



ENN-Ensemble based Neural Network method for Diabetes Classification

G L Aruna Kumari, Padmaja P, Jaya Suma G

Abstract: *Diabetes is considered as one of the most chronic disease which has serious impact on human health and leading cause of mortality worldwide. The early prediction of diabetes can help clinicians to provide a better diagnosis to the patients. Recently, computed aided diagnosis systems have gained attention due to significant growth in data mining, and machine learning. Several approaches are present based on the machine learning techniques but due to poor classification performance and computational complexity, it becomes difficult to utilize for real-time applications. Ensemble classification approaches have reported a noteworthy improvement in diabetes classification but desired accuracy is still a challenging task. Hence, in this work we introduce a combined hybrid approach called as ENN-Ensemble based neural network approach for diabetes classification. In this approach, a feature selection process is presented using neighboring search technique; the selected features are processed through the feature ranking model to generate the efficient feature subset for better classification accuracy. Finally, these features are learned and classified using neural network classifier. The experimental study shows that the proposed approach achieves better accuracy when compared with the existing techniques.*

Keywords: *Diabetes classification, diabetes mellitus, neural network, ensemble learning.*

I. INTRODUCTION

Diabetes mellitus is a group of chronic disease and pose several severe threats to human health. The characteristic of diabetes is that the elevated blood glucose level crosses the limit of normal glucose level. Generally, the diabetes is caused due to inoperative insulin secretion and impaired biological effects[1]. This glucose imbalance may lead towards several chronic health issues such as blindness, heart disease, kidney failure and death. The International Diabetes Federation reported that currently more than 382 million people are affected with diabetes in the worldwide and it is expected that this mark will increase up to 592 million diabetes patients by 2035. The diabetes mellitus is categorized into 2 types as Type-1 and Type-2 diabetes [2]. Type-1 diabetes affects children where patient's body is not capable of using insulin properly. Type-1 diabetes can be characterized increased thirst, high blood glucose levels and frequent urination. Similarly, Type-2 diabetes affects adults where body fails to produce sufficient amount of insulin. Type-2 is associated with occurrence of hypertension, obesity, dyslipidemia, and arteriosclerosis etc. [3].

Due to the unhealthy lifestyle, the diabetes has become a common disease in human being. Thus, the diagnosis of

diabetes is considered as an important task. In medical field, the diabetes diagnosis depends on the glucose levels, glucose tolerance, and fasting glucose levels. Diabetes is a chronic health related issue and can be controlled with the help of healthy lifestyle, food habits and exercise. However, till now, there is no permanent solution for diabetes diagnosis. In order to diagnose the diabetes, the patient requires special care, and deep knowledge of patient's medical history. Moreover, the medical tests are expensive and time consuming which can lead to tiring the patient. Thus, automated medical tests and diagnosis is always preferred. In order to improve the diagnosis, machine learning has gained attraction from research community. Moreover, in automated system, the computer aided diagnosis (CAD) systems are developed that use certain parameters to recognize the specific pattern. In the medical field, the pattern recognition and data mining play pivotal role to predict the diseases. Generally, researchers have focused on the classification based data mining schemes for early prediction of different types of diseases such as breast cancer [4], heart disease [5], Liver disorder [6], Thyroid [7], Lymphography [8], Parkinson's [9] and diabetes [10] etc. Significant amount of works have been carried out in these fields. The diabetes is one of the most chronic disease and several techniques have been introduced for early prediction of diabetes such as combined support vector machine (SVM) and Naïve Bayes statistical modeling [11], type-2 diabetes prediction using detection tree classifier [12], SVM [13], SM-RuleMiner [14] and boosting algorithms [15] etc..

Recently, ensemble classification based approaches are introduced to overcome the classification error related issues in the data mining related applications. In [16] authors introduced an ensemble classifier for software defect prediction. Ensemble classifiers are also used for multi-class imbalanced data classification [17] and time-series forecasting applications [18-19]. Limited work have been carried out for diabetes prediction using ensemble learning such as voting based classifier [20], SVM based ensemble learning [21] and k-nearest neighbors, naïve Bayes, decision tree, Support Vector Machine, fuzzy decision tree, artificial neural network, and logistic regression based learning approaches [22]. However, these techniques have gained attention but suffer from the poor classification accuracy and computational complexity. Hence, there is a need to introduce a novel ensemble approach to improve the diabetes classification accuracy.

In this work, we focus on the ensemble learning approach to predict the diabetes based on the data mining concept. In order to achieve this, the proposed ensemble approach uses kNN, decision tree, SVM and random forest. The main contributions of this work are as follows:

Revised Manuscript Received on February 05, 2020.

G L Aruna Kumari, Dept of CSE Gitam institute of Technology, Gitam University Visakhapatnam, India

Dr Padmaja P, Professor and Head, Dept. of Information Technology Anil Neerukonda Institute of Technology & Sciences, Visakhapatnam, India

Dr. Jaya Suma G, Professor and Head, Department of Information Technology JNTUK-University College of Engineering, Vizianagaram, India

Rest of the article is organized as follows: In section II literature related to several state-of-art techniques of data mining for diabetes prediction are discussed, section III suggests proposed solution for diabetes classification, section IV presents experimental and comparative study using proposed approach, and finally, section V gives conclusive remarks about the proposed approach.

1. Literature review

During last decade, the data mining and computer aided diagnosis has gained huge attraction due to machine learning and pattern recognition techniques. This section briefly describes these data mining approaches for diabetes prediction. Kandhasamy et al. [23] presented a comparative study for diabetes classification using different classifiers such as J48 Decision Tree, K-Nearest Neighbors, and Random Forest, and Support Vector Machines. The comparative study is divided into two phases as data with noise and data without noise. Comparative study shows that the Random forest classifier achieves better performance when compared with other techniques.

Pradhan et al. [24] used machine learning approach for diabetes prediction. In this work, authors incorporated genetic algorithm to design a new classifier model for diabetes prediction. In this approach, various selection methods are used during genetic programming where tournament selection achieves better performance.

Zhu et al. [25] introduced an ensemble approach to predict the type-2 diabetes. This approach is designed with the help of different types of classifiers such as support vector machine (SVM), Naïve Bayes (NB), C4.5, logistic regression (LR) and neural network (NN). With the help of this authors introduced a dynamic weighted voting scheme called multiple factors weighted combination for classifiers. According to this process, the local, global and diverse k-NN accuracies are considered and finally a voting scheme is applied to construct the final classifier.

Anto et al. [26] proposed least square support vector machine (LS-SVM) approach to construct the medical decision support system. Moreover, this model uses simulated annealing approach for optimization process. The feature selection task is carried out using Fisher score (FS). During classification phase, Radial Basis function is used for training the SVM classifier. Choubey et al. [26] presented naïve Bayes classification approach to predict the diabetes. In this work, authors suggested that the feature selection plays important role hence, authors used Genetic algorithm as feature selection. Chen et al. [27] presented a hybrid approach for diabetes classification. In the hybrid approach, the K-Means and decision tree approaches are combined. The K-means is used for reducing the data and J48 decision tree is used for classification.

Geman et al. [29] developed hybrid neuro-fuzzy approach for diabetes classification. The proposed neuro-fuzzy classifier focuses on the reduction of fuzzy rule to improve the classification accuracy and computational cost. The feature set is constructed with the help of Diabetes Pedigree Function which helps to get the patient's genetic relationship. Moreover, it constructs the robust membership functions. Similarly, Mohapatra et al. [30] used deep learning approach for diabetes classification.

Several techniques have been introduced for diabetes prediction and classification using machine learning techniques. The feature selection plays important role in these techniques but computational complexity and time are the crucial parameters for these techniques. The ensemble classification schemes show a significant improvement in this field but achieving the desired accuracy remains a tedious task for researchers.

II. PROPOSED MODEL

This section presents a proposed ensemble solution for diabetes classification which achieves better accuracy. In this approach a neural network is used as base learner, optimized pruning neighboring search for search method and information gain for feature ranking. In this proposed model features are ranked using information gain and the ranked features are used to train neural network model and generate the predictions for test data.

1.1. Optimized pruning neighboring feature selection

In order to reduce the dimensionality related issues feature selection process is introduced and a novel method presented for neighboring feature pruning based on the sequential forward search method. The proposed approach follows bottom-up search method which initializes from the empty set and selects the optimal features using evaluation functions. In each iteration, the features are selected from the remaining features and a feature set is constructed to minimize the classification error. According to this process, we have an objective function J and our main aim is to maximize the objective function. The objective function depends on the feature subset F_θ . k Numbers of features are selected to maximize the classification accuracy. This approach starts with an empty feature subset F_θ and selects the k features. Denote X_i as the random feature variable and Y is their corresponding label. The features which are useful for maximizing the objective function, are considered as selected features. This process is repeated until the cardinality condition is satisfied.

Algorithm 1

Input: PIMA Data set

Output: (i) Best neural network model with single attribute

(ii) Best neural network model with two attributes.

(iii) Best neural network model with three attributes.

and so on until

(iv) Best neural network model with n attributes.

Step 1: Maximize the objective function as

$$X_j = \arg \max_{X_j \in F} J(X_j, Y, F_\theta)$$

Step 2: Discard the selected features from the original set $F_\theta \leftarrow F \setminus X_j$

Step 3: Update the current feature set based on the objective function as $F_\theta \leftarrow F_\theta \cup X_j$

Step 4: Repeat the process until $|F_\theta| = k$

1.2. Feature ranking

The selected features using optimized approach as discussed in previous subsection are processed through the feature ranking process to rank the various features. In this work, information gain based feature ranking process is used where features are ranked using filtering module. This ranking is performed based on the entropy of the feature. Let us consider that a training set is given as S_x where x_i is the feature vector of i^{th} instance. The information gain of the given feature set can be computed as:

$$IG(S_x, x_i) = H(S_x) - \sum_{v=values(x_i)} \frac{|S_{x_i=v}|}{|S_x|} H(S_{x=v}) \quad (1)$$

with entropy

Where $H(S) = -p_+(S) \log_2 p_+(S) - p_-(S) \log_2 p_-(S)$ where $-p$ and $+p$ denotes the probability of class of feature to be in negative or positive category.

1.3. Neural Network classification

In this work, we use neural network based model to learn and classify the features. The neural network architecture is depicted in figure 1. According to this process, the input variables are processed through various layers and the corresponding class of these attributes is identified. The architecture shows a single hidden layer model and called as feed-forward neural network.

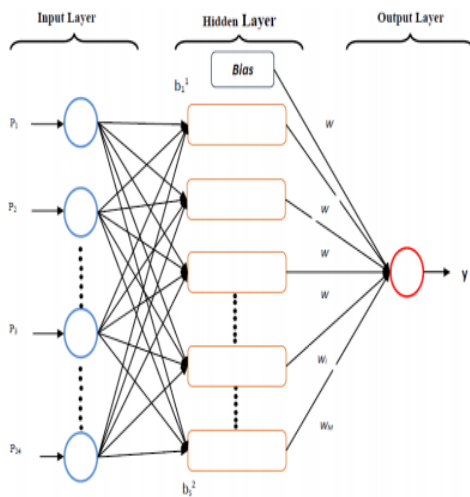


Fig.1. Neural network architecture

In these networks, each single hidden layer is connected to the previous layers with the help of weight. The knowledge process of these networks is stored in the form of neural weights. During the training process, these weights are adjusted by the network to achieve the desired outcome. Consider the Pima Indian diabetes data(PID) is the input given to the network as $p = (p_1, \dots, p_8)$ with 8 attributes and an output which denotes the predicted class. These inputs are associated with the N number of neurons in the hidden layer with the help of weights as $w_{kj}, k, j = 1, \dots, N$. These weights are specific to each individual variable to neuron. The mapping of hidden layer and output can be given as:

$$HiddenLayer\ n_k^{(1)} = \sum_{j=1}^R w_{kj}^{(1)} p_j + b_k^1 \quad (2)$$

Similarly, the output layer's mapping can be given as:

(3)

In our work, we have considered multiple neurons hence for N neurons, the biases are given as $b_1^{(1)}, b_2^{(1)}, \dots, b_N^{(1)}$. Before applying any activation function, the value of neuron k is denoted as $b_1^{(1)} + \sum_{j=1}^M w_{kj} p_j$ where M denotes the attributes. Then the activation function can be applied and the neuron value can be transformed as $f(b_1^{(1)} + \sum_{j=1}^M w_{kj} p_j)$. After applying the activation function, these values are gathered at the output layer as $\sum_{k=1}^N w'_k f_k(b_1^{(1)} + \sum_{j=1}^M w_{kj} p_j) + b_2$. Finally, these values are obtained using a function $g(\cdot)$ as $g[\sum_{k=1}^N w'_k f_k(\cdot) + b^{(2)}]$ which is considered the estimated output value for the given attribute set.

<p>Algorithm 2 Input : PID data set Output: Neural Network classifier model Step 1: Calculate information gain for each attribute</p> $IG(S_x, x_i) = H(S_x) - \sum_{v=values(x_i)} \frac{ S_{x_i=v} }{ S_x } H(S_{x=v})$ <p style="text-align: center;">with entropy</p> <p>Step 2: Generate ranking based on gain value. Step 3: Define the cutoff ranking value Step 4: With the resultant features obtained in step 3 generate a Neural Network model</p>
<p>Algorithm 3 Ensemble learning Input : Test data Output: Prediction of unknown sample Step 1: Apply test data for every model generated in algorithm1 Step 2: Apply the same test data of step1 to algorithm2 Step 3: Calculate the majority of predicted class using voting. Step 4: Assign the majority voted class label to unknown sample.</p>

III. RESULTS AND DISCUSSION

In this section, we present a detailed discussion about experimental analysis of proposed Ensemble approach for diabetes prediction. The proposed data mining approach is based on the ensemble classification where we have considered neural network as base learner. The neural network is trained using Bayesian Regulation back propagation. In order to construct an ensemble classifier Forward search optimization is incorporated and information gain for search method and feature ranking, respectively. The experimental study is carried out on the PIMA Indian Diabetes data which has total 768 data samples with 8 numerical attributes per data sample. These data are shown in table 1.

Table 1 Pima Indian Diabetes Dataset details

ID	Attribute	Range
1	No. of times pregnant	0-9
2	Plasma glucose concentration	0.-199
3	Diastolic blood pressure (mm Hg)	0.-122
4	Triceps skin fold thickness (mm)	0-99
5	2-h serum insulin (μU/ml)	0-846
6	Body mass index (kg/m2)	0-67.1
7	Diabetes pedigree function	0.078-2.42
8	Years of age	21-81
9	Class	Diabetes/Non-Diabetes

In order to evaluate the performance, we consider 70% data for training and 30% data for testing i.e. 537 samples for training and 231 samples for testing. Other implementation details are shown in table 2.

Table 2 Implementation parameters

Implementation Parameter	Considered Parameter
Base Learner	Neural network
Feature Ranking	Information gain
Training Ratio	70%
Cross validation	10
Testing ratio	30%
Hidden Layer Size	10
Number of Epochs	1000

Based on these parameters we measure the performance of proposed model of diabetes classification. The performance measurement parameters are described below.

(a) Performance Measurement

In order to measure the performance, we consider the confusion matrix analysis. First of all, we have obtained the confusion matrix where the measurement of correct and incorrect classification as given in table 3.

Table 3 Confusion Matrix

Actual class	Predicted class	
	Healthy	Diabetes
Healthy	True Positive	False Negative
Diabetes	False Positive	True Negative

This confusion matrix helps us to compute total accuracy, precision, specificity, sensitivity and F-measure of the proposed approach.

Accuracy is a measurement of rate of correct classification which is denoted by *Acc*. It is computed by taking the ratio of correct prediction and total number of prediction. It can be expressed as:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Another parameter is known as sensitivity analysis of the model. This is the measurement of true positive rate which can be computed by identifying the correctly classified non-Diabetes modules. This can be expressed as

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

Next parameter is computed as true negative rate which shows the measurement of correct classified Diabetes software modules and can be expressed as:

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

Then, we compute Precision of the proposed approach. It is computed by taking the ratio of True Positive and (True and False) positives.

$$P = \frac{TP}{TP + FP} \quad (4)$$

Finally, F-measure is computed which is the mean of precision and sensitivity performance. It is expressed as:

$$F = \frac{2 * P * Sensitivity}{P + Sensitivity} \quad (5)$$

(b) Performance analysis and comparison

In this section, we present the comparative experimental analysis in terms of different measurement metrics as mentioned in previous subsection. The obtained confusion matrix is shown in table 4.

Table 4 Confusion Matrix for PIMA Dataset

Actual class	Predicted class	
	Healthy	Diabetes
Healthy	183	7
Diabetes	14	334

$$Acc = \frac{183 + 334}{183 + 14 + 7 + 334}$$

According to the confusion matrix, we obtain the classification accuracy as 96.09%. Other performance measurement parameters for each class are presented in table 5.

Table 5 detailed performance for PIMA dataset

Class	TP	FP	Precision	Recall	F-Measure	ROC
Class-1	0.9632	0.0402	0.9289	0.9632	0.9457	0.9244
Class-2	0.9598	0.0368	0.9795	0.9598	0.9695	0.9244

The obtained performance is compared with existing data mining techniques in terms of accuracy, sensitivity, and specificity. The comparative analysis is presented in table 6.

Table 6 Comparative analysis

Algorithm	Accuracy	Sensitivity (%)	Specificity (%)
ID3 [14]	71.79	73.26	68.12
C4.5 [14]	74.20	82.60	58.58
CART [14]	70.70	73.80	64.92
SM-RuleMiner [14]	89.87	94.60	80.11
Proposed Model	96.09	96.31	95.97

Similarly, the performance of proposed model is compared with several types of meta-



heuristic rule based classification algorithms for 10 cross-validation. The comparison of proposed approach with meta-heuristic approach is given in table 7. The comparative analysis scheme is adopted from [1].

Table 7 comparison with meta-heuristic approaches

Meta-Heuristic approach	Accuracy (%)	Sensitivity (%)	Specificity (%)
ACO [31]	84.24	84.13	85.86
Modified ABC with all features [32]	82.68	81.91	83.07
Modified ABC with feature selection [32]	84.21	83.45	84.60
ABC with all features [32]	79.61	78.54	80.18
ABC with feature selection [32]	81.40	80.76	82.88
PSO [33]	82.32	84.60	80.76
SM-RuleMiner [14]	89.87	94.60	80.11
Proposed Approach EFRNN	96.09	96.31	95.97

Furthermore, the accuracy performance is compared with state-of-art techniques such as adaptive fuzzy [5], Hybrid SVR [6], Rule mining with POA [7], Modified PSO [8], Re-RX with J48graft [9], Re-RX with C4.5 [9], QFAM-GA [10] and DNN-SAE [11]. The comparative performance is presented in table 8.

Table 8 Accuracy comparison

Method	Accuracy
adaptive fuzzy [34]	89.80
Hybrid SVR [35]	86.13
Rule mining with POA [36]	79.06
Modified PSO [37]	85.19
Re-RX with J48graft [38]	83.83
Re-RX with C4.5 [38]	80.83
QFAM-GA [39]	91.91
DNN-SAE [40]	86.26
SM-RuleMiner [14]	89.87
Proposed Model	96.09

IV. CONCLUSION

In this paper, main focus is on the development of machine learning based approach for early prediction of diabetes using computer aided diagnosis systems. Several techniques have been presented in this field but achieving accuracy and computational complexity remain challenging tasks. Therefore ensemble classifier model is introduced where feature selection, feature ranking and neural network based scheme is ensemble to learn the feature efficiently and classify the patterns. The experimental study shows that the proposed approach achieves better performance in terms of classification accuracy.

REFERENCES

- Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting diabetes mellitus with machine learning techniques. *Frontiers in genetics*, 9.
- Wu, H., Yang, S., Huang, Z., He, J., & Wang, X. (2018). Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked*, 10, 100-107.
- Robertson G., Lehmann E. D., Sandham W., Hamilton D. (2011). Blood glucose prediction using artificial neural networks trained with the AIDA diabetes simulator: a proof-of-concept pilot study. *J. Electr. Comput. Eng.* 2011:681786 10.1155/2011/681786
- Lu, J., Hales, A., Rew, D., Keech, M., Fröhlingsdorf, C., Mills-Mullett, A., &Wette, C. (2015, September). Data mining techniques in health informatics: a case study from breast cancer research. In *International Conference on Information Technology in Bio-and Medical Informatics* (pp. 56-70). Springer, Cham.
- Tayefi, M., Tajfard, M., Saffar, S., Hanachi, P., Amirabadizadeh, A. R., Esmaily, H. ...&Ghayour-Mobarhan, M. (2017). Hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm. *Computer methods and programs in biomedicine*, 141, 105-109.
- Dixon, S., & Yu, X. H. (2015, August). Liver disorder detection based on artificial immune systems. In *2015 11th International Conference on Natural Computation (ICNC)* (pp. 743-748). IEEE.
- Pan, Q., Zhang, Y., Zuo, M., Xiang, L., & Chen, D. (2016, December). Improved ensemble classification method of thyroid disease based on random forest. In *2016 8th International Conference on Information Technology in Medicine and Education (ITME)* (pp. 567-571). IEEE.
- Tanveer, M., Gautam, C., &Suganthan, P. N. (2019). Comprehensive evaluation of twin SVM based classifiers on UCI datasets. *Applied Soft Computing*, 83, and 105617. doi:10.1016/j.asoc.2019.105617.
- Miljkovic, D., Aleksovski, D., Podpečan, V., Lavrač, N., Malle, B., &Holzinger, A. (2016). Machine learning and data mining methods for managing Parkinson's disease. In *Machine Learning for Health Informatics* (pp. 209-220). Springer, Cham.
- Chen, W., Chen, S., Zhang, H., & Wu, T. (2017, November). A hybrid prediction model for type 2 diabetes using K-means and decision tree. In *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)* (pp. 386-390). IEEE.
- Tafa, Z., Pervetica, N., &Karahoda, B. (2015, June). An intelligent system for diabetes prediction. In *2015 4th Mediterranean Conference on Embedded Computing (MECO)* (pp. 378-382). IEEE.
- Habibi, S., Ahmadi, M., &Alizadeh, S. (2015). Type 2 diabetes mellitus screening and risk factors using decision tree: results of data mining. *Global journal of health science*, 7(5), 304.
- Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., &Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal*, 15, 104-116.
- Cheruku, R., Edla, D. R., &Kuppili, V. (2017). SM-RuleMiner: Spider monkey based rule miner using novel fitness function for diabetes classification. *Computers in biology and medicine*, 81, 79-92.
- Chen, P., & Pan, C. (2018). Diabetes classification model based on boosting algorithms. *BMC bioinformatics*, 19(1), 109.
- Laradji, I. H., Alshayeb, M., &Ghouti, L. (2015). Software defect prediction using ensemble learning on selected features. *Information and Software Technology*, 58, 388-402.
- Haixiang, G., Yijing, L., Yanan, L., Xiao, L., &Jinling, L. (2016). BPSO-Adaboost-KNN ensemble learning algorithm for multi-class imbalanced data classification. *Engineering Applications of Artificial Intelligence*, 49, 176-193.
- Tang, L., Wu, Y., & Yu, L. (2018). A non-iterative decomposition-ensemble learning paradigm using RVFL network for crude oil price forecasting. *Applied Soft Computing*, 70, 1097-1108.
- Chen, J., Zeng, G. Q., Zhou, W., Du, W., & Lu, K. D. (2018). Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. *Energy conversion and management*, 165, 681-695.
- Husain, A., & Khan, M. H. (2018, April). Early Diabetes Prediction Using Voting Based Ensemble Learning. In *International Conference on Advances in Computing and Data Sciences* (pp. 95-103). Springer, Singapore.
- Yang, Z., Zhou, Y., & Gong, C. (2019, March). Diagnosis of Diabetes Based on Improved Support Vector Machine and Ensemble Learning. In *Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence* (pp. 177-181). ACM.
- El-Sappagh, S., Elmogy, M., Ali, F., Abuhmed, T., Islam, S. M., &Kwak, K. S. (2019). A



- Comprehensive Medical Decision–Support Framework Based on a Heterogeneous Ensemble Classifier for Diabetes Prediction. *Electronics*, 8(6), 635.
23. Kandhasamy, J. P., &Balamurali, S. (2015). Performance analysis of classifier models to predict diabetes mellitus. *Procedia Computer Science*, 47, 45-51.
 24. Pradhan, M., &Bamnote, G. R. (2015). Design of Classifier for Detection of Diabetes Mellitus Using Genetic Programming. In *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014* (pp. 763-770). Springer, Cham.
 25. Zhu, J., Xie, Q., &Zheng, K. (2015). An improved early detection method of type-2 diabetes mellitus using multiple classifier system. *Information Sciences*, 292, 1-14.
 26. Anto, S., Chandramathi, S., &Aishwarya, S. (2016). An expert system based on LS-SVM and simulated annealing for the diagnosis of diabetes disease. *International Journal of Information and Communication Technology*, 9(1), 88-100.
 27. Choubey, D. K., Paul, S., Kumar, S., & Kumar, S. (2017, February). Classification of Pima indian diabetes dataset using naive bayes with genetic algorithm as an attribute selection. In *Communication and Computing Systems: Proceedings of the International Conference on Communication and Computing System (ICCCS 2016)* (pp. 451-455).
 28. Chen, W., Chen, S., Zhang, H., & Wu, T. (2017, November). A hybrid prediction model for type 2 diabetes using K-means and decision tree. In *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)* (pp. 386-390). IEEE.
 29. Geman, O., Chiuchisan, I., &Todorean, R. (2017, June). Application of Adaptive Neuro-Fuzzy Inference System for diabetes classification and prediction. In *2017 E-Health and Bioengineering Conference (EHB)* (pp. 639-642). IEEE.
 30. Mohapatra, S. K., Nanda, S., &Mohanty, M. N. (2018, April). Diabetes Detection Using Deep Neural Network. In *International Conference on Soft Computing Systems* (pp. 225-231). Springer, Singapore.
 31. M.F. Ganji, M.S. Abadeh, A fuzzy classification system based on ant colony optimization for diabetes disease diagnosis, *Expert Syst. Appl.* 38 (12) (2011)
 32. F. Beloufa, M. Chikh, Design of fuzzy classifier for diabetes disease using modified artificial bee colony algorithm, *Comput. Methods Prog. Biomed.* 112 (1) (2013) 92–103
 33. M.E.H. Daho, N. Settouti, M.E.A. Lazouni, M.A. Chikh, Recognition of diabetes disease using a new hybrid learning algorithm for nefclass, in: *Systems, Signal Processing and their Applications (WoSSPA)*, 2013 Proceedings of the 8th International Workshop on, IEEE, 2013, pp. 239–243
 34. B. Dennis, S. Muthukrishnan, Agfs: adaptive genetic fuzzy system for medical data classification, *Appl. Softw. Comput.* 25 (2014) 242–252
 35. M.H. Zangoeei, J. Habibi, R. Alizadehsani, Disease diagnosis with a hybrid method svr using nsga-ii, *Neurocomputing* 136 (2014) 14–29.
 36. S. Kiziloluk, B. Alatas, Automatic mining of numerical classification rules with parliamentary optimization algorithm, *Adv. Electr. Comput. Eng.* 15 (4) (2015) 17–24.
 37. H.R. Sahebi, S. Ebrahimi, I. Ashtian, Afuzzy classifier based on modified particle swarm optimization for diabetes disease diagnosis, *Adv. Comput. Sci.: Int. J.* 4 (3) (2015) 11–17
 38. Y. Hayashi, S. Yukita, Rule extraction using recursive-rule extraction algorithm with j48graft combined with sampling selection techniques for the diagnosis of type 2 diabetes mellitus in the pima indian dataset, *Inform. Med. Unlocked* 2 (2016) 92–104
 39. F. Pourpanah, C.P. Lim, J.M. Saleh, A hybrid model of fuzzy artmap and genetic algorithm for data classification and rule extraction, *Expert Syst. Appl.* 49 (1) (2016) 74–85
 40. Kannadasan, K., Edla, D. R., &Kuppili, V. (2018). Type 2 diabetes data classification using stacked autoencoders in deep neural networks. *Clinical Epidemiology and Global Health*.



Dr. P. Padmaja is presently working as Professor & Head, Department of Information Technology, Anil Neerukonda Technology & Sciences affiliated to Andhra University. She published more than 30 research papers. She guided a number of students for their Ph.D and M.Tech degrees in Information Technology. Her current research interests are Data mining, machine learning.



Dr. G. Jaya Suma is presently working as Professor & Head, Department of Information Technology, JNTUK-University College of Engineering Vizianagaram, Andhra Pradesh. She is Convener for women empowerment & grievance cell from 2017. She published more than 45 research papers in National and International journals with high reputation. Her research interests are Data mining, machine learning, soft computing, mobile computing.

AUTHORS PROFILE:



G.L. Aruna Kumari is presently working as Assistant Professor in the department of Computer Science and Engineering, GIT, GITAM University, Visakhapatnam. She presented research papers in national and international conferences and journals of good repute. She guided several students for Project work in department of Computer Science Engineering. Her current research interests include data mining and machine learning.