

Diabetes Prediction in Healthcare

Tushar Jain, Rushikesh Gholap, Sushmitha Rajeswari Muppa and Vedavarshita Nunna

Abstract

As the prevalence of diabetes continues to rise globally, early prediction and intervention have become imperative for effective healthcare management. This research explores the application of Multi Layer Perceptron classifiers to predict diabetes, aiming to identify the most accurate and robust approach. The study evaluates the performance of various optimization tricks on Multi Layer Perceptron, on a comprehensive dataset of health parameters. The research involves the analysis of a diverse range of features, such as age, body mass index, blood pressure, and cholesterol levels, to train and test the classifier. Results indicate that individual classifiers exhibit varying degrees of accuracy, sensitivity, and specificity. To enhance predictive capabilities, Xavier Initialization and Adam update methods is proposed. The resulting accuracy is then calculated based on this architecture. The research evaluates the proposed architecture's performance against other standalone classifiers and ensemble did earlier, considering metrics such as precision, recall, F1-score. Furthermore, the study provides insights into the critical features contributing to diabetes prediction. This contributes to the understanding of the underlying factors influencing the disease and facilitates personalized healthcare strategies. In conclusion, this research highlights the importance of employing Multi Layer Perceptron in diabetes prediction, showcasing their ability to enhance accuracy and reliability. The findings contribute valuable insights to the field of healthcare analytics, emphasizing the potential for deep learning to revolutionize early detection and intervention in diabetes, ultimately improving patient outcomes.

1 Background and Related Work

Diabetes mellitus has emerged as a major global health concern, with an escalating prevalence that poses significant challenges to healthcare systems worldwide. The chronic nature of diabetes and its associated complications necessitate proactive and personalized healthcare strategies. Early prediction of diabetes is crucial for timely intervention, enabling healthcare providers to implement preventive measures and improve patient outcomes. Traditional methods of diabetes prediction often rely on clinical risk factors and demographic information. However, with the advent of deep learning techniques, there is an opportunity to harness the power of computational models to analyze complex datasets and extract valuable insights. Deep learning classifiers, such as Multi Layer perceptron with Xavier initialization and Adam Update, have demonstrated promise in predicting diabetes based on diverse sets of health parameters.

The study on *Diagnosis of Diabetes Mellitus Using Gradient Boosting Machine*[1] explores applying Gradient Boosting for diabetes prediction. It combines weak models, like decision trees, iteratively correcting errors, contributing to the effectiveness of ensemble methods in diabetes diagnosis. In *Performance enhancement of diabetes prediction by finding optimum K for KNN classifier with feature selection method*,[2] the focus is on optimizing K-Nearest Neighbor (KNN) for diabetes prediction. The study explores finding the optimum K value and incorporates feature

selection for accurate diabetes prognosis, providing insights into optimizing KNN for improved classification accuracy. The study *Machine Learning Technique to Prognosis Diabetes Disease: Random Forest Classifier Approach*[3] applies Random Forest Classifier for diabetes prognosis. This ensemble approach, using multiple decision trees, enhances predictive accuracy, emphasizing the potential of Random Forest Classifier as a robust forecasting model for diabetes. *Nadeem et al.*[4] introduce a fusion-based ML method for diabetes onset prediction, combining Support Vector Machines (SVM) and Artificial Neural Networks (ANN). Achieving a classification accuracy of 94.67%, it surpasses other models, showcasing the effectiveness of combining SVM and ANN for diabetes prediction.

In summary, these studies collectively highlight the versatility of machine learning techniques in improving predictive models for diabetes mellitus, contributing to the ongoing advancement of predictive analytics in healthcare.

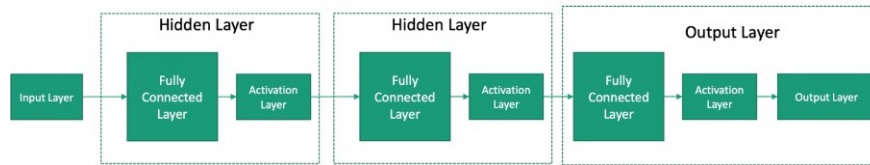


Figure 1: Basic MLP

2 Deep Learning architecture

2.1 Xavier Initialization

Xavier Initialization, also known as Glorot Initialization, aims to tackle the problem of vanishing or exploding gradients during neural network training. Xavier Initialization proposes a method for initializing the weights of a neural network layer such that the variance of the outputs of each layer remains constant during forward and backward propagation. Xavier Initialization sets the weights of each layer using a specific distribution based on the number of input and output units of the layer, aiming to keep the signal and gradients flowing through the network in a stable manner.

2.2 Adam Update

Adam (Adaptive Moment Estimation) is an optimization algorithm that combines the advantages of both AdaGrad and RMSProp, aiming to provide adaptive learning rates and momentum for efficient optimization. Adam addresses the limitations of traditional stochastic gradient descent (SGD) by introducing adaptive learning rates and momentum. Adam maintains two moving average estimates for each parameter: the first moment (mean) and the second moment (uncentered variance) of the gradients. It then computes the updates based on these estimates and incorporates bias correction to account for the initialization of these estimates. Beta1 controls the exponential decay rate for the moving average of past gradients. A value close to 1 (e.g., 0.9) means that the algorithm will consider a larger portion of past gradients, effectively increasing the momentum. This helps to stabilize and smoothen the optimization process, especially in cases where the gradients exhibit high variance or noise. Beta2 controls the exponential decay rate for the moving average of past squared gradients. A value close to 1 (e.g., 0.999) means that the algorithm will consider a larger

portion of past squared gradients, providing a more stable estimate of the variance.

A higher value of β_1 tends to increase the influence of past gradients on the optimization process, which can lead to a more conservative adaptive learning rate adjustment. This can be beneficial in scenarios where the optimization landscape is noisy or contains sharp fluctuations. β_2 influences the scaling of the adaptive learning rate by affecting the estimation of the second moment (variance). A higher value of β_2 results in a slower decay of the squared gradients' moving average, leading to a smaller effective learning rate adjustment for parameters with large variance.

A higher value of β_2 can make Adam more robust to outliers or noisy gradients by providing a smoother estimate of the gradient variance. This can help prevent overly aggressive updates in the presence of occasional large gradients.

3 Activations

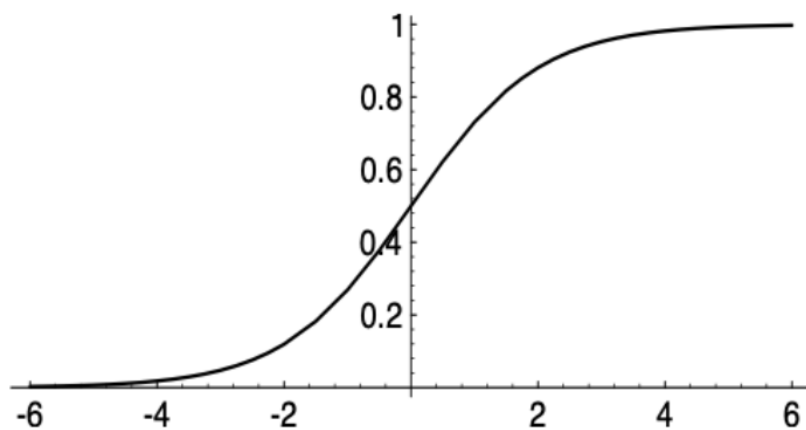
3.1 Sigmoid

Sigmoid is a widely used activation function that squashes the input values to the range $[0, 1]$, producing a smooth S-shaped curve. It is commonly used in binary classification tasks and as the output activation in the final layer of a neural network for binary classification.

Range: Sigmoid outputs values between 0 and 1, making it suitable for representing probabilities in binary classification tasks.

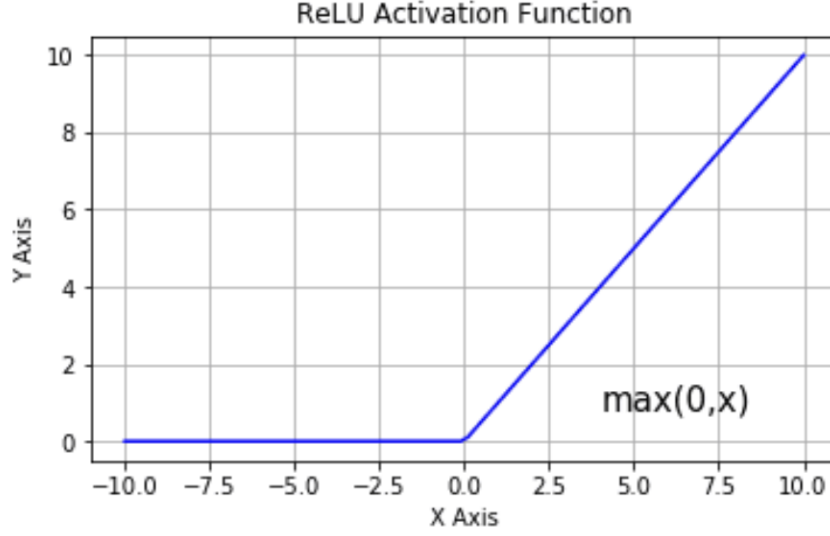
Smoothness: Sigmoid is smooth and continuously differentiable, facilitating gradient-based optimization methods such as gradient descent.

Monotonicity: Sigmoid is monotonically increasing, ensuring that increasing the input value results in an increasing output value.



3.2 Rectified Linear Unit

ReLU is an activation function commonly used in neural networks, particularly in deep learning architectures. It introduces non-linearity by outputting the input directly if it is positive, and zero otherwise.



Sparsity: ReLU activation introduces sparsity by zeroing out negative values, which can help prevent the vanishing gradient problem and accelerate convergence during training.

Computational Efficiency: ReLU is computationally efficient to compute and has a simple derivative, making it well-suited for deep learning architectures with large-scale datasets.

Improved Training: ReLU has been shown to alleviate the vanishing gradient problem associated with activation functions like sigmoid and tanh, enabling deeper networks to be trained more effectively.

4 Effects of error function

4.1 Mean Squared Error (MSE)

MSE is a commonly used metric for regression problems that measures the average squared difference between the predicted values and the actual values. A lower MSE indicates that the model's predictions are closer to the actual values, while a higher MSE suggests larger discrepancies between the predicted and actual values. It penalizes larger errors more severely due to the squaring operation, making it sensitive to outliers in the data.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

4.2 Log Loss

Log Loss, also known as Binary Cross-Entropy, is a common metric used for binary classification problems, measuring the difference between the predicted probabilities and the actual binary labels. Log Loss for binary classification is defined as:

$$LogLoss = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

it measures the accuracy of the predicted probabilities. It penalizes incorrect predictions more severely, especially when the predicted probability diverges significantly from the true label. Also minimizing Log Loss directly corresponds to maximizing the likelihood of the observed labels under the predicted probabilities.

4.3 Cross-Entropy

Cross-Entropy is a generalization of Log Loss to multi-class classification problems, measuring the difference between the predicted class probabilities and the true class labels. Cross-Entropy for multi-class classification is defined as:

$$Cross - Entropy = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m y_{ij} \log(\hat{y}_{ij})$$

It penalizes deviations between the predicted class probabilities and the true class labels, providing a measure of the model's classification accuracy across multiple classes. minimizing Cross-Entropy leads to maximizing the likelihood of the observed class labels under the predicted probabilities.

5 Evaluation of results

Preceding model input, the dataset undergoes cleaning, and data values are normalized. Random under-sampling for the majority class achieves a balanced dataset exceeding 20,000 rows. A 4:1 ratio splits the dataset into training and validation sets. Post-training, classifiers undergo evaluation through various metrics using the confusion matrix (Fig. 2). TP (true positive) signifies correctly predicted positive instances, TN (true negative) for negatives, FP (false positive) for incorrect positive predictions, and FN (false negative) for inaccuracies in negative predictions. Evaluation metrics encompass accuracy, sensitivity, specificity and Precision-Recall curve. Multi Layer Perceptron with 4 layers, and 15, 32, 16 neurons in first, second and third hidden layers respectively outperforms other diabetes models, attaining the highest accuracy (90.79%). Nonetheless, accuracy alone is insufficient for comprehensive model assessment and prediction.

		Actual Values					
		Positive (1)	Negative (0)				
Predicted Values	Positive (1)	TP	FP				
	Negative (0)	FN	TN				

Confusion Matrix:						
		precision	recall	f1-score	support	
	0	0.88	0.91	0.90	1719	
	1	0.91	0.88	0.89	1681	
accuracy				0.90	3400	
macro avg		0.90	0.89	0.89	3400	
weighted avg		0.90	0.90	0.89	3400	

Figure 2: Confusion matrix and best classifier metrics

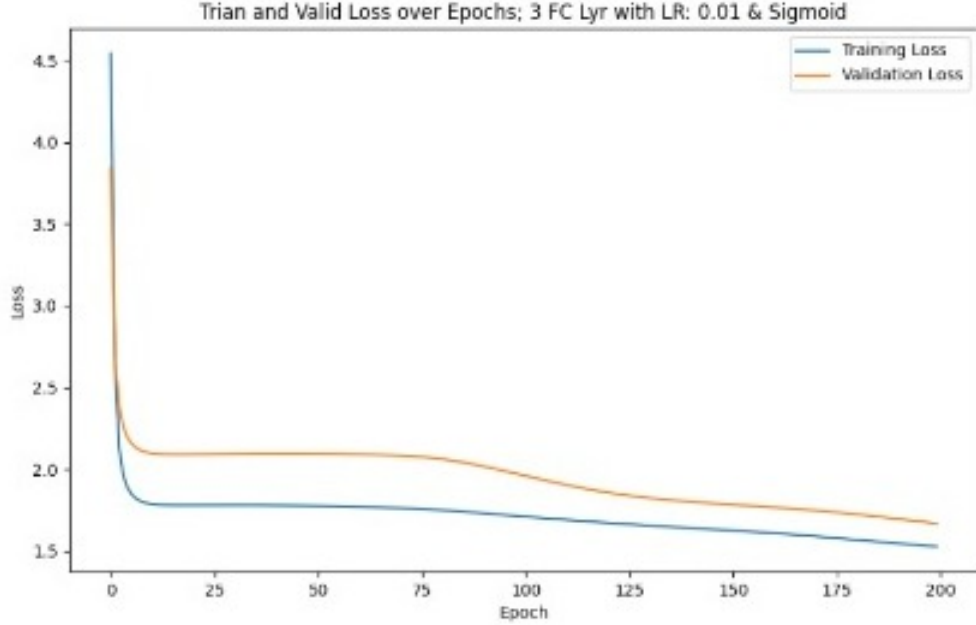


Figure 3: Loss vs No. of Epochs for best model

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

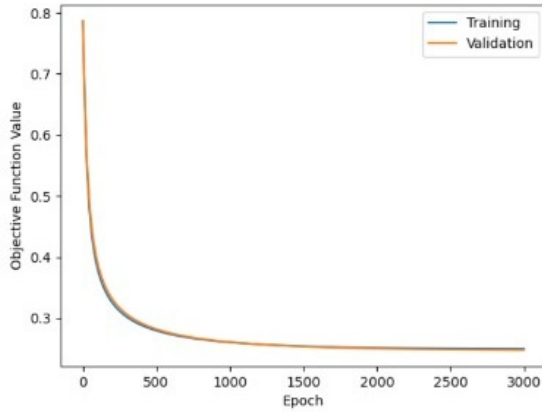
The prediction of false positives can result in misdiagnosis and the inefficient utilization of healthcare resources. Enhancing the precision of diagnostic models serves as a crucial solution to mitigate this issue. Precision specifically measures the accuracy of positive predictions, achieved by tallying correctly predicted positive samples (TP) and dividing by the total number of positive predictions (TP + FP), whether correct or incorrect. This model achieves a precision of (91%).

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

Recall, also known as sensitivity, akin to precision, seeks to determine the proportion of accurately detected true positives. It achieves this by dividing the correctly predicted positive samples (TP) by the total number of positives, whether correctly or incorrectly predicted as positive (TP, FN). In the area of medical practice, avoiding false negatives as much as possible is of paramount importance. The recall of this model is around (88%). Recall assesses the number of accurately predicted positive instances out of all potential positive predictions. Avoiding false negatives is crucial, as overlooking the presence of the disease can lead to delayed treatment and real harm.

$$\text{F-measure} = 2(\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

A robust diabetes detection system needs to minimize both missed diagnoses and misdiagnoses. However, accuracy and specificity often present conflicting performance metrics. The F-measure



	precision	recall	f1-score	support
0	0.87	0.90	0.88	1677
1	0.90	0.87	0.88	1723
accuracy			0.88	3400
macro avg	0.88	0.88	0.88	3400
weighted avg	0.88	0.88	0.88	3400

Figure 4: Metrics for second best model

takes into account both precision and recall, ranging from 0 to 1. A maximum value of 1 indicates perfect precision and recall, while a minimum value of 0 suggests that either precision or recall is zero. Our MLP model attains the highest F1 score at 90%, showcasing superior performance across all evaluation metrics. Our second best model architecture has an F1-score of 88% as mentioned in Figure 4. It is a 4 layered network with 10 neurons each and trained for 1500 epochs.

6 Analysis and Conclusions

Our research’s primary contribution lies in creating a Multi Layer Perceptron models for early diabetes detection. We investigated and experimented with various activation layers, objective/error functions, gradient optimization techniques to predict diabetes likelihood, by achieving the highest accuracy at 90%. The study focused on evaluating diabetes prediction using crucial features. Leveraging the advanced classification capabilities of deep learning algorithms, our model holds substantial potential to assist medical practitioners significantly in the diagnosis process.

Multi Layer Perceptron, a robust technique in deep learning, has been applied to enhance the predictive performance of models in the context of diabetics prediction datasets. In the pursuit of refining diagnostic accuracy, multiple diverse models were employed to generate individual predictions. Subsequently, the MLP with 4 layers, learning rate 0.1, ADAM optimization and Xavier initialization has achieved significant accuracy, even better than the previously explored machine learning models like Decision Trees, Naive Bayes, KNN and Ensemble voting as shown in Figure 5. Through the utilization of deep learning, this research endeavors to harness the strengths of various models, ultimately culminating in a consolidated prediction mechanism that excels in its ability to discern patterns and nuances within the intricate landscape of diabetics prediction.

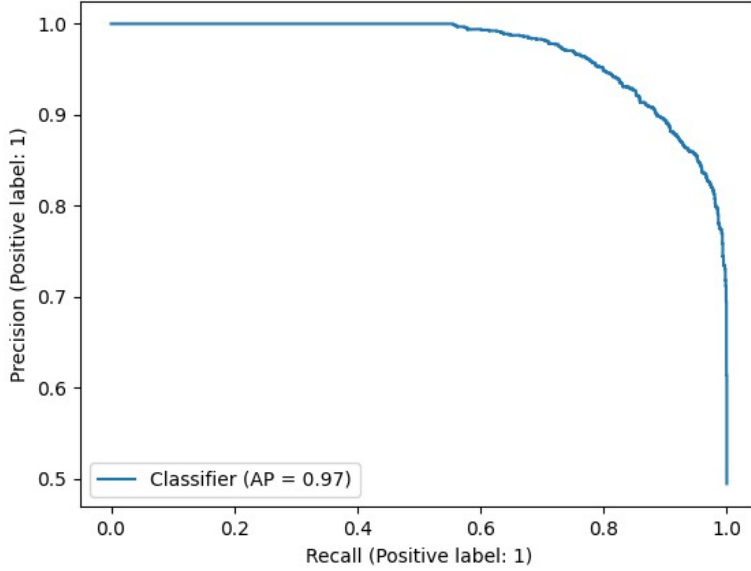


Figure 5: Precision-Recall Curve

Classifier	Accuracy	Precision	Recall	F1-Score
DT	81.78	82	82	82
NB	80.72	81	81	81
LR	81.15	81	81	81
KNN	83.34	84	83	83
Ensemble Voting	84.79	85	85	85

Figure 6: Classifier metrics for non-Deep Learning models

7 Future Work

The integration of temporal analysis into healthcare data studies, particularly in diabetes prediction, is crucial for recognizing the dynamic nature of such data. This approach involves examining how predictive models’ performance evolves over time, discerning patterns, fluctuations, and variations in accuracy. Understanding how models adapt to evolving trends is pivotal for long-term efficacy, ensuring relevance and effectiveness in real-world applications. Temporal analysis not only captures nuanced changes in health parameters but also enhances models’ robustness. By aligning with the evolving nature of healthcare data, this approach addresses the dynamic landscape of diabetes prevalence and risk factors. Incorporating temporal aspects enriches models’ adaptability, contributing to sustained accuracy and utility in healthcare decision-making.

Moreover, exploring advanced deep learning architectures like recurrent neural networks (RNNs) and attention mechanisms is essential. These techniques help model temporal dependencies in patient data, while integrating multimodal data sources such as genetic, lifestyle, and medical records

Gender	Gender refers to the biological sex of the individual, which can have an impact on their susceptibility to diabetes. (Female - 59%, Male - 41%)
Age	Age is an important factor as diabetes is more commonly diagnosed in older adults. Age ranges from 0-80 in our dataset.
Hypertension	Hypertension is a medical condition in which the blood pressure in the arteries is persistently elevated. It has values a 0 or 1 where 0 indicates they don't have hypertension and for 1 it means they have hypertension.
Heart_disease	Heart disease is another medical condition that is associated with an increased risk of developing diabetes. It has values a 0 or 1 where 0 indicates they don't have heart disease and for 1 it means they have heart disease.
Smoking_history	Smoking history is also considered a risk factor for diabetes and can exacerbate the complications associated with diabetes. In our dataset we have 5 categories i.e not current, former, No Info, current, never and ever.
BMI	BMI (Body Mass Index) is a measure of body fat based on weight and height. Higher BMI values are linked to a higher risk of diabetes. The range of BMI in the dataset is from 10.16 to 71.55. BMI less than 18.5 is underweight, 18.5-24.9 is normal, 25-29.9 is overweight, and 30 or more is obese.
HbA1c_level	HbA1c (Hemoglobin A1c) level is a measure of a person's average blood sugar level over the past 2-3 months. Higher levels indicate a greater risk of developing diabetes. Mostly more than 6.5% of HbA1c Level indicates diabetes.
Blood_glucose_level	Blood glucose level refers to the amount of glucose in the bloodstream at a given time. High blood glucose levels are a key indicator of diabetes.
Diabetes	Diabetes is the target variable being predicted, with values of 1 indicating the presence of diabetes and 0 indicating the absence of diabetes.

Figure 7: DataSet

enhances predictive accuracy and personalized treatment strategies. Dynamic Exploration, analyzing predictive model performance over time, is crucial for understanding their evolving efficacy in diabetes prediction. This holistic approach considers the complexity of healthcare data and leverages cutting-edge techniques to improve predictive capabilities and inform personalized interventions. By combining temporal analysis with advanced deep learning architectures and multimodal data integration, researchers can develop more accurate and effective predictive models for diabetes management and healthcare decision-making.

References

- [1] Diagnosis of Diabetes Mellitus Using Gradient Boosting Machine <https://www.mdpi.com/2075-4418/11/9/1714>
- [2] Performance enhancement of diabetes prediction by finding optimum K for KNN classifier with feature selection method <https://ieeexplore.ieee.org/abstract/document/9214129>
- [3] Machine Learning Technique to Prognosis Diabetes Disease: Random Forest Classifier Approach https://link.springer.com/chapter/10.1007/978-981-16-2164-2_19
- [4] An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators <https://www.sciencedirect.com/science/article/pii/S2772442522000582#b7>