Classification and cost benefit Analysis of Diabetes mellitus Dominance

Noman Sohail[†], Ren Jiadong[†], Muhammad Musa Uba[†], Muhammad Irshad[†] and Ayesha Khan[††]

[†]School of Information science and Technology. Yanshan University, Qinhuangdao, Hebei, China
[††]School of Life Sciences. Islamia University. Bahawalpur. Pakistan

Summary
The aim of research work is towards the prevalence of diabetes mellitus to improve classification accuracy and cost/benefit predictions on real-life dataset. This paper aims the real-life diabetic patients from the country 'Nigeria' (an African state) to test the experiment.Our proposed methodology consist of two parts; weka classification for the accuracy measurement ratios by applying 150 machine learning classifiers and compare the results with previous work, secondly' the improvement in clustering with positive and negative instances ratios in terms of initial care predictions by cost benefit analysis.For our Diabetes dataset of 281 instances with 100 attributes, auto-weka has performed 2,440 evaluations with the error rate of 0.014 % by the time limit of 120 minutes. We have achieved the accuracy ratio of 99.64% as compared to 92.6% by "AdaboostM1" classification in mean time of 0.015 seconds to build the model and properly classified instance ratio of 86% by improving k-means.One of the benefits of our proposed model is that it avoids deleting the original data, which ensures the high quality in experiments. And the other benefit is, model can be applied to the other datasets to attain the best accuracy and cost analysis results.

Key words:
Classification; data analysis; Diabetes mellitus; decision making; predictions.

1. Introduction

According to the doctors [1], diabetes is mostly referred as diabetes mellitus; which is spreading vastly around the globe. Around 425 million people had diabetes in the world by the survey report of IDF International Federation in '2015' and this ratio can be expected to the 700 million by 2040.Diabetes is commonly known as the group of metabolic diseases [2], which indicates the high blood pressure and sugar level for the prolonged period. The symptoms can cause other harshly problems, if it left untreated on early stages. The most commonly known complications occur in the body is hyper osmolar, diabetic keto acidosis, hyperglycemic and also death of human body [3].
William's wrote in his book "Williams textbook of Endocrinology" [4] that around 385 million people were affected with diabetes in 2013 and this ratio can go far, if it has not been treated well and can even cause death, which could be so serious.

I. In Type 1 diabetes [5], body refuses to produce insulin. Patients with this type of diabetes need to take the insulin injections every day to keep alive the body cells.
II. In Type 2 diabetes [6], the body does not produce enough amount of insulin for functioning the cells and sometimes body does not take insulin's injection well.
III. Gestational occurs in women's [7], mostly the pregnant females has this case but it disappears' when the baby born.

There are many complications embroils in diabetes [8] such as; heart attack, kidney failure, strokes, vision disability, dental problems, damage of nerves, foot problems, eye damage, frequent urination, increase in thirst, hunger, weight gain, unusual weight loss, fatigues, cuts that do not heal easily, male sexual dysfunctions, tingling in body, numbness, cardiovascular diseases, cholesterol problem in body, hypoglycemia, hyperglycemia, irresistibility, skin infections, sweating, shivering, weakness, dizziness, headache, heartbeat problems, nausea, coma, vomiting, spider cobwebs, heat/cold intolerance, palpitation and tremor and more' which we have used in our dataset attributes. The research survey of NIDDK [9] in 2015' says, 30 million people of US have diabetes', which calculated as the 9.5% of the total population. From 1 in 4 of them does not know that they have disease. Diabetes affects 1 in 4 people's over the age of 60's and about 95% cases been noticed of Type 2 in adults. Medical specialists can determine diabetes by three tests;A1C [10], FPG [11] and OGTT [12].
The objective of this research is to mine the Diabetes mellitus on platform of data mining by using machine-learning algorithms to train the machine and by testing set using auto-weka classification method. For model our work, we have used the real-life diabetes patient's data.
The remaining part of paper is organized into 6 sections. Section 2 discusses the related work. Section 3 model the real-life data collection process and explains the methodology adopted to perform the experiment. Section

Manuscript received October 5, 2018
Manuscript revised October 20, 2018

4 examines the driven results. Section 5 carries discussion of proposed results. Finally conclusion ends up the paper on basis of outcomes.

2. Related work

In recent years, data mining techniques are widely using by researchers with increasing frequency to predict different diseases [13]. By the study of different research papers, we came to know that this field has been supported by different mysteries and models from decades like machine learning [14], artificial intelligence [15], statistics [16], probabilities [17] and predictive/descriptive models [18] based on supervised and unsupervised learning. In healthcare predictive models are more common than descriptive because the datasets are normally based on supervised learning.Mostly datasets used by different researchers were taken from the 'Pima Indians Diabetes Dataset' [19] from the UCI machine-learning database. For frequent pattern mining of data,' the most commonly used algorithms are Apriori and Fp Growth [20].The risk factors of T2DM has examined in [21] by approaching both above algorithms and proposed the receiver operating characteristics to verify the results of experiments. Prevention of T2DM [22] should be straight from individuals and developed a model for diabetes on android-based application. Diabetes type of Gestational on pregnant women's has been studied in [23] and discovered data mining techniques for baby's birth outcome. A hybrid prediction model has introduced in [24] by using k-means clustering and C4.5 for chosen class label of data and aimed to built a classifier model with 93% accuracy of classification. MLP Multilayer perceptions with neural networks to compare the accuracy against J48 and ID3 algorithms with the results of 90% compared to 81.8% [25]. Prediction accuracy of 89.93% [26] on diabetes dataset by applying Artificial Meta plasticity on multilayer perceptron. The preprocessing and parameter techniques could produce the meaning full results with better predictions and accuracies [27]. An Indian researcher [28], has developed an android-based solution application to reduce down the lack of diabetes mellitus knowledge [29]. Two brother's [30] improved the k-means classifier [31] on initial clusters by using noise data filters [32].

However the prediction and accuracies were not good enough for the real life patient's of diabetes but still the researchers are trying best to develop new models and techniques to classify the data [13] but still, a room is there always for improvements. By the contribution of all authors, we have gathered the real-life data from "Nigeria" with utmost 300 patients from interval 2016 to 2018. We have used dataset to compare the classification accuracies with the previous task and analyze the cost benefit ratio of data for initial care.

3. Methods

Our proposed methodology consist of three parts; weka classification [33], the improvement in clustering and finally the cost benefit analysis for the initial care.

3.1 Real-life data

Questionnaire was designed by consulting different medical specialists to collect the data from patients, which has been asked by different diabetes patients since '2016' to '2018' in Nigeria 'an African state struggling for the life necessities [34]'. In busy life, people forget to take care of health in proper manners like daily life exercise to keep body fit, healthy diet and more. The average of people even don't know about diseases and struggling for the life necessitates [28].

3.2 Data preprocessing

Preprocessing method improves the prediction results [35]. In other words, it plays a major role in the model. Weka tool contains many filters and attributes for preprocess of data to train the machine [33]. In our model, we have used some methods to optimize the dataset by analyzing each attribute and its relation [4] by changing the numeric attribute to nominal, where value 0 represents as "No" and 1 as "Yes". Hence the complexity of data has reduced. Secondly, the missing values have removed. If we observe the past research from related work, for experiments and better accuracies, the values has been changed by researchers like wise for blood pressure and Body mass index' values cannot be 0, but the previous research indicates it. But we have used the real values for our attributes.

After the above statements, dataset has changed the values into 0 and1 by (Eq. 1), where x' is the average value and "s"is standard deviation.

$$Values = \frac{value - x'}{s} \tag{1}$$

3.3 Data mining platform

Data mining platform called 'weka'has a classifier method 'auto-weka' that performs the selection of combined algorithms and hyper parameter optimization over the regression and classification algorithms implementation to find the better accuracy. Auto-weka explores hyper parameters of settings on defined dataset and gives the recommended method of classification with

the high accuracy by evaluating all classifiers and algorithms in desired time period and helps user by giving good generalization performance by saving effort in applying of different algorithms one by one. It has been available on weka package manager since '2017'as a classification algorithm also but not much worked carried out by using this strategy on diabetes datasets. It has two ways of performance; GUI graphic user interface weka panel and the second by choosing it from classifier list in Classify panel. Auto-weka performs statistically rigorous evaluation internally by 10 fold cross-validation. Auto-weka basically has few options, which normally leaves as default but for better accuracy and results two options are really important; one is 'Time limit' and second is 'Memory limit'. By default the time limit is 15 minutes but big data needs more time to classify in better way, and for our data, we have choose the time limit as 180 minutes as 3 hours; which is better enough to classify the dataset of 100 attributes with 281 instances. On other side the memory limit is in megabytes, which is increased accordingly to the size of data and large dataset requires more memory. Auto-weka provides the status of numbers of evaluated configurations and the estimated error of the best configuration in the status bar. On our dataset classification, auto-weka has performed 2440 evaluations with the error rate of 0.014 % in 180 minutes as mentioned in experimental results.Which shows the best classifier is "AdaboostM1" with classification accuracy of 99.64% as compared to 92.6% as mentioned above on artificial dataset. We have discussed it more in proposed results 'section-4'.

3.4 K-mean clustering

Clustering aims to partitioning the observations into distance clusters in means of observation within the same cluster [28].The process steps of k-means are measure as the calculation distance between each object and cluster centers. It cluster every object to the nearest cluster by the distance accordingly by (Eq.2) [28] and it recalculate the centers to verify the changes by (Eq.3) [36].The proposed analysis results of diabetes mellitus clusters with its subsets Type 1, Type 2 and Gestational, "which we have labeled in our dataset as NID non-insulin dependent, IND insulin dependent and GTD Gestational diabetes" are presented in result section.

$$s_i^t = \{\forall_j, 1A_j A_k X_p :\| X_p - m_i^t \|^2 \le \| X_p - m_j^t \|^2 \forall_j, 1 \le j \le k\} \forall_j, 1A_j A_k \quad (2)$$

$$m_i^{t+1} = \frac{1}{|s_i^t|} \sum x_j \in s_i^t X_j \quad (3)$$

The values of clusters are 0 and 1 indicating negative and positive class. Method for clustering in weka [33] depends on the seeds input value, which is generated randomly in initial stage, but for the better improvement user have to set the cluster seeds value according to the experience, which effects directly to the results of cluster analysis. In our model experiment, we have calculated the mean rate by removing incorrectly clusters using the formula as mentioned in (Eq.4) and used the proximate rate and corresponding seed for every level and obtained 80% correctly classified patients, which we used as input for the logistic regression model.

$$Rate = \frac{\mathrm{Re}\, mainingdata}{Sum} \quad (4)$$

4. Proposed results

The graphical visualization is good and easy for understanding and weka helps us to visualize the proposed model results graphically. This section will show the entire proposed model results based on graphical visualization.

4.1 Cross validation

Cross validation helps to improve the model results. The 10-fold cross validation technique has been used for better predictions. We have divided our dataset in to 10 samples. Each sample had to go from the process of retained as a validation data, where the rest 9 samples acted as a training data. This process was vice versa for 10 times, that's why it is call 10-fold cross validation. The advantage gained by this process step is that it cut down the bias association with random sampling methods.

4.2 Kappa stats

Kappa statistics is also known as inter rater reliability to test frequently, the importance lies in the facts that it presents the extended facts about the collection of data in the study are correct for variable measurements. It compares the proposed model results with the randomly generated classified methods. While there have been introduced various methods to measure it but he method we have chosen in our proposed model is based on the kappa stats values between 0 and 1, where the values 0 presents as invalid and 1 as the expected effects of the proposed model. In our model, the kappa result is 0.9916, which indicates that the proposed model has good

consistency. The equation works for the kappa statistics in our model is mentioned in (Eq.5-7).

$$K = \frac{P_a - P_e}{1 - P_e} \tag{5}$$

$$P_a = \frac{T_p + T_n}{N} \tag{6}$$

$$P_e = \frac{(T_P + F_N) * (T_P + F_P) * (T_N + F_N)}{N^2} \tag{7}$$

4.3 Classification accuracy

Diabetes mellitus dataset of 281 instances with 100 attributes, auto-weka has performed 2,440 evaluations with the error ratio of 0.014 by time limit of 120 minutes. The experimental results are declared in (Table.1) along the clustering instance ratio and confusion instances by improved k-means. Which shows the best classifier by "auto-weka" on weka (version 3.9.2) is 'adaboostM1' (out of utmost 150 installed classifiers) with classification accuracy of 99.64% as compared to 92.6% (from previous related work) and means root error is 0.05% along estimated error rate of 0.0%. After the successful evaluation of 120 minutes, auto-weka took 0.015 seconds to build the model.

Table 1: Stipulated into two parts; Proposed classification results of auto-weka by "AdaBoostM1" classifier, showing the arguments, error rates, accuracy ratio by class and other parameter details. Secondly the detail of improved k-means clustered instance result with tested_positive and tested_negative ratios on Diabetes type wise with absolute mean percentage along cluster confusion instances

Precision	**Recall**	**F-measure**	**MCC**	**ROC**	**PRC**
0.996	0.996	0.996	0.992	1.000	1.000
Accuracy of classified instances			280		99.64%
True and False positive rates			0.996		0.008
Kappa stats analysis ratio			0.9916		
Mean absolute error rate			0.005%		

Value	**Count**	**Ratio**	**Clusters by class & Diabetes type**				
			TN	**TP**	**NID**	**GTD**	**IND**
0	138	49	47	91	128	7	3
1	143	51	40	103	128	7	8

Confusion matrix	**A**	**B**
Predicted positive	86	1
Predicted negative	0	194

The cost benefit principles span the area of diabetes from the patients to analyze the record statistics of insulin level or with other treatment terminologies. In other words, analysis finds the quantifies and adds all the positive factors regarding benefits than it identifies, quantifies and subtracts all the negative ratios, which is called "Cost". The difference between these two terms are whether the planned action is advisable or not for the patients. Cost benefit analysis can helps the medical specialists to analyze the patient's ratio with their medical stats and stages. It works with the predictions of tested positive and negative ratios. In our proposed model, it has a good prediction percentage as shown in (Fig.1), where prediction for the positive analysis is 68.68% and negative analysis is 30.96% with the total Gain of 119.51% and 0.02% with threshold.

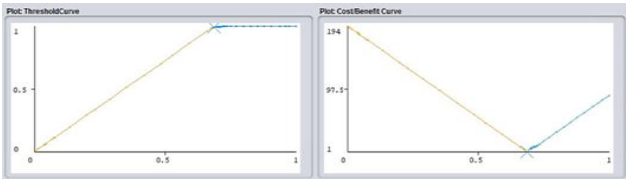

Fig. 1 Proposed the Cost/Benefit curve visualization with threshold percentage sample size by improved clusters vary from 0 to 1 and mean gain from clusters from 1 to 194 improved instances on Y-axis

5. Discussion

In the above study, our proposed model reveled the best accuracy results based on precision, kappa statistics, recall, MCC with ROC area and cost analysis ratio'as discussed above grounded on the real-life diabetes dataset from 'Nigeria', where the basic needs of life is challenge and population is migrating towards the other regions of globe. The basic needs compulsory for life like food, water, shelter, proper sleep, safety, clothing and many more. In 'Nigeria' people are rushing to achieve the life goals to live properly, where they forget to take proper care of health' which causes so many cases and the major one comes in present time is Diabetes mellitus. In this section, we will compare our proposed method results with the work of other researchers in prediction of patient's initial care.

As we discussed the researchers work in the 'related work' section and presented their work with the accuracy ratios of their experiment models, which was based on the artificial datasets and the accuracies were been improved by changing the values according to their experiences. With their work, we have compared our proposed model of results with the accuracy ratio of up to 99% by improving the clustering instance and results are shown in (Table.2), which is based on the real-life diabetes dataset without changing the values.

Table 2: Recap of researchers classification accuracy compared with our proposed model with references

Classification methods	Accuracy	Reference
Proposed Model	99.6441	This paper
MLP	73.8	[25]
Discrim	77.5	[37]
Logdisc	78.2	[38]
KNN	94.2	[24]
Logistic	85.35	[39]
BayesNet	74.7	[40]
NaiveBayes	76.3	[38]
Random Forest	76.6	[39]
LogitBoost	93.93	[39]
J48	98.1	[41]
SGD	76.6	[39]
SMO	77.26	[40]
ANN	89.84	[40]
RBF	75.7	[38]
FCM	94.7	[41]
AdaBoost.M1	92.6	[39]

6. Conclusion and future work

In conclusion, this paper has aimed to establish a prediction model for diabetes mellitus prevalence. We have compared the classification analysis with previous work and proposed the cost benefit ratio for initial care circumstances, the classification has approached with auto-weka in comparison of utmost 150 classifiers [33] and the improvement in clustering by k-means algorithm to have fruitful results for cost benefit analysis.One of the benefits of our proposed model is that it avoids deleting the original data, which ensures the high quality in experiment. And the other benefit is, our model can apply to the artificial data from different UCI databases [19]. The main problem was solved, which was to improve the accuracy of prediction analysis and making the model to be adapted by different datasets.

As the researchers are working more on improvement of weka and more plugins are coming on the way for classifiers. Our future work has aimed to maintain and improve our model by adopting new plugins for our proposed model on Meta data.

Acknowledgment

This research was funded by the scientific research projects of the "Yanshan University, Qinhuangdao, Hebei, China".

Author's contribution

Noman Sohail has performed the experiments, analyzed the results, and wrote the main manuscript with the supervision of Professor Ren Jiadong. Musa Uba and Muhammad Irshad have provided the technical support and data collection. Ayesha Khan has revised the main manuscript and contribute in discussion analysis.

Conflict of interest

The authors state that they have no conflict of interest to declare.

Ethical approval

All study procedures were performed according to the 1964 Helsinki declaration and ethical approvals was obtained from the Yanshan University ethical board.

References

[1] Whiting DR, Guariguata L, Weil C, Shaw J. IDF Diabetes Atlas: Global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes Res Clin Pract. 2011 Dec;94(3):311–21.

[2] American Diabetes Association AD. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2014 Jan;37 Suppl 1(Supplement 1):S81-90.

[3] Wolfsdorf JI, Glaser N, Agus M, Fritsch M, Hanas R, Rewers A, et al. Diabetic Ketoacidosis and Hyperglycemic Hyperosmolar State: A Consensus Statement from the International Society for Pediatric and Adolescent Diabetes. Pediatr Diabetes [Internet]. 2018 Jun;0(ja). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/pedi.12701.

[4] Melmed S, Polonsky KS, Larsen PR, Kronenberg H. Williams textbook of endocrinology. 13th ed. F J, editor. Elsevier; 2016. 722-1799 p.

[5] Australia H. Type 1 diabetes. 2018;

[6] Mostafa SA, Coleman RL, Agbaje OF, Gray AM, Holman RR, Bethel MA. Modelling incremental benefits on complications rates when targeting lower HbA 1c levels in people with Type 2 diabetes and cardiovascular disease. Diabet Med. 2018 Jan;35(1):72–7.

[7] Jones KE, Yan Y, Colditz GA, Herrick CJ. Prenatal counseling on type 2 diabetes risk, exercise, and nutrition affects the likelihood of postpartum diabetes screening after gestational diabetes. J Perinatol. 2018 Apr;38(4):315–23.

[8] Abdel-Moneim A, Bakery HH, Allam G. The potential pathogenic role of IL-17/Th17 cells in both type 1 and type 2 diabetes mellitus. Biomed Pharmacother. 2018 May;101:287–92.

[9] For Disease Control C. National Diabetes Statistics Report, 2017 Estimates of Diabetes and Its Burden in the United States Background. 2017.

[10] M.N. Sohail et al. Forecast Regression analysis for Diabetes Growth: An inclusive data mining approach. Int. J. Adv. Res. Comput. Eng. Technol.7, 715–721 (2018).

[11] Irshad Muhammad et al. An Indoor LOS & NON-LOS Propagation analysis using Visible Light Communication. Int. J. Adv. Res. Comput. Eng. Technol.7, 727–732 (2018).

[12] Musavir Bilal, B. Weihong, M.N. Sohail, M. Irshad, S.M. Zaheer, M. M. U. Magnetic fluid filled to detect the magnetic field based on Fabre-Perot cavity tube. Int. J. Adv. Res. Comput. Eng. Technol. 7, 722–725 (2018).

[13] Sohail MN, Jiadong R, Uba MM, Irshad M. A Comprehensive Looks at Data Mining Techniques Contributing to Medical Data Growth: A Survey of Researcher Reviews. In Springer, Singapore; 2019 [cited

2018 Aug 25]. p. 21–6. Available from: http://link.springer.com/10.1007/978-981-10-8944-2_3
[14] Irshad M, Liu W, Wang L, Shah SBH, Sohail MN, Uba MM. Li-local: green communication modulations for indoor localization. In: Proceedings of the 2nd International Conference on Future Networks and Distributed Systems - ICFNDS '18 [Internet]. New York, New York, USA: ACM Press; 2018. p. 1–6. Available from: http://dl.acm.org/citation.cfm?doid=3231053.3231118
[15] Parapuram GK, Mokhtari M, Hmida J Ben. Prediction and Analysis of Geomechanical Properties of the Upper Bakken Shale Utilizing Artificial Intelligence and Data Mining. Prediction and Analysis of Geomechanical Properties of the Upper Bakken Shale Utilizing Artificial Intelligence and Data Minin. 2017;
[16] Muhammad MU, Asiribo ; O E, Sohail ; M N. Application of Logistic Regression Modeling Using Fractional Polynomials of Grouped Continuous Covariates. In: Edited Proceedings of 1st International Conference [Internet]. 2017. p. 144–7. Available from: http://nss.com.ng/sites/default/files/P144-147_NSS-CP-2017-033.pdf
[17] Kruse C. The New Possibilities from "Big Data" to Overlooked Associations Between Diabetes, Biochemical Parameters, Glucose Control, and Osteoporosis. Curr Osteoporos Rep. 2018 Jun;16(3):320–4.
[18] Lashari SA, Ibrahim R, Senan N, Taujuddin NSAM. Application of Data Mining Techniques for Medical Data Classification: A Review; Application of Data Mining Techniques for Medical Data Classification: A Review.
[19] Pima Indians Diabetes Database | Kaggle.
[20] Mittal S, Mirza S, mittal S, zaman M. DESIGN AND IMPLEMENTATION OF PREDICTIVE MODEL FOR PROGNOSIS OF DIABETES USING DATA MINING TECHNIQUES. Int J Adv Res Comput Sci. 9(2).
[21] Bennett CM, Guo M, Dharmage SC. HbA 1c as a screening tool for detection of Type 2 diabetes: a systematic review. Diabet Med. 2007 Apr;24(4):333–43.
[22] Chao C-Y, Wu J-S, Yang Y-C, Shih C-C, Wang R-H, Lu F-H, et al. Sleep duration is a potential risk factor for newly diagnosed type 2 diabetes mellitus. Metabolism. 2011 Jun;60(6):799–804.
[23] Godwin M, Muirhead M, Hyunh J, Helt B, Grimmer J. Prevalence of gestational diabetes mellitus among Swampy Cree women in Moose Factory, James Bay. CMAJ. 1999;160(9).
[24] Patil BM, Joshi RC, Toshniwal D. Hybrid prediction model for Type-2 diabetic patients. Expert Syst Appl. 2010 Dec;37(12):8102–8.
[25] Ahmed M, Afzal H, Majeed A, Khan B. A Survey of Evolution in Predictive Models and Impacting Factors in Customer Churn. Adv Data Sci Adapt Anal. 2017 Jul;09(03):1750007.
[26] Marcano-Cedeño A, Torres J, Andina D. A Prediction Model to Diabetes Using Artificial Metaplasticity. In Springer, Berlin, Heidelberg; 2011. p. 418–25.
[27] Romero P, Obradovic Z, Dunker AK. Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. Genome Informatics. 1997;8:110–24.
[28] Fallah M, Niakan Kalhori SR. Systematic Review of Data Mining Applications in Patient-Centered Mobile-Based Information Systems. Healthc Inform Res. 2017 Oct;23(4):262.
[29] Hospital in Chhattisgarh - Best Hospital in Raipur Chhattisgarh, India | Fortis Healthcare.
[30] Joshi K. Survey on Different Enhanced K-Means Clustering Algorithm. Int J Eng Trends Technol. 2015;27(4).
[31] Wu Y, Ding Y, Tanaka Y, Zhang W. Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. Int J Med Sci. 2014;11(11):1185–200.
[32] Chandrakar O, Saini JR. Development of Indian Weighted Diabetic Risk Score (IWDRS) using Machine Learning Techniques for Type-2 Diabetes. In: Proceedings of the Ninth Annual ACM India Conference on - ACM COMPUTE '16. New York, New York, USA: ACM Press; 2016. p. 125–8.
[33] Witten. Weka - Data Mining with Open Source Machine Learning Software in Java [Internet]. weka. 2016. Available from: https://www.cs.waikato.ac.nz/ml/weka/
[34] Stewart F. Country experience in providing for basic needs. Finance Dev. 1979 Dec;16(4):23–6.
[35] Mirza S, Mittal S, Zaman M. Decision Support Predictive model for prognosis of diabetes using SMOTE and Decision tree. Vol. 13, International Journal of Applied Engineering Research. 2018.
[36] Iyer A, Jeyalatha S, Sumbaly R. Diagnosis of diabetes using classification mining techniques. 2015 Feb;5(1):1–14.
[37] Xiong X, Qiao S, Li Y, Zhang H, Huang P, Han N, et al. ADPDF: A Hybrid Attribute Discrimination Method for Psychometric Data With Fuzziness. IEEE Trans Syst Man, Cybern Syst. 2018;1–14.
[38] Lakshmi K, Ahmed DI, Siva Kumar G. IJSRSET32184139 | A Smart Clinical Decision Support System to Predict diabetes Disease Using Classification Techniques. 2018 IJSRSET |. 2018;4.
[39] Chen P, Pan C. Diabetes classification model based on boosting algorithms. BMC Bioinformatics. 2018 Dec;19(1):109.
[40] Shah A, Lynch S, Niemeijer M, Amelon R, Clarida W, Folk J, et al. Susceptibility to misdiagnosis of adversarial images by deep learning based retinal image analysis algorithms. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE; 2018. p. 1454–7.
[41] Bhuvaneswari G, Manikandan G. A novel machine learning framework for diagnosing the type 2 diabetics using temporal fuzzy ant miner decision tree classifier with temporal weighted genetic algorithm. Computing. 2018 Aug;100(8):759–72

Noman Sohail is holding a B.Sc degrees (with Honors) in "Computer Networks" and M.Sc in Technology Management from "University of East London, UK" and currently pursuing PhD with project involves "Data Mining & Analysis in Health care and Fiber Sensor Technology".

Ren Jiadong holds the Bachelor and Masters degree from "Harbin Institute of Technology, China". Currently he's working as a full time professor with projects involved "Data mining, Data analysis and Networks security".

Musa Uba holds the Bachelor and Masters degree in Statistics from "Kano university of Sciences & Technology, Nigeria "and" Ahmadu Bello University Zaria, Nigeria". Currently he's enrolled in PhD with "Information Technology projects".

Muhammad Irshad holds M.Sc degree from "University of Punjab, Lahore, Pakistan". Currently he's doing PhD with project involves "Wireless Communication Networks and data analysis".

Ayesha Khan has received her B.Sc and M.Sc in Zoology. Currently she is pursuing her Ph.D in Zoology from Islamia University Bahawalpur under the School of Life Sciences.