International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Existential Methods on Diabetes Detection using Machine Learning

Vaishali Yogesh Baviskar

Abstract: ***Nowadays, a lot of research is going on in healthcare. One of the significant diseases increased all over the world is Diabetes Mellitus (DM). In this paper, the literature review is done on diabetes prediction using Machine Learning and Deep Learning techniques. Various ML algorithms are used using PIDD (Pima Indian diabetes dataset), and improved k- means using logistic regression among all algorithms achieved the highest accuracy. DL algorithms like CNN and LMST used in diabetic retinopathy images.***

Keywords: SVM, NN, Naïve Bayes, KNN, Diabetes

I. INTRODUCTION

Recently diabetes is the major cause of death for all humans. In 2000, 171 million people were predicted, which can increase by 2040 up to 642 million all over the world. This increase in figure needs to pay attention to this disease. Many healthcare institutions across the globe spend billions of dollars on diabetes healthcare. Diabetes patients are categorized into four types as Type1 diabetic, pre-diabetic, Type 2 diabetic, and Gestational. Type 1 occurred due to a lack of insulin in youngsters and grownups. Pre-diabetic is the phase before Type2 and Gestational diabetes occurs in ladies during pregnancy. The diagnosis levels of all these patients can be done on various blood glucose sugar level tests. A1C means higher glucose levels test is done to detect Type1 and pre-diabetes diagnosis. Fasting glucose test is done to detect Type1, Pre diabetes and Type2 diagnosis. OTG- Oral glucose test is done to diagnose pre diabetes, Type2 and gestational disease. High level of glucose can affect on human health and leads to severe conditions like loss of vision, Kidney Neuropathy, Liver problems, Heart problems, and foot issues. Due to high sugar levels, diabetes retinopathy is required to diagnose, which can further cause for vision loss and night blindness.

II. RELATED WORK

A. Machine Learning Algorithms

Authors have shown an analysis by using a Decision tree, K-nearest neighbor, random forest, and support vector machine classifiers. Before preprocessing J48 showed the highest efficiency while after preprocessing, KNN and random forest showed the highest accuracy. An analysis is done on PIDD before and after preprocessing [1].

In [2], the authors proposed a framework to predict disease using machine learning and deep learning techniques on the PIDD dataset. Artificial Neural network (ANN) has got the highest accuracy as deep learning technique, and the Random Forest technique has got the highest precision in machine learning techniques.

In [3], the authors compared multiple prediction models using health checkups, and insurance claims data. Yearly health checks up and health insurance dataset from japan is used. XGBoost algorithm is used to predict Type2 diabetes and has got the highest accuracy.

In [4], the authors discussed diabetic research on 1) prediction and diagnosis, 2) Diabetic complications, 3) Genetic background, and environment and 4) Health care and management. Various methodologies are used as feature extraction and reduction using LDA (Linear Discriminant analysis) - MWSVM (Morlet Wavelet Support vector machine) for diabetes diagnosis, Ant colony classification used set of fuzzy rules to extract features, multivariate regression using support vector regression, fuzzy ontology-based case reasoning, multilayer classification and rotation forest on various datasets like clinical and biological datasets, gut microbiota, Electronic measurements of saliva, demographic, Anthropometric, diagnostic and clinical laboratory measurements and it is observed that SVM has got the highest accuracy among all classifiers. They have also discussed on macrovascular and microvascular diabetic complications, for these researchers used temporal data mining and machine learning algorithms for risk stratification.

In [5], the authors derived a set of predictive models of type2 diabetes complications based on electronic health record, and model validation is done. To deal with missing values and class imbalance in RF, Stepwise feature selection is made with the logistic regression. Various Classification models are used like Logistic regression, Naïve Bayes, Support vector machine and random forest. A risk score is scored based on the temporal threshold, complications, and onset date registered. Clinical Historical dataset for more than ten years has taken from the hospital of Pavia, Italy is considered. The final model taken has got an accuracy of 83%.

Methods used are center, profiling, predictive models training, predictive models construction and predictive models validation. In center profiling optimize features are selected to do an initial analysis.

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
Mrs. Vaishali Y. Baviskar, Assistant Professor, G. H. Raisoni Institite of Enginerring and Technology, Wagholi, Pune, India.

In predictive models training, focused on microvascular complications like nephropathy, neuropathy and retinopathy. In predictive models construction, after the first visit, the patient is predicted whether he will develop microvascular complications or not ?, got the best results for retinopathy and neuropathy cases.

In [6], the authors predicted incident of diabetes using medical records of cardiorespiratory fitness. Methodologies like Data preprocessing, features selection, multiple linear regression,

information gain ranking is done using a Decision tree, Naïve Bayes, Logistic regression and random forest on the Henry ford fit dataset (Patients who underwent treadmill stress). To handle imbalanced datasets, the Synthetic minority over-sampling technique (SMOTE) was used. Combined three classifiers i.e. RF, NB, and LMT, and has got an accuracy of 92% achieved higher accuracy of 3.04% compared to other researcher's prediction model. Efficiently predicted cardiorespiratory fitness data using ensemble machine learning and SMOTE methods.

In [7], authors did an enhancement in prediction model and have got an accuracy of 95% on PIDD and other two datasets as Donated by Dr. Schorling from the Department of Medicine of the University of Virginia School of Medicine and collected an online questionnaire. Improved K – means algorithm is used to remove incorrectly clustered data and to get an optimized dataset where preprocessing is done, and Logistic Regression is used to classify remaining data i.e., whether a person has diabetes or not. Data mining toolkit is used where preprocessing, classifying, ranking algorithms, and the visual interface is done. 10 fold cross -validation is used so that it reduces the bias associated with the random sampling method. The model is evaluated by the confusion matrix. The Mathews correlation coefficient (MCC) is used are the measure of binary classification. Kappa statistics is used to test the consistency of the model.

In[8], Dimensionality is reduced using Principal Component Analysis and minimum redundancy maximum relevance(mRMR) to avoid redundant features Decision Tree – C4.5, Random forest, and the Neural Network used as classifiers, 5 fold and 10- fold cross validation method is used on PIDD and physical examination clinical dataset received in Luzhou, China. The Random forest has got the highest accuracy of 80%. RF is a multifunctional machine learning method and plays a significant role in the ensemble machine learning method. Compare to PCA, mRMR has got the best efficiency.

In [9], the authors developed a deep learning model for retinopathy detection. Various color retina images dataset was taken from the kaggle website. Feature extraction is done using the Convolutional Neural network, and SVM, and KNN classifiers are used for detection. The proposed CNN model achieved good results in discovery. The proposed model used regression activation mapping (RAM) to get more accurate results.

In [10], the authors designed a prediction model. They selected significant attributes by giving sequences to each attribute and used classifiers like J48, Random Forest, Naïve Bayes, MLP, KNN, and Neural network. PIDD dataset is used and based on the best attribute selection, the result of classification techniques are improved.

Naïve Bayes has shown the average accuracy which is the highest among all of 82.30%. Features are mapped effectively from low to high dimensions.

B. Datasets, Evaluation matrix and features

The evaluation metrics used for Diabetes detection in table 1. In this table, various datasets used for diabetes detection using machine learning techniques and various preprocessing algorithms are given. Most of the predictions are done on PIDD dataset which is publicly available. Also, accuracy comparison is done for various machine learning algorithms.

In table 2, Classifiers used are mentioned with accuracy. Here, various datasets used for classifiers are given. Out of the datasets taken, some are publicly available on kaggle and some datasets taken are from donated by hospitals, which were authenticated and approved by senior doctors.

Table- I: Comparison of Dataset, Algorithms and Metrics Used

Authors	Algorithms Used	Dataset used	Details
J.Pradeep Kandhasamy, S. Balamurali[1]	J48 Decision tree, K-nearest neighbor, random forest and support vector machine	PIDD	768 patients with eight attributes-number of times pregnant, glucose level, diastolic blood pressure, triceps skinfold thickness, serum insulin, BMI, diabetes pedigree function, age and class
Neha Sharma, Ashima Singh[4]	Artificial Neural network (ANN)	PIDD	768 patients with 8 attributes-number of times pregnant, glucose level, diastolic blood pressure, triceps skin fold thickness, serum insulin, BMI, diabetes pedigree function, age and class
Masatoshi Nagata, Koichi Takai et al.[9]	XGBoost algorithm, LSTM algorithm based on RNN and LILR,	Yearly health checkup and health insurance dataset from japan	Record of 40,000 people aged 20 to 64 years. – profile information(age, sex), Lab test results (e.g., body mass index, blood pressure, HbA1c), and a health questionnaire (e.g., smoking, alcohol intake, exercise level). overall 33 health checkup items
Arianna Dagliati, Simone Marini, Lucia Sacchi et al.[35]	Logistic regression, Naïve Bayes, Support vector machine and random forest	Clinical Historical dataset for more than 10 years has taken from hospital of Pavia, Italy is taken	943 records – Demographic(age, gender, time to diagnosis), clinical data(BMI, Hbalc, lipid profile, smoking habit),administrative data(antihypertensive therapy)

Manal Alghamdi, Mouaz AI-Mallah et al. [36]	Decision tree, Naïve Bayes, Logistic regression and random forest	Henry ford fit dataset (Patients who underwent tread mill stress)	32,555 patients record with 26 attributes - age, resulting heart rate, metabolic equivalent, resting systolic blood pressure, resting diastolic blood pressure, Sedentary lifestyle, black, obesity, hypertension, %HR achieved, Hypertidimedia, Aspirin, family history of premature coronary rtery disease, coronary artery disease, nitrate use, diuretic use, beta blocker use, sex, smoking, Plavix use, angiotensin, angiotensin receptors blockers use, other hypertension medication use, prior cerebrovascular accident, congestive heart failure, calcium channel blocker
Han Wu, Shengqi Yang et al.[37]	Improved K – means algorithm, Logistic Regression	1)PIDD 2) Donated by Dr. Schorling from the Department of Medicine of the University of Virginia School of Medicine 3) collected from online questionnair e	1)768 patients with 8 attributes-number of times pregnant, glucose level, diastolic blood pressure, triceps skin fold thickness, serum insulin, BMI, diabetes pedigree function, age and class 2)1046 records with 19 attributes- Total cholesterol, stabilized glucose, high density lipo protein, (HDL),Cholesterol HDL ratio, glycosylated hemoglobin, age, gender, height, weight, systolic blood pressure, diastolic blood pressure, waist-hip ratio 3) 384 instances collected with 14 attributes – age, gender, pregnant, family doctor, BMI, sleep time, sleep quality, snoring, diuresis, hunger, smoking and drinking, blood pressure, blood glucose, OGTT
Quan Zu, Kaiyang ku et al., 2018[38]	Decision Tree – C4.5, Random forest and Neural Network	1) PIDD 2) physical examination clinical dataset received Luzhou, China.	1) 768 patients with 8 attributes-number of times pregnant, glucose level, diastolic blood pressure, triceps skin fold thickness, serum insulin, BMI, diabetes pedigree function, age and class 2)68994 healthy people and diabetes patients data with 14 attributes- age, pulse rate, breathe, Left systolic pressure(LSP), right systolic pressure(RSP), left diastolic pressure(LDP), right diastolic pressure(RDP), height, weight, physique index, fasting glucose, waistline, low density lipoprotein(LDL),high density lipoprotein(HDL)
Zhiguang Wang, Jianbo Yang[39]	Convolutiona l Neural network and SVM and KNN classifiers are used for detection	Various color retina images	dataset from kaggle website. 35,126 images
N. Sneha, Tarun Gangil[40]	like J48, Random forest, Naïve Bayes, MLP, KNN and Neural network	PIDD	768 data items with 15 attributes-age, gender, plasma glucose fasting, plasma glucose post prandial, pregnancy, blood glucose level, blood pressure, skin thickness, insulin, BMI (body mass index), DPF , serum creatinine, serum sodium, serum potassium and HBAIC

The features extracted with the method used is mentioned in Table II.

Table- II: Comparative Analysis of Various Datasets, Accuracy and Classifiers Used

Classifier	Dataset	Accuracy	Reference
Gradient Boosting Machine	CPSSN (Canadian patients dataset)	84.7	Hang Lai, Huaxiong Huang et al.[2], 2019
Logistic Regression	CPSSN (Canadian patients dataset)	84.0	Hang Lai, Huaxiong Huang et al.[2], 2019
Neural network with 10- fold cross validation	PIDD	85.24	Raghavendra S, Santosh Kumar J, Raghavendra B. K.[3], 2019

RF	Luzhou	78.52	Quan Zu, Kaiyang ku et al.[38], 2018
J48	Luzhou	78.06	Quan Zu, Kaiyang ku et al.[38], 2018
RF	PIDD	76.04	Quan Zu, Kaiyang ku et al.[38], 2018
J48	PIDD	72.75	Quan Zu, Kaiyang ku et al.[38],2018
Naïve Bayes	PIDD	76.30	Dipti Sisodia, Dileep Sisodia [5], 2018
SVM	PIDD	65.10	Dipti Sisodia, Dileep Sisodia[5],2018
Decision Tree	PIDD	73.82	Dipti Sisodia, Deileep Sisodia[5], 2018
Support vector machine	PIDD	98%	Gandhi [46],2014
Artificial Neural Network	Tabriz, Iran	97.44	Heydari [47], 2013
Random Forest	Iris Image	89.66	Samant[48], 2017
Fuzzy Logic	UCI	78.00	Ephizbah [49], 2011

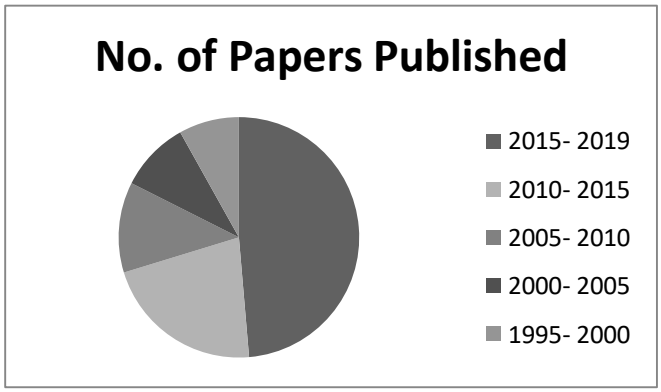

Figure 1. No. of Papers Published Yearwise

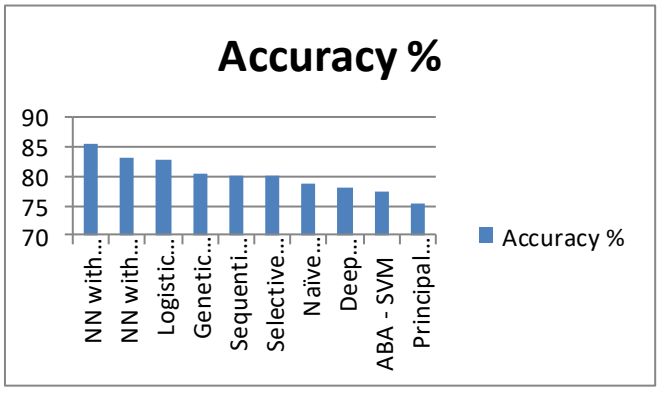

Figure 2 . Accuracy Comparison of Machine Learning Techniques on PIDD Dataset

III. PROPOSED SYSTEM DESIGN

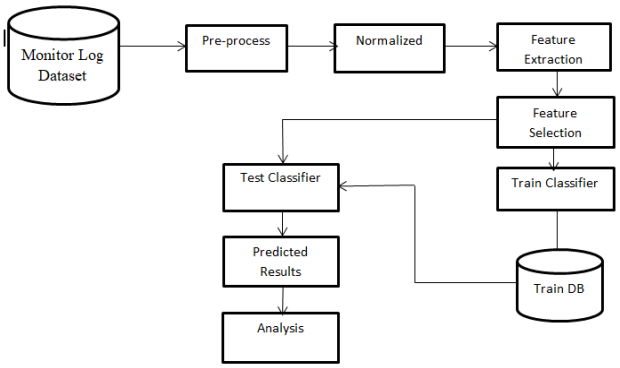

Figure 3 : Proposed system architecture

IV. ALGORITHMS

Q- Learning Algorithm

Input: inp[1…..n] all input parameters which is generated by sensors, Threshold group TMin[1…n] and TMax[1…n] for all sensor.

Output : Trigger executed on appliances, Buzzer execution and GPS message.

Step 1 : Read all records from database (R into DB)

Step 2 : Parts [] ← Split(R)

Step 3 : $\text{CVal} = \sum_{k=0}^{n} \textbf{Parts[k]}$

Step 4 : check (Cval with Respective threshold of TMin[1…n] and TMax[1…n] **)**

If(true) execute trigger on respective output appliances.

Else Continue;

Step 5 : T← get current state with timestamp

Step 6 : if(T.time > Defined Time)

Active GPS for messaging or on buzzer

Else continue.

Step 7 : end for

Step 8: return DB

Linear Regression Phase

Input: User input file data record which contains {symptoms, disease} segment from train database .

Output: Projected weight

Step 1: Read R {current input from sensor} from current parameters.

Step 2: Map with train features with each sample.

Step 3 : calculate average weight of train DB with same evidences

$$Weight[i \ldots .n] = \sum_{k=0}^{n} (Sc)$$

Step 4 : optimized all n instances and select top k instances top[k]

Step 5: Return top[k].disease

V. RESULST AND DISCUSSION

The proposed implementation has done with open source environment, in python. Synthetic dataset has used for training as well as testing with cross fold validation with 5 fold, 10 fold and 15 fold respectively.

Various machine learning algorithms has used to evaluate the performance analysis of entire execution. Below figure 2 shows classification of proposed system with various sample inputs.

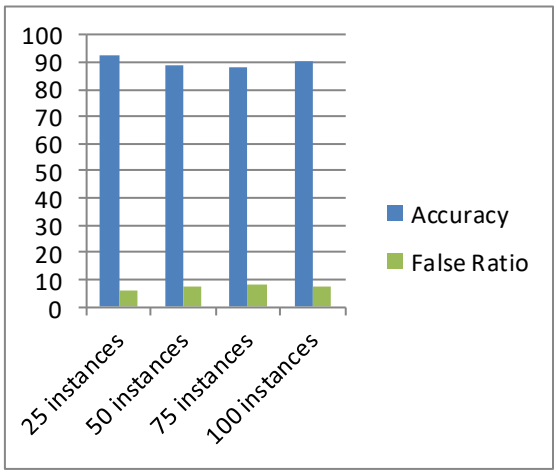

Figure 4: System accuracy of proposed system with false ratio.

The second experiment analysis has done to detect the efficiency of proposed system, the Figure 3 shows the detail description.

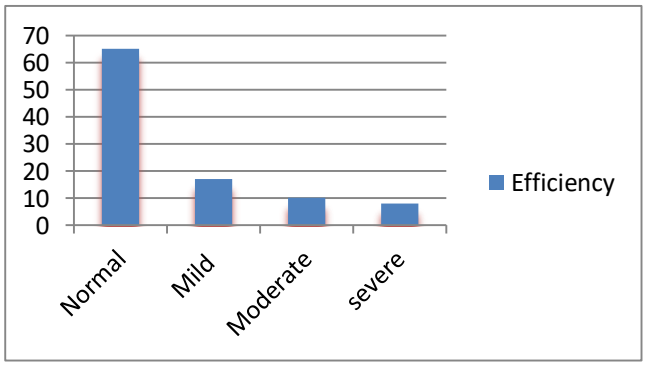

Figure 5: System accuracy of proposed system with false ratio.

VI. CONCLUSION

It is observed that, diabetes detection is done in the most of the areas like retinopathy, neuropathy, nephropathy, cardiovascular patients. Also, prediction can be used for various health claim insurance areas considering historical datasets. Most of the Machine Learning techniques are used on PIDD dataset by extracting various features. Furthermore research is required for detecting and diagnosing this devastating disease.

REFERENCES

1. J. Pradeep Kandhasamy, S. Balamurali, "Performance Analysis of Classifier Models to Predict Diabetes Mellitus", Elseiver – Science direct ", 2015
2. Hang Lai, Huaxiong Huang, Karim Keshavjee, Aziz Guergachi and Xin Gao, "Predictive models for diabetes mellitus using machine learning techniques", BMC Endocrine Disorders (2019) 19:101 https://doi.org/10.1186/s12902-019-0436-6
3. Raghavendra S, Santosh Kumar J, Raghavendra B. K., "Performance Evaluation of Machine Learning Techniques in Diabetes Prediction" , International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-3, February 2019
4. Neha Sharma, Ashima Singh, "Diabetes Detection and Prediction Using Machine Learning: A Survey", Springer – Nature ", ICAICR 2018
5. Dipti Sisodia, Dileep Singh Sisodia, " Prediction of diabetes using Classification algorithms", Elseiver, Procedia Computer Science, Volume 132, 2018 Pages 1578-1585.
6. S. K. Wasan and V. Bhatnagar and H. Kaur, "The Impact of Data Mining Techniques on Medical Diagnostics," Data Science Journal, vol. 5, pp.119-126, 2006
7. M. K. Bodla and S. M. Malik and M. T. Rasheed and M. Numan and M. Z. Ali and J. B. Brima,"Logistic Regression and Feature Extraction Based Fault Diagnosis of Main Bearing of Wind Turbines," IEEE 11th International Conference on Industrial Electronics and Applications (ICIEA), pp. 1628-1633, 2016
8. C. P. Prathibhamol and K. V. Jyothy and B. Noora,"Multi Label Classification Based on Logistic Regression (MLC-LR)," International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 2708-2712, 2016
9. Masatoshi Nagata, Kohichi Takai, Keiji Yasuda , Panikos Heracleous, Akio Yoneyama" Prediction Models for Risk of Type-2 Diabetes Using Health Claims", Proceedings of the BioNLP 2018 workshop, pages 172–176 Melbourne, Australia" , July 19, 2018
10. B. K. Raghavendraand J. B. Simha,"Performance Evaluation of Logistic Regression and Neural Network Model with Feature Selection Methods and Sensitivity Analysis on Medical Data Mining," International Journal of Advanced Engineering Technology, vol. 2, no. 1, pp. 289-298, 2011
11. C. Cortes and V. Vapnik, "Support Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273-297, 1995
12. T. K. Ho, "The Random Subspace Method for Constructing Decision Forests," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 20, no. 8, pp. 832-844, 1998
13. W. G. Touw and J. R. Bayjanov and L. Overmars and L. Backus and J. Boekhorst and M. Wels and S. A. F. T. V. Hijum, "Data Mining in the Life Sciences with Random Forest: a Walk in the park or lost in Jungle?," Briefings in Bioinformatics, vol. 14, no. 3, pp. 315-326, 2012
14. M. L. Raymer and T. E. Doom and L. A. Kuhn and W. F. Punch, "Knowledge Discovery in Medical and Biological Datasets Using a Hybrid Bayes Classifier/Evolutionary Algorithm,"IEEE Transactions on Systems, Man and Cybernatics, vol. 33, no. 5, pp. 802-813, 2003
15. M. B. Dowlatshahi and M. Rezaeian, "Training Spiking Neurons with Gravitational Search Algorithm for Data Classification," IEEE 1st International Conference on Swarm Intelligence and Evolutionary Computation, pp. 53-58, 2016.
16. E. Tuba and M. Tuba and D. Simian,"Adjusted Bat Algorithm for Tuning of Support Vector Machine Parameters," IEEE Congress on Evolutionary Computation, pp. 2225-2232, 2016
17. R. Bruni and G. Bianchi, "Effective Classification Using a Small Training Set Based on Discretization and Statistical Analysis," IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 9, pp. 2349-2361, 2015
18. P. Sykacek and S. Roberts, "Adaptive Classification by Variational Kalman Filtering," Advances in Neural Information Processing Systems (NIPS), pp. 737-744, 2002
19. F. J. Li and Y. H. Qian and J. T. Wang and J. Y. Liang, "Multigranulation Information Fusion: A Dempster-Shafer Evidence Theory Based Clustering Ensemble Method," Proceedings of IEEE International Conference on Machine Learning and Cybernetics (ICMLC), vol. 1, pp. 58-63, 2015
20. S. J. Perantonis and V. Virvilis, "Input Feature Extraction for Multilayered Perceptrons Using Supervised Principal Component Analysis," Neural Processing Letters, vol. 10, no. 3, pp. 243-252, 1999
21. C. A. Ratanamahatana and D. Gunopulos, "Feature Selection for the Naïve Bayesian Classifier Using Decision Trees," Applied Artificial Intelligence (AAI), vol. 17, no. 5-6, pp. 475-487, 2003
22. F. Divina and E. Marchiori, "Knowledge-Based Evolutionary Search for Inductive Concept Learning," Knowledge Incorporation in Evolutionary Computation, Springer, vol. 167, pp. 237-253, 2005
23. G. T. Hilda and R. R. Rajalaxmi, "Effective Feature Selection for Supervised Learning Using Genetic Algorithm," IEEE 2nd International Conference on Electronics and Communication Systems (ICECS 2015), pp. 909-914, 2015

24. Blayvas and R. Kimmel, "Machine Learning via Multiresolution Approximations," IEICE Transaction on Information System, vol. E86-D, no. 7, pp. 1172-1180, 2003
25. Watkins and J. Timmis and L. Boggess, "Artificial Immune Recognition System (AIRS): An Immune Inspired Supervised Learning Algorithm," Genetic Programming and Evolvable Machines, vol. 5, no. 3, pp. 291-317, 2004
26. S. Dora and S. Sundaram and N. Sundararajan,"A Two Stage Learning Algorithm for a Growing-Pruning Spiking Neural Network for Pattern Classification Problems,"International Joint Conference on Neural Networks (IJCNN), pp. 1-7, 2015
27. S. Raghavendra and M. Indiramma, "Performance Evaluation of Logistic Regression and Artificial Neural Network Model with Feature Selection Methods Using Cross Validation Sample and Percentage Split on Medical Datasets," International Conference on Emerging Research in Computing, Information, Communication and Applications, vol. 2, 2014
28. S. Raghavendra and M. Indiramma, "Classification and Prediction Model Using Hybrid Technique for Medical Datasets," International Journal of Computer Applications, vol. 127, no. 5, pp. 20-15, 2015
29. S. Raghavendra and M. Indiramma, "Hybrid Data Mining Model for the Classification and Prediction of Medical Datasets," International Journal of Knowledge Engineering and Soft Data Paradigms," vol. 5, no. 3/4, pp. 262- 284, 2017
30. F. G. Woldemichael and S. Menaria, "Prediction of Diabetes Using Data Mining Techniques," 2nd International Conference on Trends In Electronics and Informatics, pp. 414-418 , 2018
31. D. K. Choubey and S. Paul and S. Kumar and S. Kumar,"Classification of Pima Indian Diabetes Dataset Using Naïve Bayes with Genetic Algorithm as an Attribute Selection," Communication and Computing Systems, Taylor & Francis Group, pp. 451-455, 2017
32. S. Wei and X. Zhao and C. Miao, "A Comprehensive Exploration to the Machine Learning Technique for Diabetes Dataset," IEEE 4th World Forum on Internet of Things, pp. 291-295, 2018
33. F. Mercaldo and V. Nardone and A. Santone,"Diabetes Mellitus Affected Patients Classification and Diagnosis through Machine Learning Techniques,"Procedia Computer Science,vol. 112, pp. 2519–2528, 2017
34. Ioannis Kavakiotis , Olga Tsave c, Athanasios Salifoglou c, Nicos Maglaveras b,d, Ioannis Vlahavas a, Ioanna Chouvarda b,d , "Machine Learning and data mining methods in diabetes research", ELSEIVER - Computational and structural biotechnology journal" , 8th January 2017
35. Arianna Dagliati, Simone Marini, Lucia Sacchi, Giulia Cogni, Marsida Teliti, Valentina Tibollo, Pasquale De Cata, Luca Chiovato, and Riccardo Bellazzi "Machine Learning Methods to Predict Diabetes Complications", Journal of Diabetes Science and Technology", 12th May 2017
36. Manal Alghamdi, Mouaz AI- Mallah , Steven Keteyian , Cliinton Brawner, Jonathan Ehrman "Predicting diabetes mellitus using SMOTE and ensemble machine learning approach : The Henry Ford Exercise (FIT) project ", PLOS one , 24 July 2017
37. Han Wu, Shengqi Yang *, Zhangqin Huang, Jian He, Xiaoyi Wang, "Type 2 diabetes mellitus prediction model based on data mining", ELSEIVER – Informatics in medicines unlocked" , 12th December 2017
38. Zou, Quan et al. "Predicting Diabetes Mellitus With Machine Learning Techniques." *Frontiers in genetics* vol. 9 515. 6 Nov. 2018, doi:10.3389/fgene.2018.00515
39. Zhuang Wang, Jianbo Yang "Diabetic Retinopathy Detection via Deep Convolutional Networks for Discriminative Localization and Visual Explanation ", 32 AAAI International Conference on Artificial Intelligence ", Loisiana , USA, June 2018
40. N. Sneha, Tarun Gangil "Analysis of diabetes mellitus for early prediction using optimal features selection ", Springer - Open – Journal of Big data " , 6th February 2019
41. Jelinek HF, Stranieri A, Yatsko A, Venkatraman S. Data analytics identify glycated hemoglobin co-markers for type 2 diabetes mellitus diagnosis. Compute Biology Med Aug 1 2016;75:90
42. Cai L, Wu H, Li D, Zhou K, Zou F. Type 2 diabetes biomarkers of human gut microbiota selected via iterative sure independent screening method. PLoS One Oct 19 2015;10(10):e0140827.
43. Lee BJ, Kim J Y, Lee BJ, Kim JY. Identification of type 2 diabetes risk factors using phenotypes consisting of anthropometry and triglycerides based on machine learning. IEEE J Biomed Health Inform Jan 2016;20(1):39–46. [Epub 2015 Feb 6]
44. Habibi S, Ahmadi M, Alizadeh S. Type 2 diabetes mellitus screening and risk factors using decision tree: results of data mining. Glob J Health Sci Mar 18 2015;7(5): 304–10
45. Shankaracharya, Odedra D, Samanta S, Vidyarthi AS. Computational intelligence based diagnosis tool for the detection of prediabetes and type 2 diabetes in India. Rev Diabet Stud Spring 2012;9(1):55–62 Epub 2012 May
46. Gandhi, K.K., Prajapati, N.B.: Diabetes prediction using feature selection and classification. Int. J. Adv. Eng. Res. Develop. (2014). https://doi.org/10.21090/ijaerd.0105110
47. Heydari, M., Teimouri, M., Heshmati, Z., Alavinia, S.M.: Comparison of various classification algorithms in the diagnosis of type 2 diabetes in Iran. Int. J. Diabetes Develop. Countries, 167–173. (2016). https://doi.org/10.15417/1881
48. Samant, P., Agarwal, R.: Diagnosis of Diabetes using computer methods: soft computing methods for diabetes detection using iris (2017). https://doi.org/10.1016/j.cmpb.2018.01.004
49. Ephzibah, E.P.: Cost effective approach on feature selection using genetic algorithms and fuzzy logic for diabetes diagnosis. Int. J. Soft Comput. (IJSC) 2, 1–10 (2011). https://doi.org/ 10.5121/ijsc.2011.2101

AUTHORS PROFILE

Mrs. Vaishali Y. Baviskar, is working as an Assistant Professor in G. H. Raisoni Institite of Enginerring and Technology, Wagholi, Pune. She has completed her B.E. in Computer Engineering from North Maharashtra University, M.E. in Computer Engineering from Savitribai Phule Pune University and currently pursuing her PhD from Bennett University, Greater Noida. Her Research area is Artificial Intelligence and Machine Learning. She has published 2 National and 6 International Journal Papers. She has an IETE Membership.