



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 6, June 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Diabetes Mellitus Prediction and Diagnosis using different Machine Learning Algorithm: A Survey

<sup>1</sup>Sanjeev Kumar Suman, <sup>2</sup>Prof. Prakash Saxena

M. Tech. Scholar, Dept. of Electronics and Communication Engineering, BIST, Bhopal, India<sup>1</sup>

Head of Dept., Dept. of Electronics and Communication Engineering, BIST, Bhopal, India<sup>2</sup>

**ABSTRACT:** Diabetes Mellitus (DM) is a chronic, lifelong metabolism disorder. It affects the ability of the body system to use the energy found in food. The improper management of the disease will lead to Heart disease, kidney disease, eye disease, nerve disease and pregnancy complications. Classification model helps physicians to improve their prognosis, diagnosis or treatment planning procedures. Big Data Analytics plays an significant role in healthcare industries. Healthcare industries have large volume databases. Using big data analytics one can study huge datasets and find hidden information, hidden patterns to discover knowledge from the data and predict outcomes accordingly. In existing method, the classification and prediction accuracy is not so high. In this paper, we have proposed a diabetes prediction model for better classification of diabetes which includes few external factors responsible for diabetes along with regular factors like Glucose, BMI, Age, Insulin, etc. Classification accuracy is boosted with new dataset compared to existing dataset. Further with imposed a pipeline model for diabetes prediction intended towards improving the accuracy of classification.

**KEYWORDS-** Diabetic dataset, Classification, Machine Learning

## I. INTRODUCTION

Diabetes has been most prevalent disease in the world which is not confined to any specific age group.[1] Threatening all age group make it more worrisome for the doctor and Government all over the globe. We need pensive attention and research to eradicate this menace from the life of people. Deficiency of insulin in human body results in excess glucose (sugar) level because insulin the main agent to regulate the sugar level in our body system.[9] The fundamental reason explored by the medical science for such phenomenon is that the absence of insulin enhances the carbohydrate rich food conversion into glucose which is accumulated in the blood , enhancing the sugar level more than normal. Diabetes is responsible for other diseases like heart disease, kidney damage, blindness, blood vessel damage or it may even cause death. Presently there is no such permanent treatment in medical science for diabetes; however it can be controlled by proper control over diet and regular physical exercise [1].

Detection of diabetes in advance and in early stages is very important for better prevention and cure. The researchers have been founded several techniques to diagnose the diabetes by contributing their continuous efforts and knowledge. There have been many techniques used in Neural Networks that were applied over diabetes diagnosis to help physicians. An uncommon approach utilized by Polat and Gunes is Principal element Associate in Nursing analysis and reconciling Neuro Fuzzy reasoning System (ANFIS). In initial stage PCA reduces options for higher accuracy and in second stage ANFIS is employed for classification of polygenic disorder knowledge set [4]. Equally LDA is employed for feature reduction of polygenic disorder knowledge with classification by ANFIS within the second stage [3]. Manjeevan Seera et al have planned 3 methods- Fuzzy Min-Max Neural network (FMM), Classification and Regression Tree (CART) and Random Forest Model (RF) over 3 totally different medical datasets- carcinoma Wisconsin, Pima Indians polygenic disorder and Liver Disorders. Among them FMM-CART-RF generates higher results compared to FMM and FMM-CART results. The conception of fuzzy modeling has been used effectively in diagnosis systems. within the work planned by Sean N. Ghazavi et al 3 strategies namely- fuzzy k-nearest neighbor rule, a fuzzy clustering-based modeling and reconciling Network-based Fuzzy reasoning System (ANFIS) were used over Wisconsin carcinoma and Pima Indians polygenic disorder dataset. The classification accuracy obtained is ninety seven.17% and 77.65% severally. many different mechanisms like k-means agglomeration and C4.5 rule are used for polygenic disorder prediction. K-means agglomeration enforced by maori hen tool is used for pattern extraction. The resultant output is given to C4.5 call tree used for classification exploitation k-fold cross validation is employed to separate samples into totally different

categories. Preprocessing of knowledge is incredibly vital to enhance the accuracy and to scale the info in same vary of values, that is achieved by Principal element Analysis. The classification of resultant knowledge is completed by Feed forward Higher Order Neural network exploitation k-fold cross validation. Support Vector Machines have proved exceptional success within the space of diagnosis. Within the work planned by V. Anuja kumari et al have classified the polygenic disorder knowledge exploitation SVM classifier with RBF kernel. The classification accuracy obtained is seventy eight.

The classification task consists in distribution instances from a given domain, delineated by a group of distinct price attributes, into a group of categories, which might be thought-about values of selected distinct target attribute known as target conception. The proper category labels area unit typically unknown, however area unit provided for a set of the domain. It is wont to produce the instance from identical domain, describes by identical set of attributes. This follows the overall assumption of inductive learning of the classification task is that the commonest mental representation [2].

The classification Techniques disperse a category to assortment of knowledge records having specific attributes and its values, therefore the classification techniques in aid is applied for nosology functions. A classification model receives a group of connected attribute values, like clinical measurements and offers a category of knowledge records as output.

In our study we tend to propose a brand new technique by adding Particle swarm improvement with Multi layer Neural Network (PSO) to classify designation of Pima Indians polygenic disorder dataset taken from UCI machine learning repository named as Particle swarm improvement Neural Network (PSO). the basic plan for this rule is that at the start stage of finding out the optimum parameter of PSO by meted out random tuples from the input dataset, then PSO is used to accelerate the coaching speed. once the fitness operate price has not amendment for a few generations, the looking method is switched to gradient downhill looking per heuristic information. Finally PSO combined with Neural Network is employed to classify the samples that be 2 classes either '0' for negative (Not Diabetic) or '1' for positive (Diabetic).

## II. LITERATURE REVIEW

**Isfafuzzaman Tasin et al. [1]**, diabetes is the most fatal and prevalent infectious disease, affecting 537 million people worldwide. Excessive body weight, abnormal cholesterol levels, a family history of diabetes, inactivity, unhealthy eating habits, and other factors are just a few of the many factors that can lead to diabetes. One of this disease's most common symptoms is increased urination. Diabetes can lead to a number of complications, including heart disease, kidney disease, nerve damage, diabetic retinopathy, and others. Be that as it may, its gamble can be decreased assuming that it is anticipated early. An automated diabetes prediction system based on a private dataset of female Bangladeshi patients and a variety of machine learning methods is presented in this paper. The creators utilized the Pima Indian diabetes dataset and gathered extra examples from 203 people from a nearby material manufacturing plant in Bangladesh. Highlight determination calculation shared data has been applied in this work. The insulin characteristics of the private dataset have been predicted using a semi-supervise model with extreme gradient boosting. The problem of class imbalance has been addressed through the use of SMOTE and ADASYN strategies. The creators utilized AI characterization strategies, that is to say, choice tree, SVM, Arbitrary Timberland, Calculated Relapse, KNN, and different outfit procedures, to figure out which calculation delivers the best forecast results. The proposed system had the highest accuracy, 0.81 F1 coefficient, and AUC of 0.84 in the XGBoost classifier with the ADASYN approach after training on and testing all classification models. In addition, the domain adaptation technique has been used to demonstrate the proposed system's adaptability. In order to comprehend how the model predicts the outcomes, the LIME and SHAP frameworks are used to implement the explainable AI strategy. At last, a site system and an Android PDA application have been created to include different elements and foresee diabetes immediately.

**Olisah et al. [2]**, diabetes mellitus is a metabolic problem portrayed by hyperglycemia, which results from the insufficiency of the body to discharge and answer insulin. While possibly not appropriately overseen or analyzed on time, diabetes can represent a gamble to crucial body organs like the eyes, kidneys, nerves, heart, and veins thus can life-compromise. The numerous long periods of examination in computational determination of diabetes have highlighted AI to as a practical answer for the expectation of diabetes. The accuracy rate, on the other hand, suggests that there is still a lot of room for improvement. In this paper, we are proposing an AI system for diabetes expectation and finding utilizing the PIMA Indian dataset and the research center of the Clinical City Emergency clinic (LMCH) diabetes dataset. We conjecture that taking on highlight choice and missing worth attribution techniques can increase the presentation of order models in diabetes expectation and determination.

**Deberneh et al. [3]**, a person who is at risk can take steps to either delay the onset of type 2 diabetes (T2D) or prevent it altogether by anticipating its occurrence. Using variables from the current year (Y), we developed a machine learning

(ML) model in this study to predict the occurrence of T2D in the following year ( $Y + 1$ ) From 2013 to 2018, a private medical institute used electronic health records to collect the dataset for this study. ANOVA tests, chi-squared tests, and recursive feature elimination methods were used to select key features prior to building the prediction model. Fasting plasma glucose (FPG), HbA1c, triglycerides, body mass index (BMI), gamma-GTP, age, uric acid, sex, smoking, drinking, physical activity, and family history were the resulting characteristics. Based on these variables, we predicted the outcome as normal (non-diabetic), prediabetes, or diabetes using ensemble machine learning algorithms like logistic regression, random forest, support vector machine, and XGBoost. The prediction model performed reasonably well in predicting the prevalence of type 2 diabetes in the Korean population, as demonstrated by the results of the experiments. The cross-validation (CV) results demonstrated that the ensemble models performed better than the single models, providing clinicians and patients with useful predictive information regarding the likelihood of developing T2D. By including more of the dataset's medical history, the prediction models' CV performance improved.

**Nikos Fazakis et al. [4]**, a steady rise has been seen in the level of older individuals who want and are as yet ready to add to society. Consequently, exiting the workforce or exit from the work market, because of wellbeing related issues, represents a significant issue. These days, because of mechanical advances and different information from various populaces, the gamble factors examination and medical problems screening are moving towards mechanization. In the setting of this work, a specialist driven, IoT empowered inconspicuous clients wellbeing, prosperity and practical capacity checking structure, engaged with AI apparatuses, is proposed. Diabetes is a high-predominance ongoing condition with hurtful ramifications for the personal satisfaction and high death rate for individuals around the world, in both created and non-industrial nations. Thus, its extreme effect on people's life, e.g., individual, social, working, can be significantly decreased assuming that early identification is conceivable, however most exploration works in this field fall flat to give a more customized approach both in the demonstrating and forecast process. Toward this path, our planned framework concerns diabetes risk expectation in which specific parts of the Knowledge Discovery in Database (KDD) process are applied, assessed and consolidated. Specifically, dataset creation, highlights choice and classification, utilizing different Supervised Machine Learning (ML) models are thought of. The outfit Weighted Voting LRRFs ML model is proposed to work on the forecast of diabetes, scoring an Region Under the ROC Curve (AUC) of 0.884. Concerning the weighted democratic, the ideal loads are assessed by their comparing Sensitivity and AUC of the ML model in light of a bi-objective hereditary calculation. Likewise, a near report is introduced among the Finnish Diabetes Risk Score (FINDRISC) what's more, Leicester risk score frameworks and a few ML models, utilizing inductive and transductive learning. The tests were led utilizing information removed from the English Longitudinal Study of Aging (ELSA) data set.

**Naveen Kishore et al. [5]**, diabetes is considered as one of the deadliest and persistent illnesses which causes an increment in glucose. Numerous entanglements happen if diabetes stays untreated and unidentified. The monotonous distinguishing measure brings about visiting of a patient to a symptomatic focus and counseling specialist. However, the ascent in AI approaches tackles this basic issue. The thought process of this investigation is to plan a model which can visualize the probability of diabetes in patients with greatest exactness. Thusly three AI characterization calculations specifically Decision Tree, SVM and Naive Bayes are utilized in this analysis to identify diabetes at a beginning phase. Analyses are performed on Pima Indians Diabetes Database (PIDDD) which is sourced from UCI AI storehouse. The exhibitions of the multitude of three calculations are assessed on different estimates like Precision, Accuracy, F-Measure, and Recall. Exactness is estimated over effectively and erroneously ordered cases. Results got show Naive Bayes beats with the most elevated precision of 76.30% relatively different calculations. These outcomes are confirmed utilizing Receiver Operating Characteristic (ROC) bends in a legitimate and methodical way.

**Chatrati et al. [6]**, for patients with diabetes and high blood pressure, home health monitoring can make it easier to monitor them remotely. Because these chronic diseases frequently result in complications that can be life-threatening if discovered at a later stage, early detection is crucial. A smart home health monitoring system that helps the patient analyze their home readings of blood pressure and glucose and notifies the healthcare provider in the event of any abnormalities is proposed in this work. Predictions of hypertension and diabetes status are made, respectively, using a combination of conditional decision-making and machine learning methods. Using the patient's glucose and blood pressure readings, the objective is to predict the patient's diabetes and hypertension status. In this case, an algorithm for supervised machine learning classification is used to train a system to predict the patient's diabetes and hypertension levels. The support vector machine classification algorithm was chosen to train the model because it was found to be the most accurate after analyzing all of the classification algorithms. This proposed work fosters an application for a home wellbeing checking framework with an easy to understand simple to-utilize graphical UI to analyze circulatory strain and

diabetes status of patients alongside sending classified cautions and constant notices to their enlisted doctor or facility all from home.

**Hasan et al. [7]**, diabetes, otherwise called persistent sickness, is a gathering of metabolic illnesses because of an elevated degree of sugar in the blood over an extensive stretch. The gamble component and seriousness of diabetes can be decreased altogether assuming the exact early forecast is conceivable. The hearty and precise forecast of diabetes is profoundly difficult because of the predetermined number of named information and furthermore the presence of exceptions (or missing qualities) in the diabetes datasets. Outlier rejection, filling the missing values, data standardization, feature selection, K-fold cross-validation, various Machine Learning (ML) classifiers (k-nearest Neighbour, Decision Trees, Random Forest, AdaBoost, Naive Bayes, and XGBoost), and Multilayer Perceptron (MLP) are all used in the framework we propose in this literature for diabetes prediction. The weights are estimated from the corresponding Area Under ROC Curve (AUC) of the ML model, and this literature also suggests that weighted ensembling of various ML models can improve diabetes prediction. The performance metric of choice is AUC, which is then maximized through the grid search method during hyperparameter tuning. Every one of the tests, in this writing, were led under similar trial conditions utilizing the Pima Indian Diabetes Dataset. With sensitivity, specificity, false omission rate, diagnostic odds ratio, and AUC of 0.789, 0.934, 0.092, 66.234, and 0.950, respectively, our proposed ensembling classifier outperforms the state-of-the-art results by 0.00% in AUC out of all the extensive experiments. Our proposed system for the diabetes expectation outflanks different techniques talked about in the article. Additionally, it may yield superior outcomes from the same dataset, which may improve diabetes prediction performance. Our diabetes prediction source code is made available to the general public.

**Cervantes et al. [8]**, support vector machines (SVMs) and how they can be used in a variety of scientific fields have received a lot of attention in recent years. SVMs are one of the most remarkable and strong order and relapse calculations in various fields of utilization. The SVM has been assuming a huge part in design acknowledgment which is a widely well known and dynamic examination region among the specialists. Other applications, such as SVM for large data sets, SVM for multi-classification, and SVM for unbalanced data sets, have been developed as a result of research in some fields in which SVMs do not perform well. Further, SVM has been incorporated with other high level strategies, for example, advance calculations, to upgrade the capacity of grouping and improve boundaries. SVM algorithms are now widely used in a wide range of scientific and engineering fields. SVMs are briefly introduced, numerous applications are discussed, and challenges and trends are summarized in this paper. SVMs' limitations will also be identified. Additional applications will be discussed alongside the SVMs' future. SVMs' applications, particularly in some fields, will also be examined.

**Pranto et al. [9]**, machine learning has a significant impact on a variety of scientific and technological fields, including medical and life sciences research. Diabetes Mellitus, more commonly referred to as diabetes, is a chronic condition characterized by the body's use of insulin and abnormally high levels of glucose sugar in blood cells. Using various Machine Learning techniques to construct a model with a few dependencies based on the PIMA dataset, this article has focused on analyzing diabetic patients and detecting diabetes. A portion of PIMA that has never been seen was used to test the model, as was a dataset from Kurmitola General Hospital in Dhaka, Bangladesh. The exploration is directed to exhibit the presentation of a few classifiers prepared on a specific country's diabetes dataset and tried on patients from an alternate country. In this study, we looked at decision trees, K-nearest neighbors, random forests, and naive bayes classifiers. The results show that random forest and naive bayes classifiers performed well with both datasets.

**Mohan et al. [10]**, diabetes is arising as a prevalent illness in the emerging nations like India. The disease is getting a lot worse and causes a lot of other problems for the body. Many variables are staying as a reason for this sickness in human body. The disease can only be controlled and cannot be cured. Support vector machine (SVM) prediction is the focus of this paper. The SVM algorithm's performance is evaluated for various readily available kernels. For the purpose of prediction, the best kernel is chosen. The proposed work is implemented using Python, and its performance is comparable to that of other algorithms.

### III. METHODOLOGY

#### Learning

The main property of an ML is its capability to learn. Learning or preparing is a procedure by methods for which a neural system adjusts to a boost by making legitimate parameter modifications, bringing about the generation of wanted reaction. Learning in an ML is chiefly ordered into two classes as [9].

- Supervised learning
- Unsupervised learning

### Supervised Learning

Regulated learning is two stage forms, in the initial step: a model is fabricated depicting a foreordained arrangement of information classes or ideas. The model developed by investigating database tuples portrayed by traits. Each tuple is expected to have a place with a predefined class, as dictated by one of the qualities, called to have a place with a reclassified class, as controlled by one of the traits called the class name characteristic. The information tuple are dissected to fabricate the model all things considered from the preparation dataset.

### Unsupervised learning

It is the kind of learning in which the class mark of each preparation test isn't knows, and the number or set of classes to be scholarly may not be known ahead of time. The prerequisite for having a named reaction variable in preparing information from the administered learning system may not be fulfilled in a few circumstances.

Data mining field is a highly efficient techniques like association rule learning. Data mining performs the interesting machine-learning algorithms like inductive-rule learning with the construction of decision trees to development of large databases process. Data mining techniques are employed in large interesting organizations and data investigations. Many data mining approaches use classification related methods for identification of useful information from continuous data streams.

### Nearest Neighbors Algorithm

The Nearest Neighbor (NN) rule differentiates the classification of unknown data point because of closest neighbor whose class is known. The nearest neighbor is calculated based on estimation of  $k$  that represents how many nearest neighbors are taken to characterize the data point class. It utilizes more than one closest neighbor to find out the class where the given data point belong termed as KNN. The data samples are required in memory at run time called as memory-based technique. The training points are allocated weights based on their distances from the sample data point. However, the computational complexity and memory requirements remained key issue. For addressing the memory utilization problem, size of data gets minimized. The repeated patterns without additional data are removed from the training data set.

### Naive Bayes Classifier

Naive Bayes Classifier technique is functioned based on Bayesian theorem. The designed technique is used when dimensionality of input is high. Bayesian Classifier is used for computing the possible output depending on the input. It is feasible to add new raw data at runtime. A Naive Bayes classifier represents presence (or absence) of a feature (attribute) of class that is unrelated to presence (or absence) of any other feature when class variable is known. Naive Bayesian Classification Algorithm was introduced by Shinde S.B and Amrit Priyadarshi (2015) that denotes statistical method and supervised learning method for classification. Naive Bayesian Algorithm is used to predict the heart disease. Raw hospital dataset is employed. After that, the data gets preprocessed and transformed. Finally by using the designed data mining algorithm, heart disease was predicted and accuracy was computed.

### Support Vector Machine

SVM are used in many applications like medical, military for classification purpose. SVM are employed for classification, regression or ranking function. SVM depends on statistical learning theory and structural risk minimization principal. SVM determines the location of decision boundaries called hyper plane for optimal separation of classes as described in figure 1.4. Margin maximization through creating largest distance between separating hyper plane and instances on either side are employed to minimize upper bound on expected generalization error. Classification accuracy of SVM not depends on dimension of classified entities. The data analysis in SVM is based on convex quadratic programming. It is expensive as quadratic programming methods need large matrix operations and time consuming numerical computations.

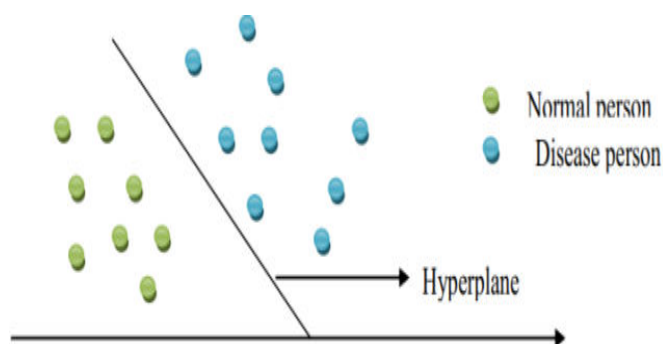


Fig. 1: Support Vector Classification

#### IV. CONCLUSION

Diabetes is metabolic disease that arises due to high blood glucose level in body. Insulin is not sufficient enough in body of diabetic patients to regulate the sugar level. Further several other diseases also arise from diabetes which is hazardous to health. It is necessary to detect such a serious health issue as early as possible. Diabetes is cause of various diseases in human body. To make diabetes diagnosis easier for Physicians, there have been several methods employed. The diabetic data set is tested with selected classification algorithm.

#### REFERENCES

- [1] Isfafuzzaman Tasin, Tansin Ullah Nabil, Sanjida Islam, Riasat Khan, "Diabetes prediction using machine learning and explainable AI techniques", *Healthcare Technology Letters*, pp. 01-10, Wiley 2022.
- [2] Olisah, C.C., Smith, L., Smith, M., "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective", *Comput. Methods Programs Biomed.*, Vol. 20, pp. 1–12, 2022.
- [3] Deberneh, H.M., Kim, I., "Prediction of type 2 diabetes based on machine learning algorithm", *Int. J. Environ. Res. Public Health*, Vol. 18, pp. 1–14, 2021.
- [4] Nikos Fazakis, Otilia Kocsis, Elias Dritsas, Sotiris Alexiou, Nikos Fakotakis, and Konstantinos Moustakas, "Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction", *IEEE Access* 2021.
- [5] Naveen Kishore G, V.Rajesh, A.Vamsi Akki Reddy, K.Sumedh, T.Rajesh Sai Reddy, "Prediction of Diabetes using Machine Learning Classification Algorithms", *International Journal of Scientific & Technology Research*, Vol. 9, No. 01, 2020.
- [6] Chatrati, S.P., Hossain, G., Goyal, A., "Smart home health monitoring system for predicting type 2 diabetes and hypertension", *J. King Saud Univ. Comput. Inf. Sci.*, Vol. 34, No. 3, pp. 862–870, 2020.
- [7] Hasan, M.K., Alam, M.A., Das, D., Hossain, E., Hasan, M., "Diabetes prediction using ensembling of different machine learning classifiers", *IEEE Access*, Vol. 8, pp. 76516–76531, 2020.
- [8] Cervantes, J., García-Lamont, F., Rodríguez, L., Lopez-Chau, A., "A comprehensive survey on support vector machine classification: Applications, challenges and trends", *Neurocomputing*, Vol. 408, pp. 189–215, 2020.
- [9] Pranto, B., "Evaluating machine learning methods for predicting diabetes among female patients in Bangladesh", *Information* Vol. 11, pp. 1–20, 2020.
- [10] Mohan, N., Jain, V., "Performance analysis of support vector machine in diabetes prediction", In: *International Conference on Electronics, Communication and Aerospace Technology*, pp. 1–3, 2020.
- [11] Muhammad Azeem Sarwar, 2Nasir Kamal, 3Wajeeda Hamid, 4Munam Ali Shah, "Prediction of Diabetes Using Machine Learning Algorithms in Healthcare", *24th International Conference on Automation & Computing*, Newcastle University, Newcastle upon Tyne, UK, 6-7 September 2019.
- [12] Rao G.A., Syamala K., Kishore P.V.V., Sastry A.S.C.S. ., "Deep convolutional neural networks for sign language recognition", *International Journal of Engineering and Technology(UAE)* ,Vol: 7, Issue 5, pp: 62-70, 2018.
- [13] Quan Zou, Kaiyang Qu, Yamei Luo, Dehui Yin, Ying Ju and Hua Tang: "Predicting Diabetes Mellitus With Machine Learning Techniques", Springer, 2018.
- [14] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, "Machine learning on big data: Opportunities and challenges," *Neuro computing*, vol. 237, pp. 350–361, May 2017.
- [15] J. B. Heaton, N. G. Polson, and J. H. Witte, "Deep learning for finance: deep portfolios," *Appl. Stoch. Model. Bus. Ind.*, vol. 33, no. 1, pp. 3–12, Jan. 2017.
- [16] Reddy S.S., Suman M., Prakash K.N. ., "Micro aneurysms detection using artificial neural networks", 2018, *Lecture Notes in Electrical Engineering* ,Vol: 434 ,Issue 3, pp: 409 to 417.



- [17] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., Chouvarda, I., "Machine Learning and Data Mining Methods in Diabetes Research", Computational and Structural Biotechnology Journal 15, 104–116, 2017.
- [18] Majid Ghonji Feshki and Omid Sojoodi Shijan, "Improving the Heart Disease Diagnosis by Evolutionary Algorithm of PSO and Feed Forward Neural Network", International paper on IEEE 2016.
- [19] L. Hermawanti, "Combining of Backward Elimination and Naive Bayes Algorithm To Diagnose Breast Cancer", Momentum, vol. 11, no. 1, pp. 42-45, 2015.
- [20] O.S. Soliman, E. Elhamd, "Classification of Diabetes Mellitus using Modified Particle Swarm Optimization and Least Squares Support Vector Machine", IEEE 2014.
- [21] K. Saxena, Z. Khan, S. Singh, "Diagnosis of Diabetes Mellitus using K Nearest Neighbour Algorithm", International Journal of Computer Science Trends and Technology (IJCSST), 2014.
- [22] L. Hermawanti, S.G. Rabiha, "Combining of Backward Elimination and K-Nearest Neighbor Algorithms To Diagnose Heart Disease", Prosiding SNST Ke-5 Fakultas Teknik Universitas Wahid Hasyim, pp. 1-5, 2014.
- [23] R.A. Vinarti, W. Anggraeni, "Identification of Prediction Factor Diagnosis of Breast Cancer Rates with Stepwise Binary Logistic Regression Method", Jurnal Informatik, vol. 12, no. 2, pp. 70-76, November 2014.
- [24] Muhammad Waqar Aslam, Zhechen Zhu and Asoke Kumar Nandi, "Feature generation programming with comparative partner selection for diabetes classification", "Expert Systems with Applications", 5402-5412, IEEE 2013.





SJIF Scientific Journal Impact Factor

Impact Factor: 8.423



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details