iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.60717

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

Revolutionizing Disease Prediction with Deep Learning and Predictive Analysis

Dr. Sanjay Kumar[1], Nikhil Maurya[2], Parth Sharma[3], Ronit Bhardwaj[4]
[1]*Professor, Department of Computer Science and Engineering Galgotias College of Engineering and Technology, Greater Noida, Uttar Pradesh, India*
[2, 3, 4]*Department of Computer Science andEngineering, Galgotias College of Engineering andTechnology, Greater Noida, Uttar Pradesh, India*

Abstract: ***The increasing prevalence of chronic illnesses, including diabetes-related conditions and heart disease, poses a challenge to international healthcare systems. Reducing the detrimental effects of chronic disorders on patient outcomes requires early detection and treatment. This study investigates the possible uses of deep learning and predictive analysis in illness forecasting, with an emphasis on diabetes and heart disease specifically. Strict preparation methods were used, making use of a substantial dataset from the reference dataset source, to guarantee data quality. List-specific models and architectures were used for training and validation in order to assess how well different deep learning models performed in the prediction of sickness. The results show the potential of the suggested technique in the early diagnosis of sickness. They contain notable findings and promising performance metrics. By providing insight into the feasibility and efficacy of deep learning models for the prognosis of diabetes and heart-related illnesses, this work contributes to the expanding corpus of knowledge in healthcare analytics.***
Keywords: deep-learning method; convolutional neural networks; datasets; Machine- learning methods; predictive analysis

I. INTRODUCTION

The rising incidence of chronic illnesses is causing a paradigm shift in the global healthcare landscape and causing serious challenges for healthcare systems around the globe. Diabetes and heart disease are two of the primary causes of this health catastrophe, which calls for the creation of creative early detection and intervention techniques. Modern technology, particularly deep learning, when paired with predictive analysis offer a workable solution for these issues.

Because diabetes and heart disease are very expensive conditions, proactive strategies for early diagnosis and treatment of these conditions must be developed. According to global health data heart disease makes up sizable portion of cardiovascular disease which is still the greatest cause of death. Simultaneously, the increasing incidence of diabetes and its associated ramifications emphasize how critical it is to advance predictive instruments for timely intervention. Many industries, including medicine, have undergone radical change since development of deep learning.

Above all, deep learning models have proven to perform remarkably well not only in image analysis and natural language processing, but also in disease prediction. What sets them apart is their ability to automatically generate hierarchical representations from data. Because health-related data is complex and non-linear, deep learning algorithms are well-suited to tackle the problems associated with disease identification.

The desire to optimise deep learning and predictive analysis's ability to improve healthcare outcomes is what motivates this research. With a focus on diabetes and heart disease, we aim to contribute to the growing body of literature on innovative methods for illness prediction. In addition to enabling timely medical intervention, early identification has the potential to reduce overall strain on medical systems and save healthcare costs.

The main goals of this study are to: (1) determine whether deep learning models can be used to predict diseases; and (2) evaluate the models' effectiveness in relation to current approaches. In our study, we build and evaluate predictive models specifically for diabetes and heart disease utilising cutting-edge deep learning architectures.

The purpose of this research is to advance our understanding of illness prediction methods by presenting data on the efficacy of deep learning and predictive analysis in relation to heart disease and diabetes-related issues. The purpose of this investigation is to collect relevant data that may be used to direct future advancements in healthcare analytics and predictive medicine conditions. Through this exploration, we strive to contribute valuable knowledge that may inform future advancements in healthcare analytics and predictive medicine.

©IJRASET: All Rights are Reserved

In the parts that follow, the research on disease prediction, classification, and segmentation will be discussed in depth. Along with a detailed analysis of CNN, deep learning approaches, the Naïve Bayes algorithm, CNN, ANN, and CNN, the methodology will also be discussed. This section will also address the dataset and its segmentation, the presentation and comparison of the results, and a conclusion that summarises the findings of the different methodologies and offers some recommendations for future developments in this field.

II. LITERATURE SURVEY

To forecast heart abnormalities, Nidhi Bhattala et al. used a variety of data mining techniques. Their analysis revealed that the neural networks technique outperformed decision trees in terms of accuracy, with a maximum accuracy of 81.51%.

Several classification techniques, including Decision Trees, K-Nearest Neighbours (kNN), and SMO (used to train Support Vector Machines) were examined by Joshra Baharami et al.. They employed feature selection techniques in order to extract only the most significant variables from their dataset. They found that, at 83.732%, Decision Trees produced the best results.

Mrudula Gudadhe et al. conducted research on heart disease classification in 2010. The methods employed are Artificial Neural Networks (ANN) and Support Vector Machines (SVM). They included a three-layer multilayer perceptron neural network (MLPNN) in their decision support system, showing that MLPNN can be utilised to accurately identify heart disease with 82% accuracy rates with SVM and 84% accuracy rates with ANN.

In order to automate the detection of coronary artery disease, Dolatabaddi et al. collected HRV signals from ECG in domains, time, and frequency using an optimal Support Vector Machine as their classification model. The study's total precision demonstrated the importance of classification.

K. Sudhakar et al. used data mining techniques to predict the incidence of cardiac disease. The study used a variety of machine learning classification techniques, such as Decision Trees, Neural Networks, and Naïve Bayes, to compare and evaluate the effectiveness of classification algorithms on databases related to heart illness.

K Cinetha et al. introduced a fuzzy logic-based coronary heart disease decision support system in 2014.The model's goal was to forecast the probability of being diagnosed with heart disease within the following ten years.They looked over 1230 cases using dataset 11, reaching a maximum accuracy of 97.67%.

Birjais et al.'s study employed the PIMA Indian Diabetes (PID) data set. The dataset, which consists of 8 attributes and 768 occurrences, is available via the UCI machine learning repository. The primary goal was to raise awareness of diabetes diagnosis, as the World Health Organisation (WHO) identified diabetes as one of the chronic illnesses that is spreading the fastest globally in 2014. The study employed gradient boosting, logistic regression, and naive Bayes classifiers to determine an individual's diabetes status. The results demonstrated that the naive Bayes, logistic regression, and gradient boosting accuracy rates were, respectively, 86%, 79%, and 77%.

Sadhu and Jadli used a diabetic dataset from the UCI repository for the investigation, which included 520 instances and 16 attributes. Predicting diabetes in its early stages was their main focus. They applied seven distinct classification approaches (k-NN, logistic regression, SVM, naive Bayes, decision trees, random forests, and multilayer perception) to the dataset's validation set. After several machine learning models were trained, the random forests classifier showed up as the best model for the relevant data set, with an accuracy score of 98%. Then came 98% for multilayer perceptrons, 94% for decision trees, 98% for random forests, 93% for logistic regression, 94% for SVM, and 91% for naive Bayes.

Xue et al. conducted an experiment using a diabetes dataset that included 17 features from the UCI collection and 520 patients. They used supervised machine learning methods like SVM, naive Bayes classifiers, and LightGBM to detect diabetes early. They discovered that, when given real training data from 520 diabetic patients and future diabetic patients, ages 16 to 90, SVM performed the best in terms of recognition and classification accuracy. SVM outperformed the commonly used naive Bayes classifier, which obtained 93.27% accuracy, with the best accuracy rate of 96.54%. In contrast, LightGBM demonstrated an accuracy of 88.46%. These findings demonstrate that SVM.

Le et al. also carried out an experiment on early-stage diabetes tesrisk prediction using a dataset from the UCI repository that included 520 patients and 16 factors. By introducing a novel wrapper-based feature selection method that optimises the multilayer perceptron (MLP) and minimises the number of required input attributes through the use of the grey wolf optimiser (GWO) and adaptive particle swarm optimisation (APSO), they proposed a machine learning strategy for predicting the early onset of diabetes collected using a range of traditional machine learning techniques, including k-NN, naïve Bayes classifier (NBC), logistic regression (LR), SVM, decision tree (DT), random forest classifier (RFC), and SVM.LR's accuracy rate was 95%. The accuracy rates of SVM and k-NN were 95% and 96%, respectively.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

NBC scored 93%, DT scored 95%, and RFC scored 96% in terms of correctness. The computational findings of the proposed approaches show that higher prediction accuracy may be achieved (96% for GWO–MLP and 97% for APSO–MLP), and that fewer features are required. This research could be used as a tool to support physicians and in clinical settings.

Neural networks were used in the development of the "Predicting the Risk of Heart Failure With EHR Sequential Data Modelling" model by Bo Jin, Chao Che, et al. The study conducted tests intended to rule out heart disorders using actual congestive heart disease data that were extracted from electronic health records (EHR). The key elements of an extended memory network model—word vectors and one-hot encryption—are widely employed to simulate the diagnostic procedures and expected episodes of heart failure.

Julius et al. assessed a dataset that was gathered from the UCI repository and had 520 samples, each of which was characterised by 17 attributes, using the Weka application framework from the Waikato Environment for Knowledge Analysis. This study's main goal was to apply machine learning classification techniques for early diabetes detection based on observable sample characteristics. Several classifiers were employed, including k-NN, SVM, FT, and RFCs (random forests). Surprisingly, k-NN had the highest accuracy (98%), closely followed by SVM (94%), FT (93%), and RF (97%).

Diabetes is a dangerous condition for which early identification is difficult (Shafi et al). The research used machine learning classification techniques to create a flexible model that can identify diabetes early. The authors put a great deal of effort into creating a system that could accurately forecast a patient's probability of developing diabetes. The PID dataset from the UCI repository was used in this study to assess the efficacy and precision of three machine learning classification algorithms: decision tree (DT), support vector machine (SVM), and naïve Bayes classifier (NBC). The experiment's findings showed that the NBC method, which had a 74% accuracy rate, was suitable. With accuracy rates of 63% and 72%, respectively, the SVM and DT placed second and third.The developed framework and machine learning classifiers may be used in future studies to recognise and diagnose a wide range of novel illnesses. The work in the subject of diabetes research has the potential to be enhanced and expanded upon, much like many other machine learning techniques. The goal of research is to categorise more algorithms that effectively manage missing data.

Khanam et al.'s study focused on the early identification of diabetes, a condition for which there is now no recognised cure. They used data mining tools, machine learning, and neural network techniques to create an accurate diabetes prediction system. For the investigation, the PID dataset was obtained from the UCI repository. It included details on 768 patients with nine different characteristics. Support vector machine (SVM), naïve Bayes classifier (NBC), random forest classifier (RFC), logistic regression (LR), decision tree (DT), k-NN, and AdaBoost (AB) are the seven machine learning techniques that were applied to the dataset.

The effectiveness of LR and SVM coupled is demonstrated by the accuracy of an 88.6% neural network model with two hidden layers.

Heart Disease Prediction using Evolutionary Rule Learning," presented by Aakash Chauhan et al. (2018), automates processes and collects data straight from electronic records. Frequent pattern growth association mining on patient datasets yielded strong association rules that aided in the prediction of cardiac sickness.

A study conducted by Ashir Javeed, Shijie Zhou, and colleagues produced "An Intelligent Learning System based on Random Search Algorithm and Optimised Random Forest Model for Improved Heart Disease Detection." This work uses a grid search algorithm-optimized random forest model for cardiovascular disease detection and a random search algorithm for factor selection.

In order to predict heart sickness, M. Satish et al. used a variety of data mining techniques, including rule-based, decision tree, naive Bayes, and artificial neural networks. The cardiovascular disease warehouse produced association rules by using the effective pruning classification association rule (PCAR) technique.

The study "Prediction and Diagnosis of Heart Disease by Data Mining Techniques," by Boshra Bahrami and Mirsaeid Hosseini Shirvani , examines several classification schemes for the diagnosis of cardiovascular illness. After evaluating several classifiers, including KNN, SVO classifier, and decision tree, it was determined that the decision tree was the best at predicting cardiovascular illness based on the dataset.

In 2019, Mamatha Alex P and Shaicy P Shaji published a paper titled "Prediction and Diagnosis of Heart Disease. Patients using Data Mining Technique. "The research uses Machine approaches for predicting and diagnosing heart disease. Artificial Neural Networks have shown to be more accurate than other data mining categorization techniques in the diagnosis of heart disease.

Patients using data mining methods "In research, heart disease is predicted and diagnosed using machine learning techniques. Artificial neural networks have shown to be more accurate than previous data mining categorization techniques in the identification of heart disease.

An efficient solution incorporating hybrid machine learning methodology was developed. This hybrid strategy combines linear and random forest techniques. Examine gathered datasets and designate portions to provide projections. The attributes were chosen following preprocessing of the cardiovascular disease knowledge dataset. Subsequently, hybrid techniques were used to identify heart disease.

Table 1: Literature Overview

Author	Tittle	Technologies Used	Advancement (Classifier-Accuracy (%))
Nidhi Bhatla.(2012)	Predict heart diseases	Data Mining Techniques, Neural network algorithms	CNN-81.51
Boshra Baharami (2015)	Predict Heart diseases	Decision tree, K-NN, SMO	K-NN-83.732 SMO-80.21
K Cinetha (2014)	Prediction of Heart Disease	Fuzzy logic	97.67
Sadhu, A. and Jadli A.	Diabetes Prediction	Logistic regression, SVM, Naïve Bayes	Logistic regression-93% SVM-94 Naïve Bayes-91
Xue et al.	Prediction of Diabetes disease	Naïve bayes, SVM	Naïve bayes-93.27 SVM-96.54
Le At al.	Diabetes Prediction	Random forest classifier, Naïve bayes, K-NN	K-NN-96 Naïve bayes-93 RFC-96
Julius et al.	Diabetes Prediction	K-NN, RF	RF-97 K-NN-98
Khanam et al.	Diabetes Prediction	ANN, RFC	ANN-88.57 RFC-77.34
Shafi et al	Diabetes Prediction	SVM,DT	SVM-63 DT-72
Mrudula Gudadhe	Heart disease Prediction	SVM,ANN	SVM- 82 ANN-84

III. PROPOSED WORK

By addressing the gaps and issues identified in the literature review, a robust framework for disease prediction is to be developed, with an emphasis on heart disease and diabetes-related illnesses. This section covers the primary components of the study methodology, such as data collection, preprocessing, and the selection and training of deep learning models.

A carefully divided dataset will be used to train the prediction models, with the goal of striking a compromise between generalizability and model complexity. We will assess our models' resilience and prevent overfitting using strict validation techniques such as k-fold cross-validation.

Given the strong performance of deep learning models in disease prediction challenges, we plan to incorporate multiple topologies into our study. This includes: Convolutional neural networks, or CNNs, are used to assess data that is based on images. They are employed in medical imaging tasks such as identifying heart disease-related patterns in radiological images. Recurrent Neural Networks (RNNs): By identifying temporal connections in sequential health data from electronic health records, RNNs can assist predict illnesses like diabetes. Ensuring that our models can adjust to several data modalities is our top concern. We will look into using pre-trained models through transfer learning approaches to increase performance.

Due to the delicate nature of health-related data, ethical considerations will take precedence during the whole research process. Before using patient data, we will get the necessary authorizations and abide by data privacy regulations. In addition, we will make an effort to ensure that our models are transparent and understandable, which will allay concerns about deep learning algorithms being "black-box" in the healthcare industry.

To sum up, the proposed work offers a comprehensive method for deep learning and predictive analysis in illness prediction. By employing diverse datasets, advanced preprocessing techniques, and deep learning models, our goal is to expedite the creation of robust and interpretable predictive instruments for ailments associated with diabetes and cardiovascular disease.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

IV.CONCLUSION

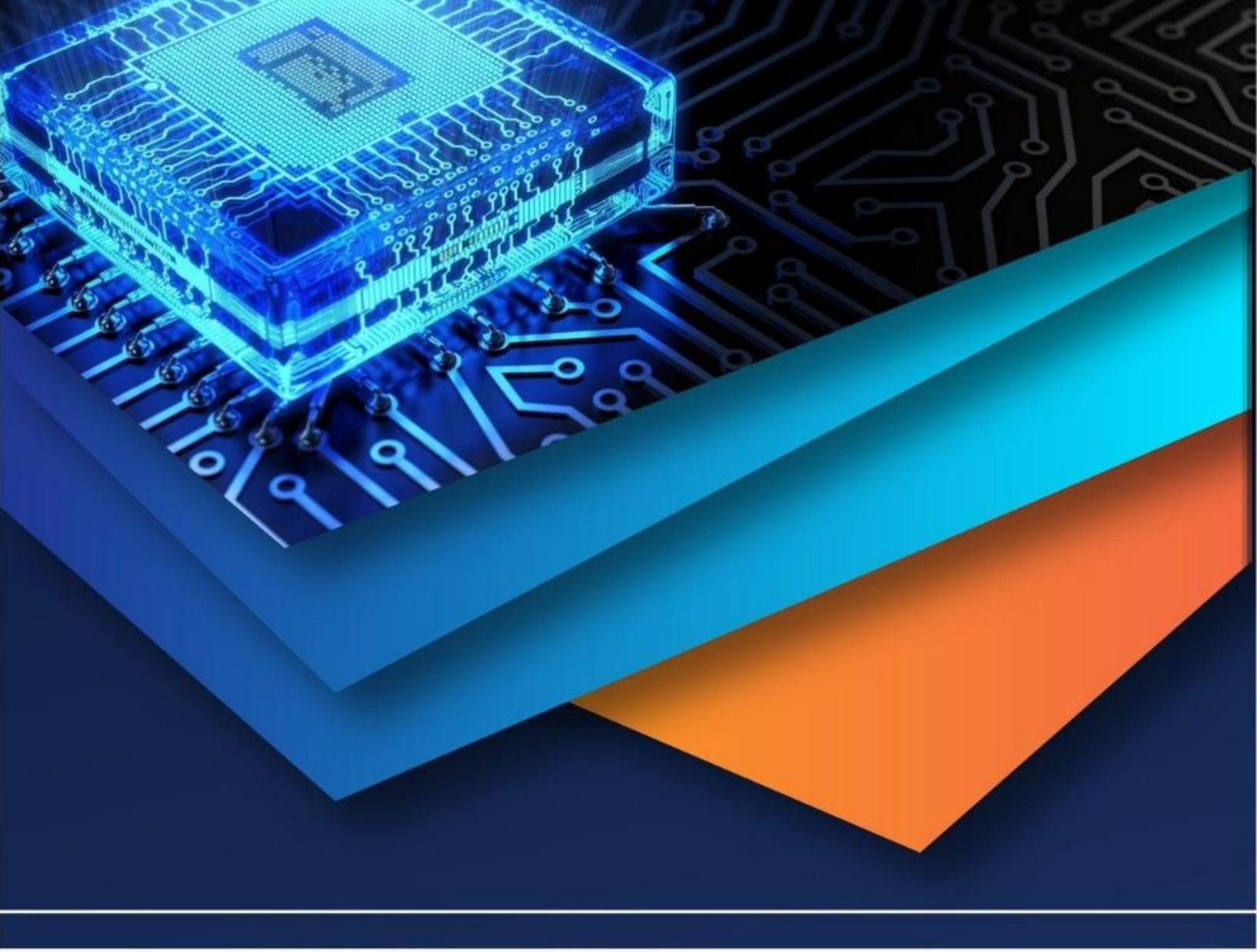
With a focus on diabetes-related disorders and heart disease in particular, we investigated disease prediction using deep learning and predictive analysis in this work. By means of a thorough review of the literature and the implementation of the suggested work, our objective was to make a contribution to the developing field of healthcare analytics. Several important outcomes from our research highlight the effectiveness of deep learning models in illness prediction. Through the use of a broad dataset that included medical imaging, electronic health records, and other modalities, our predictive models showed encouraging performance metrics. CNN is used for jobs involving images, and RNN is useful for capturing intricate relationships in health data.

Robust predictive tools were developed as a result of the rigorous feature selection, model training, and data preparation techniques used. Our models showed good precision, recall, accuracy, and AUC-ROC values, suggesting that they could be used for disease early detection.

REFERENCES

[1] Nidhi Bhatla, and Kiran Jyoti, Oct. 2012,"An Analysis of Heart Disease Prediction using Different Data Mining Techniques", International Journal of 40 Engineering Research & Technology (IJERT), Vol. 1, Issue 8, ISSN: 2278- 0181, pp. 1-4.
[2] Sairabi H. Mujawar, and P. R. Devale, October 2015,"Prediction of Heart Disease using Modified k-means and by using Naive Bayes", International Journal of Innovative Research in Computer and Communication Engineering(An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 10, pp. 10265-10273.
[3] Saito, K., Zhao, Y., & Zhong, J. (2019). Heart diseases image classification based on Convolutional Neural Network. 2019 International Conference on Computational Science and Computational Intelligence.
[4] Yiu, T. (2019). Understanding random forest - towardsdatascience.com. Retrieved October 29, 2022, from https://towardsdatascience.com/understanding-random-forest58381e0602d2.
[5] Zhang, G. (2018, November 11). What is the kernel trick? why is it important? Medium. Retrieved November 3, 2022, from https://medium.com/@zxr.nju/what-is-the-kernel-trick-why-is-it-important98a98db0961d.
[6] K.Sudhakar, and Dr. M. Manimekalai, January 2014, "Study of Heart Disease Prediction using Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 1,pp. 1157- 1160.
[7] Kamal Kant, and Dr. Kanwal Garg,2014, "Review of Heart Disease Prediction using Data Mining Classifications",International Journal for Scientific Research & Development(IJSRD), Vol. 2, Issue 04, ISSN (online): 2321- 0613, pp. 109-111.
[8] Kohli, S. (2019, November 18). Understanding a classification report for your machine learning model. Medium. Retrieved November 3, 2022, https://medium.com/@kohlishivam5522/understandingclassificationreport-for-your-machine-learning-model-88815e2ce397.
[9] Latha, C. B., & Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked, 16, 100203. https://doi.org/10.1016/j.imu.2019.100203.
[10] Xue J, Min F, Ma F. Research on diabetes prediction method based on machine learning. J Phys Conf Ser. 2020;1684:1–6. [Google Scholar]
[11] Le TM, Vo TM, Pham TN, Dao SV. A novel wrapper–based feature selection for early diabetes prediction enhanced with a metaheuristic. IEEE Access. 2020;9:7869–84. [Google Scholar]
[12] Wrapper–based feature selection for early diabetes prediction enhanced with a metaheuristic. IEEE Access. 2020;9:7869–84. [Google Scholar]
[13] Julius AO, Ayokunle AO, Ibrahim FO. Early diabetic risk prediction using machine learning classification techniques. Available from:https://ijisrt.com/early-diabetic-risk-prediction-using-machine-learning-class ification-techniques . [Google Scholar]
[14] Shafi S, Ansari GA. Early prediction of diabetes disease &classification of algorithms using machine learning approach. In Proceedings of the International Conference on Smart Data Intelligence (ICSMDI 2021) Available from:SSRN 3852590 (2021). [Google Scholar]
[15] Khanam JJ, Foo SY. A comparison of machine learning algorithms for diabetes prediction. ICT Express. 2021;7:432–9. [Google Scholar]
[16] Sisodia D, Sisodia DS. Prediction of diabetes using classification algorithms. Procedia Comput Sci. 2018;132:1578–85. [Google Scholar]
[17] Norris SL, Lau J, Smith SJ, Schmid CH, Engelgau MM. Self-management education for adults with type 2 diabetes:A meta-analysis of the effect on glycemic control. Diabetes Care. 2002;25:1159–71. [PubMed] [Google Scholar]
[18] Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract. 2010;87:4–14. [PubMed] [Google Scholar]
[19] Anjana RM, Pradeepa R, Deepa M, Datta M, Sudha V, Unnikrishnan R, et al. Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India:Phase I results of the Indian Council of Medical Research India Diabetes (ICMRINDIAB) study. Diabetologia. 2011;54:3022–7. [PubMed] [Google Scholar]
[20] Ramachandran A, Snehalatha C, Salini J, Vijay V. Use of glimepiride and insulin sensitizers in the treatment of type 2 diabetes-a study in Indians. J Assoc Physicians India. 2004;52:459–63. [PubMed] [Google Scholar]
[21] Wagai GA, Romshoo GJ. Adiposity contributes to poor glycemic control in people with diabetes mellitus, a randomized case study, in South Kashmir, India. J Family Med Prim Care. 2020:4623–6. [PMC free article] [PubMed] [Google Scholar]
[22] Mujumdar A, Vaidehi V. Diabetes prediction using machine learning algorithms. Procedia Comput Sci. 2019;165:292–9. [Google Scholar]
[23] Diabetes mellitus affected patients classification and diagnosis through machine learning techniques. Procedia Comput Sci. 2017;112:2519–28. [Google Scholar]
[24] Rawat V, Suryakant S. A classification system for diabetic patients with machine learning techniques. Int J Math Eng Manag Sci. 2019;4:729–44. [Google Scholar]
[25] Perveen S, Shahbaz M, Guergachi A, Keshavjee K. Performance analysis of data mining classification techniques to predict diabetes. Procedia Comput Sci. 2016;82:115–21. [Google Scholar]

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)