Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

Health Screening Analysis Using Machine Learning

Raju Sharma, Parul Sharma, Shivangi Pandit, Punit Mittal

Department of computer science and information technology.

Meerut Institute of Engineering and Technology

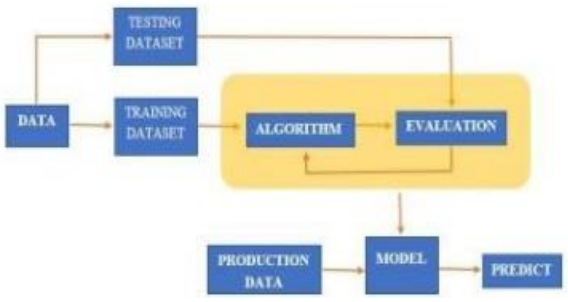
Abstract:- Diabetes Mellitus is a disease in which blood sugar levels rise due to the body's inability or inability to produce insulin. Factors such as age, sedentary lifestyle, genetic predisposition, poor nutrition and high blood pressure have caused the epidemic to become more common in affected people. The healthcare industry manages large data warehouses, making big data analytics an important tool for uncovering insights, identifying patterns, relationships, and predictive knowledge. This article introduces a diabetes testing model that uses machine learning algorithms to improve accurate classification. Explore various algorithms to determine the best classification model. The proposed model aims to facilitate the provision of medical services through early detection of individuals at risk. This study highlights the importance of advanced clinical screening to improve decision-making processes and improve patient outcomes.

Keywords*:* diabetes, medicine, machine learning, data analysis.

1. Introduction

Technological developments in the last decade have affected every aspect of our lives, including the healthcare sector. Diabetes Mellitus is an endocrine disease caused by metabolic diseases and long-term problems affecting the body such as the eyes, kidneys, nerves and nerves. People with diabetes are at high risk for other health problems, such as heart disease, kidney problems, eye diseases and nerve damage. Uncontrolled blood sugar can also cause poor circulation, affect the delivery of nutrients to the wound, and delay healing. There are two types of diabetes: type I and type II. Type I diabetes, [1] formerly known as juvenile diabetes, is often diagnosed in children and adolescents. Also known as insulin-dependent diabetes (IDDM),) is the most common form and is characterized by hyperglycemia, insulin resistance, and relative insulin deficiency. This type occurs when the body cannot use insulin effectively. Forecasting involves using multiple machine learning algorithms, data mining techniques, and statistical analysis to use current and historical data to gain insight and predict future events. Machine learning is a branch of artificial intelligence (AI) that automates and enhances computer-based learning. This analysis highlights the importance of health screening, its impact on public health.This review focuses on the importance of health management through routine screening to improve the health of individuals and

communities. For this purpose it provides an overview. [2] Regular health checks have many benefits and improve overall health. Initially, immediate recognition of health problems allowed for effective intervention and management, often stopping the spread of disease. This strategy improves health outcomes as well as reduces overall health care costs. Health screenings also provide people with important information about their health, allowing them to make lifestyle decisions and reduce risk. It encourages people to take personal responsibility for their own health, thus promoting a culture of health protection. Additionally, the analysis results in the effectiveness of public health measures through early detection of health hazards, ultimately improving population events. More importantly, health screening is beneficial to both people and society.[3] As the data is prepared, various machine learning algorithms will be examined and compared to determine the best model for predicting diabetes and heart disease. Analysis algorithms include decision trees, random forests, support vector machines, neural networks, and ensemble learning. An interpretive model will also be considered to The project will ultimately create a powerful predictive model that can accurately measure a person's risk of diabetes and heart disease. The performance and interpretation of the model will be analyzed comprehensively and validly using independent data or through competition. The findings of the project have the potential to have significant impacts on healthcare and public health strategies. Early identification of groups at higher risk for diabetes and heart disease can facilitate timely intervention, [4]lifestyle changes and personalized care plans. Additionally, understanding the program can guide physicians in allocating resources, disease control, and preventive measures. In summary, the project aims to develop predictive models for diabetes and heart diseases using machine learning algorithms. The program focuses on drug prevention efforts and improving health outcomes for individuals at risk for chronic disease through the use of primary data and advanced analytical methods.



Literature Review

Health screening is an important part of maintaining health and is designed to identify diseases or risk factors before the patient becomes ill. Many studies have investigated the effectiveness of screening for a variety of conditions, including cancer, heart disease, diabetes, and infectious diseases. One of the main tasks of research in health screening is to test the sensitivity and specificity of screening tests. Sensitivity refers to the test's ability to accurately identify individuals with the disease,[22] while specificity refers to its ability to identify individuals

without the disease. Measuring sensitivity and specificity is important to reduce false positives and negatives, which can lead to unnecessary interventions or missed diagnoses, respectively. In addition, the study evaluated the overall impact of the health screening program in reducing morbidity and mortality rates. with different diseases. Cost-effectiveness analysis helps policymakers allocate health care resources less efficiently and prioritize screening programs that provide the greatest health benefits at a reasonable price.

.Diabetes prediction

Information on[19] diabetes health screening is comprehensive and covers all aspects important for effective disease management and prevention. Most diabetes screenings aim to identify people with type 2 diabetes or undiagnosed diabetes to initiate early intervention and prevent diabetes related complications. [3] Research in this area examines the effectiveness of screening, the impact of screening on health outcomes, ethical considerations, and costs. This study evaluated the characteristics of different diabetes tests, including fasting blood glucose, oral glucose tolerance test, and glycosylated hemoglobin. These tests are designed to determine the sensitivity, specificity, and reliability of each test to accurately identify people who have or are at risk of diabetes.There is evidence that screening combined with treatment strategies can improve glycemic control, reduce the incidence of diabetes complications, and improve the quality of life of people with diabetes. [7] Decision making plays an important role in diabetes screening, especially on issues such as informed consent, patient confidentiality and equality in examination. Researchers emphasize the importance of providing individuals with clear information about the purpose, risks, and benefits of screening to facilitate decisionmaking. Additionally, efforts to ensure equitable access to care, especially among the underserved, are critical to addressing disparities in the many benefits of diabetes. Intelligent Traffic [24]Management Systems , The transition to more intelligent traffic management systems began with the integration of sensor technologies and real-time data analysis. Researchers and practitioners explored adaptive traffic signal control systems that rely on data collected from cameras, vehicle detectors, and other sources. While these early efforts marked progress, they often struggled to adapt adequately to the intricacies of urban traffic.

.Heart disease prediction

The literature on cardiovascular disease screening includes many studies designed to identify individuals at risk for cardiovascular disease, such as heart disease, heart attack, and stroke. The basis for clinical evaluation data is its effectiveness in detecting early signs of heart disease and its effects on health. The study investigated a variety of tests, including electrocardiogram (ECG),[20] stress test, echocardiography, and blood calcium scores, to determine sensitivity, specificity, and predictors in identifying individuals with cardiovascular disease.Ethical judgment is important in cardiovascular screening, especially regarding patient privacy, consent and test equity. [9] Individuals should be provided with accurate information about the purpose, benefits and risks of screening to

facilitate shared decisionmaking and support patient care, the researchers said.Cost effectiveness analyzes play an important role in the cardiovascular disease evaluation literature as they inform resource allocation decisions and guide drug use. Economic evaluation of screening includes evaluation of costs associated with screening, [18]diagnosis, and subsequent treatment, as well as savings from preventing cardiovascular disease and reducing healthcare utilization. In summary, information on health screening for heart disease highlights the importance of early diagnosis, risk assessment, and the ethics of screening programs to reduce the burden of cardiovascular disease and improve patient outcomes. Continued research and collaboration among medical professionals, policymakers, and community stakeholders are essential to improve screening strategies, address emerging issues, and improve cardiovascular health across many cultures.

2. Methodology

3.1. Working

Health screening activities focused on heart disease and diabetes typically involve several things:

Data collection: Collect relevant information from individuals. Health screening, including demographic information, medical history, lifestyle (including diet and exercise), and health indicators of heart disease and diabetes (such as high blood pressure, diabetes, cholesterol levels).

Risk Assessment: Use risk assessment tools and algorithms to evaluate a person's risk of heart disease and diabetes based on the data they collect. [23]These tools may include scores such as the Framingham Heart Disease Risk Score or the American Diabetes Association Diabetes Risk Assessment Tool.

Data Analysis: Analyze collected data to identify patterns, associations, and risk factors for heart disease and diabetes. This may include statistical analysis, machine learning algorithms, or other data analysis techniques to uncover insights from the data.

Predictive Modeling: Develop predictive models to estimate the incidence of heart disease or diabetes and other health indicators based on individual risk factors.[16] Advice and interventions: Provide advice and interventions based on individual risk and health. This may include lifestyle changes (such as diet, exercise), medication management, and referral to a specialist for further evaluation and treatment. Track and trace: Continuous monitoring for tracking and follow-up.

Overall, the goal of the Health Research Clear is to identify people at risk of heart disease and diabetes early and proactively. Prevents heart diseases and diabetes. Prevent or manage these conditions and ultimately improve health outcomes and quality of life

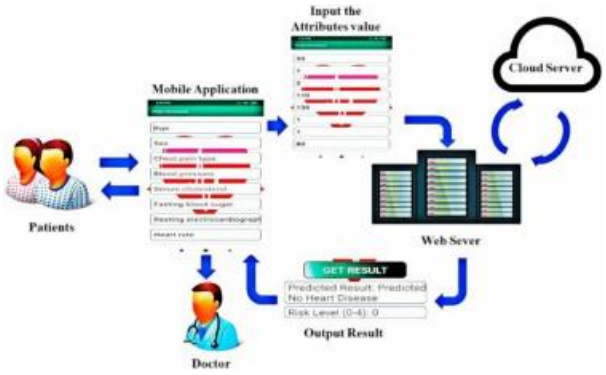


Fig2: - Architecture Diagram

3.2. Algorithm

In this research endeavor, a diverse array of machine learning algorithms was employed, accompanied by various data analysis techniques to assess the dataset's balance and structural characteristics concerning click predictions.

3.2.1 Logistic regression:

Logistic regression is a widely accepted method in classification and plays an important role in this study. As a supervised learning algorithm, logistic regression uses techniques to predict binary outcomes (such as “yes” or “no,” “true” or “false,” “0” or “1”). [11]This model establishes a relationship between inputs and binary outcomes by representing the logistic curve as a sigmoid function. This sigmoid function maps the input value to a range between 0 and 1, specifies the probability of the binary result being "1" given the given property. To make predictions, the logistic regression model uses a logistic curve to calculate the probability of a new binary entry being a "1". If the probability is greater than the threshold (usually 0.5), the model predicts “1”; otherwise it predicts "0". Logistic regression is valued for its simplicity and efficiency in classifying tasks and is widely used in many areas such as predicting customer churn, testing, and spam detection. $\sigma(x) = 1+e^{-x}$

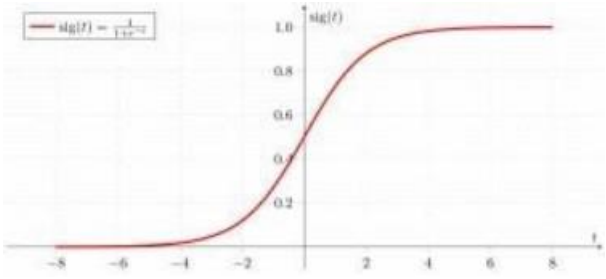


Fig3: Logistic Regression Segmoid Function

3.2.2 VM:

Support Vector Machine (SVM) has become one of the most important supervised learning algorithms known for its strong ability in classification and text classification . Although widely used, SVM is mainly used to solve

classification problems in machine learning. [4]The purpose of SVM revolves around creating a good line or decision boundary that can divide the space into different classes, thus facilitating a new classification of data for the future. This well defined boundary is called a hyperplane and is created by SVM by selecting point clouds or vectors. In health screening,[9] SVM can be used to classify patients into different risk groups based on characteristics obtained from medical examinations, image data, or a patient letter. By accurately classifying patients, SVM can help doctors make decisions about diagnosis, treatment planning, and prevention.

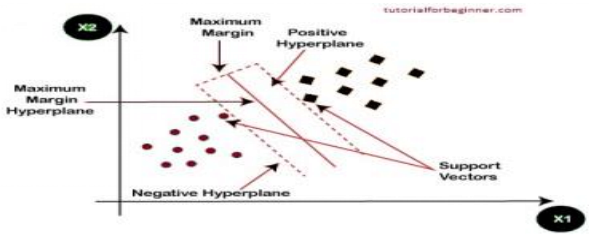


Fig4: Support Vector Machine(SVM) Classification

There are many lines or arbitrary boundaries that can separate groups in N dimensional space. But the task is to determine the best decision for the distribution of data points. This consensus boundary is represented as a large plane in the context of Support Vector Machine (SVM).

Support vectors:

These are the data points closest to the plane in the support vector machine (SVM). They play an important role in determining the position and orientation of the hyperplane, influencing the edges, that is, the distance from the hyperplane and the closest data to each class. The SVM algorithm optimizes the hyperplane based on these underlying supports to ensure class separation at a particular location.

3.2.3 Decision trees:

Decision trees stand out as various training approaches suitable for both classification and regression problems, especially focusing on the solution of classification problems. As a distributed system with a tree model, its internal nodes define the properties of the dataset, the decision rules are influenced by the tree branches, and each leaf of the leaf represents the trunk for the difference. The decision process has many branches that help in decision making; The leaf represents the final result. [17] Make decisions or run tests based on dataset characteristics, Create a tree-like representation. The decision tree process begins with a root node and branches to create a tree structure. Creation of decision trees, including the use of the CART algorithm

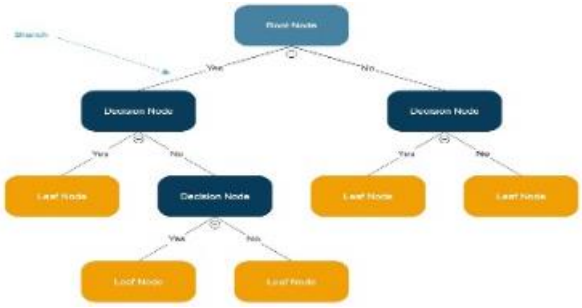


Fig:5 Decision Tree

3.2.3 Random Forest:

Random Forest is a combination of machine learning algorithms designed for classification and regression techniques involving multiple decision trees. Each tree is trained by selecting from data and its predictions are combined to create the final prediction. While aggregation in classification is done by majority vote, in regression it is based on average estimation. It is worth noting that one of the advantages of random forest is good training of big data and good data management, which has a lot of power.

4. Result

The effectiveness of machine learning algorithms in predicting diabetes depends on many factors, including the data used, specific applications, previously performed procedures, and metrics used to evaluate results. The performance of diabetes models is often measured by metrics such as accuracy, sensitivity, specificity, and area under the receiver operating curve (AUC-ROC).

Accuracy measures the overall accuracy of the prediction, while sensitivity and specificity measure the model's ability to detect positive and negative events, respectively. [21]The results reveal important information about the participants' overall health, highlighting risk factors such as hypertension, obesity, dyslipidemia and mental illness.

After this, a diagnostic procedure was developed that included physical examination, biochemical examination and health questionnaire.[5] Participants underwent a series of tests to evaluate their cardiovascular health, including blood pressure measurement, body mass index, lipid profile, blood sugar test and nutritional assessment.

A psychological assessment was also conducted through a standardized questionnaire to assess psychological well-being and stress levels. Carefully collect and analyze collected data using statistical software to identify relationships and trends in the data.

Accuracy table:

ALGORITHM	ACCURACY
• Random Forest	90 %
• Logistic Regression	73 %
• XGBoost	88 %
• SVM	74 %
• KNN	89 %

5. Conclusion:

This research explores the application of different machine learning algorithms to datasets, with a specific focus on classification tasks. After a rigorous analysis, we found random forest to be the most accurate algorithm with90% classification accuracy. This study includes a comprehensive comparison of the accuracy of machine learning algorithms, emphasizing the evaluation of confusion matrices to reduce false positives

The surprising accuracy of the random forest algorithm demonstrates its ability to disperse good events in the data. The importance of reducing negative outcomes in the confusion matrix is based on the urgent need to reduce the risk associated with misclassification of people who may have diabetes.

As an avenue for future research, there is a good way to expand the research to determine whether people without diabetes will develop diabetes next year. This predictive model can provide insight into early intervention and prevention strategies.

By exploring the predictive power of machine learning models in predicting future diabetic patients, this research may contribute to diabetes prevention.

In summary, this study not only provides insight into the comparison of machine learning algorithms, it also lays the foundation for future research on diabetes management and prevention. Findings highlight the importance of continuing research on the use of machine learning for healthcare analytics.

3. References

[1] Mark A Atkinson, PhD, Prof, George S Eisenbarth, MD, Prof, and Aaron W Michels, MD ‘Type 1 diabetes’.

[2] Abdulfatai B.Olokoba,1,* Olusegun A. Obateru,2 and Lateefat B. Olokoba3 ‘Type 2 Diabetes Mellitus: A Review of Current Trends.

[3] Rabina , Er. Anshu Chopra ‘DIABETES PREDICTION BY SUPERVISED AND UNSUPERVISED LEARNING WITH

SELECTION’.

[4] Santi Wulan Purnami, Jasni Mohamad Zain For medical diagnosis using a new smooth SVM", Communications in Computer and Information Science, Vol 88, Part 1, pp.15- 27,2010.

[5] Aiswarya Iyer, S. Jeyalatha and Ronak Sumbaly, "Diagnosis of Diabetes Using Classification Mining Techniques", Knowledge International Journal of Data Mining & Management Process (IJDKP) Vol.5, No.1, January 20Humar Research 931 - 932,

[6] Abdullah, R.M., Tarik A., Rashid, S.M.A., Abstract, 2016. "An Intelligent Approach for Diabetes Classification, Prediction and Description". Advances in Intelligent Systems And Computing. [7] B.M. Patil, R.C. Joshi and Durga Toshniwal, "Association Rule for Classification of Type-2 Diabetic Patients", ICMLC '10 Proceedings of the 2010 Second International Conference on Machine Learning and Computing, February 09 -11, 2010

[8] Nai-Arun, N., Sittidech, P., 2014. "Ensemble Learning Model for Diabetes Classification". Advanced Materials Research 931 - 932,

[9] Humar Kahramanli and Novruz Allahverdi "Design of a Hybrid System for the Diabetes and Heart Disease", Expert Systems with Applications: An International Journal, Volume 35 Issue 1-2, July, 2008.

[10] American Diabetes Association. "Standards Of Medical Care in Diabetes—2020 Abridged for Primary Care Providers." Diabetes Care, Vol. 43, Issue 6, June 2020.

[11] World Health Organization. "Global Report on Diabetes." Geneva, Switzerland, 2016.

[12] Li, Rui et al. "Prediction of Type 2 Diabetes Mellitus With EHR Data Using a Machine Learning Approach." Computational and Methods in Medicine, Volume 2015.

[13] Duffy, David et al. "Machine learning to predict the occurrence of diabetes complications."Journal of Diabetes and its Complications, Vol. 30, Issue 1, January 2016.

[14] Jia, Lei et al. "Machine Learning in Medicine: A Practical Introduction." The Lancet Digital Health, Vol. 2, Issue 8, August 2020.

[15] Lee, Joon et al. "Application of Machine Learning to Predict the Development of Type 2 Diabetes: A Systematic Review." BMC Medical Informatics and Decision Making, Vol. 19, Issue1, January 2019.

[16] Kavakiotis, Ioannis et al. "Machine Learning and Data Mining Methods in Diabetes Research." Computational and Structural Biotechnology Journal, Vol. 15, August 2017.

[17] Anjana, Ranjit Mohan et al. "Machine Learning Predictions of Diabetes Complications." BMC Medical Informaticsand Decision Making, Vol. 19, Issue 1, January 2019.

[18] Cho, Yangseok et al. "Application of Machine Learning Approaches in the Prediction of Metabolic Syndrome." International Journal of Medical Informatics, Vol. 118, January 2018.

[19] Géron, Aurélien. "Hands-On Machine Learning with ScikitLearn, Keras, and TensorFlow." O'Reilly Media, 2019.

[20] Gandomi, Amir et al. "Deep Learning Applications in Civil Engineering." Archives Of Computational Methods in Engineering, Vol. 25, Issue 1, January 2018.

[21] Krawczyk, Bartosz. "Learning from Imbalanced Data: Open Challenges and Future Directions." Progress in Artificial Intelligence, Vol. 5, Issue 4, December 2016.

[22] K. Rajesh and V. Sangeetha, “Application of Data Mining Methods and Techniques for Diabetes Diagnosis”, International Journal Of Engineering and Innovative Technology Volume 2, Issue 3, September 2012.

[23] Zhang, Wei et al. "A Machine Learning Framework for Predicting Type 2 Diabetes Mellitus." IEEE Access, Vol. 9, February 2021.

[24] Das, Suman et al. "Prediction of Type 2 Diabetes Using Ensemble Learning Techniques." Informatics in Medicine Unlocked, Vol. 20, March 2020.

[25] Miftahutdinov, Zulfat et al. "Machine Learning Approaches to Predicting the Risk Of Type 2 Diabetes Mellitus." Journal of Healthcare Engineering, Volume 2020.