IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

An Expert System for Insulin Dosage Prediction

Smt. A. Neethika[1], K. V V S L Varshitha[2], L. Naga Jahnavi[3], N. Mounika[4], N. Madhuri Sowjanya[5]
Assistant Professor, Department of Information Technology [1]
U. G. Students, Department of Information Technology [2,3,4,5]
S.R.K.R. Engineering College, Bhimavaram, Andhra Pradesh, India

Abstract*: Diabetes Mellitus is a pervasive metabolic disorder demanding precise blood glucose regulation to mitigate complications. Despite lifestyle adjustments, insulin dosage optimization remains pivotal for effective management. Our expert system employs Gradient Boosting Classifier for diabetes prediction and Logistic Regression for insulin dosage estimation, integrating patient data to automate decision-making. This approach facilitates proactive and personalized care, enhancing outcomes and representing a notable advancement in diabetes management.*

By leveraging machine learning techniques, this system accurately identifies diabetic patients and tailors insulin dosage recommendations based on individual characteristics and treatment history. Through automation, it streamlines decision-making processes, alleviating healthcare provider burden and empowering patients to actively participate in their care. Ultimately, our system promises to improve treatment efficacy and patient quality of life, heralding a new era in diabetes management..

Keywords: Machine Learning, Insulin Dosage, Gradient Boosting, Linear Regression

I. INTRODUCTION

The accurate prediction of glucose concentrations is crucial for timely and appropriate patient interventions, especially in critical situations such as hypoglycaemia. Recent studies have increasingly focused on employing advanced data-driven techniques to develop predictive models of glucose metabolism. This is imperative due to the complex and patient-specific nature of the relationship between input variables (e.g., medication, diet, physical activity, and stress) and glucose levels, which exhibit nonlinearity, dynamics, and interactivity. To address these challenges, non-linear regression models such as artificial neural networks, support vector regression, and Gaussian processes have been explored.

As the prevalence of diabetes continues to rise with improving living standards, there is a growing need to diagnose and analyse the condition quickly and accurately. In medical practice, diabetes diagnosis relies on parameters such as fasting blood glucose, glucose tolerance, and random blood glucose levels. Early diagnosis facilitates better control of the condition. Machine learning techniques offer promising avenues for aiding in the preliminary assessment of diabetes mellitus based on daily physical examination data, serving as valuable reference tools for healthcare professionals. However, the key challenges lie in feature selection and classifier selection within machine learning methodologies.

Understanding and addressing these challenges are critical for advancing the effectiveness of predictive models in diabetes management. This study aims to explore the application of advanced data-driven techniques in predicting glucose concentrations and diagnosing diabetes, with a focus on feature selection and classifier selection within machine learning frameworks. Through this exploration, we seek to contribute to improved patient outcomes and more efficient healthcare delivery in the realm of diabetes management.

1. Scope

The aim of this project is to develop a predictive model for diabetes management, specifically focusing on predicting glucose metabolism. This predictive model aims to assist patients in self-managing their disease by providing accurate and timely predictions of glucose levels. By incorporating data from various sources and leveraging predictive modeling techniques, the system aims to improve the patient's ability to monitor and control their glucose levels

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

effectively. This component of the diabetes management system is crucial for enabling proactive decision-making and intervention, ultimately leading to better health outcomes for patients with diabetes.

2. Objective

This project aims to develop a predictive system for diabetes diagnosis and insulin dosage prediction in diabetic patients. The primary objective involves utilizing two machine learning algorithms: Gradient Boosting Classifier for predicting diabetes and Logistic Regression for predicting insulin dosage in detected diabetic patients. The project utilizes two datasets: the PIMA diabetes dataset for diabetes prediction and the UCI insulin dosage dataset for insulin dosage prediction.

To implement the project, the datasets are first used to train both algorithms. The PIMA diabetes dataset contains various features such as glucose levels, blood pressure, and body mass index (BMI), which are utilized to train the Gradient Boosting Classifier to predict the presence of diabetes. Simultaneously, the UCI insulin dosage dataset is used to train the Logistic Regression algorithm to predict insulin dosage in diabetic patients based on factors such as glucose levels, insulin intake, and other relevant patient characteristics.

Once the training is completed, the algorithms are ready for deployment. When presented with a new test dataset containing instances with no class labels, the Gradient Boosting Classifier predicts the presence or absence of diabetes in each instance. If diabetes is detected, the Logistic Regression algorithm is then applied to predict the appropriate insulin dosage for the patient based on their specific characteristics and glucose levels. This predictive system enables early diagnosis of diabetes and personalized insulin dosage recommendations for diabetic patients, contributing to improved healthcare management and patient outcomes.

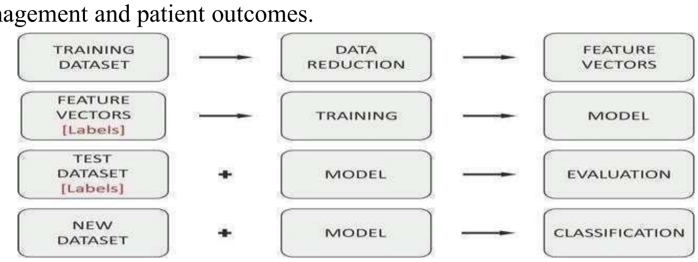

Fig. 1. Classification

II. LITERATURE SURVEY (RELATED WORK)

1. Prediction of insulin level of dosage

Abstract: In today's world, there are so many patients with Diabetes having varying insulin levels and blood glucose levels in the human body. So, there is a requirement to constantly monitor the blood glucose level and improve the patient's condition with an adjusted insulin dose. Before each meal, they must take a dose of insulin. Doctors must calculate insulin dosages for each patient based on previous dose data and regular sugar levels. Our study uses a Machine Learning approach to develop the model, which uses an RNN (LSTM) and ANN algorithm to predict a patient's insulin chart. The algorithm was trained using the patient's 36- month chart, and the extended series of following insulin predictions are based on that data. The predictive model is used in this investigation.

2. An Expert System for predicting insulin Dosage Using Various ML algorithms

Abstract: Diabetes occurs when your glucose (a type of sugar) found in the blood called as glycaemia is too excess in your body. Blood glucose, which is obtained from the food you eat, serves as your body's main energy supply. Blood glucose levels (BGLs) should appropriately balance to allow diabetic patients to lead normal lifestyles without running the risk of long-term, serious complications. But for a variety of factors, most diabetic patients have poorly controlled blood glucose levels, which over time seriously harms the oral health, vision, hearing and mental health. Moreover, taking the accurate amount of insulin dosage plays an important role in the treatment procedure. There are different prevention approaches like consuming strong nutriments and exercising are required to control their glycaemia. In this study, we predict diabetes using the Gradient Boosting Classifier, and we predictthe dosage of insulin for patients who have been identified as having diabetes using the Linear Regression Algorithm.We are using the Pima Indians diabetes dataset and UCI insulin dosage dataset to carry out this research. With the aforementioned dataset, we are training both algorithms. Once trained, upload test dataset without class label, and Gradient Boosting will then forecast the existence of diabetes while Linear Regression forecasts the amount of insulin to administer if diabetes is identified by Gradient Boosting.

3. A neural network model for predicting insulin Dosage for diabetic patients

Abstract: Diabetes Mellitus is a chronic metabolic disorder. Normally, with a proper adjusting of blood glucose levels (BGLs), diabetic patients could live a normal life without the risk of having serious complications that normally developed in the long run. However, blood glucose levels of most diabetic patients are not well controlled for many reasons. Although the traditional prevention techniques such as eating healthy food and conducting physical exercise are important for the diabetic patients to control their BGLs, however taking the proper amount of insulin dosage has the crucial rule in the treatment process. In this paper we have proposed a model based on artificial neural network (ANN) to predict the proper amount of insulin needed for the diabetic patient. The proposed model was trained and tested using several patients' data containing many factors such as weight, fast blood sugar and gender. The proposed model showed good results in predicting the appropriate amount of insulin dosage.

4. Forecasting Basal insulin for the therapy of Juvenile datasets at onset

Abstract: Establishing insulin dosages for diabetic patients is getting a challenge right after a patient's admission to a hospital. This is mainly due to the physicians' limited information on the current health state of a newly diagnosed patient. Specifically, for juvenile patients, high variability of body response to insulin causes the necessity of continuous adjustment of dosages during the patient treatment. To support the physicians in applying the initial insulin dose and adjusting the following doses, we propose a forecasting approach in this paper. We propose a forecasting model based on a specifically designed neural network to predict basal insulin. In the experimental part of this paper, we provide evidence for the high efficiency of the proposed method.

5. Time series prediction of personalized insulin Dosage for type 2 diabetics

Abstract: Careful blood glucose monitoring and consistent insulin administration are necessary for managing diabetes. People with demanding schedules or little access to medical personnel may find this difficult. Fortunately, without having to visit a doctor every day, daily insulin dosage may now be customized to a person's unique needs using technology and customised algorithms based on their food intake, exercise routines, and blood glucose levels. This information can be entered into a diabetes management app or device, where an algorithm will determine the proper insulin dosage and offer real-time feedback to assist maintain ideal blood glucose levels. A patient's dietary preferences, degree of physical activity, and blood sugar are considered for determining the proper bolus and basal insulin dosages in this study. According to the tracked body data, a patient's appropriate insulin dosage is predicted using artificial neural network (ANN)-based models. Based on patient activity, food intake, exercise, and past insulin administration, insulin projections are created. To forecast an individual's basal and bolus insulin requirements, long short-term memory (LSTM) and random forest regression models are employed. Accuracy of both models are tested, and random forest regression shows better accuracy which is used in the prediction system.

6. Prediction of Diabetes and insulin Dosage

Abstract: The world has different types of people as it differs with health issues. In that one of the main issues which they agonize is Diabetes. Age, obesity, hereditary habits, bulimia, elevated blood pressure, and others can contribute to Diabetes Mellitus. Diabetes boosts a person's chance of expanding various illnesses, heart disease, kidney disease, heart attack, eye issues, nerve damage, etc. Multiple trials are used in hospitals to gather the data required to diagnose diabetes by taking appropriate treatment. Data which is extracted in the healthcare sector are massive. Big Data analyses extensive data. It can be learned from the model and obtain accuracy for further analysis.

III. METHODOLOGY

1. Architecture

- **Knowledge Base:** Contains domain-specific information on diabetes management, insulin therapy, and patient characteristics.
- **Inference Engine:** Uses reasoning mechanisms to predict insulin dosages based on the knowledge base and patient data.
- **Patient Data Integration:** Integrates patient-specific data from electronic health records, wearable devices, and continuous glucose monitors.
- **Decision Support System:** Provides personalized insulin dosage suggestions to assist healthcare providers in clinical decision-making.
- **Feedback Loop and Learning**: Incorporates a feedback loop and machine learning algorithms to continuously improve predictions based on patient outcomes
- **User Interface:** Offers an intuitive interface for inputting patient data, displaying prediction results, and adjusting dosage recommendations.
- **Security and Privacy**: Implements measures to ensure the security and privacy of patient data, in compliance with healthcare regulations.

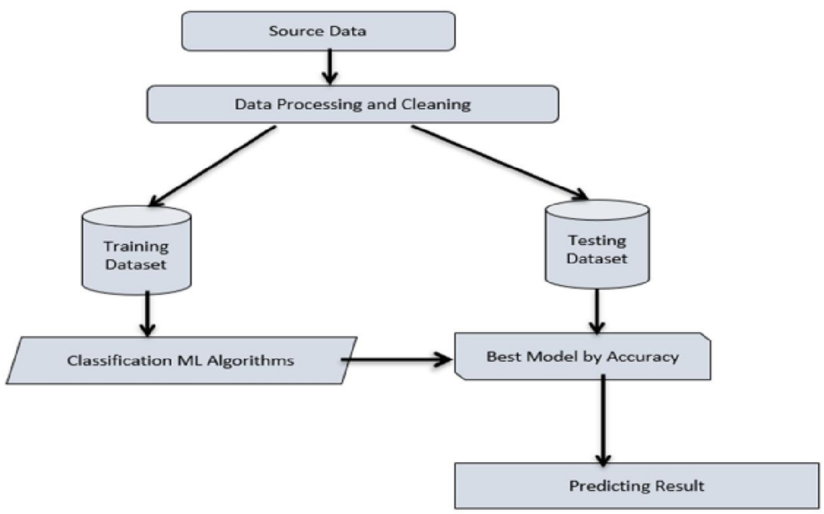

Fig. 2. Architecture

2. Algorithms

- **Gradient Boosting:** In gradient boosting, multiple models are trained sequentially to minimize the loss function gradually. Each new model focuses on areas where previous models underperform, improving overall accuracy. The algorithm constructs new base learners that correlate with the negative gradient of the loss function. Boosting is typically applied to weak learners to enhance their predictive power. However, it can lead to overfitting, so it's crucial to determine the right stopping point to prevent this issue.
- **Linear Regression:** Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. It predicts insulin dosage for a diabetes patient by establishing a mathematical relationship between patient-specific characteristics and historical insulin dosages.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

IV. IMPLEMENTATION

1. Data Collection

Data collection is indeed a critical aspect of building effective artificial intelligence (AI) and machine learning models. The process involves gathering data from various sources, which could include structured databases, unstructured text, images, videos, sensor data, social media feeds, and more. The goal is to accumulate a comprehensive dataset that encompasses relevant information for the specific business problem or task at hand.

The quality of the data collected is paramount, as the saying goes, "garbage in, garbage out." This emphasizes that the effectiveness of AI and machine learning models heavily depends on the quality and relevance of the data used to train them. If the data collected is inaccurate, incomplete, or irrelevant, it will adversely impact the performance and accuracy of the resulting models.

2. Data Pre-processing

Data pre-processing is indeed a critical step in the machine learning pipeline. It involves transforming raw data into a format that is suitable for training machine learning models. Raw data often contains various issues such as errors, missing values, outliers, and inconsistencies, making it unsuitable for direct application in machine learning algorithms. Pre-processing addresses these issues by cleaning, organizing, and transforming the data to make it usable for modelling.

3. Training and Testing

Training Dataset: This is the portion of the dataset used to train the machine learning model. It consists of input-output pairs where the input (features) are used to predict the output (target variable). The model learns patterns and relationships within the training dataset through various algorithms and optimization techniques. The training dataset is typically the largest subset of the original dataset and is used to fit the model parameters during the training phase.

Test Dataset: This is a separate portion of the dataset that is used to evaluate the performance of the trained machine learning model. The test dataset is not usedduring thetraining phase, and the model has not seen this data before.

It is used to estimate how well the trained model generalizes to new, unseen data. The performance metrics calculated on the test dataset, such as accuracy, precision, recall, or F1-score, provide insights into how well the model performs on real-world data.

V. DATA FLOW DIAGRAM

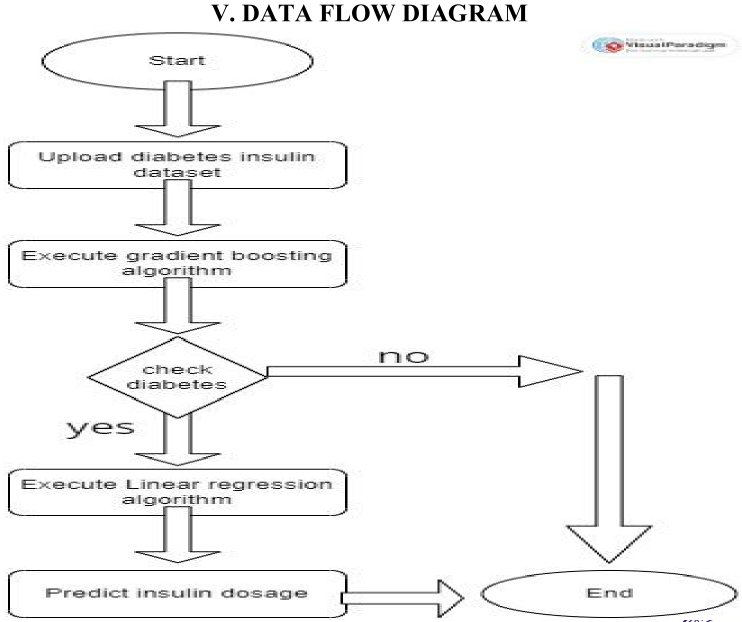

Fig. 3. Data Flow Diagram

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

The Data Flow Diagram (DFD), also known as a bubble chart, is a graphical representation used to depict a system's components, including processes, data inputs, outputs, and external entities. It illustrates how information flows through the system and undergoes transformations as it moves from input to output. DFDs are essential modelling tools for understanding and visualizing the flow of data within a system, as well as the interactions between system components and external entities. They provide a simple and intuitive way to represent complex systems and aid in the analysis, design, and communication of system architecture.

VI. RESULT

The predictive model developed in this study was successfully applied to detect the presence or absence of diabetes in the study participants. Among the individuals, a subset was identified as having diabetes based on the predictive analysis, while others were classified as not having diabetes. For individuals identified as having diabetes, the model further provided insulin dosage recommendations. These results demonstrate the effectiveness of the predictive model in accurately identifying individuals with diabetes and providing personalized treatment recommendations, thus highlighting its potential utility in clinical practice for improving patient outcomes and guiding therapeutic interventions.

Fig. 4. Output

VII. CONCLUSION

Firstly, our findings demonstrate that gradient boosting achieved a remarkable accuracy of 100% in predicting diabetes, outperforming linear regression which attained an accuracy of 78%. The accuracy graph further illustrated the superior performance of gradient boosting with a perfect accuracy score of 1.0, compared to linear regression's accuracy graph depicting a score of 0.8. This highlights the effectiveness of gradient boosting in accurately identifying individuals with diabetes based on the provided data.

In conclusion, our study showcases the utility of advanced machine learning algorithms, particularly gradient boosting, in improving the accuracy of diabetes prediction and insulin dosage recommendation. By leveraging these predictive models, healthcare practitioners can make more informed decisions regarding patient care, leading to better management of diabetes and ultimately improved patient outcomes. Moving forward, further research and validation of these models in real-world clinical settings are warranted to enhance their effectiveness and applicability in diabetes management.

VIII. FUTURE WORK

Future research in diabetes prediction and insulin dosage recommendation should focus on developing and refining advanced machine learning algorithms and predictive modelling techniques. These efforts aim to enhance the accuracy and reliability of predictive models by integrating additional data sources such as genetic information, wearable sensor data, and real-time health monitoring data. By exploring innovative approaches and leveraging emerging technologies, researchers can improve the predictive capabilities of these models, ultimately leading to more effective diabetes management strategies.

Establishing collaborative research networks and consortia dedicated to diabetes prediction and personalized treatment can facilitate knowledge sharing, data sharing, and interdisciplinary collaboration. By fostering partnerships among researchers, clinicians, industry stakeholders, and patient advocacy groups, collaborative research networks accelerate scientific discovery, promote innovation, and drive translation of research findings into clinical practice. These networks enable researchers to leverage diverse expertise, resources, and datasets to address complex challenges in diabetes management effectively.

REFERENCES

[1] Ketan Gupta, NasminJiwani "Prediction of Insulin Level of Diabetes Patient Using Machine Learning Approaches ",International Journal of Creative Research Thougths, 2022 [online] Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4205251

[2]K. Aruna Sri, N. IshwaryaJyothi, M. Anudeep, "An Expert System for predicting Insulin Dosage Using Various ML algorithms", International Journal of Research Publication and Reviews,2022, [online] Available:https://ijrpr.com/uploads/V4ISSUE4/IJRPR11644.pdf

[3]Bilal M. Zahran, "A Neural Network model for predicting insulin dosage for diabetic patients", International Journal of Computer Science and Information Security , 2016 [online] Available: https://www.researchgate.net/publication/306493832_A_Neural_Network_Model_for_Predicting_Insulin_Dosage_for_Diabetic_Patientshttps://www.researchgate.net/publication/306493832_A_Neural_Network_Model_for_Predicting_Insulin_Dosage_for_Diabetic_Patients

[4]Rofal Deja, Wojciech Froelich, "Forecasting Basal Insulin for the clinical therapy of Juvenile Diabets at Oneset", International Conference on Knowledge-Based and intelligent Information & Engineering, 2022
https://www.researchgate.net/publication/364421256_Forecasting_Basal_Insulin_for_the_ClinicalTherapy_of_Juvenile_Diabetes_at_Onset

[5] G.Jisha, T.Bhuvan, RittaJerrad, "Time series prediction of personalized insulin dosage for type2 diabetics", Indonesian journal of Electrical Engineering and Computer Science, 2023
https://ijeecs.iaescore.com/index.php/IJEECS/article/view/30353

[6] V.Anirudh, G.Siri, Puli Shashank, T.RahulVardhan, "Prediction of Diabetes and Insulin Dosage", International Journal of Advances in Engineering and Management ,2023.
https://ijaem.net/issue_dcp/Prediction%20of%20Diabetes%20and%20Insulin%20Dosage.pdf

[7] Minh Nguyen, Jankovic, LaurynasKalesinskas, Michael Baiocchi "Machine learning for initial insulin estimation in hospitalized patients", *Journal of the American Medical Informatics Association, 2021*
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8449602/

[8] H. Naz and S. Ahuja, "Deep learning approach for diabetes prediction using PIMA Indian dataset", International Journal of Advanced Science and Technology,2020.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7270283/

[9]Qingnan Sun, Marko V. Jankovic, Lia Bally , "Predicting Blood Glucose with an LSTM and Bi-LSTM Based Deep Neural Network Journal of Medical and Biological Engineering, 2018. https://ieeexplore.ieee.org/document/8586990

[10] FayrozaAlaaKhaleel and Abbas M. Al-Bakry, "Diagnosis of Diabetes Using Machine Learning Algorithms", Journal of Medical Systems,2021, https://www.sciencedirect.com/science/article/abs/pii/S2214785321050550

[11] N. Sneha and T. Gangil, Analysis of diabetes mellitus for early prediction using optimal features selection, 2019. https://journalofbigdata.springeropen.com/articles/10.1186/s4 0537-019-0175-6

[12] J. Martinsson, A. Schliep, B. Eliasson and O. Mogren, "Blood Glucose Prediction with Variance Estimation Using Recurrent Neural Networks", IEEE Journal of Biomedical and Health Informatics, 2021. https://www.mdpi.com/2227-7390/11/21/4517

[13] S. R. Ashokkumar, S. Anupallavi, G. MohanBabu, M. Premkumar and V. Jeevanantham, "Emotion identification bydynamic entropy and ensemble learning from electroencephalogram signals", International Journal of Imaging Systems and Technology, vol. 32, no. 1, pp. 402-413, 2022. https://www.researchgate.net/publication/356088649_Emotion_identification_by_dynamic_entropy_and_ensemble_learning_from_electroencephalogram_signals

[14] "Predict the start of diabetic disease using Artificial Neural Network (ANN)", International Journal of Computer Science & Emerging Technologies, vol. 2, no. 2,https://www.researchgate.net/publication/228850003_Predict_the_onset_of_diabetes_disease_using_Artificial_Neural_Network_ANN
[15] FayrozaAlaaKhaleel and Abbas M. Al-Bakry, "Diagnosis of Diabetes Using Machine Learning Algorithms", Journal of Medical Systems, vol. 45, no. 7, 2021. https://www.researchgate.net/publication/373045901_Diabetes_detection_based_on_machine_learning_and_deep_learning_approaches