e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707051 |

An Automation for Diabetes and Reverse Diabetes Prediction using Efficient Ml Algorithms

Ramya Shree R[1], Bhavani R[2]

PG Student, Department of Computer Applications, The National Institute of Engineering, Mysuru, India[1]

Assistant Professor, Department of Computer Applications, The National Institute of Engineering, Mysuru, India[2]

ABSTRACT: Diabetes, a chronic metabolic disorder, causes sustained high blood glucose due to insufficient insulin production or cellular resistance. This project's aim is to address the gap in real-time applications for early diabetes prediction and dietary recommendations. We're developing an app that predicts diabetes early, identifies potential reversals using machine learning algorithms like Random Forest, KNN, or Decision Tree, and provides personalized dietary plans. Built with Microsoft tools such as Visual Studio and SQL Server, this real-time medical system will be a valuable asset for hospitals and doctors.

KEYWORDS: Model, Random Forest, K-Nearest Neighbour, Diabetes Mellitus, Machine Learning

I.INTRODUCTION

Diabetes mellitus, a leading cause of mortality, encompasses a group of metabolic disorders affecting millions worldwide. Early detection is crucial due to the severe complications associated with the condition. Numerous studies, often utilizing the Pima Indian diabetes dataset, have explored diabetes identification, typically focusing on complex techniques without thoroughly comparing common methods. Characterized by elevated blood sugar levels, diabetes leads to symptoms such as increased urination, thirst, hunger, and weight loss. Without ongoing treatment, it can result in life-threatening complications.

Diagnosis generally involves measuring a 2-hour post-load plasma glucose level of at least 200 mg/dL, underscoring the importance of timely identification to prevent serious health outcomes. Automation for diabetes prediction using efficient machine learning algorithms is essential. This real-time application would significantly assist hospitals and healthcare providers in managing patients more effectively. Our proposed system aims to enhance disease prediction processes, enabling healthcare professionals to deliver superior patient care.

We intend to develop a real-time application for early diabetes detection and management using machine learning algorithms such as Random Forest, K-Nearest Neighbour (KNN), and Decision Tree. By leveraging these algorithms, our system will predict diabetes early, identify potential reversals, and provide personalized dietary plans. This will address the current gap in real-time applications for early prediction and dietary recommendations, providing a valuable asset for hospitals and doctors. Built with Microsoft tools like Visual Studio and SQL Server, this system will streamline diabetes management and improve patient outcomes by facilitating timely and accurate diagnosis and treatment.

II.LITERATURE REVIEW

[1] "Diabetes Prediction Using Machine Learning Model", Viswanatha,Ramachandra, Dhanush Murthy, Thanishka

The paper "Diabetes Prediction Using Machine Learning Model" explores predictive models for early diabetes detection. Authors analyze datasets from PIMA Indians and rural African Americans, primarily using logistic regression. Techniques like feature selection and ensemble methods improve accuracy, peaking at 78% and 93% for datasets 1 and 2, respectively.

[2] "Research on Diabetes Prediction Method Based on MachineLearning",Jingyu Xue, Fanchao Min, Fengying Ma

Diabetes mellitus (DM) is a metabolic disease marked by elevated blood sugar levels. Types include type 1 and type 2, with a rising incidence of type 1 among young people. Early detection is crucial due to its chronic nature and potential complications. This study employs supervised machine learning, particularly Support Vector Machine (SVM), for

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707051 |

prediction.

[3] "Machine learning for diabetes clinical decision support:a review",AshwiniTuppad, ShantalaDevi Patil

Type 2 diabetes is now considered an epidemic, despite being non-communicable, due to its widespread prevalence and complex nature. Lifestyle factors play a significant role, with machine learning (ML) increasingly used to alleviate its burden. This paper reviews ML applications in diabetes prevention and management, covering risk assessment, diagnosis, and prognosis. It identifies gaps in current ML methodologies and highlights the need for future research to addressthese shortcomings.

[4] "Clinical Decision Support System for Diabetic Patients by Predicting Type 2 Diabetes Using Machine Learning Algorithms", Rakibul Islam, Azrin Sultana, Md.Nuruzzaman Tuhin, Md.Sazzad Hossain Saikat, Mohammad Rashedul Islam.

This study employs various ML algorithms on the Pima Indian Diabetes dataset to predict type 2 diabetes, achieving over 90% accuracy with decision tree and histogram-based gradient boosting. A clinical decision support system (CDSS) is implemented, aiding physicians and patients in real-time diagnosis and providing analytical suggestions for improved medical care.

[5] "Pima Indians diabetes mellitus classification based on machinelearning (ML) algorithms", Victor Chang, Jozeene Bailey, Qianwen Arie, Zhili Sun

The paper proposes an e-diagnosis system for type 2 diabetes using interpretable ML models on IoMT. Three models—Naïve Bayes, random forest, and J48 decision tree—are trained andtested on the Pima Indians dataset in R. Naïve Bayes performs well with refined feature selection for binary classification, while random forest excels with more features.

Literature Summary/Gap Analysis

Many research efforts have focused on predicting diabetes using efficient machine learning algorithms, with some presenting concepts and others implementing solutions. Several papers have used tools like Python, R, and Weka, often relying on static datasets without real-time capabilities or reverse diabetes prediction. Typically, patients undergo manual diagnoses by doctors, a process requiring significant expertise, time, equipment, and expense.

Our proposed system offers a novel approach by predicting and reversing diabetes using machine learning algorithms with dynamic datasets. This GUI-based application for hospitals simplifies usage for doctors and patients, a unique development in the field. It leverages machine learning models for real-time predictions, a feature not previously realized. With a simple button click, the system predicts both diabetes and reverse diabetes, utilizing a broader range of parameters and larger datasets to enhance accuracy and reliability.

III.METHODOLOGY

Medical datasets containing disease information are processed using machine learning algorithms, particularly supervised learning algorithms, to analyse these datasets. The goal is to achieve high accuracy and efficiency. The application of these sophisticated algorithms results in enhanced outcomes, improving the overall effectiveness of disease prediction and management.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707051 |

Diabetes and Reverse Diabetes Prediction Process

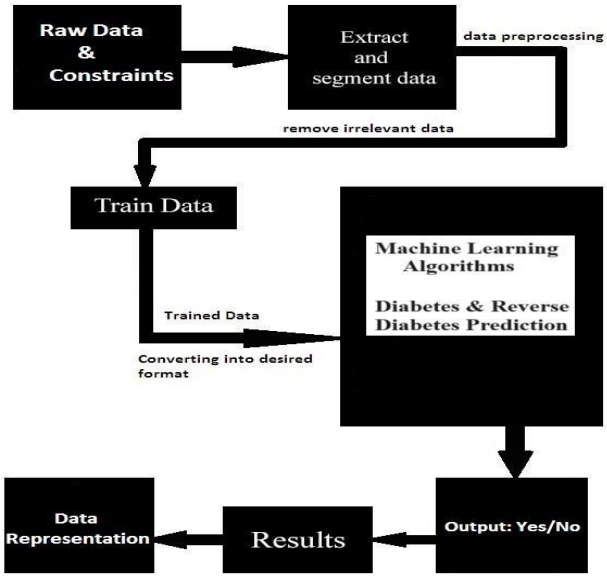

Fig 1. Diabetes and Reverse Diabetes Prediction Process

Step 1: Initially training datasets required to process and predict the results are collected from many online sources such as kaggle.com, dataworl.com, data.gov.in, githib.com etc.… Raw data collected from these sources.

Step 2: Once datasets collected, we need to understand the data and preprocessing of raw data is required. In data preprocessing, we fix missing values and removes irrelevant data and extracts the relevant data. Some unwanted data such as serial no, patient Id, mobile, address etc.… will be removed and only data required to process will be fetched.

Step 3: After data preprocessing, the preprocessed data is now inputted to machine learning algorithms. We use supervised learning algorithms to build machine learning models to predict diabetes disease and reverse diabetes disease.

Step 4: Algorithms such as KNN algorithm or Decision tree algorithm and Naïve bayes algorithmused to train the model and diabetes and reverse diabetes disease prediction is done.

Step 5: First machine model predicts patient diabetes disease (YES or NO), then if patient is classified to YES, then reverse diabetes will be predicted using machine learning model.

Step 6: To test the machine learning model, we divide the training datasets into training and testing in the ratio 90:10 and model evaluation is done and accuracy is calculated.

Step 7: Final outputs displayed on GUI. Data visualization is done.

Step 8: Proper treatments will be given to the patients to cure the disease at earliest.

Operation of K-Nearest Neighbors (KNN)

The working of the K-Nearest Neighbors algorithm is as follows:

1. A random seed is chosen to facilitate the selection of a subset of data points from the training dataset, maintaining the original distribution of classes. This sampling process ensures that the dataset remains representative while reducing computational load.

2. In a dataset with M input attributes, the KNN algorithm considers all M attributes to compute distances between data points. Unlike random forest, KNN does not select a random subset of attributes but instead uses all available input variables to ensure comprehensive distance measurement.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707051 |

3. For each data point, the KNN algorithm identifies the k-nearest neighbors based on a distance metric such as Euclidean distance. The class of the new data point is then determined by majority vote from its k-nearest neighbors. This process is repeated for each data point to classify the entire dataset.

Operation of Random Forest

The working of random forest algorithm is as follows.

1.A random seed is chosen to facilitate the random sampling of a subset of data points from the training dataset. This ensures that the original distribution of classes within the dataset is maintained during the sampling process. From this selected dataset, a random subset of attributes is chosen based on user-defined criteria, not all input variables are considered to avoid excessive computation and reduce the risk of overfitting.

2.In a dataset with M input attributes, a random selection process is employed to choose R attributes for each decision tree, where R is a value less than M.

3.Utilizing this subset of attributes, the decision tree model is constructed by iteratively identifying the best split at each and every node based on the Gini index value. This method or recursive process continues until a stopping criterion is reached, indicating that the resulting leaf nodes are too small to further subdivide.

IV.RESULTS

KNN Algorithm Results

We have developed a real-time application aimed at benefiting society, leveraging Microsoft technologies. Our project focuses on utilizing diabetes datasets trained using the K-nearest neighbour (KNN) algorithm, yielding highly promising results. Our KNN algorithm implementation is designed to accommodate dynamic datasets efficiently. With our proprietary KNN library, we have achieved an impressive accuracy rate of approximately 92.2%. Moreover, our prediction process operates swiftly, typically completing within 1500 milliseconds.

Constraint	KNN Algorithm
Accuracy	92.2 %
Time (milli secs)	1606
Correctly Classified (precision)	92.2 %
Incorrectly Classified (Recall)	7.8 %

Fig.2 KNN algorithm

RF Algorithm Results

Constraint	RF Algorithm
Accuracy	91.18 %
Time (milli secs)	2606
Correctly Classified (precision)	91.18 %
Incorrectly Classified (Recall)	8.82 %

Fig.3 RF Algorithm Result

V. CONCLUSION

Diabetes stands as a major contributor to global mortality rates. Detecting this condition early can significantly impact outcomes, prompting the building of machine learning models. Our system is dedicated to pinpointing and potentially reversing diabetes by analysing specific parameters. It aids healthcare professionals in forecasting diabetes during its nascent stages, facilitating prompt and tailored interventions. Leveraging a variety of ML methods enhances the accuracy of predictions. Through automation, our system employs efficient data science or machine learning algorithms, notably utilizing the effective supervised learning technique Naive Bayes. This approach efficiently processes medical data to produce predictive insights.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707051 |

REFERENCES

1. "Diabetes Prediction Using Machine Learning Model", Viswanatha, Ramachandra, Dhanush Murthy, Thanishka
2. "Research on Diabetes Prediction Method Based on Machine Learning", Jingyu Xue, Fanchao Min, Fengying Ma
3. "Machine learning for diabetes clinical decision support:a review", Ashwini Tuppad, ShantalaDevi Patil
4. "Clinical Decision Support System for Diabetic Patients by Predicting Type 2 Diabetes Using Machine Learning Algorithms", Rakibul Islam, Azrin Sultana, Md.Nuruzzaman Tuhin, Md.Sazzad Hossain Saikat, Mohammad Rashedul Islam.
5. "Pima Indians diabetes mellitus classification based on machinelearning (ML) algorithms", Victor Chang, Jozeene Bailey, Qianwen Arie, Zhili Sun.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com