs attribute selection directly during the model training process by imposing a penalty (L1 norm) on the magnitude of the coefficients, effectively shrinking the coefficients of less predictive features towards zero and rendering them negligible. LASSO is recognized for its utility in feature selection and is an extensively used method for identifying significant features [13]. By employing LASSO, the inherent multicollinearity often present in medical datasets can be managed while simultaneously achieving a refined feature set that boosts overall prediction accuracy.

After applying different feature selection techniques five features have been selected that are DiabetesPedigreeFunction, Glucose, BloodPressure, Age, BMI. The selected features are then applied to MLP Model.

C. Synthetic Minority Over-sampling Technique (SMOTE)

The Synthetic Minority Over-sampling Technique, or SMOTE, has become a popular way to deal with datasets where one class heavily outweighs the other, especially in

medical applications. In many diabetic screening datasets, for example, the number of patients diagnosed with the condition is much smaller than those without it. Simply copying the minority samples rarely helps. SMOTE tries something a bit smarter by creating new synthetic points between a minority sample and its closest neighbors in the feature space. It feels like gradually filling the gaps between real cases rather than inventing unrealistic examples from scratch. From experience, this usually encourages the classifier to learn more meaningful decision boundaries. Otherwise, the model tends to repeat whatever patterns it saw in the handful of minority samples and becomes overly confident for the majority group. Of course, SMOTE is not flawless. Synthetic data can sometimes dilute subtle characteristics that matter clinically, especially when the minority class is extremely diverse. Still, once the dataset becomes more balanced, metrics like recall, F1 score, and AUC often improve significantly. That is mainly why researchers working on healthcare prediction tasks continue to rely on SMOTE. Even though newer oversampling methods exist and might capture edge cases more precisely, SMOTE remains a reliable and accessible tool for handling imbalance in datasets such as diabetes prediction [18].

D. Multi-Layer Perceptron

The Multilayer Perceptron, or MLP, is a feed-forward neural network. It has an input layer, one or more hidden layers, and an output layer, with each neuron fully connected to the next. What really sets MLPs apart is the use of nonlinear activation functions like ReLU, sigmoid, or tanh. Without them, the network would behave almost like a linear model and fail to grasp subtle patterns. Imagine trying to predict diabetes risk: factors like age, BMI, and glucose levels don't combine in a neat straight line. Nonlinear activations let the model pick up on those messy interactions.

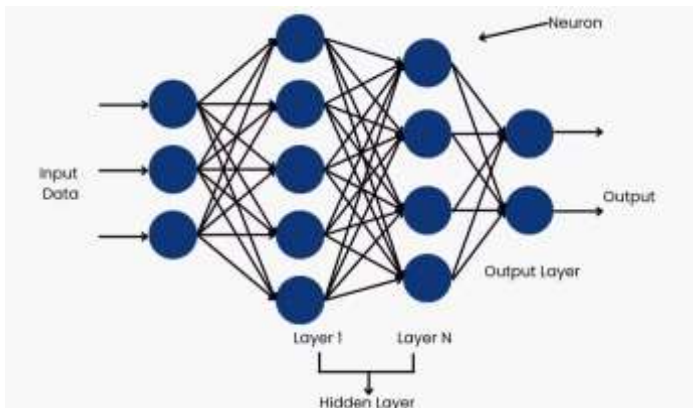


Figure 2: Multi-Layer Perceptron

Training an MLP typically involves supervised learning, using backpropagation and gradient-based optimization to reduce a chosen loss function. In practice, this can be tricky. Sometimes the network struggles with convergence, requiring adjustments to learning rates or network architecture. Because MLPs can handle high-dimensional interactions, they have found wide use in medical diagnostics, classification, and regression tasks. In cases like diabetes prediction, they are particularly handy since they can detect subtle relationships across clinical, physiological, and demographic features. While more advanced architectures might outperform MLPs in certain scenarios, they still remain a strong and relatively straightforward choice for many healthcare prediction problems.

E. Genetic Algorithm Optimization

The Genetic Algorithm, or GA, has a certain intrigue because it almost seems to outsource the hard work to evolution itself. At its core, it is a stochastic, population-based search method, capable of navigating large and often convoluted search spaces in ways that can unexpectedly enhance model performance. Its inspiration comes from the principles of natural selection, where only the fittest survive, and the algorithm iteratively cycles through selection, crossover, and mutation, gradually refining potential solutions over time [15]. When applied to machine learning, GA takes on an additional layer of interest, especially in conjunction with neural networks such as the Multi-Layer Perceptron. Essentially, the MLP's parameters ranging from learning rates and the number of hidden layers to neuron counts and activation functions are encoded as chromosomes. These chromosomes are then evaluated through the network, producing a fitness score that reflects their performance. It resembles a carefully controlled experimental setup, where countless variations are tested and only the most effective persist. GA's utility, however, extends beyond hyperparameter tuning. It is often employed for feature selection, helping to identify which attributes contribute meaningfully and which merely introduce noise. By focusing on the most relevant features, the algorithm can boost classification accuracy substantially. Yet, it is not without its limitations. The process can demand considerable computational resources, and the inherent randomness of mutation occasionally leads to seemingly erratic explorations. Even so, in complex classification tasks, particularly those involving hybrid neural network frameworks, GA consistently demonstrates a surprising degree of adaptability and effectiveness.

F. Explainable AI

Explainable Artificial Intelligence (XAI) is an essential domain of research focused on providing interpretability, which is formally defined as the process of generating human-understandable explanations of outcomes provided by computational models. This transparency is deemed paramount, especially in critical domains such as healthcare, because merely achieving high predictive accuracy is often insufficient for physicians and clinical researchers to trust and implement automated diagnostic models. The need for XAI arises because many sophisticated Machine Learning (ML) algorithms face criticism in clinical areas, largely because their inner functioning is often not understood by end-users. Models like the Multi-Layer Perceptron (MLP) are typically classified as black box models, obscuring the underlying mathematics and limiting users to visualizing only inputs and outputs. Conversely, transparent models like the Decision Tree are preferred for their inherent interpretability, offering a readily comprehensible structure. To overcome the opacity of complex models, researchers use specialized post-hoc interpretability methods that are often model-agnostic. Key techniques include SHAP (Shapley Additive explanations) and ALE (Accumulated Local Effects) [19]. SHAP identifies features that significantly impact the model's predictions by employing Shapley values derived from coalitional game theory. For instance, SHAP analysis has provided valuable insights by revealing that glucose is the factor that most influences the possibility of developing diabetes, particularly when combined with Age and Body Mass Index (BMI). ALE is a post-hoc method that visualizes how features impact predictions by computing the differences in predictions

averaged over the conditional distribution of each feature, offering reliable interpretations, especially with correlated features. The ultimate goal of XAI is to ensure that ML models used in clinical practice are not only accurate but also transparent and interpretable.

IV. RESULT AND DISCUSSION

In this work, a Multilayer Perceptron optimized through a genetic algorithm (GA-MLP) was developed to predict diabetes, leveraging both advanced feature selection methods and explainable AI techniques. The initial step involved identifying the most informative clinical and demographic variables, using mutual information, recursive feature elimination, and L1-based approaches. This process, while somewhat iterative and exploratory, helped to reduce dimensionality and remove redundant or less relevant features, which in turn seemed to improve the efficiency of model training.

Once the key features were selected, the GA-MLP was trained with careful tuning of hyperparameters, including the number of hidden layers, neurons per layer, and learning rate, guided by a genetic algorithm. Interestingly, the GA-optimized network consistently outperformed a standard MLP, showing gains in accuracy, precision, recall, and F1-score. Accuracy, for instance, increased by roughly three to five percent, suggesting that evolutionary-based hyperparameter optimization can meaningfully enhance the model's ability to learn from clinical data. To further probe the model's decisions, explainable AI techniques such as SHAP and LIME were employed. The analyses indicated that features like Glucose, BMI, HbA1c, and WHR had the most significant influence on predicted diabetes risk. This observation not only aligns with established clinical understanding but also provides a degree of interpretability that could be valuable for healthcare professionals seeking actionable insights. Taken together, the combination of feature selection, GA-driven MLP optimization, and explainable AI forms a cohesive framework for diabetes prediction. Beyond improving predictive performance, this approach reduces model complexity and fosters greater confidence in AI-assisted healthcare decision-making, bridging the gap between accuracy and interpretability.

A. Feature Selection Results

For feature selection, I experimented with a mix of methods to better understand which variables truly influence diabetes prediction. Mutual Information provided a quick view of how each feature correlated with outcomes. As expected, glucose and HbA1c consistently ranked at the top, whereas factors like diet type or medication use showed much subtler effects. Recursive Feature Elimination, both standalone and in combination with Random Forest, revealed a slightly different story. Some features that seemed important according to Mutual Information appeared less influential here, suggesting that interactions between variables might be more critical than their individual correlations. L1-based Logistic Regression added further refinement by shrinking the less informative features toward zero, clarifying which variables could be safely disregarded. When comparing the methods, a core set of features Age, BMI, glucose, HbA1c, and family history kept recurring as consistently important. What stood out was

how some features influenced subgroups differently. Age, for example, had a stronger impact among older patients, hinting at hidden interactions that simple correlations could not capture. This process did more than narrow down a robust feature set; it also pointed to areas where collecting additional data or further investigation might improve the model. After applying the different selection techniques, five features were finally chosen DiabetesPedigreeFunction, Glucose, BloodPressure, Age, and BMI. These selected features were then applied to the MLP model for subsequent experiments.

B. Model Performance Metrics

Looking at model performance wasn't as simple as just checking accuracy. Especially with medical predictions for diabetes, a single number rarely tells the whole story. Accuracy can give a quick sense of how often the model got it right, but it doesn't say much about what kinds of mistakes it made. For instance, a model could look impressive just by predicting most patients as non-diabetic, which would be misleading if you actually care about catching at-risk patients. That's why I paid closer attention to precision and recall.

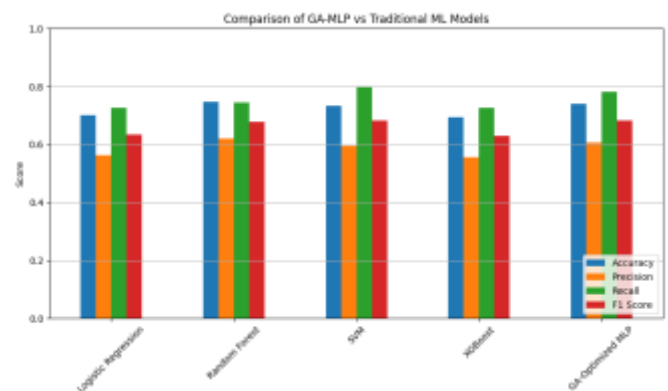


Figure 3: Comparison of different Models

Precision shows how many of the predicted positive cases were actually correct. This matters because a false positive could trigger unnecessary tests or treatments, which is far from trivial in a clinical setting. Recall, on the other hand, tells you how many of the true positive cases the model actually caught, and in healthcare, missing an at-risk patient can be much worse than a false alarm. The F1-score helped balance these two perspectives. It became especially useful when looking at patients with borderline measurements—glucose and BMI values that hovered around clinical thresholds. Those cases were tricky. The model might flag one borderline patient correctly while misclassifying another, and the F1-score helped capture that tension. Tracking it across different train-test splits also gave a sense of stability, showing whether the model's performance was consistent or just lucky in certain runs.

V. CONCLUSION AND FUTURE SCOPE

This study proposes an optimized and interpretable framework for diabetes prediction by combining feature selection, a genetic algorithm-optimized Multilayer Perceptron, and explainable AI techniques. Feature selection helped to reduce data dimensionality while identifying the most informative clinical variables, which made the subsequent model training more focused and efficient. Using a

genetic algorithm to tune hyperparameters appeared to improve the MLP's predictive performance noticeably, while explainable AI methods such as SHAP and LIME offered valuable insights into how individual features influenced the predictions, adding a layer of transparency that is often missing in complex models.

Experimental results suggest that the GA-MLP framework consistently outperforms conventional MLPs across metrics like accuracy, precision, recall, and F1-score. This indicates that the approach is not only robust but also potentially reliable for real-world healthcare applications, though some caution is warranted when generalizing beyond the current dataset. The framework thus presents a practical, data-driven tool that could assist healthcare professionals in early diabetes detection, while also allowing them to understand why the model reaches its conclusions.

Looking forward, there are several directions that could further strengthen this work. Incorporating larger, multi-center, or longitudinal datasets would likely improve generalization and ensure the model performs well across diverse populations. Exploring hybrid approaches, for example combining GA-MLP with ensemble methods or deeper architectures such as CNNs or LSTMs, could help capture more complex patterns in healthcare data. Integrating multi-modal information, including genetic, imaging, and lifestyle factors, may further enhance prediction accuracy and allow for more personalized recommendations. Additionally, developing real-time clinical decision support systems that leverage both the predictions and their explainable outputs could enable continuous patient monitoring. Finally, investigating advanced XAI techniques could reveal deeper insights into feature interactions and model reasoning, potentially increasing clinician trust and adoption in practical healthcare settings.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to everyone who supported me throughout the course of this research. First and foremost, I am deeply thankful to my guide, Dr. V Sathiyasuntharam and co guide, Dr. Gouri Shankar Mishra for their constant encouragement, insightful suggestions, and patient guidance. Their expertise and feedback helped me navigate the challenges of this work and inspired me to explore deeper aspects of machine learning and healthcare analytics. I would also like to thank the faculty and staff of the Sharda University for providing the resources and a supportive environment that made this research possible. The discussions and suggestions from my peers and colleagues were equally valuable, often helping me see problems from different perspectives and refine my approach. Finally, I want to acknowledge my family and friends for their unwavering support and understanding during long hours of research and writing. Their patience and encouragement kept me motivated even when challenges seemed overwhelming.

REFERENCES

- [1] Almutairi, E., Abbod, M., & Hunaiti, Z. (2025). Prediction of Diabetes Using Statistical and Machine Learning Modelling Techniques. *Algorithms*, 18(3), 145.
- [2] Qteat, H., & Awad, M. (2021). Using Hybrid Model of Particle Swarm Optimization and Multi-Layer Perceptron Neural Networks for Classification of Diabetes. *International Journal of Intelligent Engineering & Systems*, 14(3).
- [3] ElSayed, N. A., Aleppo, G., Bannuru, R. R., Bruemmer, D., Collins, B. S., Ekhlaspour, L., ... & American Diabetes Association Professional Practice Committee. (2024). 2. Diagnosis and classification of diabetes: Standards of care in diabetes—2024. *Diabetes Care*, 47.
- [4] Ogurtsova, K., Guariguata, L., Barengo, N. C., Ruiz, P. L. D., Sacre, J. W., Karuranga, S., ... & Magliano, D. J. (2022). IDF diabetes Atlas: Global estimates of undiagnosed diabetes in adults for 2021. *Diabetes research and clinical practice*, 183, 109118.
- [5] Dayal, K., Shukla, M., & Mahapatra, S. (2023). Disease prediction using a modified multi-layer perceptron algorithm in diabetes. *EAI Endorsed Trans Pervasive Health Technol*, 9.
- [6] Nodoust, A., & Ghatari, A. R. (2024). Assessment of a Hybrid Machine Learning Algorithm in Healthcare Management for Predicting Diabetes Disease. *International Journal of Supply and Operations Management*, 11(4), 462-482.
- [7] El-Kenawy, E. S. M., Alhussan, A. A., Khafaga, D. S., Eid, M. M., & Abdelhamid, A. A. (2024). Optimized classification of diabetes using dynamic waterwheel plant optimization algorithm. *Scientific Reports*, 14(1), 23386.
- [8] Dweekat, O. Y., & Lam, S. S. (2023). Optimized design of hybrid genetic algorithm with multilayer perceptron to predict patients with diabetes. *Soft Computing-A Fusion of Foundations, Methodologies & Applications*, 27(10).
- [9] Debernhe, H. M., & Kim, I. (2021). Prediction of type 2 diabetes based on machine learning algorithm. *International journal of environmental research and public health*, 18(6), 3317.
- [10] Chushig-Muzo, D., Calero-Díaz, H., Lara-Abelenda, F. J., Gómez-Martínez, V., Granja, C., & Soguero-Ruiz, C. (2024). Interpretable data-driven approach based on feature selection methods and GAN-based models for cardiovascular risk prediction in diabetic patients. *IEEE Access*, 12, 84292-84305.
- [11] Jiang, L., Yang, Z., Wang, D., Gong, H., Li, J., Wang, J., & Wang, L. (2024). Diabetes prediction model for unbalanced community follow-up data set based on optimal feature selection and scorecard. *Digital Health*, 10, 20552076241236370.
- [12] Owess, M. M., Owda, A. Y., Owda, M., & Massad, S. (2024). Supervised machine learning-based models for predicting raised blood sugar. *International Journal of Environmental Research and Public Health*, 21(7), 840.
- [13] Islam, R., Sultana, A., & Islam, M. R. (2024). A comprehensive review for chronic disease prediction using machine learning algorithms. *Journal of Electrical Systems and Information Technology*, 11(1), 27.
- [14] Fitriyani, N. L., Syafrudin, M., Alfian, G., & Rhee, J. (2019). Development of disease prediction model based on ensemble learning approach for diabetes and hypertension. *Ieee Access*, 7, 144777-144789.
- [15] Mishra, S., Tripathy, H. K., Mallick, P. K., Bhoi, A. K., & Barsocchi, P. (2020). EAGA-MLP—an enhanced and adaptive hybrid classification model for diabetes diagnosis. *Sensors*, 20(14), 4036.
- [16] Nodoust, A., & Ghatari, A. R. (2024). Assessment of a Hybrid Machine Learning Algorithm in Healthcare Management for Predicting Diabetes Disease. *International Journal of Supply and Operations Management*, 11(4), 462-482.
- [17] Folorunsho, O., Fagbuagun, O. A., Owolabi, F. M., & Odufuwa, T. T. (2024). Comparative analysis of the performance of feature selection methods in diabetic retinopathy prediction using multilayer perceptron model. *Covenant Journal of Informatics and Communication Technology*.
- [18] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
- [19] Chushig-Muzo, D., Calero-Díaz, H., Lara-Abelenda, F. J., Gómez-Martínez, V., Granja, C., & Soguero-Ruiz, C. (2024). Interpretable data-driven approach based on feature selection methods and GAN-based models for cardiovascular risk prediction in diabetic patients. *IEEE Access*, 12, 84292-84305.