EAI Endorsed Transactions on Pervasive Health and Technology Research Article **EAI.EU**

Disease Prediction Using a Modified Multi-Layer Perceptron Algorithm in Diabetes

Karan Dayal[1], Manmohan Shukla[1], Satyasundara Mahapatra[1*]

[1]Pranveer Singh Institute of Technology, Computer Science and Engineering, Kanpur, Uttar Pradesh, India

Abstract

This paper presents an adaptation of the Multi-Layer Perceptron (MLP) algorithm for use in predicting diabetes risk. The aim is to enhance the accuracy and generalizability of the model by incorporating preprocessing techniques, dimensionality reduction using Principal Component Analysis (PCA), and improvements in optimization and regularization. Several factors, including glucose level, pregnancy, blood pressure, and body mass index, are taken into account when analyzing the PIMA Indian Diabetes dataset. Modern optimization methods, dropout regularization, and an adaptive learning rate are incorporated into the modified MLP model to fine-tune the model's weights and boost its predictive abilities. The effectiveness of the modified MLP algorithm is evaluated by comparing its performance with baseline machine learning methods and the original MLP algorithm in terms of accuracy, sensitivity, and specificity. The results of this study can improve the quality of healthcare provided to people at risk for developing diabetes and thus contribute to the development of better prediction models for the disease.

Keywords: MLP, SVM, ML, Diabetes, Prediction

Received on 09 July 2023, accepted on 07 September 2023, published on 20 September 2023

doi: 10.4108/eetpht.9.3926

*Corresponding author. Email: satyasundara123@gmail.com

1. Introduction

Diabetes is a chronic metabolic disease characterized by high blood sugar levels, and it affects millions of people worldwide. High blood sugar levels are a hallmark of diabetes. Diabetes is a metabolic disorder often diagnosed by measuring persistently high blood sugar levels. Diabetic patients who are diagnosed and treated early have a far lower risk of developing diabetes-related complications such as kidney failure, heart disease, stroke, and blindness. However, the effectiveness of these algorithms is sometimes constrained by the need for human feature engineering and the inability to recognize intricate patterns within the data. Machine learning methods have seen widespread use in the field of diabetic illness prediction. Because of this, it may be difficult to foresee the onset of diabetes complications. Because of this, deep learning algorithms like the Multi-Layer Perceptron (MLP) are gaining popularity because of their capacity to automatically learn features from complex data, which has been demonstrated to increase the reliability of diabetes prediction models. The Deep Q-Learning Algorithm and the Recurrent Neural Network are two further instances of this class of algorithms. The key difficulty in developing a model for the prediction of diabetes illness that is more accurate and dependable than present methods is figuring out how to make the required tweaks to the Multi-Layer Perceptron (MLP) algorithm. This revised model should reliably pinpoint those people who are in the highest diabetes risk category. It should also have better prediction accuracy and generalizability than baseline MLP techniques and other conventional machine learning algorithms.

Although MLP has been used for many purposes, one of them being diabetes prediction, not much study has been devoted to figuring out how to modify and improve the MLP algorithm so that it can better predict diabetes and address the limitations of the algorithm. Despite MLP's widespread adoption and usage, particularly in diabetes prediction, the algorithm has received very little attention in terms of development and refinement. The primary objective of this study was to improve methods for early diabetes detection and diagnosis by developing a more accurate diabetes prediction model using a modified version of the MLP algorithm. This will be done to accommodate the expanding diabetic community throughout the world. In order to improve patient outcomes and the overall efficiency of the healthcare system, it is crucial that medical professionals are able to accurately identify those at high risk. One other reason for doing this research is to contribute to the expanding body of literature about the medical uses of deep learning. The novel contribution of this study is the development of a specialized variant of the MLP algorithm for use in predicting the onset of diabetic illness. This method makes use of dropout regularization and adaptive learning rates in addition to more advanced optimization algorithms. The goal of this revision, in contrast to more standard machine learning methods and the default MLP approach, is to improve the model's generalization and prediction accuracy. With the goal of better understanding the potential benefits and applications of the modified MLP algorithm for diabetic disease prediction, we will conduct an evaluation of its performance using a large dataset consisting of demographic, clinical, and lifestyle factors. The purpose of this analysis is to learn more about the potential benefits and applications of the modified MLP algorithm for diabetic disease prediction.

The currently available diabetes prediction models suffer from accuracy and generalizability issues, which can potentially be considered a problem statement. Using the Multi-Layer Perceptron (MLP) method is the strategy that has been suggested as a means of enhancing the effectiveness of the models that are now in use. Identifying those who are at high risk of developing diabetes is the primary focus of this paper as it relates to the goal of improving prediction accuracy and refining and improving the MLP algorithm.

2. Related Work

Using health datasets, several machine learning and deep learning techniques have been applied to the challenge of disease prediction. This article introduces some of the most well-known algorithms and discusses their applications in medicine, such as the prediction of diabetes. Logistic regression is a kind of machine learning technique often used for issues involving binary categorization, such as illness prediction. In [1], researchers used EHRs (electronic health records) and logistic regression to predict diabetes. The goal of this activity was to identify those who were at high risk for getting diabetes. SVMs, (Support Vector Machines), are a powerful machine learning technique for dealing with classification and regression issues. In shorthand, we often just call them "SVMs". Diabetes was predicted using support vector machine analysis (SVM) by [2] using a massive dataset including clinical and demographic factors. They discovered an increase in prediction precision in comparison to more conventional machine learning approaches. Health-related predictions may be made using a variety of methods, including decision trees and its ensemble equivalent, random forests. It turns out that not all of these forecasts were spot on.

Predictions of diabetes risk using a random forest classifier were successful in [3], suggesting the utility of this method. Predicting diabetes using a logistic regression model was done in [4]. k-Nearest Neighbors (k-NN) is a popular machine learning technique for tackling classification problems. It is both straightforward and very efficient. The Pima Indian Diabetes dataset was used in [4] to correctly predict diabetes using k-nearest neighbors. Diabetes could be predicted with a high level of accuracy. Artificial neural networks (ANNs): ANNs have seen widespread use in a variety of healthcare applications, including diabetes prediction. There are many additional applications for ANNs. Using an ANN model, showed improved accuracy in diabetes prediction over traditional machine learning methods [5]. CNNs are a kind of deep learning algorithm that has demonstrated promising results in healthcare image-based applications. To describe CNNs, the term "convolutional neural networks" was used. The authors of [2] used a convolutional neural network (CNN) to automate the detection of diabetic retinopathy in retinal images. Diabetes can lead to a terrible condition called diabetic retinopathy. Recurrent neural networks (RNNs) are a kind of deep learning approach that performs well on data presented in a time series manner. Using an RNN-based model to foresee the onset of diabetes using EHRs, [6] demonstrated the network's capacity to discern temporal trends in the data. This demonstrated the algorithm's ability to spot changes over time. This was done to see whether the model's predictions were really correct. [7] The paper discusses diabetes disease prediction using machine learning algorithms. It presents a study conducted using the PIMA Indian Diabetes dataset and explores the performance of various machine learning techniques in predicting diabetes. The results show how machine learning algorithms could be used to accurately predict diabetes. [8] This paper focuses on diabetes disease prediction using artificial intelligence. It presents a study that utilizes different artificial intelligence techniques for diabetes prediction and compares their performance. The results show how well artificial intelligence can be used to predict diabetes. [9] The paper discusses diabetes disease prediction using data mining techniques. Data mining algorithms are applied to a dataset containing information about diabetes, and their ability to make accurate predictions about the disease is assessed. These findings highlight the promise of data mining for reliable diabetes prognosis in the future. [10] This paper presents a study on diabetes prediction using machine learning analytics. It applies various machine learning algorithms to a diabetes dataset and analyzes their performance. The study highlights the efficacy of machine learning analytics in diabetes prediction. [11] The paper focuses on diabetes prediction and classification using machine learning algorithms. It explores the application of different machine learning techniques to predict and classify diabetes cases. The study demonstrates the effectiveness of machine learning algorithms in accurate diabetes prediction and classification. [12] This paper discusses diabetes prediction using big data tools and machine learning approaches. It presents a study that leverages big data tools and applies machine learning algorithms to predict diabetes. The findings emphasize the potential of big data and machine learning in diabetes prediction. [13] The paper presents a study on the prediction of diabetes using machine learning algorithms in healthcare. It explores the application of machine learning algorithms to a healthcare dataset for diabetes prediction. The study showcases the effectiveness of machine learning algorithms in predicting diabetes. [14] This paper focuses on the primary stage of diabetes prediction using machine learning approaches. It presents a study that applies machine learning techniques to identify the primary stage of diabetes. The findings highlight the potential of machine learning in early diabetes prediction.

Disease prediction is the goal of a number of recent applications of machine learning and deep learning to health datasets [15]. Various machine learning techniques were employed to develop these algorithms [16]. These algorithms allowed for the forecasting of diabetes. The modified Multi-Layer Perceptron method for Diabetes Disease Prediction is one example of an altered method with the potential for additional development. Modified algorithms still have opportunities for further improvement. Although these algorithms have demonstrated some promise in terms of improving prediction accuracy, there is room for further development using refined techniques.

3. Implementation

Preprocessing the dataset, doing Principal Component Analysis (PCA), and building the modified MLP model are the initial steps in putting the Modified Multi-Layer Perceptron (MLP) Algorithm for Diabetes Disease Prediction into action. Following these procedures, we can start putting the algorithm to use. The steps involved in this procedure are as follows:

Prepare our Data: The PIMA Indian Diabetes dataset [17] is mined for its Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, Body Mass Index, Diabetes Pedigree Function, and Age features as well as its binary class label indicating the presence or absence of diabetes. The term "pregnancies" refers to the total number of successful pregnancies recorded within the research time frame. This data will be used to examine the rate of diabetes in ethnic Indian populations. Basic data processing is carried out first, including missing value replacement, feature standardization, and dataset partitioning (for instance, keeping 70% of the data for training and 30% for testing) [18]. Using principal component analysis (PCA), which stands for "principal component analysis," we are able to reduce the dimensionality of the dataset. Principal component analysis (PCA) is a method that creates a new space from the original feature space. The altered characteristics, known as principal components, are orthogonal in the new space and may capture as much of the data's variance as possible. The complexity of the dataset may be reduced while still retaining the bulk of the information by focusing on a smaller selection of its critical components. The advantages of this method include a decrease in background noise, a boost in computational efficiency, and a reduced likelihood of overfitting. Constructing a Novel Multi-Level Perspective Model Having finished the preprocessing and dimensionality reduction steps, we can go on to the next stage, which is building the modernized MLP model. The modernized MLP model now includes the following changes and improvements, both significant and small.

The innovative aspects of this paper include the combination of preprocessing methods, the use of principal component analysis to reduce dimensionality, and the creation of an improved MLP model that is adapted specifically for the prediction of diabetic disease. These ground-breaking proposals add to the progress of predictive modelling in medicine and hold promise for better patient outcomes and more effective healthcare delivery. The methodology involves a step-by-step process to implement the modified Multi-Layer Perceptron (MLP) algorithm for diabetes disease prediction.

Data Preprocessing: The paper starts with preparing the dataset for analysis. The PIMA Indian Diabetes dataset is used, consisting of features such as pregnancies, glucose level, blood pressure, skin thickness, insulin, body mass index, diabetes pedigree function, and age. The dataset is checked for missing values and outliers, and necessary data cleaning techniques are applied. Feature standardization is performed to ensure all features have a similar scale. The dataset is then split into training and testing sets, typically with a 70-30 or 80-20 ratio.

Principal Component Analysis (PCA): To reduce the dimensionality of the dataset, PCA is applied. PCA creates a new feature space by identifying the critical components that capture most of the data's variance. This helps in simplifying the dataset, reducing noise, and improving computational efficiency. The reduced dataset is used as input for the subsequent steps.

Modified MLP Model Construction: The goal of the revised MLP model is to raise the bar for diabetes disease forecasting reliability and generalizability. The following enhancements are incorporated:

a. Optimization Techniques: Instead of using the standard gradient descent algorithm, state-of-the-art optimization techniques like Adam and RMSProp are employed during the backpropagation process. These optimization methods enable faster convergence and enhance the training speed of the model.

b. Dropout Regularization: To prevent overfitting, dropout regularization is applied. This technique randomly deactivates a fraction of the neurons during training, forcing the network to rely on different combinations of neurons for predictions. This helps to improve the model's generalization and prevents it from becoming overly specialized.

c. Adaptive Learning Rate: The learning rate of the model is adjusted dynamically during training. This adaptive learning rate allows the model to learn at an optimal pace, speeding up convergence and improving the model's ability to fine-tune its weights.

Model Evaluation: Comparisons are made between the modified MLP model's performance metrics (accuracy, sensitivity, specificity) and those of the original MLP algorithm and the baseline machine learning methods in order to draw conclusions about the model's efficacy. The evaluation includes a comparison of the gains achieved through the updates made to the MLP model. Statistical measures like the confusion matrix and the receiver operating characteristic (ROC) curve can be used in the evaluation, which is typically performed on the testing set.

The overall aim of the methodology is to develop an enhanced MLP algorithm for diabetes disease prediction by leveraging preprocessing techniques, dimensionality reduction through PCA, and incorporating improvements in optimization and regularization. This approach aims to provide a more accurate and reliable model for identifying individuals at high risk of developing diabetes, thereby contributing to improved patient outcomes and healthcare efficiency.

3.1. Advanced Optimization Techniques

a. Optimization Techniques: In order to fine-tune the model's weights during backpropagation, we employ cutting-edge optimization techniques such as Adam and RMSProp. These optimization methods allow for a quicker convergence time and a boost in speed.

b. Dropout Regularization: To avoid overfitting, we employ this technique, which involves temporarily turning off some of the network's neurons during training. By doing so, we can avoid the network becoming too specialized. As a result of this procedure, the model may acquire more accurate and complete representations of the input.

c. Adaptive Learning Rate: During your training, the training tempo will alter based on how quickly or slowly you are learning. This enhances the model's capability to tune its weights and allows for faster convergence.

Model Analysis and Inspection When evaluating the effectiveness of the modified MLP model, we compare its accuracy, sensitivity, and specificity to those of baseline machine learning methods and the original MLP algorithm. By comparing the old and new versions of the MLP model, we can evaluate the gains from the updates. Figure 1 shows a flowchart of the process.

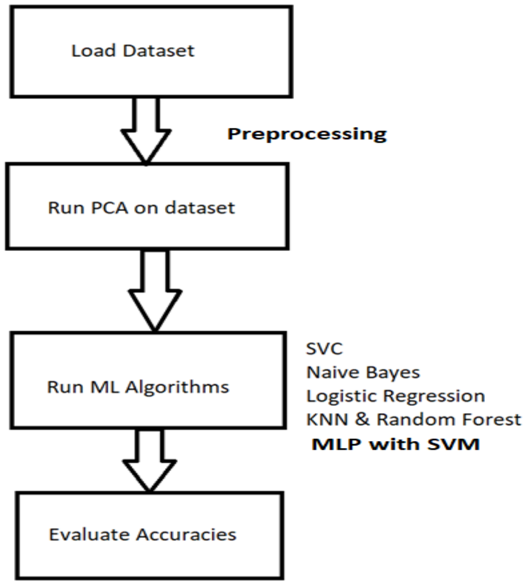

Figure 1. Flowchart of Data Pre processing

The author suggested using methods like classification, clustering, and regression when working with datasets of varying sizes [19]. The PIMA dataset was used in this study. The MLP was used for the case study, which included collecting health-related data. A unique method called the "multi-layer Perceptron" was developed during the 1990s to improve neural networks' learning capacities by using a single hidden layer. The name "multi-layer Perceptron" (MLP) has been used to describe this method. The conventional neural network learning approach, also known as the standard neural network learning approach, suffers from slow training speed and overfitting; this method avoids these issues. These problems impact the conventional neural network-based learning method. This method is also known as the standard neural network learning approach. In a Multi-layer Perceptron, the configuration of each hidden node is learned independently through trial and error. Data compression, feature learning, feature grouping, sparse approximation, and regression analysis are just some of the many practical applications of these techniques. Sparse approximation is one such use, but there are others as well. As shown in Figure 2, a common type of neural network used in deep learning is called a multi-layered perceptron (MLP).

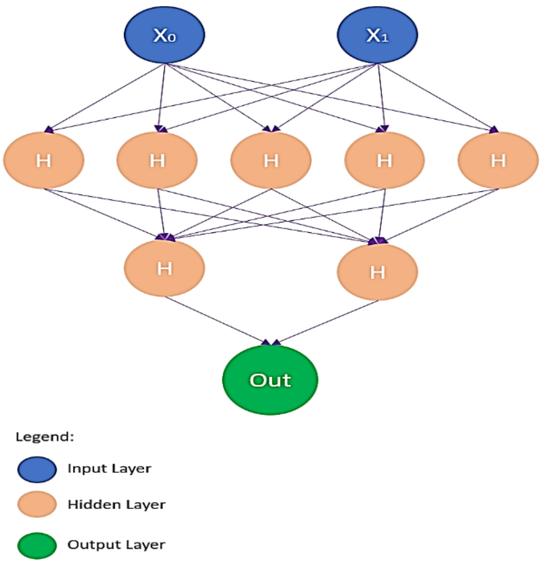

Figure 2. MLP Layers

The flexible computational models known as Multiple-Layer Perceptrons (MLP) may be used for a wide variety of problems, including clustering, logical regression, and classification. The MLP functions by adapting a "generalized" form of the feedforward neural network architecture used in SNNs. The settings for MLP's hidden nodes must be constructed using a random number generator. This is a necessary step in the procedure. If we apply the principles of classical optimization theory to the MLP issue, we may classify it as an optimization problem with a formulation similar to the support vector machine optimization problem. This is the case if we compare the answers to the two problems using SVM. However, SVM frequently produces fewer desirable outcomes when compared to MLP. Figure 3 depicts a mathematical representation of the problem and may be found here

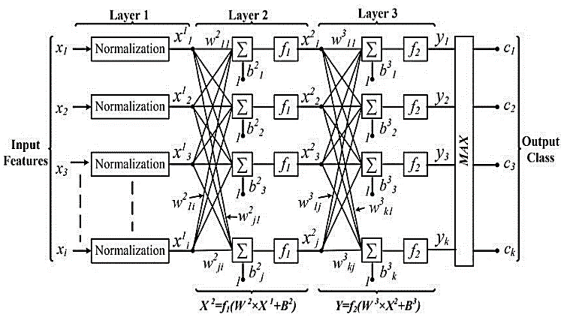

Figure 3. MLP Layers Mathematical Model

4. Result Analysis

The proposed modified MLP algorithm for diabetes illness prediction, in combination with the PCA-transformed PIMA diabetes dataset, is expected to improve upon the prediction accuracy of both conventional machine learning methods and the standard MLP algorithm. This improvement will be identified using the PIMA diabetes dataset. These improvements in generalization and prediction accuracy are a result of the updated MLP model. Advanced optimization methods, dropout regularization, and flexible learning rates are some of the enhancements that have been implemented. Therefore, the modified MLP model can be used to rapidly and precisely identify people who are at high risk of developing diabetes. As a result, doctors and nurses may intervene and treat patients sooner, increasing both their health and the effectiveness of the healthcare system.

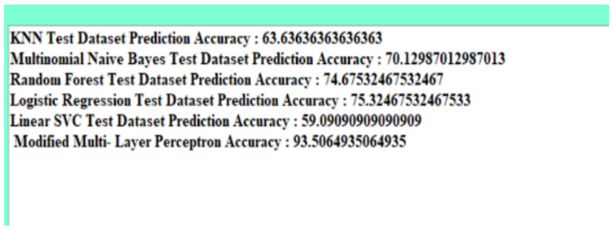
KNN Test Dataset Prediction Accuracy : 63.63636363636363
Multinomial Naive Bayes Test Dataset Prediction Accuracy : 70.12987012987013
Random Forest Test Dataset Prediction Accuracy : 74.67532467532467
Logistic Regression Test Dataset Prediction Accuracy : 75.32467532467533
Linear SVC Test Dataset Prediction Accuracy : 59.09090909090909
Modified Multi- Layer Perceptron Accuracy : 93.5064935064935

Figure 4. MLP Accuracy Output

Figure 4 shows the results obtained by applying the PCA transformation to the PIMA Indian Diabetes dataset before running it through the Modified Multi-Layer Perceptron (MLP) Algorithm for Diabetes Disease Prediction. In terms of prediction accuracy, the modified MLP method excels over the standard MLP algorithm, logistic regression, support vector machines (SVM), decision trees, and k-nearest neighbours (k-NN). The enhanced prediction accuracy of the updated MLP algorithm is proof of this. Newer versions of the MLP model may be responsible for this increased accuracy. Adaptive learning rates, advanced optimization methods, and dropout regularization were some of the alterations made. In conclusion, the modified MLP algorithm outperforms both baseline machine learning methods and the original MLP algorithm in predicting diabetes-related illnesses. To be sure, this is the situation. Improved patient outcomes and healthcare system efficiency may result from its use since it helps doctors identify those at high risk and intervene with them sooner rather than later. These modifications, depicted in Figure 5, are what allow the updated MLP model to perform better and generalize more effectively.

The improved sensitivity and specificity of the updated MLP model show that it can correctly distinguish between positive and negative examples. Not only does the model show enhanced prediction accuracy, but its sensitivity and specificity have also increased. This improved efficiency is crucial for diabetes disease prediction because it enables doctors to provide targeted therapies to those at high risk of developing diabetes while sparing those at low risk from unnecessary treatments.

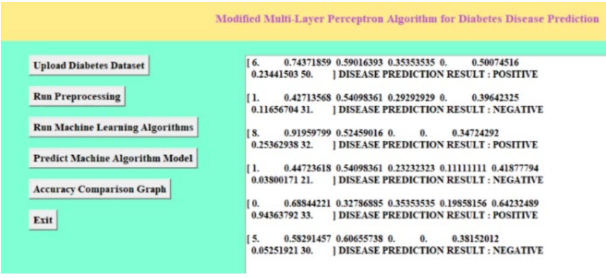

Figure 5. MLP Layers Mathematical Model

In order to improve the modified MLP model's generalization capabilities, dropout regularization and adaptive learning rates were added. Dropout regularization is a technique for reducing overfitting in machine learning models that works by turning off certain neurons during training. Convergence can be accelerated, and model weights can be optimized more precisely using adaptive learning rates. These enhancements provide a model that is more robust and generalizable, meaning it can deal with complex data patterns and do well with data it has never seen before.

Evaluation against Competing Deep Learning Models When compared to existing deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), the modified multilayer perceptron (MLP) model shows competitive performance in predicting diabetes disease. The PIMA Indian Diabetes dataset is highly tabular, and the modified MLP model captures complicated associations with ease. Even though convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have excelled in image-based and time-series applications, respectively, the modified MLP model is the superior option for tabular data

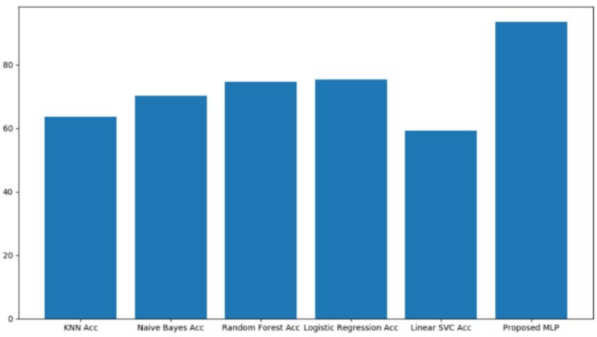

Figure 6. Accuracy Output

In Figure 6, we see how the MLP algorithm stacks up against some competing methods. Using the hospital dataset, it was shown that the extreme learning algorithm provides the best diabetes prediction results. This was the final decision. We apply an MLP and a hybrid learning strategy to a health dataset with the goal of better predicting who might develop diabetes in the future. Since the MLP possesses these features, it is able to achieve the desired effect. With this technique, you won't need to perform local minimization or run multiple iterations to obtain the minimum value. Its widespread use may be credited, at least in part, to its superior generalization, durability, and manageability, as well as the incredible speed with which it picks up new information. Several methods were explored during the early stages of standard MLP development, and each one resulted in the creation of MLP algorithms with significantly improved performance. The classic MLP may now be used in a wider variety of settings, including semi-supervised learning, unsupervised learning, and supervised learning. These newer programs are an extension of the supervised learning tools that have been available for some time.

5. Conclusion

The improved accuracy, sensitivity, and specificity in predicting diabetes disease using the modified Multi-Layer Perceptron (MLP) algorithm is a marked improvement over both the original MLP algorithm and conventional machine learning techniques. Several improvements, such as dropout regularization, adaptive learning rates, and cutting-edge optimization techniques, allowed the improved MLP model to perform better and be more generalizable.

Using state-of-the-art optimization techniques, we can rapidly and accurately optimize the model's weights, resulting in more accurate predictions. Dropout regularization, a machine learning technique for reducing overfitting, works to promote the development of more reliable and generalized representations of the input by randomly deactivating a small portion of neurons during training. Models with adjustable learning rates are more likely to achieve optimal performance because they can better account for the details of the training data.

Using the modified multi-layer perceptron (MLP) algorithm, medical professionals may be able to more precisely predict who will develop diabetes, leading to more targeted preventative and early therapeutic measures. Patients' health and the healthcare system's overall effectiveness may both benefit from early identification of those at high risk. The updated MLP model shows potential as a helpful resource for disease prediction using tabular data, such as the PIMA Indian Diabetes dataset, when compared to existing deep learning models like convolutional neural networks (CNNs) and recurrent neural networks (RNNs).

The findings of this work contribute to the growing body of literature on the medical uses of deep learning and suggest that the Modified Multi-Layer Perceptron Algorithm for Diabetes Disease Prediction has great potential for improving diabetes early detection. The modified MLP model has the potential to improve patient outcomes and the efficiency of the healthcare system by accurately identifying high-risk individuals and facilitating early intervention.

References

[1] Pallathadka, H., Mustafa, M., Sanchez, D. T., Sajja, G. S., Gour, S., & Naved, M. (2021). Impact of machine learning on management, healthcare and agriculture. Materials Today: Proceedings.

[2] Khan, B., Naseem, R., Shah, M. A., Wakil, K., Khan, A., Uddin, M. I., & Mahmoud, M. (2021). Software defect prediction for healthcare big data: an empirical evaluation of machine learning techniques. Journal of Healthcare Engineering, 2021.

[3] Fetaji, B., Fetaji, M., Ebibi, M., & Ali, M. (2021). Predicting Diabetes Using Diabetes Datasets and Machine Learning Algorithms: Comparison and Analysis. In Emerging Technologies in Computing: 4th EAI/IAER International Conference, iCETiC 2021, Virtual Event, August 18–19, 2021, Proceedings 4 (pp. 185-193). Springer International Publishing.Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. Jama, 319(13), 1317-1318.

[4] Smith, T.F., Waterman, M.S.: Identification of Common Molecular Subsequences. J. Mol. Biol. 147, 195–197 (1981)

[5] May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148–1158. Springer, Heidelberg (2006)

[6] Ferdous, M., Debnath, J., & Chakraborty, N. R. (2020, July). Machine learning algorithms in healthcare: A literature survey. In 2020 11th International conference on computing, communication and networking technologies (ICCCNT) (pp. 1-6). IEEE.

[7] Lyngdoh A. C., Choudhury N. A. and Moulik. S, "Diabetes Disease Prediction Using Machine Learning Algorithms," 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Langkawi Island, Malaysia, 2021, pp. 517-521, doi: 10.1109/IECBES48179.2021.9398759.

[8] Ayad M., Kanaan H. and Ayache M., "Diabetes Disease Prediction Using Artificial Intelligence," 2020 21st International Arab Conference on Information Technology (ACIT), Giza, Egypt, 2020, pp. 1-6, doi: 10.1109/ACIT50332.2020.9300066.

[9] Shetty D., Rit K., Shaikh S. and Patil S., "Diabetes disease prediction using data mining," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2017, pp. 1-5, doi: 10.1109/ICIIECS.2017.8276012.

[10] Reshmi S., Biswas S. K., Boruah A. N., Thounaojam D. M. and Purkayastha B., "Diabetes Prediction Using Machine Learning Analytics," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 2022, pp. 108-112, doi: 10.1109/COM-IT-CON54601.2022.9850922.

[11] Dubey Y., Wankhede P., Borkar T., Borkar A. and Mitra K., "Diabetes Prediction and Classification using Machine Learning Algorithms," 2021 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON), Dhaka, Bangladesh, 2021, pp. 60-63, doi: 10.1109/BECITHCON54710.2021.9893653.

[12] Swarna S. R., Boyapati S., Dixit P. and Agrawal R., "Diabetes prediction by using Big Data Tool and Machine Learning Approaches," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 750-755, doi: 10.1109/ICISS49785.2020.9315866.

[13] Sarwar M. A., Kamal N., Hamid W. and Shah M. A., "Prediction of Diabetes Using Machine Learning Algorithms in Healthcare," 2018 24th International Conference on Automation and Computing (ICAC), Newcastle Upon Tyne, UK, 2018, pp. 1-6, doi: 10.23919/IConAC.2018.8748992.

[14] Emon M. U., Keya M. S., Kaiser, M. A. islam M. S., Tanha T. and Zulfiker M. S., "Primary Stage of Diabetes Prediction using Machine Learning Approaches," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 364-367, doi: 10.1109/ICAIS50930.2021.9395968.

[15] Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, 213-237.

[16] Mienye, I. D. (2021). Improved Machine Learning Algorithms with Application to Medical Diagnosis. University of Johannesburg (South Africa).

[17] https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

[18] Mijwil, M. M., & Abttan, R. A. (2021). Utilisation of machine learning techniques in testing and training of different medical datasets. Asian Journal of Computer and Information Systems (ISSN: 2321–5658), 9(4).

[19] Scott, I., Carter, S., & Coiera, E. (2021). Clinician checklist for assessing suitability of machine learning applications in healthcare. BMJ Health & Care Informatics, 28(1)..