International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Type-2 Diabetes Prediction System Using Deep Neural Network Model

Mohannad Alseraiy[1], Dr. Raed Alsini[2]

Submitted:12/03/2024 **Revised**: 27/04/2024 **Accepted**: 04/05/2024

Abstract: Recently, diabetes is the most common chronic disease in the Kingdom of Saudi Arabia, affecting a high percentage of the population. In general, the diabetic disease may be the cause of kidney failure, blindness, and lower-limb amputations. Therefore, the early diagnose of diabetic disease is an essential task to save human lives. On the other hand, the revolution of Artificial Intelligence (AI) approaches has played significant roles in diverse aspects, where Machine Learning (ML) and Deep Learning (DL) methods present a central role in the early prediction of diabetic disease in early stages. This paper aims to research the recent developed AI-based type-2 diabetes prediction systems and analyse their efficiency in terms of classification accuracy. In addition, an efficient AI-based system is presented, analysed, and discussed, for the purpose of diagnosing type-2 diabetes among adults in early stages, based on the employment of an efficient Neural Network (NN) model. The developed system has been validated using two different diabetes datasets for the purpose of assessing the model's efficiency.

Keywords: *Type 2 diabetes, Diabetes early prediction, Machine Learning (ML), Deep Learning (DL).*

1. Introduction

Diabetes disease is considered as a common condition that disturb people of all ages. Diabetes is a disease that happens when the glucose (blood sugar) is too high, where it develops when the pancreas does not generate enough insulin amount. In general, there are several forms of diabetes disease, where the most common forms include: type 1 and type 2 diabetes [1].

Type 1 diabetes is an autoimmune disease where the immune system attacks and terminates insulin generating cells in the pancreas with unknown reasons. Usually, type 1 diabetes is diagnosed in children and young adults, however it can develop at any age, whereas type 2 diabetes is the most common type of diabetes and usually affect adults, where the body does not generate sufficient insulin or the body's cells do not respond to the insulin [2].

In general, type 2 diabetes is a disease that occurs because of the problem in the way the body regulates and uses glucose as a fuel, where the long-term condition results in a high amount of sugar circulating in the blood. Eventually, high blood glucose levels can lead to disorders of the circulatory, and then nervous and immune systems [3]. Usually, type 2 diabetes used to be more known as adult-onset diabetes, however, both type 1 and type 2 diabetes can start during childhood and adulthood. However, type 2 diabetes is more common in older adults [4].

[1] *Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia*
[2] *Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia*

Recently, the diabetes disease is considered as among the most critical diseases where a high ratio of people surfer from this disease. Unfortunately, people with diabetic disease have a high risk of several diseases, for instance: kidney disease, stroke, heart disease, eye problem, and nerve damage. The current practices in hospitals involve several blood tests for the purpose of identifying the possibility of having diabetic disease, and then an appropriate treatment is offered [5].

Over the last few decades, the area of Artificial Intelligence (AI) has gained popularity in several sectors, including the healthcare, where AI has been involved in the management of chronic diseases. Recently, a large number of research studies and case studies are being conducted in the area of healthcare [6].

The early prediction of diabetes disease is considered as significant factor as this disease might be detected to prevent and avoid further complications. However, the traditional methods for diagnosing type-2 diabetes include using the Glycated Haemoglobin (A1C) test, which measures the average blood sugar level for the recent (2-3) months. In addition to that, the random blood sugar test can be also used to diagnose the existence of type 2 diabetes. Another way is known as fasting blood sugar test, where a blood sample is taken after the person has not eaten for the last 6 hours. However, the employment of such methods requires the collection of a blood sample for processing and analysis, and this is considered as time-consuming and impractical for some people.

Hence, the adoption of AI technology can enhance the management of diabetes in the following ways: the early detection of the diabetic disease, the prediction of blood-glucose level, and a reliable collection of diabetic patient's data and henceforth improving the communication between the doctors and diabetic patients [7]. Therefore, this paper aims to develop an efficient AI-based approach for diagnosing type-2 diabetes among adults, through designing an accurate classification model and validate the developed diabetes prediction system using a real diabetes dataset. A Neural Network (DNN) model will be designed and implemented for the purpose of diagnosing type 2 diabetes among adults.

Recent research works have discussed the area of developing an efficient early prediction diabetes systems based on the adoption of large datasets and AI-based approaches, that are surveyed in [8]. However, most of the existing works didn't obtain accurate classification accuracy. For instance, ML-based approaches are efficient, however, they are susceptible to errors when the employed dataset is quite small. In addition, the results generated from ML approaches are difficult to be interpreted. Therefore, an alternative approach is employed named as Neural Network (NN) approaches, however, NN-based approaches are complicated in terms of structure, and requires an intensive knowledge in the design and development of NN systems.

Therefore, this paper aims to investigate, design, and develop an efficient diabetes prediction system based on the employment of a reliable diabetic dataset and the implementation of an efficient deep neural network classification approach. The research objectives associated with the project aims, are as follows:

1. Research the recent developed diabetes prediction approaches that adopt ML and DL methods.
2. Discuss and analyse the results obtained from the recent developed diabetes predictions systems.
3. Investigate and analyse the available diabetes dataset for adults and choose the most reliable diabetes dataset.
4. Design an efficient a DNN model for the purpose of predicting the diabetes of adults.
5. Validate the efficiency of the designed DNN model through conducting several experiments using different diabetes datasets, and assess the developed DNN model through several evaluation metrics, including: confusion matrix, accuracy, precision, and recall.

The rest of this paper is organized as follows: Section 2 reviews the recent developed AI-based type-2 diagnosing systems, where Section 3 discusses the proposed AI-based type-2 diabetes diagnosing system. In Section 4, the experimental setup is discussed, including the development environment and the selected dataset. Section 5 presents and discusses the experimental results obtained from the developed ML and NN models. A discussion is conducted in Section 6, whereas Section 7 concludes the work presented in this paper, and shows the future works.

2. Related works

Type-2 diabetes prediction systems have received considerable attention recently with the aim of classifying the existence of type-2 diabetes among adults, using the employment of deep neural network and machine learning models. This section discusses the recent developed type-2 diabetes classification systems for adults.

The work presented in [10] involves the design and development of diabetes prediction model for classifying type-2 diabetes prediction using additional external factors, including: Glucose, GMI, Age, Insulin, … etc. This work involves the employment of 12 different machine learning models. According to the obtained results, the logistic regression obtained the best classification accuracy with 96%. Authors of [11] developed an enhanced logistic regression model, where the obtained results are compared to 3 ML models. In [12], authors proposed a Decision Support System (DSS) for diabetes prediction based on the employment of ML techniques. The developed system was validated using Pima dataset with accuracy of 76,81%, 65.38%, and 83.76% for the DL, SVM, and RF respectively.

The work presented in [13] involves the design and development of an end-to-end remote monitoring approach for automated diabetes risk prediction and management tasks, through the employment of personal health devices, smart wearable and smartphones. A Support Vector Machine (SVM) approach has been developed for diabetes risk prediction using the Pima Indian Diabetes dataset. The obtained results are as follows: accuracy of 83.20%, sensitivity of 87.20%, and specificity score of 79%. Authors of [14] developed a diagnosis approach using 11 ML models for the purpose of diagnosing type-2 diabetes among adults. The developed system has been validated using two different diabetes datasets, where the developed system achieves an efficient classification accuracy of 98.95% using a tailored RF classification approach.

In [15], authors applied a deep learning model that combines the strength of a generated linear model along with several features and a deep feed-forward neural network to enhance the prediction accuracy of type-2 diabetes mellitus. The classification accuracy for the proposed type-2 diabetes prediction was 84.28%, sensitivity of 31.17, and specificity of 96.85%. The work presented in [16] involves the investigation of employing two ML models (KNN and Naïve Bayes) for the purpose of accomplishing the task of diabetes prediction. Authors revealed that the Naïve Bayes classification model for diabetes prediction achieves better results than employing the KNN model.

Authors of [17] proposed an efficient framework for diabetes prediction through employing outlier rejection, data standardization, feature selection, and K-fold cross-validation. Authors revealed that the XGBoost offers the best performance in terms of diabetes prediction accuracy with an average accuracy of 94.6%. In [18], authors employed decision tree, random forest, and neural network for the purpose of predicting diabetes mellitus. A dataset is the hospital physical examination data obtained from Luzhou, China, which consists of 14 different attributes. A 5-fold cross validation method has been employed to examine the above three models. Authors of [19] employed the logistic regression classification model to identify the risk factors for diabetes disease. Four classification models have been employed The logistic regression model demonstrated that 7 factors out of 14 are the risk factors for diabetes. The combination of logistic regression and random forest classifiers perform better results.

The work presented in [20] involves the design and development of a diabetes prediction system through employing three different machine learning models: decision tree, support vector machine, and Naïve Bayes. The results obtained from this study showed that the NB outperforms the SVM and DT with a classification accuracy of 76.30%. In [21], authors employed diverse ML algorithms for the purpose of early prediction of diabetes type-2, through the employment of Pima dataset. According to the obtained results the DL algorithm offers the best classification accuracy with a rate of 98.07%. Authors of [22] proposed a novel type-2 diabetes prediction model based on data mining techniques, where the developed prediction model consists of an improved K-means algorithm and logistic regression algorithm. The obtained results showed an enhancement of 3.04% in classification accuracy compared to the previous research works. Moreover, authors employed two other diabetes datasets to confirm the efficiency of the proposed approach.

The work presented in [23] involves a comparative analysis for employing several ML and DL approaches for the purpose of type-2 diabetes early prediction. A diabetes dataset from the UCI repository has been adopted which consists of 17 attributes including the class label. According to the obtained results, the XGBoost classifier achieved the best classification accuracy with approximately 100.0%. In [24], authors proposed a strategy for the diagnosis of type-2 diabetes through the employment of deep neural network by training its attributes in five-fold and ten-fold cross-validation way. An enhanced SVM and DNN system was proposed in [25], for the purpose of diabetes type-2 prediction purposes. The proposed approach employed the DNN to obtain the input from the output of the enhanced SVM model. The developed system has been validated using the popular Pima dataset with a classification accuracy of 98.45%.

3. A NN-based Diabetes Diagnosing System

This paper involves the design and development of a NN approach for the purpose of predicting the existence of type-2 diabetes among adults, through adopting real diabetic datasets. This section discusses the main methods and development environments which have been employed in order to develop an efficient type-2 diabetes prediction system. In addition, the employed dataset is discussed and analysed. Moreover, the development environment including the programming tools and APIs are presented.

The presented diagnosing type-2 diabetes system involves a set of tasks that need to be addressed in order to achieve the design and development of a type-2 diabetes classification system. At the first stage, we consider reliable and accurate type-2 diabetes datasets for the purpose of training and testing the developed NN prediction model. In general, the datasets contain missing and incomplete data, therefore, it is important to preprocess the data in the selected dataset in order to obtain an efficient classification accuracy. The preprocessing methods involve either recovery of incomplete data or eliminating of missing data, according to the significant of the data attributes. For instance, some attributes might be recovered, whereas others are hard to be recovered, therefore, an elimination method is processed in this case.

In the next stage, several ML models will be adopted, trained, and tested for the purpose of comparing the obtained results with the results obtained from the developed NN model, and to validate the efficiency of the developed diagnosing system. In addition, the obtained results are analysed, compared, and discussed.

4. Experimental Analysis

This section discusses the main tools and packages that have been adopted to develop the type-2 diabetes diagnosing system. In addition, the system design is discussed in details, and the employment of several ML models are taken place.

4.1. Diabetes Datasets

In this paper, we employed two different diabetes datasets for the purpose of training and testing the proposed NN model. First, PIMA dataset, which was gathered by the National Institute of Diabetes and Digestive and Kidney Disease. The main objective of PIMA dataset was to diagnostically predict whether the person has a diabetic disease or not. PIMA dataset consists of several medical predictors including: pregnancies, glucose, blood pressure, skin thickness, insulin, BMI, Diabetes prediction function, and age. In addition, this dataset contains a single label which is the outcome attribute.

The main problem with the original PIMA dataset is that the numbers of healthy records are much greater than the diabetic records as presented in Figure 1 (A), and hence the PIMA is considered as unbalanced dataset, and this may lead to inefficient classification accuracy. Therefore, an efficient data processing method needs to be applied in order to solve out the problem of unbalancing dataset.

Second, the National Health and Nutrition Examination Survey (NHANES) dataset [27] was adopted. NHANES dataset consists of 4,896 records with 7 attributes: Region, Gender, BMI, Diet, BP, Smoking, and Status. In addition, the distribution of normal and diabetic records in the NHANES dataset is presented in Figure 1 (B). Table 1 presents general statistics (dataset properties) on the Pima and NHANES datasets.

Table 1: General statistics on Pima and NHANES datasets

Parameter	PIMA	NHANES
Number of records	768	4,896
Number of diabetic persons	268	990
Number of healthy persons	500	3,906
Number of attributes	9	7
Number of features	8	6
Number of labels	1	1

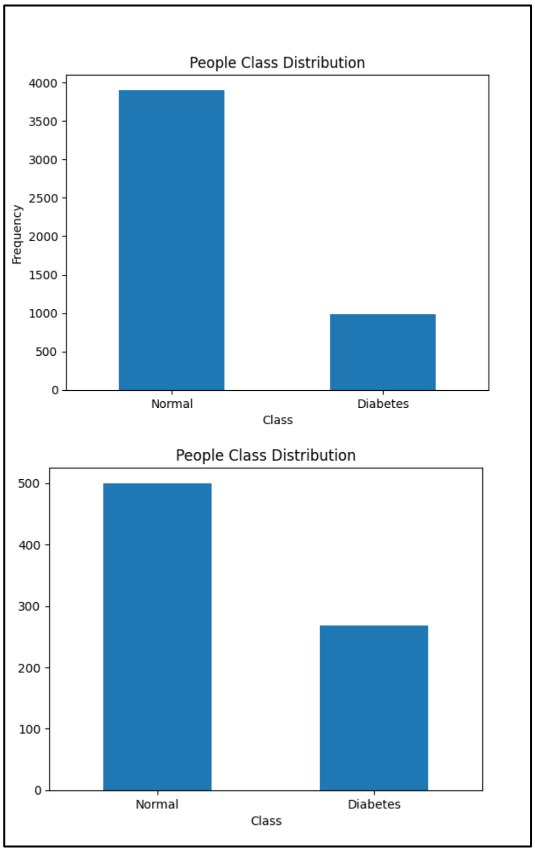


Fig 1: The distribution of normal and diabetic records in the original NHANES (up) and PIMA (down)

4.2. Tailored ML Models for Type-2 Diabetes Prediction

This subsection discusses the employment of several ML models for the purpose of diagnosing the type-2 diabetes among adults. The employed ML models have been tuned in order to present better classification accuracy. Therefore, 7 ML models have been implemented, tuned, trained, and tested, as follows:

1. **Support Vector Machine (SVM):** SVM offers reliable accuracy for binary classification scenarios, where the SVM works by classifying the elements of the dataset into two groups, through finding out an optimal line that maximizes the distance among each class in an N-dimensional space.
2. **Decision Tree (DT):** DT model involves creating a classification model through building a decision tree that consists of a set of nodes, where each node represents a test on an attribute.
3. **Gradient boosting (GB):** GB model relies on the assumption that the best possible next model when integrated with the previous model, will minimize the total prediction error.
4. **Light Gradient Boosting (LightGB):** LightGB is a fast, high performance, distributed gradient boosting classifier based on the concept of decision trees.
5. **Random Forest (RF):** RF is a common ML model that integrates the output of the multiple decision trees to reach to a single result. In general, RF adds randomness to the model while growing the RF tree.
6. **CatBoost:** is a ML model that is based on the concept of supervised learning, where CatBoost learns from labelled data, when the training data includes both the input data and the desired output.
7. **Naïve Bayes (NB):** NB is a probabilistic ML model where the classifier is a latent variable which is probabilistically relevant to the practical variables.

4.3. Customized Neural Network (NN) Diagnosing System

This section discusses the development of an efficient NN model for the purpose of diagnosing type-2 diabetes among adults. Then, the developed DNN model is trained and tested using the selected dataset in order to validate the efficiency of the developed NN model. Then, the obtained results are verified, however, if the obtained results are inaccurate, then, the development process of the pre-designed DNN model will take place.

Therefore, the proposed NN model is depicted in Figure 2. A set of health information is needed (features) in order to feed the customized NN model, and then perform a diagnosing operation based on a pretrained diabetes dataset. The obtained result (label) will be in the form of two classes (diabetic person or non-diabetic person).

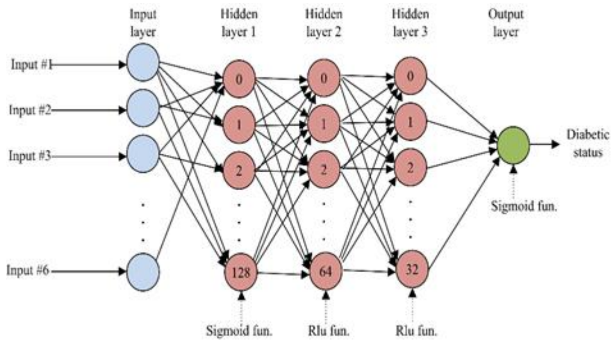


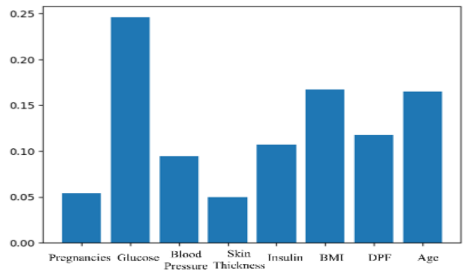
Fig 2: The proposed type-2 diabetes prediction using the proposed NN model

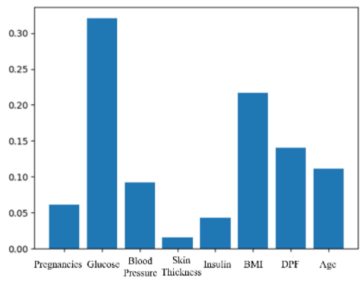
4.4. PIMA Dataset Augmentation

For the purpose of enhancing the accuracy of classification approaches, an efficient augmentation method has been adopted for the purpose of enlarging the PIMA dataset in a right manner, where the number of diabetes records has been increased, in order to offer reliable classification accuracy.

4.5. Feature Extraction

The feature extraction approach involves measuring the significant of the input features for the purpose of enhancing the classification accuracy of the diagnosing system. Therefore, 3 different feature extraction algorithms have been adopted for the purpose of estimating the significant of the input features, as follows: DT, RF, and XGBoost, where these models offer significant feature extraction tasks. Figure 3 shows the significance of each feature for the DT, RF, and XGBoot models.





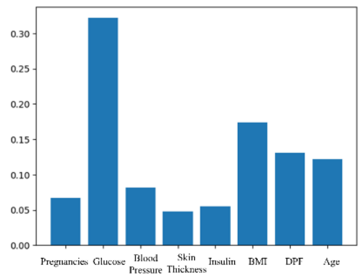


Fig 3: Feature significant weights using the DT (up), RF (middle), and XGBoost (down) models

As presented above, the Glucose, BMI, DPF, and Age input features are the most significant features in the PIMA dataset. Therefore, these input features will be considered for the purpose of training the tailored ML models and the designed DNN model. Therefore, a new dataset has been obtained with the most significant features in order to prepare for implementing several ML models.

5. Experimental Results

This section discusses the obtained results from the implementation of several tailored ML models and the designed NN (d-NN) model. Therefore, for evaluation purposes, several parameters have been considered for the purpose of validation the ML and NN models, as follows:

1. Classification accuracy: this refers to how often the classifier is correct in overall.
2. Precision score (Prc): this includes the total number of cases (healthy and diabetic cases) that were correctly classified among that class.
3. Recall score (Rec): this indicates the ability of the classifier to correctly find all positive instances (diabetic cases).
4. F1-score: is the weighted harmonic mean of precision and recall.
5. Misclassification rate (MCR): this refers to how often the classification model is wrong.

First, the classification accuracy has been obtained for each tailored ML model. The classification accuracy is considered as a significant evaluation metric for assessing the efficiency of any ML and DL model. Table 2 presents the above evaluation metrics for all the employed AI models including the tailored ML models and the designed NN model. As presented in the table below, the RF classification model achieves the best classification accuracy (87.54%) among all the ML models, whereas the SVM classification model offers the worst classification accuracy (71.38%). However, the classification accuracy of the p-NN classification model offers the best classification accuracy (92.78%).

Table 2: Classification accuracy for 8 AI models

Tailored ML model	Accuracy	Precision	Recall	F1-score	MCR
SVM	71.38	72.00	71.50	71.50	27.60
DT	80.80	81.00	80.50	80.50	19.19
GB	81.48	82.00	81.50	81.50	18.51
LightGB	81.18	86.50	85.00	85.00	14.81
RF	87.54	88.00	87.50	87.50	12.45
CatBoost	83.16	83.50	87.50	83.00	16.83
KNN	80.47	71.97	75.83	73.84	26.93
d-NN	92.78	87.75	94.03	88.54	09.90

Second, precision evaluation metric is assessed. The classification accuracy is a significant evaluation metric, however, other parameters such as precision is considered as an important evaluation metric. The NN classification models offers the best precision score (88.00%) among all the employed ML models, however, the designed NN models achieved less precision score with 87.75%.

Third, the recall evaluation metric has been estimated for the employed AI models. In this scenario, the recall is a significant evaluation metric, since it measures the classifier's ability to correctly find all the diabetic cases in the employed dataset. The RF and CatBoost achieved the best recall score (87.5%) among all the ML models, whereas the proposed NN model accomplished better results for the recall score (94.03%), and this indicates that the d-NN model offers significant enhancement over the employed ML models.

Fourth, the F1-score is considered, where F1-score estimates the weighted harmonic mean between both the precision and recall. Again, the RF offered the best F1-score results (87.50%) among all the employed ML models, however, the d-NN model achieved the best classification accuracy (88.57%).

Fifth, the miss-classification scores are evaluated for all AI models, where it refers to the total number of wrong classification cases and it presents a significant evaluation metric. The d-NN achieved the best miss-classification score (9.9%) among all the employed ML models.

In addition to the above evaluation metrics, the confusion matrix has been adopted to assess the efficiency of each ML and NN models. The confusion matrix offers an efficiency evaluation method to assess the classification accuracy for each single class.

The confusion matrixes for the 8 AI models were analysed, however, in this section, we present the confusion matrix for the NN model. The confusion matrix of the NN classifier is presented in Table 3, where the TN, TP, and FP values are the best among all the aforementioned models. However, the FN value is equal to the value obtained in the LightGB and RF classifiers.

Table 3: The confusion matrix for testing the proposed NN model

N = 297	Predicted 'No'	Predicted 'Yes'	
Actual No	TN = 128	FP = 18	146
Actual Yes	FN = 9	TP = 142	151
	137	160	

*N: is the total number of testing records

Moreover, for validating the developed NN model, we employed the NHANES dataset to assess the efficiency of the developed diabetes prediction system. After the employment of the developed NN model, the obtained classification accuracy was almost 82.46%, whereas the obtained precision score was equal to 70.54%. On the other hand, the recall score was equal to 69.03%, and the F1-score was equal to 62.25%. Finally, the misclassification rate has been obtained with a value of 17.81%. Moreover, the confusion matrix for the proposed NN model using the NHANES dataset is presented in Table 4.

Table 4: The confusion matrix for testing the proposed NN model using NHANES dataset

N = 1469	Predicted 'No'	Predicted 'Yes'	
Actual No	TN = 1089	FP = 84	1173
Actual Yes	FN = 209	TP = 87	296
	1298	171	

6. Discussion

The problem of type-2 diabetes prediction has been widely studied recently with the aim to developing an efficient classification system (with reliable classification accuracy and recall) to classify an adult of having type-2 diabetes or not. The existing developed type-2 diabetes prediction systems have achieved reasonable classification accuracy through adopting different datasets. However, there is still high demand to obtain high classification accuracy for type-2 with diabetes.

The PIMA dataset has been selected in several research works, however, PIMA is unbalanced dataset, where the number of health records are much greater than the diabetic ones. Hence, this affects the classification accuracy for the AI-based diagnosing systems. However, for the purpose of developing an efficient diagnosing system, an augmentation approach has been adopted to increase the size of the PIMA dataset (enlarge the number of diabetic records), where an efficient dataset has been obtained.

On the other hand, the feature extraction plays a significant role when developing an accurate classification system. The adopted PIMA dataset contains unreliable data values, which drastically affects the classification accuracy and minimizes the diagnosing system's efficiency. Therefore, in this paper, three different feature extraction methods have been adopted to extract the most significant features in the PIMA dataset. According to the obtained results, the most significant features involve: glucose, BMI, DPF, age, blood pressure, and insulin. Therefore, a new dataset has been obtained in order to improve the classification accuracy of the developed diabetes diagnosing system.

ML-based diabetes diagnosing systems offer reasonable classification accuracy as presented earlier in the previous section. However, the tailored ML models offer limited recall, precision, and miss-classification scores. Therefore, the designed NN model achieves better classification accuracy, recall, f1-score, and miss-classification scores, as the d-NN model involves reliable structure which allows the NN model to be tailored with the PIMA dataset.

Moreover, the developed NN model has been validated using an additional diabetes dataset (NHANES) for the purpose of assessing the efficiency of the developed diabetes prediction system. The obtained results are promising in terms of classification accuracy, where an efficient classification accuracy has been obtained.

The work presented in [28] employed the NHANES dataset to assess the efficiency of diabetes prediction using a customized decision forest ML model. This study achieved a reasonable classification accuracy of diabetes prediction with a result of 82.0%. However, in this paper, we slightly enhanced the classification accuracy using the same dataset (NHANES) through the adoption of the developed NN model.

Unlike the work presented in [28], the developed NN model in this paper has been validated using two different diabetes datasets, where each dataset was divided into two subsets: training and testing. The obtained classification accuracy (82.46%) is more accurate than the decision forest model presented in [28].

7. Conclusion and Future work

The area of type-2 diabetes prediction has received considerable attention recently due to its importance in the medical sector. This paper discussed the recent developed diabetes prediction systems based on the employment of AI algorithms, and compares these systems in terms of their efficiency and reliability. Moreover, 7 different ML models have been tailored to solve the problem of diagnosing type-2 diabetes among adults, where ML models offer reasonable classification accuracy. A NN model has been proposed in this paper to obtain efficient classification accuracy and diagnose the existence of type-2 diabetes among adults with efficient classification results. The developed NN-based diabetes prediction system achieved an efficient classification in terms of accuracy using two different diabetes datasets. For future works, additional records (for diabetic and non-diabetic persons) need to be collected and perform further validation for the proposed NN model.

References

[1] Redondo, M.J., Hagopian, W.A., Oram, R., Steck, A.K., Vehik, K., Weedon, M., Balasubramanyam, A. and Dabelea, D., 2020. The clinical consequences of heterogeneity within and between different diabetes types. *Diabetologia*, *63*, pp.2040-2048.

[2] Salsali, A. and Nathan, M., 2006. A review of types 1 and 2 diabetes mellitus and their treatment with insulin. *American journal of therapeutics*, *13*(4), pp.349-361.

[3] Robert, A.A. and Al Dawish, M.A., 2020. The worrying trend of diabetes mellitus in Saudi Arabia: an urgent call to action. *Current diabetes reviews*, *16*(3), pp.204-210.

[4] Al Dawish, M.A. and Robert, A.A., 2021. Diabetes Mellitus in Saudi Arabia: challenges and possible solutions. In *Handbook of healthcare in the Arab world* (pp. 1083-1100). Cham: Springer International Publishing.

[5] Pecoits-Filho, R., Abensur, H., Betonico, C.C., Machado, A.D., Parente, E.B., Queiroz, M., Salles, J.E.N., Titan, S. and Vencio, S., 2016. Interactions between kidney disease and diabetes: dangerous liaisons. *Diabetology & metabolic syndrome*, *8*(1), pp.1-21.

[6] Yu, K.H., Beam, A.L. and Kohane, I.S., 2018. Artificial intelligence in healthcare. *Nature biomedical engineering*, *2*(10), pp.719-731.

[7] Mobasseri, M., Shirmohammadi, M., Amiri, T., Vahed, N., Fard, H.H. and Ghojazadeh, M., 2020. Prevalence and incidence of type 1 diabetes in the world: a systematic review and meta-analysis. *Health promotion perspectives*, *10*(2), p.98.

[8] Sai, P.M.S. and Anuradha, G., 2020, March. Survey on Type 2 diabetes prediction using machine learning. In *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 770-775). IEEE.

[9] Aggarwal, K., Mijwil, M.M., Al-Mistarehi, A.H., Alomari, S., Gök, M., Alaabdin, A.M.Z. and Abdulrhman, S.H., 2022. Has the future started? The

current growth of artificial intelligence, machine learning, and deep learning. *Iraqi Journal for Computer Science and Mathematics*, *3*(1), pp.115-123.

[10] Mujumdar, A. and Vaidehi, V., 2019. Diabetes prediction using machine learning algorithms. *Procedia Computer Science*, *165*, pp.292-299.

[11] Krishnamoorthi, R., Joshi, S., Almarzouki, H.Z., Shukla, P.K., Rizwan, A., Kalpana, C. and Tiwari, B., 2022. A novel diabetes healthcare disease prediction framework using machine learning techniques. *Journal of Healthcare Engineering*, *2022*.

[12] Yahyaoui, A., Jamil, A., Rasheed, J. and Yesiltepe, M., 2019, November. A decision support system for diabetes prediction using machine learning and deep learning techniques. In *2019 1st International informatics and software engineering conference (UBMYK)* (pp. 1-4). IEEE.

[13] Ramesh, J., Aburukba, R. and Sagahyroon, A., 2021. A remote healthcare monitoring framework for diabetes prediction using machine learning. *Healthcare Technology Letters*, *8*(3), pp.45-57.

[14] Alhmiedat, T. and Alotaibi, M., 2022. The Investigation of Employing Supervised Machine Learning Models to Predict Type 2 Diabetes Among Adults. *KSII Transactions on Internet & Information Systems*, *16*(9).

[15] Nguyen, B.P., Pham, H.N., Tran, H., Nghiem, N., Nguyen, Q.H., Do, T.T., Tran, C.T. and Simpson, C.R., 2019. Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records. *Computer methods and programs in biomedicine*, *182*, p.105055.

[16] Febrian, M.E., Ferdinan, F.X., Sendani, G.P., Suryanigrum, K.M. and Yunanda, R., 2023. Diabetes prediction using supervised machine learning. *Procedia Computer Science*, *216*, pp.21-30.

[17] Hasan, M.K., Alam, M.A., Das, D., Hossain, E. and Hasan, M., 2020. Diabetes prediction using ensembling of different machine learning classifiers. *IEEE Access*, *8*, pp.76516-76531.

[18] Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y. and Tang, H., 2018. Predicting diabetes mellitus with machine learning techniques. *Frontiers in genetics*, *9*, p.515.

[19] Maniruzzaman, M., Rahman, M.J., Ahammed, B. and Abedin, M.M., 2020. Classification and prediction of diabetes disease using machine learning paradigm. *Health information science and systems*, *8*, pp.1-14.

[20] Sisodia, D. and Sisodia, D.S., 2018. Prediction of diabetes using classification algorithms. *Procedia computer science*, *132*, pp.1578-1585.

[21] Naz, H. and Ahuja, S., 2020. Deep learning approach for diabetes prediction using PIMA Indian dataset. *Journal of Diabetes & Metabolic Disorders*, *19*, pp.391-403.

[22] Wu, H., Yang, S., Huang, Z., He, J. and Wang, X., 2018. Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked*, *10*, pp.100-107.

[23] Refat, M.A.R., Al Amin, M., Kaushal, C., Yeasmin, M.N. and Islam, M.K., 2021, October. A comparative analysis of early-stage diabetes prediction using machine learning and deep learning approach. In *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)* (pp. 654-659). IEEE.

[24] Ayon, S.I. and Islam, M.M., 2019. Diabetes prediction: a deep learning approach. *International Journal of Information Engineering and Electronic Business*, *12*(2), p.21.

[25] Nagaraj, P. and Deepalakshmi, P., 2021. Diabetes Prediction Using Enhanced SVM and Deep Neural Network Learning Techniques: An Algorithmic Approach for Early Screening of Diabetes. *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, *16*(4), pp.1-20.

[26] Zhao, Q., Zhu, J., Shen, X., Lin, C., Zhang, Y., Liang, Y., Cao, B., Li, J., Liu, X., Rao, W. and Wang, C., 2023. Chinese diabetes datasets for data-driven machine learning. *Scientific Data*, *10*(1), p.35.

[27] Centers for Disease Control and Prevention. (Sep. 2009). *National Center for Health Statistics (NCHS). National Health and Nutri- tion Examination Survey*. Accessed: May 28, 2020. [Online].

[28] Syed, A.H. and Khan, T., 2020. Machine learning-based application for predicting risk of type 2 diabetes mellitus (t2dm) in saudi arabia: A retrospective cross-sectional study. *IEEE Access*, *8*, pp.199539-199561.