ISSN 2278-3091
Volume 9, No.4, July – August 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse379942020.pdf
https://doi.org/10.30534/ijatcse/2020/379942020

Logistic regression and Random forest-based hybrid classifier with recursive feature elimination technique for diabetes classification

Aruna Kumari G L [1], Dr Padmaja P [2], Dr Jaya Suma G [3]

[1]Dept of CSE,Gitam institute of Technology,Gitam University,
Visakhapatnam, India, agorli@gitam.edu
[2] Dept.of Information Technology,Anil Neerukonda Institute of Technology &Science,India,padmaja.it@anits.edu.in
[3] Department of Information Technology,JNTUK-University College of Engineering,
Vizianagaram,India,gjsuma.it@jntukucev.ac.in

ABSTRACT

Diabetes mellitus is a chronic metabolic ailment being considered as one of the deadliest diseases in the world. Millions of cases and deaths have been reported due to diabetes. Initial forecast of diabetes condition helps in reducing the death rate significantly that happen due to it. Current advancements in biomedical techniques have facilitated to store the electronic health record datasets which can be analyzed for better diagnosis. In order to explore these datasets, data mining techniques are considered as promising techniques which examines the data rapidly and provides a desired outcome. In this work, the data mining technique was adopted for diabetes classification using machine learning techniques. The proposed approach comprises of several steps such as missing value imputation, attribute selection and classification which are performed using mean missing value imputation, logistic regression &Recursive Feature Elimination for feature selection and random forest for classification, respectively. The evaluation results demonstrate that projected methods achieve97.39% prediction accuracy which shows a significant improvement in contrast to prevailing current approaches.

Key words: diabetes classification, missing value imputation, feature selection, Random forest classification

1. INTRODUCTION

Diabetes mellitus is basically a term for a group of diseases which gets occurred due to metabolic disorders and represented by the high level of blood sugar [1]. It is also known as diabetes. The inappropriate diagnosis and care of diabetes can cause serious health issues such as kidney failure, cardiovascular disease, heart attack, stroke, blood vessels, arterial disease, and damage to the nerves [2,3,4]. A study presented in [5] reported that about 122M population was suffering with diabetes in 1980s. Later, in 2014, this number has increased and reached about 422 million people [5] and it is estimated that this figure will reach about 642 million diabetes patients in 2040 [6]. Moreover, about 1.6 million death count has been reported due to diabetes [7]. Hence, it becomes a challenging task for biomedical researchers to prevent the increase of diabetes. The best way to reduce the death and diabetes patient counts is to predict it in early-stage and get the appropriate diagnosis. Thus, the early prediction is an important task that can help to reduce the diabetes patient count.

Generally, diabetes is categorized into three main categories as, type-I, type-II and gestational diabetes (GD). The types-I diabetes normally affects the youngsters with less than 30 years of age. The early signs of Type-1D are realized as weight loss, polyuria, hunger, thirst, and vision change. The type-2D affects the individuals aged over 45 years and the persons associated with hypertension, obesity, arteriosclerosis and dyslipidemia [8]. The third type of diabetes mostly affects pregnant women. The early prediction of diabetes is a tedious task because the medical data are nonlinear, unstructured and complex in nature. Currently, research community has focused on development of the automated systems for diabetes prediction. Machine learning (ML) and Data Mining (DM) based techniques have gained attraction to handle these types of data to learn the patterns for early prediction. Data mining is used to mine the probable information from a huge data corpus while making use of certain algorithms. Currently, data mining techniques are embraced globally in diverse real time applications such as business and medical field. Current advancements in computer technology have facilitated several advantages to the medical treatments. Nowadays, huge amount of data is recorded and stored in medical institutions which are considered as useful resource for disease research. the data mining technique can process large amount of historical medical data and diagnosis rules can be built to improve the diagnosis performance. The data mining and machine learning techniques are widely adopted in diabetes

classification such as Rashid et al. [9] used Artificial Neural Network (ANN) and Decision Tree Classifier (DT) for diabetes classification, neuro-fuzzy classifier [10], convolution neural networks [11], AdaBoost[12], improved KNN [13], decision tree & Random Forest [14],SVM, k-nearest neighborhood, Naïve Bayes, logistic regression and many more [15]. The performance of these techniques depends on the attributes. The selection of optimal attributes is an important task. Several techniques are present based on feature selection scheme such as rough set [16], genetic algorithm [17], missing value imputation & F-score based feature selection [18]. These techniques of feature selection are used to choose substantial features which helps to improve the classification accuracy performance. Due to the significant improvement in the classification accuracy performance, and main focus on the machine learning based technique and introduce a novel classification algorithm for diabetes classification. However, the existing techniques suffer from various issues such as computational complexities, and poor classification accuracy.

1.1. Brief description and contributions of proposed approach

The proposed model contains several stages such as missing value imputation, attribute selection, random forest classification and performance measurement. The first phase is to perform missing value imputation where mean modeling is used for estimating the missing value. The missing value helps to improve the classification accuracy. The next phase is to perform attribute selection where logistic regression model and recursive feature elimination method are used. Further, employed random forest classifier to classify the data with the help of decision trees. Finally, measured the efficiency of proposed approach in terms of classification accuracy, precision, recall and F1-score.

This article is arranged in the following manner: the 2nd section of this paper briefs out some of the latest and standard techniques in the area of diabetes classification using machine learning methods, 3rd section of the article describes proposed scheme to enhance diabetes classification accuracy, subsequently, 4th section evaluates the efficiency of proposed scheme and compares with standard existing methods, lastly,5th section lists out the concluding remarks and give directions for the future works.

2. LITERATURE SURVEY

Here, review of the contemporary techniques of diabetes classification which includes techniques for feature selection and classification.

Rashid et al. [9] presented data mining-based model for diabetes classification which uses two sub-modules to establish the relationship between diabetes and blood sugar rate. The first one makes use of ANN to predict the rate of fasting blood sugar (FBS), the second sub-module establishes the relation between FBS and symptoms of patient's healthiness by using decision tree classifier. Gemanet al. [10] presented a hybrid ANFIS (Adaptive Neuro-Fuzzy Inference System)to classify diabetes patients by making use of diabetes pedigree function for fuzzy rules. Zhong et al. [13] focused on prediction of Gestational diabetes mellitus using machine learning by using improved KNN and improved Back Propagation (BP) neural network. Sejdinović et al. [19] used ANN for classification of pre diabetes and type-II diabetes. Choubey et al. [20] presented feature selection-based strategy where genetic algorithm (Genetic Algorithm) is applied for attribute selection which reduces 4 attributes among 8 attributes. In next phase, Radial Basis Function Neural Network (RBF NN) is applied to classify these attributes. Similar to this, Akyol et al. [22] presented feature selection scheme combined with ensemble learning method for diabetes classification. In the first phase, weighting methods are applied for feature selection. Later, AdaBoost, gradient boosted trees and Random forest-based ensemble methods are applied to classify the data. Ndaba et al. [21] introduced an improved GRNN (Generalized Regression Neural Network)to predict T-2 diabetes. Their technique enhances K-means clustering to generate the centroids of clusters which are used to train the network. Deep learning-based techniques are also widely adopted in various real-time machine learning based systems. Kannadasan et al. [23] focused on the type-II diabetes classification using deep learning scheme with stacked auto encoders. The auto encoders analyzes the features and softmax layer performs data classification. To further improve the performance, a backpropagation model is incorporated in a supervised manner which is used for fine tuning the network. Prabhu et al. [27] presented deep learning model for type-II diabetes classification. According to this process, the diabetes data is pre-processed and normalized later deep belief neural network is applied for data classification. In some case, the imbalanced data and missing values in the recorded data degrades the classification performance. To overcome this issue, Wang et al. [24] presented a combined approach where first of all, Naïve Bayes method is applied to identify and impute the missing values. Later, class imbalance problem is solved by using an adaptive synthetic sampling method (ADASYN), finally, random forest (RF) classifier is applied to classify the data and generate the prediction. Choubey et al. [25] used feature selection scheme using PSO-SVM and later fuzzy decision tree classifier is applied for diabetes classification. Lukmanto et al. [18] used F-score feature selection and fuzzy support vector machine for classification. SVM is used for training the dataset which helps to generate the fuzzy rules. With the help of these rules, the fuzzy inference classifies the selected attributes. Erdem et al. [26] focused on evolutionary computation approach and presented multi-objective genetic algorithm with symbolic regression to find best solution according to the formulated problem. Finally, majority voting scheme is applied for classification. Zheng et al. [28] focused on Type 2 Diabetes Mellitus and developed a semi-automated machine framework to improve the classification performance. Several classifiers are evaluated such as KNN, Naïve Bayes (NB), SVM,Decision Tree (DT), Random Forest and Logistic Regression (LR).Polat et al. [31] used

PCA for dimension reduction and combination of ANN and Fuzzy logic as adaptive neuro-fuzzy inference for prediction of class. Kahramanli et al. [32] presented a novel methodology using ANN and fuzzy Neural Network. Temurtas et al. [33] presented a comparative research on PIMA Indian diabetes classification techniques. This work uses multilayer NNthat is trained using Levenberg–Marquardt (LM) scheme and a probabilistic NN. Caliskan et al. [34] presented deep learning-based strategy where internal parameter spaces are divided into different partitions and each partition is optimized individually using L-BFGS optimization. This deep learning scheme uses auto encoder with soft max classifier [35].

3. PROPOSED MODEL

Previous sections present various aspects of diabetes and recent techniques of diabetes classification using data mining and machine learning techniques. However, the existing techniques suffer from various issues and challenges. Hence, to overcome these issues, and presented a novel methodology to classify the diabetes disease using data mining techniques. Our proposed method uses missing value imputation, feature selection and random forest-based classification approach [36-39]. The complete overview of proposed model is presented in below given figure 1.

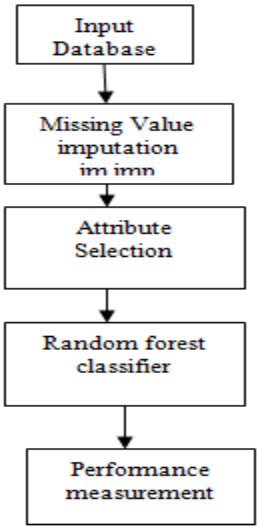

Figure 1: Process of proposed approach of diabetes classification

3.1 Missing value imputation

The missed or absent values in the dataset lead towards the poor classification or misclassification. In this work, mean imputation method is used where each missing attribute is substituted with the average of other known values of that particular column. Let is the missing attribute value of instance which can be assigned as:

Here ,I(complete), is the array of indexes which aren't absent in Xi, and nI(complete) denotes the sum of instances in which jth attribute isn't absent. Here presented the outcome of mean imputation in pima Indian diabetes dataset. selected original input as 5x8 table where 5 instances and their 8 attributes are considered. Below given table 1 gives information about abbreviations and table 2 shows the input for imputation process.

Table 1: Abrivations

Parameter	Symbol
No. of times pregnant	Np
Plasma glucose concentration	Glu
Diastolic blood pressure (mm Hg)	Dbp
Triceps skin fold thickness (mm)	Tskft
2-h Serum insulin (µU/ml)	2-h Si
Body mass index (kg/m2)	BMI
Diabes pedigree function	Dpf
Years of age	Ya
Identification of Type 2 Diabetes	It2b

Table 2: Input data for missing value imputation

Np	Glu	Dbp	Tskft	2-h Si	BMI	Dpf	ya
5	139	64	35	140	28.6	0.411	26
1	96	122	0	0	22.4	0.207	27
10	101	86	37	0	45.6	1.136	38
0	141	0	0	0	42.4	0.205	29
0	125	96	0	0	22.5	0.262	21

In this table, the 0 value indicates the missing value. This data in process through the mean missing value imputation which gives the output as presented in table 3.

Table 3: Output of missing value imputation

Np	Glu	Dbp	Tskft	2-h Si	BMI	Dpf	ya
5	139	64	35	140	28.6	0.411	26
1	96	122	29	149.61	22.4	0.207	27
10	101	86	37	149.61	45.6	1.136	38
4.45	141	71.625	29	149.61	42.4	0.205	29
4.45	125	96	29	149.61	22.5	0.262	21

With the help of this, the missing value problem is solved and also it helps to improve the classification accuracy.

3.2 Feature Selection

Feature selection plays an important role in the field of DM and ML by reducing the computational complexity and selecting the significant features to improve the classification performance. In this work, used a combined model of logistic regression and recursive feature elimination for attribute selection. Regression is one of the

Table4: Attributes details of PIMA Indian diabetes dataset

Attribute number	Name of attribute	Description	Minimum	Maximum	Mean	Standard deviation
1	Np	SC	0	17	3.8	3.4
2	Glu	SC	0	199	120.9	32.0
3	Dbp	SC (mm Hg)	0	122	69.1	19.4
4	Tskft	SC mm	0	99	20.5	16.0
5	2-h Si	SC in (mu U/ml)	0	846	79.8	115.2
6	BMI	SC (wt in kg/(ht in m)^2)	0	67.1	32.0	7.9
7	Dpf	SC	0.078	2.42	0.5	0.3
8	ya	SC	21	81	33.2	11.8
9	It2b	Y=1 N=0	NA	NA	NA	NA

machine learning technique which helps to associate the relation between dependent and target variables [40, 41]. Generally, the regression models are categorized as linear, polynomial and logistic regression. The linear and polynomial models use numeric variables whereas logistic regression uses categorical variables. In this work, we have diabetes classes as the categorical for hence, use of logistic regression model to construct the feature selection model. Let us consider that the dataset contains N points where each point contains a set of Iinput variables as x_(1,i),x_(2,i)…x_(m,i) which are independent variables and known as attributes, and binary output which contains dependent variables known as target class. Used a sigmoid function to normalize the values in the range of [0,1]. The logistic function can be defined as:

$$\sigma(t) = \frac{1}{1 + e^{-t}} \qquad (2)$$

The dataset contains several attributes, however each attribute doesn't have a significant role to classify the data accurately is listed in table 4. Hence, removing the features with low importance helps to improve the accuracy and reduces complexity. In this work, used Recursive feature elimination (RFE) technique, which uses logistic regression model for fitting the attributes and removes the features which shows low importance. This process is repeated until the entire attribute list is processed. According to the elimination, the features are ranked for further processing. The attributes which are having higher value, are selected. Below given table 5 shows attribute selection process for the different number of attributes.

Table 5: Attribute selection and ranking

Number of attributes	Selected attributes	Feature Ranking
2	1,7	1,4, 3, 5,7,2,1, 6
3	1,6,7	1, 3, 2, 4, 6, 1, 1, 5
4	1,3,6,7	1,2,1,3,5,1,1,4
5	1,2,3,6,7	1,1,1,2,4,1,1,3
6	1,2,3,4,6,7	1,1,1,1,3,1,1, 2

These selected attributes are processed through the random forest classifier to classify the diabetes data.

3.3 Random Forest classification

In this section, presented the random forest classifier modeling for dataset classification. Random forest is considered as a promising ensemble learning scheme in the arena of pattern recognition and ML. This scheme is associated with tree classifiers. However, the tree classifiers suffer from the higher variance issues, hence, the minor variation in training data may lead towards the significant change in the constructed decision tree because of hierarchal nature of trees. Let us consider that a learning set is presented as $L = ((M_1, N_1) \ldots\ldots$ contains number of vectors, in which signifies the observations or training data and which signifies the class labels. During classification, the data is mapped according to their class labels as with the help of trees. The main aim of random forest to construct the binary sub-trees with the help of the training bootstrap sample from learning dataset. The considered random forest method uses bagging and random feature selection models from Breiman's and Ho's concept to

Algorithm 1: Random forest algorithm classifier

Step 1:*Start*
Step 2: *Create classifiers*
Step 3:*For i=1 till c*
Arbitrarily sample the training samples X with replacement for generate the trained dataset
Construct a decision node in the tree as N_i *which contains the training samples*
Step 4: *check the number of classes in* N_i
Step 5 : *if N contains only one class then*
Return
Else
Step 6: *randomly select certain number of features from N*
Step 7: *generate the T number of child nodes as* $N,N_1 \ldots N_f$ *the selected features f*
Step 8: *For i=1 to f*
Map the contents of N_i *to* X_i *where* X_i

denotes the total instances in N
Build decision tree
End for
End if
***Step 9** End.*

Construct the ensemble classifier. The random forest method selects the features and constructs the decision node for database training. During the training process, the training set is generated by considering random examples from the training set for each classifier. In this way, all classifiers training sets are aggregated and a final classifier is constructed. Below given figure presents algorithm steps for the random forest classifier.

4. RESULTS AND DISCUSSION

Here, presented the experimental outcomes and comparative analysis using proposed classification approach. The proposed approach is implemented using Python 3.7on Intel Core i5 3.40GHz CPU with 8GB RAM on windows 10 OS. The performance of proposed scheme is evaluated against existing methods as mentioned in [23] in terms of classification accuracy. The performance measurement metrics are described in the following subsection. The dataset is obtained from the publicly available UCI repository which has overall 768 attributes with diabetes and non-diabetes classes. Below given table 1 describes the dataset details such as attribute details, mean and standard deviation values.

4.1 Evaluation metrics

This section describes the performance measurement parameters such as classification accuracy, precision, recall and F1-score. These metrices are obtained with the help of Confusion Matrix (CM) which contains the instance counts of correctly classified samples and incorrectly classified. Below given table 6 shows a two-class CM.

Table 6: Confusion matrix

Actual class	**Predicted class**	
	Healthy	Diabetes
Healthy	TP	FN
Diabetes	FP	TN

With the help of this matrix, compute classification accuracy which is a measurement of rate of correct classification. It is measured by taking the ratio of correctly classified instances and total instances. This can be expressed as:

$$Acc = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}$$

Later, compute precision matrix with the help of true positive and false positive instances. This computation can be expressed as:

.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Similarly, recall value is computed using true positive and false negative values which is computed as:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Finally, F1-score is calculated that is the average of precision and sensitivity outcomes. It is measured as:

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall}$$

4.2 Efficiency measurement and comparative analysis

The performance of proposed approach is measured into two test cases as training data and testing data. Later obtained all performance metrics for both training and testing datasets. The CM of training dataset is presented in below given table 3. With the help of above matrix table, attain the accuracy performance as 98.88% and other performance parameters as given in table 4. Similarly, obtained the CM values for testing data as presented in below table 5. This matrix is measured for 30% testing dataset. With the help of this confusion matrix, compute other performance measurement parameters such as precision, recall, F1- score and accuracy as presented in table 7.

Table 7: Confusion matrix for training dataset

Actual class	Predicted class	
	Healthy	Diabetes
Healthy	348 (TP)	1 (FN)
Diabetes	6 (FP)	182 (TN)

Table 8: Statistical performance parameters for training dataset.

Class	Precision	Recall	F1-score	Accuracy
0	0.96	1	0.98	98.88%
1	1	0.92	0.96	

Table 9: Confusion matrix for testing dataset

Actual class	Predicted class	
	Healthy	Diabetes
Healthy	151 (TP)	0 (FN)
Diabetes	6 (FP)	73 (TN)

Table 10: Statistical performance for training dataset.

Class	Precision	Recall	F1-score	Accuracy
0	0.96	1	0.98	97.39
1	1	0.92	0.96	

According to proposed experiment, obtained classification accuracy as 98.88%, and 97.39% for training and testing dataset. compare the performance of the proposed approach with existing techniques as presented in below given table 8, table 9 table 10 .

Table.11. Comparative analysis in terms of classification accuracy

Authors	Method Name	Accuracy
Kayaer et al. [29]	General Regression Neural Network	80.21%
Mohamadi et al. [30]	Simulated Annealing	75.71 ± 4.41
Polat et al. [31]	Multilayer Perceptron + Back Propagation	75.8 ± 6.2
	Smart	76.8
	LDA	77.5
	Quadratic Discriminant Analysis	59.5
	SNBa	75.4
	DIPOL	77.6
	Semi-NB	76.0 ± 0.8
	OCN2	65.1 ± 1.1
	MML Tree	75.5 ± 7.8
	k-Nearest Neighbour	71.9
	MML	75.5 ± 6.3
	IB3	71.7 ± 5.0
	Least Square -Support Vector Machine	78.21
	General Discriminant Analysis -LS-SVM	79.16
Kahramanli et al. [32]	Logdisc	77.7
	Back Propagation	75.2
	k-Nearest Neighbour	76.7 ± 4.0
	ASR	74.3
	SSV Decision Tree	73.7 ± 4.7
	FDA	76.5
	LFC	75.8
	Hybrid system	84.2
	Kohonen	72.7
Temurtas et al. [33]	Multi-Layer Neural Network with LM	79.62
	MLNN with levenberg-marquardt	82.57
	Deep Neural Network L-BFGS	77.09
Caliskan et al. [34]	AparseAutoencoder□based DNN	86.26
Proposed Approach	Logistic Regression and Random Forest	97.39

The comparative analysis shows that the proposed approach achieves better performance when compared with the state-of-art techniques shown in table 11 and figure 2.

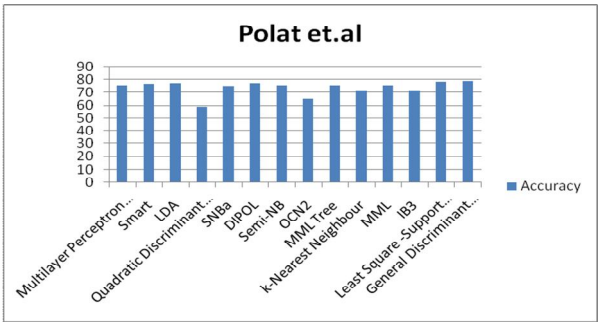

Figure 2. a

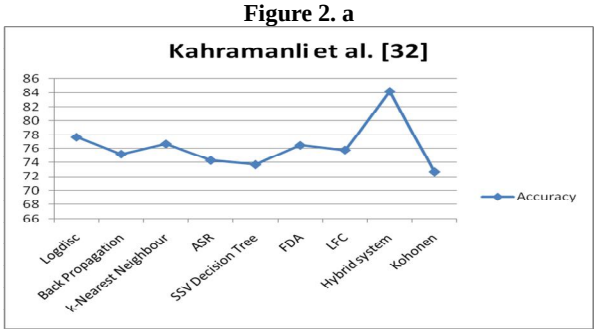

Figure 2. b

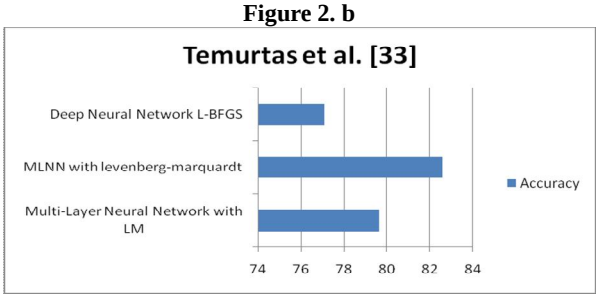

Figure 2. c

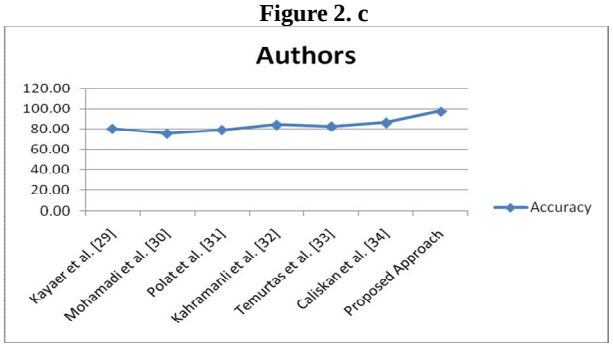

Figure 2. d

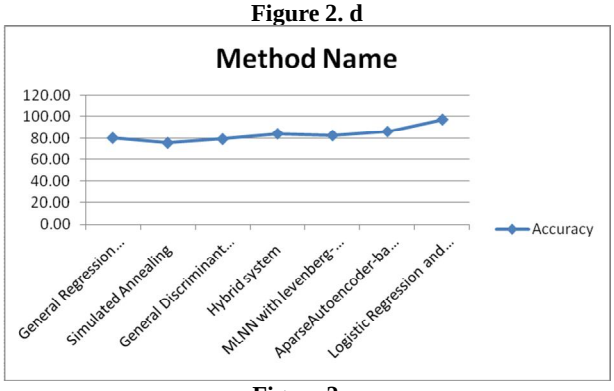

Figure 2. e

5. CONCLUSION

In this work, mainly focused on diabetes classification using DM and ML approaches. Several techniques were presented in the past for it but the existing techniques suffer from the performance issues due to missing values in the diabetes data, poor selection of attributes and learning error. To overcome these issues, introduced three different techniques for each phase such as missing value imputation using mean value, logistic regression with Recursive Feature Elimination for attribute selection and random forest for classification. The proposed method is assessed on PIMA Indian diabetes and implemented using Python 3.7. The experimental results demonstrate a significant improvement in the classification accuracy.

REFERENCES

1. American Diabetes Association. (2017). 2. **Classification and diagnosis of diabetes. Diabetes care**, 40(Supplement 1), S11-S24.
2. Hippisley-Cox, J., & Coupland, C. (2016). **Diabetes treatments and risk of amputation, blindness, severe kidney failure, hyperglycaemia, and hypoglycaemia: open cohort study in primary care**. *bmj*, *352*, i1450.
3. Mohammedi, K., Woodward, M., Hirakawa, Y., Zoungas, S., Williams, B., Lisheng, L., ... &Marre, M. (2016). **Microvascular and macrovascular disease and risk for major peripheral arterial disease in patients with type 2 diabetes. Diabetes care**, 39(10), 1796-1803.
4. Jende, J. M., Groener, J. B., Rother, C., Kender, Z., Hahn, A., Hilgenfeld, T., ... & Pham, M. (2019). **Association of serum cholesterol levels with peripheral nerve damage in patients with type 2 diabetes. JAMA network open,** 2(5), e194798-e194798.
5. NCD Risk Factor Collaboration. (2016). **Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19· 2 million participants. The Lancet**, 387(10026), 1377-1396.
6. Zheng, Y., Ley, S. H., & Hu, F. B. (2018). **Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nature Reviews Endocrinology**, 14(2), 88.
7. Bharath, C., Saravanan, N., &Venkatalakshmi, S. (2017). **Assessment of knowledge related to diabetes mellitus among patients attending a dental college in Salem city-A cross sectional study.** *Brazilian Dental Science, 20*(3), 93-100.
8. 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**. Journal of Electrical and Computer Engineering, 2011.
9. Rashid, T. A., Abdullah, S. M., & Abdullah, R. M. (2016). **An intelligent approach for diabetes classification, prediction and description**. In Innovations in Bio-Inspired Computing and Applications (pp. 323-335). Springer, Cham.
10. Geman, O., Chiuchisan, I., &Toderean, 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.
11. Kogias, K., Andreadis, I., Dalakleidi, K., & Nikita, K. S. (2018, July). **A two-level food classification system for people with diabetes mellitus using convolutional neural networks.** In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 2603-2606). IEEE.
12. Wang, Y., Liu, S., Chen, R., Chen, Z., Yuan, J., & Li, Q. (2017). **A novel classification indicator of Type 1 and Type 2 diabetes in China. Scientific reports,** 7(1), 1-7.
13. Zhong, W., Jiang, S., Wu, W., Peng, D., Xu, T., Wang, J., & Wang, G. (2019, October). **Gestational Diabetes Mellitus Prediction Based on Two Classification Algorithms. In 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics** (CISP-BMEI) (pp. 1-7). IEEE.
14. Hebbar, A., Kumar, M., & Sanjay, H. A. (2019, July). **DRAP: Decision Tree and Random Forest Based Classification Model to Predict Diabetes**. In 2019 1st International Conference on Advances in Information Technology (ICAIT) (pp. 271-276). IEEE.
15. Choudhury, A., & Gupta, D. (2019). **A survey on medical diagnosis of diabetes using machine learning techniques. In Recent Developments in Machine Learning and Data Analytics** (pp. 67-78). Springer, Singapore.
16. Cheruku, R., Edla, D. R., Kuppili, V., &Dharavath, R. (2018). Rst-batminer**: A fuzzy rule miner integrating rough set feature selection and bat optimization for detection of diabetes disease.** Applied Soft Computing, 67, 764-780.

17. Vaishali, R., Sasikala, R., Ramasubbareddy, S., Remya, S., &Nalluri, S. (2017, October). **Genetic algorithm-based feature selection and MOE Fuzzy classification algorithm on Pima Indians Diabetes dataset**. In 2017 International Conference on Computing Networking and Informatics (ICCNI) (pp. 1-5). IEEE.
18. Lukmanto, R. B., Nugroho, A., & Akbar, H. (2019). **Early Detection of Diabetes Mellitus using Feature Selection and Fuzzy Support Vector Machine**. *Procedia Computer Science*, *157*, 46-54.
19. Sejdinović, D., Gurbeta, L., Badnjević, A., Malenica, M., Dujić, T., Čaušević, A., ... &Mehmedović, L. D. (2017). **Classification of prediabetes and type 2 Diabetes using Artificial Neural Network.** In CMBEBIH 2017 (pp. 685-689). Springer, Singapore.
20. Choubey, D. K., & Paul, S. (2017). GA_RBF NN**: a classification system for diabetes. International Journal of Biomedical Engineering and Technology**, 23(1), 71-93.
21. Ndaba, M., Pillay, A. W., &Ezugwu, A. E. (2018, May). **An improved generalized regression neural network for type ii diabetes classification**. In International Conference on Computational Science and Its Applications (pp. 659-671). Springer, Cham.
22. Akyol, K., &Şen, B. (2018). **Diabetes mellitus data classification by cascading of feature selection methods and ensemble learning algorithms.** Int. J. Modern Educ. Comput. Sci, 6, 10-16.
23. Kannadasan, K., Edla, D. R., &Kuppili, V. (2019). **Type 2 diabetes data classification using stacked autoencoders in deep neural networks.** Clinical Epidemiology and Global Health, 7(4), 530-535.
24. Wang, Q., Cao, W., Guo, J., Ren, J., Cheng, Y., & Davis, D. N. (2019). DMP_MI: **an effective diabetes mellitus classification algorithm on imbalanced data with missing values.** IEEE Access, 7, 102232-102238.
25. Choubey, D. K., Paul, S., Bala, K., Kumar, M., & Singh, U. P. (2019**). Implementation of a hybrid classification method for diabetes.** In Intelligent Innovations in Multimedia Data Engineering and Management (pp. 201-240). IGI Global.
26. Erdem, M. B., Erdem, Z., &Rahnamayan, S. (2019, August). **Diabetes Mellitus Prediction Using Multi-objective Genetic Programming and Majority Voting**. In 2019 14th International Conference on Computer Science & Education (ICCSE) (pp. 953-958). IEEE.
27. Prabhu, P., &Selvabharathi, S. (2019, July). **Deep Belief Neural Network Model for Prediction of Diabetes Mellitus.** In 2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC) (pp. 138-142). IEEE.
28. Zheng, T., Xie, W., Xu, L., He, X., Zhang, Y., You, M., ... & Chen, Y. (2017). **A machine learning-based framework to identify type 2 diabetes through electronic health records**. *International journal of medical informatics*, *97*, 120-127.
29. Kayaer, K., & Yildirim, T. (2003, June). **Medical diagnosis on Pima Indian diabetes using general regression neural networks.** In Proceedings of the international conference on artificial neural networks and neural information processing (ICANN/ICONIP) (Vol. 181, p. 184).
30. Mohamadi, H., Habibi, J., Abadeh, M. S., &Saadi, H. (2008). **Data mining with a simulated annealing based fuzzy classification system. pattern recognition**, 41(5), 1824-1833.
31. Amiripalli, S. S., & Bobba, V. (2019). **Trimet graph optimization (TGO) based methodology for scalability and survivability in wireless networks**. International Journal of Advanced Trends in Computer Science and Engineering, 8(6), 3454-3460. doi:10.30534/ijatcse/2019/121862019.
32. Amiripalli, S. S., & Bobba, V. (2019). **An Optimal TGO Topology Method for a Scalable and Survivable Network in IOT Communication Technology**. Wireless Personal Communications, 107(2), 1019-1040.z
33. Amiripalli, S. S., & Bobba, V. (2019). **Impact of trimet graph optimization topology on scalable networks**. Journal of Intelligent & Fuzzy Systems, 36(3), 2431-2442.
34. Amiripalli, S. S V. Bobba, "**A Fibonacci based TGO methodology for survivability in ZigBee topologies".** INTERNATIONAL JOURNAL OF SCIENTIFIC &TECHNOLOGY RESEARCH, 9(2), pp. 878-881. 2020.
35. Amiripalli, S. S., Kumar, A. K., & Tulasi, B. (2016, February**). Introduction to TRIMET along with its properties and scope**. In AIP Conference Proceedings (Vol. 1705, No. 1, p. 020032). AIP Publishing LLC.
36. Amiripalli, S. S., & Bobba, V. (2020). **An Optimal Graph based ZigBee Mesh for Smart Homes.** Journal of Scientific & Industrial Research Vol. 79, April 2020, pp. 318–322

37. Amiripalli, S. S., Kollu, V. V. R., Jaidhan, B. J., Srinivasa Chakravarthi, L., & Raju, V. A. (2020). **Performance improvement model for airlines connectivity system using network science.** International Journal of Advanced Trends in Computer Science and Engineering, 9(1), 789-792. doi:10.30534/ijatcse/2020/113912020
38. Polat, K., &Güneş, S. (2007). ulti. Digital Signal Processing, 17(4), 702-710.
39. Kahramanli, H., &Allahverdi, N. (2008). **Design of a hybrid system for the diabetes and heart diseases**. *Expert systems with applications*, *35*(1-2), 82-89.
40. Temurtas, H., Yumusak, N., &Temurtas, F. (2009). **A comparative study on diabetes disease diagnosis using neural networks.** *Expert Systems with applications*, *36*(4), 8610-8615. https://doi.org/10.1016/j.eswa.2008.10.032
41. Caliskan, A., Yuksel, M. E., Badem, H., &Basturk, A. (2018). **Performance improvement of deep neural network classifiers by a simple training strategy.** *Engineering Applications of Artificial Intelligence*, *67*, 14-23.