An expert system for diabetes prediction using auto tuned Multi-layer Perceptron

Maham Jahangir
Department of Computer Science
University of Management & Technology,
Sialkot, Pakistan
maham.jahangir@skt.umt.edu.pk

Mehreen Ahmed
National University of Sciences & Technology,
Islamabad, Pakistan
mahreenmcs@gmail.com

Hammad Afzal
National University of Sciences & Technology,
Islamabad, Pakistan
hammad.afzal@mcs.edu.pk

Khawar Khurshid
National University of Sciences & Technology,
Islamabad, Pakistan
khawar.khurshid@seecs.edu.pk

Raheel Nawaz
Manchester Metropolitan University, United Kingdom
r.nawaz@mmu.ac.uk

Abstract—**Medical Expert Systems is an active research area where data analysts and medical experts are continuously collaborating to make these systems more accurate and therefore, more useful in real life. Recent surveys by World Health Organization indicated a great increase in number of diabetic patients and the deaths that are attributed to diabetes each year. Therefore, early diagnosis of diabetes is a major concern among researchers and practitioners. The paper presents an application of *automatic multilayer perceptron (AutoMLP)* which is combined with an outlier detection method *Enhanced Class Outlier Detection using distance based algorithm* to create a novel prediction framework. AutoMLP is auto-tunable and performs parameter optimization automatically on the run during training process, which otherwise requires human intervention. Our framework performs outlier detection during pre-processing of data. A series of experiments are performed publicly available dataset: UCI (Prima Indian) and system achieved an accuracy of 88.7% which bests the highest reported results.**

Keywords—machine learning, diabetes prediction, medical expert system, multi layer perceptron

I. INTRODUCTION

Data mining using advanced pattern recognition and classification has played major role in development of expert systems in medical domain. Medical Expert Systems is an active area of research where data analysts and experts in machine learning (ML) are continuously striving to make them more accurate. The improved diagnostic systems with better performance save time of the medical practitioners. Moreover, these systems assist doctors and physicians in their clinical routine. According to recent statistics by World Health Organization (WHO), 422 million adults have diabetes and 1.5 million deaths are directly attributed to diabetes each year.[1] The figures indicate that diabetes is a major problem faced by the world today. Therefore, there is an immense need for supporting the medical decision-making process so that diabetes can be detected at an early stage. The three common types of diabetes include Type I (the human body fails to produce insulin), Type II (the cells fail to use insulin) and Gestational (high blood sugar level during pregnancy) [1]. ML algorithms have successfully been applied for diagnosis of various diseases like heart, diabetes, cancer, hepatitis etc [2-8]. After reviewing literature it could be seen that neural network gave quite significant results as compared to others. Various architectures of neural networks have been employed by different researchers in different medical diagnosis [9-11]. Many researchers propose that ANN is more flexible in modeling and gives reasonable results in accuracy prediction[12]. Shanker [13] showed in his research that ANN is a better approach than other classifiers. However, the network topology has to be decided before training ANN. The major problem faced in neural networks is parameter optimization. Parameter optimization involves selecting number of hidden layers, neurons, number of epochs and learning rate while defining the network topology of neural network. This problem is solved by AutoMlp which is a small ensemble of multilayer perceptrons and is auto tunable. It adjusts the parameters automatically. This paper proposes a novel decision support framework that combines pre-processing techniques with AutoMLP to provide a hybrid prediction model. The system uses Enhanced Outlier Detection using Distance Based Class Outlier factor as pre-processing of dataset. The results

---

1. http://www.who.int/diabetes/en/

demonstrate that the proposed approach outperformed other reported techniques and realized the highest accuracy of 88.7%.

The structure of the paper is as follows. The systematic review of literature along with dataset description is presented in Section 2. Section 3 provides the proposed framework, followed by the explanation of experiments in Section 4 along with discussion on results. Conclusions and future implications are discussed in Sections 4.

II. LITERATURE REVIEW

Researchers have been applying different ML techniques for diabetes prediction to improve the accuracy of healthcare systems. This research mainly focuses on Pima Indians Diabetes Dataset (PIDD). PIDD is collected from UCI ML repository. It contains the records of females of at least 21 years of age from the Pima Indian heritage. The dataset, number of instances, number of attributes, prevalence of diabetes, and features are listed in Table 1.

TABLE 1: DATASET DESCRIPTION WITH ATTRIBUTES

No. of Instances	No. of Attributes	Prevalence of diabetes	Features
768	8	34.89%	Number of times pregnant, Plasma glucose concentration, 2 hours in an oral glucose tolerance test, Diastolic blood pressure (mm Hg) , Triceps skin fold thickness (mm), 2-Hr serum insulin (mu U/ml), BMI, Diabetes pedigree function, Age

ZhilbertTafa et al. [8] in 2015 proposed a joint implementation of SVM and Naïve Bayes on **Kosovo diabetes dataset.** The split ratio for training and testing set used was 50:50%. Two ML classifiers are individually trained on the training set. The accuracies of 95.52 % and 94.52% are reported for SVM and Naïve Bayesian respectively. Similarly, in 2010, BayuAdhi Tama et al. [14] performed a series of experiments to predict diabetes on a private dataset. The SVM outperformed other classifiers as well as ensemble based methods. An average accuracy of **96.49%** using hold out and 10- fold cross validation is reported by the authors. In both researches, private datasets are used therefore their results are not comparable with any other research.

In 2013, Kumari et al. [15] applied SVM with RBF kernel on PIDD. They achieved an accuracy of 78%, 80% sensitivity and 76.5%specificity. Aibinu et al. [16] suggested the application of CVNN (complex valued neural network) and RVNN (real valued neural network) to PIDD for the prediction of diabetes in 2010. The normalization techniques used are: z-, min-max, complex and unitary data normalization. Two types of biomedical data are studied and

evaluated by authors in their research: Real Valued Data (RVD) and Complex Valued Data (CVD). Complex data normalization was used to convert RVD to CVD. Then the application of the suggested ANN based autoregressive model classification technique was evaluated with different activation functions, learning rate, number of neurons in the hidden layer and the number of epoch. The accuracy ranged from 80% to 81% when model was tested using different parameters.

In 2011, Aibinu et al. extended their work in the study [17]. This time they proposed the application of CVNN combined with complex-valued pseudo autoregressive (CAR) using split weights and adaptive coefficients; thus forming CVNN-based CAR model. The CAR coefficients are obtained from the weights and adaptive coefficients of a trained network. The authors found out that any increase in the number of training epoch and number of neurons in the hidden layer plays no significant role on the accuracy of the system. An accuracy of 81.28% is reported using PIDD.

In 2011 AlJarullah[18] applied decision tree for the prediction of diabetes on PIDD. The pre-processing techniques studied and evaluated for this research are attribute identification and selection, handling missing values, and numerical discretization. The dataset is trained using J48 algorithm using 10 fold cross validation. The accuracy achieved as a result of this model was 78.17%.

In 2012 Xue-HuiMeng[19] did a comparative analysis of three models for diabetic prediction. Among the three, C5.0 outperformed others giving best accuracy. The experiments are carried out on a private dataset collected from China The decision tree model achieved an accuracy of 77.87% with a sensitivity of 80.68% and specificity of 75.13%.

Ms. K Sowjanya et al. in 2015 [20] presented a comparative analysis of four ML algorithms in their research. The three possible values for class variable are High, Medium and Low; representing the degree of risk of suffering with diabetes within a test subject. During pre-processing the missing values are replaced by median and the units of measurement were standardized. The results showed that J48 gave better results than other three algorithms giving sensitivity (0.890), specificity (0.928) and ROC areas of 0.928. In the same year Purushottam et al. [21] achieved an accuracy up to 81.27% using rules extracted from C4.5 on PIDD.

In 2014 [22] Nnamoko et al. proposed a meta model combination of individual classifiers to improve accuracy of diabetes prediction. The dataset of PIDD is used to conduct experiments**.** The authors explained that prediction can be improved by combining the individual classifiers. First of all Synthetic Minority Over-Sampling Technique (SMOTE) is used in pre-processing stage to increase the minority class. The dataset is trained on five different learning algorithms: Sequential Minimal Optimization (SMO), Radial Basis Function (RBF), C4.5, Naïve Bayes and RIPPER. C4.5 produced greatest accuracy of 77.9% and aROC of 83.1%., whereas RBF gave the lowest prediction accuracy i.e 73.6%

and aROC of 80.5%. In the next step, the authors trained a meta model i-e combiner of best and first individual classifier with simple logistic regression algorithm. The outputs of RBF and C4.5 were used as additional inputs to the meta model. The accuracy achieved by meta model was **77.0% and aROC of 84.9%**

In 2014 Lin Li [23] diagnosed diabetes using weight adjusted voting approach by training proposed model on PIDD**.** The author used an ensemble of SVM, ANN and naïve bayes to predict diabetes. During pre-processing phase the records with biologically impossible values are removed. Wrapper method is used for feature selection with five features for classification instead of nine. Weight adjustment approach is used to combine the results of individual classifiers. Based on this approach they attained an accuracy of 77.0%, specificity 86.8% and sensitivity 58.3%. A summary of techniques applied on PIDD reported in literature are presented Table 2.

Sabariah et al. proposed the combination of Random Forest (RF) and Classification and Regression Tree (CART) for early detection of diabetes in 2014 [24]. In this research, the authors used different number of trees and candidate attributes splitter to get the optimal results. Moreover they analyzed and reported the most relevant attributes in predicting the disease. RF is built from a series of decision trees. One of these trees is CART. CART is modified when used with RF. This modification basically is determination of candidate attributes that will be used as a splitter. The findings of this research after series of experiments are: 50 numbers of trees and 3 number of attributes splitter with 83.8% average accuracy. The attributes: heredity, age, and body mass index are regarded as most important and relevant attributes for the early detection of disease**.**

Another ensemble-based research is carried out by R. Ali et al. [25] in 2014. In this study, the authors conducted experiments to classify the dataset into the respective type of diabetes. Adaboost M1 algorithm incorporated with random committee was used. Random tree is used as a base classifier in random committee. The algorithm repeatedly runs random tree over various distribution of training diabetes data and combines the outputs in a single random committee classifier. The final output is the average of the results generated by individual random tree classifiers.

TABLE 2: SUMMARY OF SELECTIVE TECHNIQUES APPLIED ON PIMA INDIANS DIABETES DATASET AND THEIR PERFORMANCES

Ref/ Year	Pre-processing Technique	Prediction Technique	Performance (Accuracy)
[28]/ 2002	None	SVM	82.29%
[29]/ 2003	None	General Regression NN (GRNN)	80.21%
[30]/ 2005	None	ANN	Correlation Coefficient =1
[31]/ 2006	None	ANN	74%
[32]/ 2009	None	LM	82.37%
[16]/ 2010	normalization and formatting of data	Complex Valued Neural Network (CVNN)	80% to 81%
[15]/ 2013	None	SVM with RBF Kernel	78% Sensitivity=80% Specificity=76.5 %.
[33]/ 2014	None	Neuro-Fuzzy Inference System	80%
[21]/ 2015	None	C4.5	81.27%

The authors are able to achieve an accuracy of 81.0% using 10-fold cross validation.

In 2015 Longfei Han et al. [26] applied an ensemble of SVM and RF on diabetes dataset collected for China Health and Nutrition Survey (CHNS). The training set is first trained on SVM by tuning parameters to get the highest accuracy, followed by extraction of rules using RF by tuning the rule induction parameters to get the best rules. These rules are then used to predict the class of each record from test data. Pre-processing techniques used are vacant data exclusion, noise data canceling and feature selection. The values for precision, recall and f-value calculated after 10-fold cross validation were 81.8%, 75.6% and 0.786 respectively.

In 2013,SubhamKhanna et al. [27] proposed an applied weight based classifier for prediction of diabetes. Binning, substitution and removal of duplicate records are used as pre-processing techniques. The performance measures used are accuracy, sensitivity, and specificity and kappa values giving values of 83.2%, 70.9%, 89.7% and 1.003 respectively.

III. PROPOSED FRAMEWORK

The proposed system consists of two stages data pre-processing which involves enhanced outlier detection followed by training automatic multi layer perceptron.

The outliers are detected from diabetes datasets using enhanced class outlier based method. This method detects outliers based on class outlier factor. Using this method, top 10 outliers are detected from the training set based on the 12 nearest neighbors and correlation similarity based distance measure. These outliers detected from the training set are removed. This outlier free diabetes dataset is then used from training classifier. A detailed illustration of outlier detection can be seen in Figure 1.

In the next step, Automatic Multilayer Perceptron is used to classify diabetic patients. An ensemble of 4 MLPs with different number of hidden units and learning rates are used. After ten training cycles the error rate is determined and worst MLP's are trained with the best ones. A complete block diagram of the proposed model can be seen in Figure 1.

In the experiments 70% of the data is used for training, 15% for validation and the other 15% for testing. The performance is evaluated on the test sets keeping the same parameters as that on validation set. It is ensured, that

testing data remains unseen and is not manipulated during experimentation in any way. The details of each step are described in this section. The details of the experimental results, discussion and comparison with other pre-processing and training algorithms are described in next section.
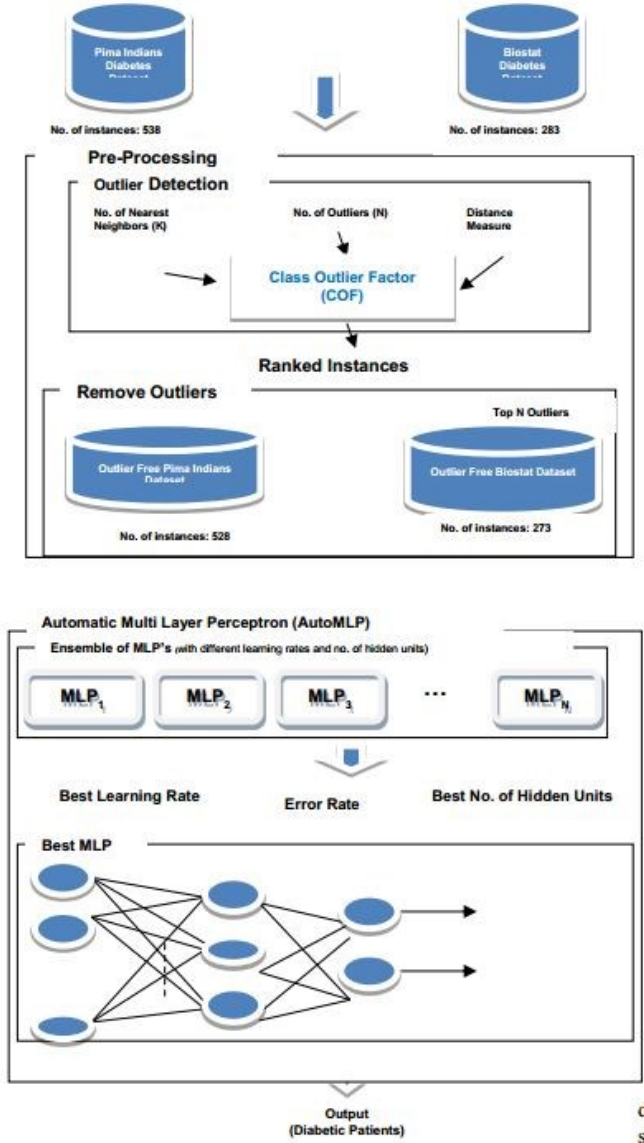

Figure 1: Proposed Predictive Framework

A. Data Pre-Processing

Each dataset has a class label with two values; either the patient is healthy or a sufferer of diabetes. The training sets are subject to pre-processing technique and then trained on AutoMLP.

Outlier Detection: Outlier detection and mining is used as a pre-processing step in this research. Outliers are defined as data points or tuples that are rare cases, exceptions, deviate in behavior from other data points or do not comply with

general behavior of data. Conventional techniques detect outliers in data irrespective of the class label i.e. the rare events or exceptions are detected with respect to whole dataset. On the other hand, in Class Outlier Mining the class label is taken into account while detecting outliers in the dataset. We have used Enhanced Class Outlier Distance Based (ECODB) algorithm which is an enhanced version of Class Outlier Distance Based (CODB) algorithm. Both CODB and ECODB have been compared with conventional techniques using a number of public datasets available in [34, 35]. The authors reported reasonable results as compared to conventional techniques.

Class outlier distance based (CODB) introduced by Hewahi and Saad[35] proposed detecting outliers based on nearest neighbors and distance based approach. It detects outliers based on class outlier factor (COF). COF depicts the degree of being a class outlier for a particular data instance. COF (degree of being class outlier) for a particular instance is based on the following key factors:

- Probability of instance class among its neighbors
- The deviation of the particular instance from the respective instances of the same class
- The distance between the particular instance and its k nearest neighbors

According to CODB algorithms class outliers from a given dataset are the instances that satisfy following conditions:

- The K-Distance of the particular instance from its K nearest neighbors is the least.
- The deviation of the particular instance from the respective instances of the same class is the greatest.
- It has different class label of its K nearest neighbors' class.

The class outlier factor for any instance (I) as per CODB concept is given as: [35]

$$COF(I) = K \times PCL(I, K) + \alpha \times \frac{1}{Deviation(I)} + \beta \times KDist(I) \quad (1)$$

Where,

PCL (*I, K*) = Probability of the class label of the instance (*I*) with respect to the class labels of its *K* Nearest Neighbors.

Deviation (*I*) = deviation of Instance (I) from the instances of the same class; computed by summing the distances between the instance *(I)* and others.

KDist(I)= summation of distance between the instance(I)and its *K* nearest neighbors

α, β = factors to control the importance and the effects of *Deviation(I)* and *KDist(I),* and they determined by trial and error.

ECODB is an enhanced version of CDOB proposed by same authors. ECODB is proposed to get rid of the hit and trial method for adjusting values of α and β. The ECODB algorithm defines Class Outlier Factor of particular instance COF (I) as:

$$COF(I) = K \times PCL(I,K) - norm(Deviation(I)) + orm(KDist(I)) \quad (2)$$

ECODB uses normalized values of deviation(I) and KDist(I) instead of α and β with a value range of [0-1]

$$norm(Deviation(I)) = \frac{Deviation(I) - MinDev}{MaxDev - MinDev} \quad (3)$$

$$norm(KDist(I)) = \frac{KDist(I) - MinKDist}{MaxKDist - MinKDist} \quad (4)$$

Where,
MaxDev= highest deviation value for top N class outliers
MinDev= lowest deviation value for top N class outliers
MaxKDist= highest KDist value for top N class outliers
MinKDist= lowest KDist value for top N class outliers

B. Training Learner

After the pre-processing step the classifier is trained on AutoMLP. AutoMLP is an auto tunable neural network introduced by Breuel et al. 2010 [36] which is feed forward neural network that automatically adjusts learning rate and number of hidden units. The parameters required by AutoMLP for training are:

- **Training cycles:** The number of maximum training cycles used for the neural network training. *Range:* integer; 1-+?; default: 10
- **Number of generations:** The number of generations for AutoMLP training. *Range:* integer; 1-+?; default: 10
- **Number of ensemble MLP's:** The number of MLP's per ensemble. *Range:* integer; 1-+?; default: 4

We used default values for all the parameters in AutoMLP. After pre-processing and training the performance of the classifier is an evaluated using validation and test set. This is described in the next sections.

IV. EXPERIMENTAL SETUP AND RESULTS

The proposed framework combines ECODB with AutoMLP as shown in Figure 1. The figure illustrates that the original medical dataset is first subjected to data transformation in which any nominal attribute data is changed to numerical data. After data transformation the training set is fed to outlier detection phase. These outliers are then removed as described earlier and illustrated in the figure. The next step is training the learner. Four MLP's are trained in proposed setup and the best MLP is replaced with the worst ones after 10 training cycles. The numbers of generations used are also 10. The network topology for the best MLP selected after training process consisted of one hidden layer and 160 nodes in hidden layer. Sigmoid is used as an activation function for adjusting weights at hidden nodes to obtain final result.

The experiments are performed using training, validation and test datasets. The split ratio is 70%: 15%: 15% for training validation and test set. The accuracies, weighted mean precision and recall obtained on test sets are calculated. Weighted mean precision is the average of precision obtained per class (two classes), similarly weighted mean recall is the average of recall calculated per class (two classes).In order to establish the supremacy of propose method over existing state-of-the-art approaches reported for prediction of diabetes, the comparison of proposed technique with the ones reported in literature is presented in Table 3. Various prediction methods comprising different classifiers including ANNs have been employed. The highest accuracies ranging from 81 to 82% have been reported using ANNs. The pre-processing techniques, prediction technique and the performance evaluators are detailed in the Table 3. It can be clearly seen that the proposed technique outperformed techniques presented in literature.

TABLE 3: COMPARISON OF PROPOSED TECHNIQUE WITH LITERATURE

Ref/ Year	Pre-processing	Prediction Technique	Accuracy (%)
[16]/ 2010	normalization formatting of data	Complex Valued Neural Network (CVNN)	81
[17]/ 2011	None	CVNN based CAR model	81.2
[37]/ 2012	normalization, discretization feature selection	NB	72.3
[38]/ 2013	None	Neuro-fuzzy Classifier	82.3 Specificity=84.6 Sensitivity=80.7
[22]/2014	(SMOTE)	Meta model of 5 classifier	77 aROC=84.9
[33]/2014	None	Neuro-Fuzzy Inference System	80%
[21]/2015	None	C4.5	81.27
[39]/2016	Feature Selection: F-score PCA, forward selection and backward elimination Outlier Detection: Grubb's test, Noise Removal: clustering,	Hierarchal Majority Voting (HMV) based ensemble technique	77.08% Sensitivity=78.9 Specificity=88.4 F-measure=83.40
2017 Proposed Technique	**Outlier Detection**	**Automatic Multilayer Perceptron**	**Accuracy=88.7 Weighted Mean Recall=88.5 Weighted Mean Precision=85.8**

V. CONCLUSION

This paper presents a novel approach to predict diabetes. It presents a comparative study conducted on PIDD dataset. The technique gave promising results for both the datasets.

Outlier detection is employed as a pre-processing step to detect and remove outliers in diabetes datasets. The systematic literature review revealed that the neural structures can be successfully used to predict diabetes. The selection of optimal number of hidden units and learning rate is a long run problem while defining the network topology for neural network architectures. For this purpose this paper presents the application AutoMLP. The proposed framework used an ensemble of four MLP's to achieve greater accuracy. The AutoMLP gave higher accuracy weighted mean recall and precision when compared with other architectures of neural network.

The proposed technique is compared with the results reported in literature. The experimental results prove that our system achieved an accuracy of 88.7% PIDD dataset which bests the highest reported accuracies. The most relevant attributes for diabetes prediction in PIDD are: plasma glucose concentration, diastolic blood pressure and number of times pregnant. In future the technique should be validated against other medical datasets like heart, hepatitis, cancer etc. The effect of outlier detection using nominal measures along with AutoMLP can be studied for other datasets.

REFERENCES

[1] M. Thirugnanam, P. Kumar, S. V. Srivatsan, and C. Nerlesh, "Improving the prediction rate of diabetes diagnosis using fuzzy, neural network, case based (FNC) approach," *Procedia Engineering,* vol. 38, pp. 1709-1718, 2012.

[2] I. Yoo, P. Alafaireet, M. Marinov, K. Pena-Hernandez, R. Gopidi, J.-F. Chang, and L. Hua, "Data mining in healthcare and biomedicine: a survey of the literature," *Journal of medical systems,* vol. 36, no. 4, pp. 2431-2448, 2012.

[3] K. Srinivas, B. K. Rani, and A. Govrdhan, "Applications of data mining techniques in healthcare and prediction of heart attacks," *International Journal on Computer Science and Engineering (IJCSE),* vol. 2, no. 02, pp. 250-255, 2010.

[4] M. Anbarasi, E. Anupriya, and N. Iyengar, "Enhanced prediction of heart disease with feature subset selection using genetic algorithm," *International Journal of Engineering Science and Technology,* vol. 2, no. 10, pp. 5370-5376, 2010.

[5] D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: a comparison of three data mining methods," *Artificial intelligence in medicine,* vol. 34, no. 2, pp. 113-127, 2005.

[6] S. Kharya, "Using data mining techniques for diagnosis and prognosis of cancer disease," *arXiv preprint arXiv:1205.1923*, 2012.

[7] G. Sathyadevi, "Application of CART algorithm in hepatitis disease diagnosis." pp. 1283-1287.

[8] Z. Tafa, N. Pervetica, and B. Karahoda, "An intelligent system for diabetes prediction." pp. 378-382.

[9] B. Apolloni, G. Avanzini, N. Cesa-Bianci, and G. Ronchini, "Diagnosis of epilepsy via backpropagation." pp. 571-574.

[10] D. G. Bounds, P. J. Lloyd, B. Mathew, and G. Waddell, "A multilayer perceptron network for the diagnosis of low back pain." pp. 481-489.

[11] L. Ohno-Machado, and M. A. Musen, "Sequential versus standard neural networks for pattern recognition: an example using the domain of coronary heart disease," *Computers in biology and medicine,* vol. 27, no. 4, pp. 267-281, 1997.

[12] J. Park, and D. W. Edington, "A sequential neural network model for diabetes prediction," *Artificial intelligence in medicine,* vol. 23, no. 3, pp. 277-293, 2001.

[13] M. S. Shanker, "Using neural networks to predict the onset of diabetes mellitus," *Journal of chemical information and computer sciences,* vol. 36, no. 1, pp. 35-41, 1996.

[14] B. A. Tama, F. Rodiyatul, and H. Hermansyah, "An Early Detection Method of Type-2 Diabetes Mellitus in Public Hospital," *TELKOMNIKA (Telecommunication Computing Electronics and Control),* vol. 9, no. 2, pp. 287-294, 2013.

[15] V. A. Kumari, and R. Chitra, "Classification of diabetes disease using support vector machine," *International Journal of Engineering Research and Applications,* vol. 3, no. 2, pp. 1797-1801, 2013.

[16] A. M. Aibinu, M. J. E. Salami, and A. A. Shafie, "Application of modeling techniques to diabetes diagnosis." pp. 194-198.

[17] A. M. Aibinu, M.-J. E. Salami, and A. A. Shafie, "A novel signal diagnosis technique using pseudo complex-valued autoregressive technique," *Expert Systems with Applications,* vol. 38, no. 8, pp. 9063-9069, 2011.

[18] A. A. Al Jarullah, "Decision tree discovery for the diagnosis of type II diabetes." pp. 303-307.

[19] X.-H. Meng, Y.-X. Huang, D.-P. Rao, Q. Zhang, and Q. Liu, "Comparison of three data mining models for predicting diabetes or prediabetes by risk factors," *The Kaohsiung journal of medical sciences,* vol. 29, no. 2, pp. 93-99, 2013.

[20] K. Sowjanya, A. Singhal, and C. Choudhary, "MobDBTest: A machine learning based system

for predicting diabetes risk using mobile devices." pp. 397-402.

[21] K. Saxena, and R. Sharma, "Diabetes mellitus prediction system evaluation using C4. 5 rules and partial tree." pp. 1-6.

[22] N. A. Nnamoko, F. N. Arshad, D. England, and J. Vora, "Meta-classification model for diabetes onset forecast: A proof of concept." pp. 50-56.

[23] L. Li, "Diagnosis of Diabetes Using a Weight-Adjusted Voting Approach." pp. 320-324.

[24] M. Mira Kania Sabariah, S. Aini Hanifa, and M. Siti Sa'adah, "Early detection of type II Diabetes Mellitus with random forest and classification and regression tree (CART)." pp. 238-242.

[25] R. Ali, M. H. Siddiqi, M. Idris, B. H. Kang, and S. Lee, "Prediction of diabetes mellitus based on boosting ensemble modeling," *Ubiquitous Computing and Ambient Intelligence. Personalisation and User Adapted Services*, pp. 25-28: Springer, 2014.

[26] L. Han, S. Luo, J. Yu, L. Pan, and S. Chen, "Rule extraction from support vector machines using ensemble learning approach: an application for diagnosis of diabetes," *Biomedical and Health Informatics, IEEE Journal of,* vol. 19, no. 2, pp. 728-734, 2015.

[27] S. Khanna, and S. Agarwal, "An Integrated Approach towards the prediction of Likelihood of Diabetes." pp. 294-298.

[28] T. Raicharoen, and C. Lursinsap, "Critical support vector machine without kernel function." pp. 2532-2536.

[29] K. Kayaer, and T. Yıldırım, "Medical diagnosis on Pima Indian diabetes using general regression neural networks." pp. 181-184.

[30] S. Farhanah, and M. Darmawaty, "Diabetes Mellitus Forecast Using Artificial Neural Networks (ANN)." pp. 135-138.

[31] W. Wettayaprasit, and U. Sangket, "Linguistic Knowledge Extraction from Neural Networks Using Maximum Weight and Frequency Data Representation." pp. 1-6.

[32] H. Temurtas, N. Yumusak, and F. Temurtas, "A comparative study on diabetes disease diagnosis using neural networks," *Expert Systems with applications,* vol. 36, no. 4, pp. 8610-8615, 2009.

[33] C. Kalaiselvi, and G. Nasira, "A New Approach for Diagnosis of Diabetes and Prediction of Cancer Using ANFIS." pp. 188-190.

[34] M. K. Saad, and N. M. Hewahi, "A comparative study of outlier mining and class outlier mining," *Computer Science Letters,* vol. 1, no. 1, 2009.

[35] N. M. Hewahi, and M. K. Saad, "Class outliers mining: Distance-based approach," *International Journal of Intelligent Systems and Technologies,* vol. 2, pp. 5, 2007.

[36] T. Breuel, and F. Shafait, "Automlp: Simple, effective, fully automated learning rate and size adjustment."

[37] Y. Guo, G. Bai, and Y. Hu, "Using Bayes Network for Prediction of Type-2 Diabetes." pp. 471-472.

[38] M. El Habib Daho, N. Settouti, M. El Amine Lazouni, and M. A. Chikh, "Recognition of diabetes disease using a new hybrid learning algorithm for NEFCLASS." pp. 239-243.

[39] S. Bashir, U. Qamar, F. H. Khan, and L. Naseem, "HMV: A medical decision support framework using multi-layer classifiers for disease prediction," *Journal of Computational Science,* vol. 13, pp. 10-25, 2016.