WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

DIABETIC INDICATOR MODEL USING MACHINE LEARNING APPROACH

Ayushi Aher [*1], Janhavi Thawre[*2], Shruti Nadekar[*3], Yashavi Telang[*4], Sakshi Khamankar[*5]

[*1,2,3,4]Student, Department of Computer Engineering, Nagpur, Maharashtra, India

[*5]Asst. Professor, Department of Computer Engineering, Nagpur, Maharashtra, India

ABSTRACT:

The Diabetic Indicator is a cutting-edge web tool that harnesses the power of machine learning to forecast the risk of diabetes at an early stage. By scrutinizing key health metrics such as blood sugar levels, body mass index (BMI), and lifestyle habits, it offers invaluable insights for proactive health management. The interface is intuitively designed, allowing users to input their data and swiftly receive predictions easily. Tailored recommendations are generated through meticulous analysis, empowering users to make informed choices about their well-being.

This proactive approach significantly contributes to overall health by enabling the early detection of potential diabetes risks. The application prioritizes user convenience, ensuring seamless access to its predictive capabilities for efficient health monitoring. With a focus on actionable insights, the Diabetic Indicator is a vital tool for individuals committed to enhancing their health and averting the onset of diabetes.

Keywords: Predictive model, Random Forest Algorithm, Real-time Monitoring, Healthcare analytics, Machine learning (ML)

INTRODUCTION :

Diabetes is a chronic condition characterized by insufficient insulin production, leading to unregulated blood sugar levels. Risk factors such as obesity, sedentary lifestyle, hypertension, and abnormal cholesterol levels can exacerbate the likelihood of developing diabetes. If left unmanaged, diabetes can result in various complications such as frequent urination, skin ailments, nerve damage, and vision impairment, potentially progressing to kidney failure and vision loss without timely intervention.

Machine Learning (ML), a subset of Artificial Intelligence (AI), is increasingly utilized across various domains, including healthcare, to analyze vast datasets and forecast outcomes. ML algorithms like classification and regression are valuable in medical diagnosis, particularly in early diabetes detection.

ML holds promise in enhancing diabetes management. It can analyze data from devices like continuous glucose monitors to predict hypoglycaemic episodes, offering real-time alerts to patients and healthcare providers, thus improving health outcomes. Moreover, ML can sift through extensive patient data to identify patterns and risk factors, enabling earlier interventions and tailored treatments for at-risk individuals.

Moreover, ML enables the analysis of extensive patient datasets to discern trends and patterns in diabetes development and progression. This identification of risk factors and early indicators empowers healthcare providers to intervene earlier and deliver targeted interventions to at-risk individuals.

In essence, machine learning empowers healthcare providers with advanced tools to better understand, predict, and manage diabetes, ultimately leading to improved patient care and outcomes.

METHODOLOGY :

In developing an automated diabetes prediction system using machine learning, a structured approach was followed. In our development of an automatic diabetes prediction system using machine learning, we adhered to a systematic methodology to ensure robustness and accuracy. We outline the step-by-step process and methodology employed in developing an automatic diabetes prediction system using machine learning techniques.

Initially, we meticulously cleaned the raw dataset, addressing any missing values and rectifying imbalances within the data. This involved a comprehensive examination of the dataset, employing suitable techniques to fill missing values and ensure its representativeness and balance.

Following data cleaning, we employed a holdout validation technique to evaluate the performance of our predictive models. This method entails randomly dividing the dataset into two distinct subsets: a training set utilized to train the machine learning models, and a separate test set designated for

assessing the models' performance. By partitioning the data in this manner, we effectively gauge the models' ability to generalize to unseen data, simulating real-world scenarios.

Our selection of algorithms was deliberate, chosen based on their appropriateness for the task and their capacity to accommodate the unique characteristics of the dataset. Each algorithm was meticulously assessed to ensure its suitability and efficacy in facilitating accurate diabetes prediction.

Analytical Framework:

Random Forest Algorithm

Random Forest is a powerful supervised learning algorithm renowned for its versatility in handling both classification and regression tasks. It operates on the principle of ensemble learning, a technique that amalgamates predictions from numerous decision trees trained on diverse subsets of the dataset. Each decision tree within the Random Forest model independently predicts outcomes based on a specific subset of features extracted from the dataset. These individual predictions are then aggregated using a majority voting mechanism to arrive at the final prediction. This ensemble approach contributes significantly to enhancing the accuracy and resilience of the model when compared to utilizing a single decision tree.

By harnessing the collective insights derived from multiple trees, Random Forest adeptly captures intricate relationships within the dataset, thereby enabling more precise predictions. Its efficacy in discerning complex patterns and delivering reliable predictions renders it a favoured choice across a spectrum of machine-learning applications where accuracy and dependability are paramount.

Due to its ability to provide accurate and reliable predictions, Random Forest is a favoured choice across various machine-learning applications where high accuracy and dependability are paramount

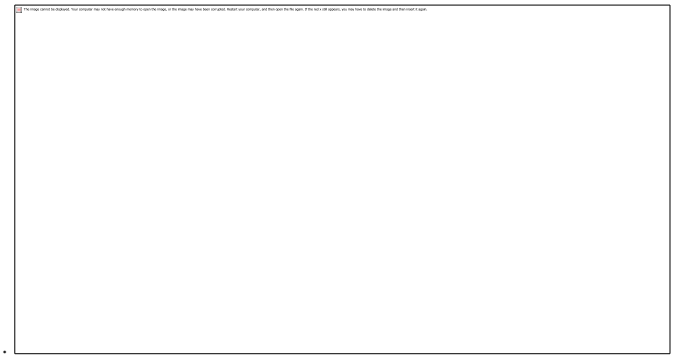


Fig 1. Random Forest Algorithm

Working process of Random Forest Algorithm:

- In the Random Forest model, a subset of data points and a subset of features are selected for constructing each decision tree. Simply put, n random records and m features are taken from the data set having k number of records.
- Individual decision trees are constructed for each sample.
- Each decision tree will generate an output.
- Final output is considered based on ***Majority Voting or Averaging*** for Classification and regression, respectively.

Development Model

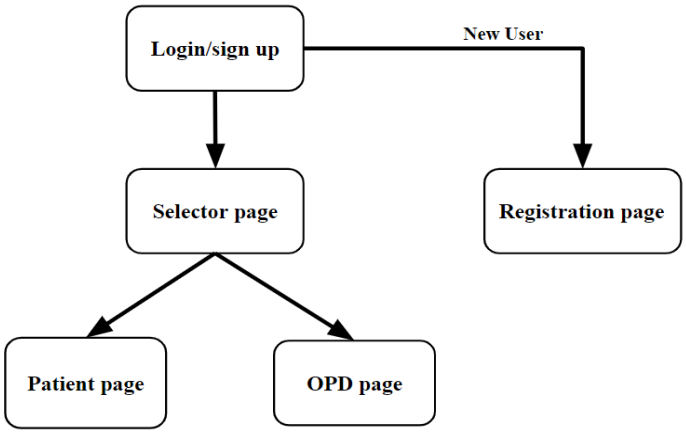


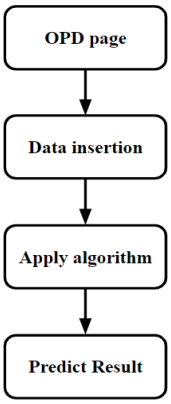
Fig 2. Working Model

In the developmental model, the workflow initiates with the user accessing the login or sign-up page. Here, individuals input their details to create an account, enabling them to proceed further. Upon successful registration, a selection page emerges, presenting users with two distinct paths to choose from, tailoring the experience to their specific needs.

On this selection page, users are prompted to specify their role and purpose within the system. They can opt to proceed as a patient seeking access to medical reports or as a healthcare provider from the outpatient department (OPD) interested in predicting diabetes for patients. This bifurcation ensures a streamlined journey for users, directing them towards functionalities most relevant to their intentions and requirements.

Once users have selected their respective paths, the workflow diverges into Phase 1 or Phase 2, aligning with the distinct objectives associated with each user category. This segmentation facilitates a targeted approach, allowing for the efficient utilization of system resources and ensuring a tailored experience that caters to the specific needs of patients and healthcare providers alike.

This meticulous structuring of the workflow ensures clarity and efficiency, enabling seamless navigation for users as they engage with the system according to their respective roles and objectives. Through this thoughtful design, the developmental model aims to optimize user experience and maximize the utility of the platform across diverse user demographics and needs.



II.2.1 Phase 1

Fig 3. Phase 1

The user initiates the process by accessing the diabetic indicator system or application. This could be through a web browser. The user selects or opts for an OPD within the system. This step signifies the user's intent to proceed with a specific task or operation within the system. After opting for the desired operation, the user is directed or navigated to the prediction page. This page likely contains the interface or form where users can input relevant data or information required for predicting their diabetic risk or any other related outcome.

At this stage, the user interface prompts the OPD personnel to enter specific values corresponding to essential health metrics. These metrics typically include blood pressure (BP), body mass index (BMI), glucose levels, and insulin levels. Additionally, the system requests the individual's name and gender to personalize the prediction process and ensure accuracy. This interactive process allows the OPD personnel to input comprehensive health data directly into the system, facilitating the generation of tailored predictions for diabetes risk or status.

In the preprocessing phase, we focus on two key tasks: outlier removal and data standardization. First, we identify and remove instances with zero values to ensure the integrity of our dataset. Then, we standardize the processed data to make it consistent and easier to work with. This preprocessing step is crucial before creating our model, as it ensures the accuracy and reliability of our classifiers.

Once the preprocessing is complete, we move on to model creation. We start by splitting our data into two subsets: a training set and a testing set. About 80% of the data is allocated for training, while the remaining 20% is set aside for testing the trained model. This division allows us to train our algorithm effectively on a substantial portion of the data while still having unseen data to evaluate its performance. By following these steps, we ensure that our model is robust and capable of making accurate predictions on new, unseen data.

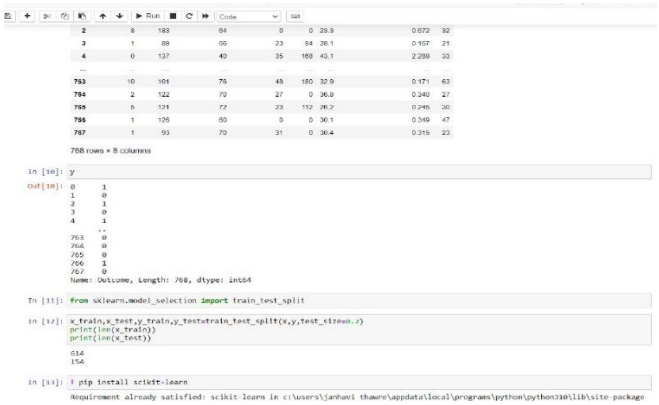


Fig 4. Data Splitting

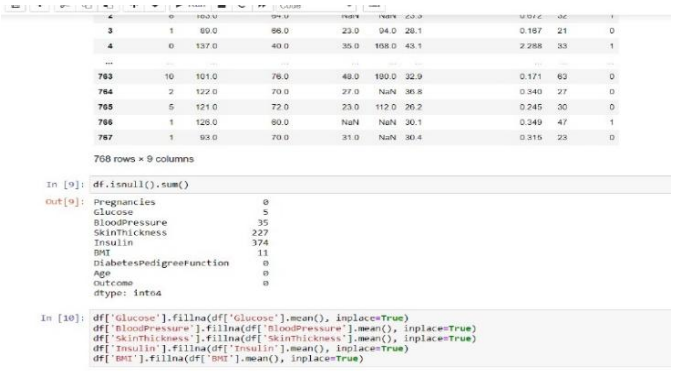


Fig 5. Missing Data Removal

After conducting data preprocessing, Python libraries such as NumPy, pandas, and scikit-learn (sklearn) are employed to handle input fields such as blood pressure (BP), body mass index (BMI), glucose levels, and insulin. These libraries assist in various tasks including data manipulation, and cleaning. Specifically, scikit-learn facilitates the implementation of both supervised and unsupervised learning algorithms, which are crucial for enhancing the accuracy of data processing.

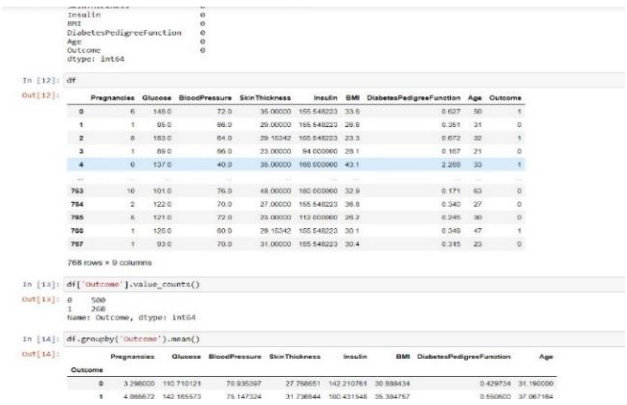


Fig 6. Applying Algorithm

In particular, the Random Forest algorithm is employed to further refine the data. This algorithm utilises decision trees and ensemble learning techniques to improve predictive accuracy. By utilising a multitude of decision trees, Random Forest reduces overfitting and provides robust predictions.

In summary, through the aid of Python libraries such as NumPy, pandas, and scikit-learn, along with the application of the Random Forest algorithm, the input data undergoes thorough processing to enhance its accuracy and suitability for analysis.

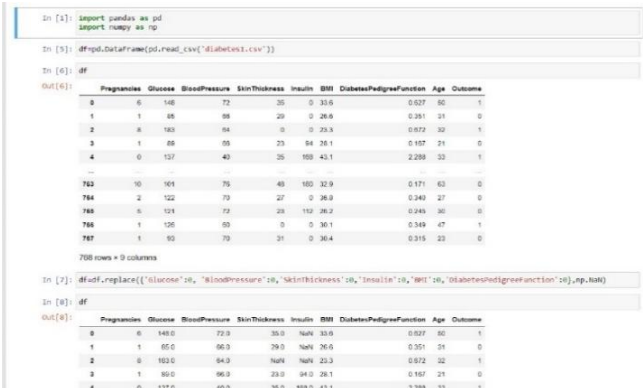


Fig 7. Using Python libraries

Fig 8. Prediction

The image depicts how diabetes predictions are made using different factors and algorithms. It shows how factors like blood pressure, body mass index (BMI), glucose levels, and insulin levels are used to predict the likelihood of diabetes. It indicates whether individuals have diabetes or not, considering certain ranges of insulin levels. This simplification aids in understanding the results more clearly. The outcome is a prediction, likely represented visually. The system aims to provide insightful assessments that can guide clinical decision-making and patient management strategies effectively.

II.2.2 Phase 2

In Phase 2 of the system, after users have identified themselves as patients and accessed the platform, they encounter a crucial stage where they specify the duration for which they require the predictive report. This essential step allows patients to tailor their experience according to their specific needs, whether they seek short-term insights or longer-term monitoring. Within this duration selection interface, patients are presented with options to choose the timeframe in either months or days, providing them with flexibility and control over the generated report's relevance to their health management goals.

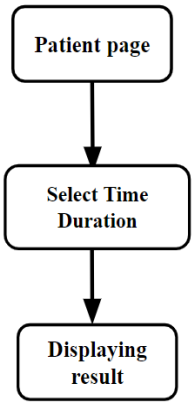


Fig 9. Phase 2

Once the duration is selected, patients proceed to receive a comprehensive report containing vital health metrics and predictive insights. This detailed document serves as a valuable resource, offering a comprehensive overview of their health status. The report includes essential information such as the patient's name, age, and a range of health indicators including insulin levels, glucose levels, body mass index (BMI), blood pressure (BP), and most importantly, the predicted diabetic range. By consolidating these key data points into a single, easily accessible report, patients are empowered to make informed decisions about their health and proactively manage their well-being with greater confidence and precision.

Through this meticulous approach to reporting, Phase 2 of the system endeavours to enhance patient engagement and promote proactive healthcare management strategies, ultimately contributing to improved health outcomes and quality of life.

RESULT :

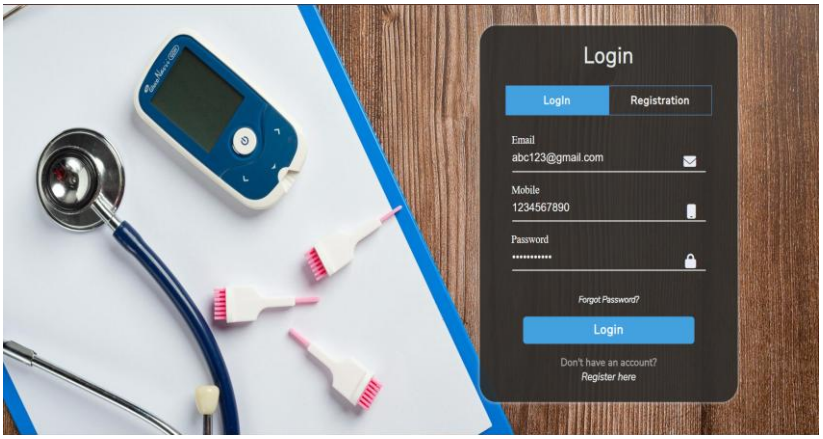


Fig 10. Login Page

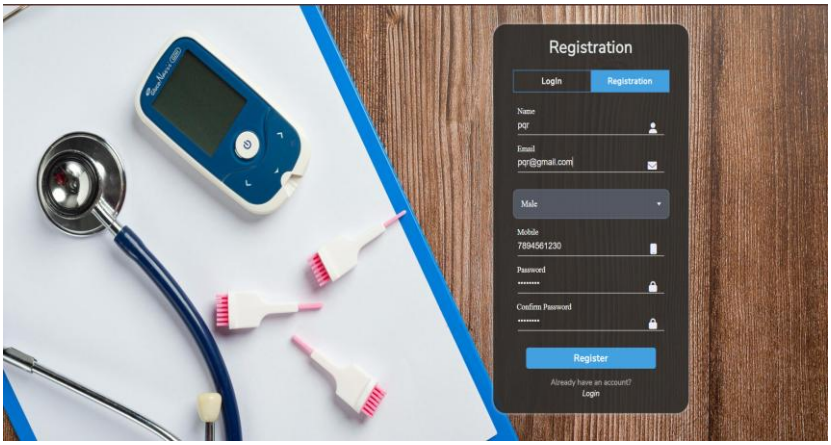


Fig 11. Registration Page

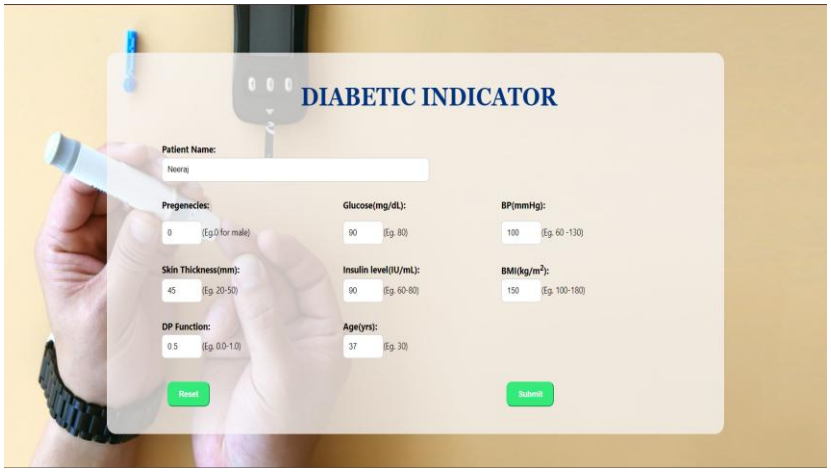


Fig 12. Data Insertion Page

Fig 13. Predicted Result

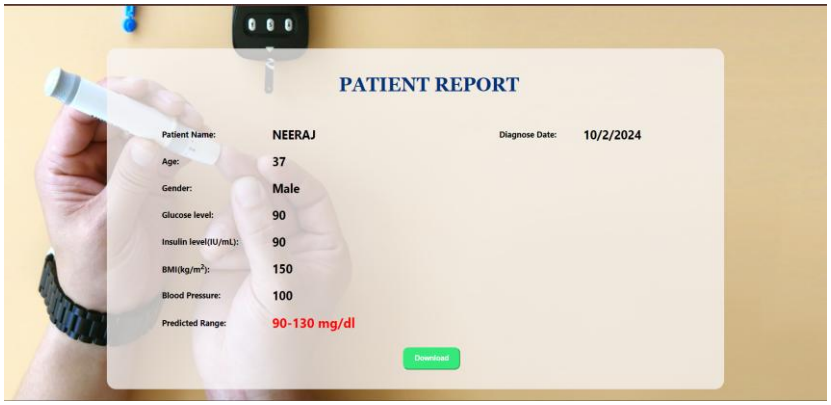


Fig 14. Patient Report

FUTURE SCOPE :

To enhance the accuracy of prediction for health-related outcomes such as diabetes, implementing advanced machine learning algorithms and incorporating additional relevant features into predictive models could be beneficial. By leveraging a diverse range of health indicators such as blood glucose levels, BMI, insulin levels, age, gender, and lifestyle factors, predictive models can generate more precise risk assessments and recommendations. Furthermore, continuous refinement and validation of these models using real-world data can help improve their accuracy over time.

Implementing a one-on-one chatbot service for providing health guidance can greatly enhance patient engagement and support. Such a service can offer personalized advice, answer queries in real time, and provide ongoing support to individuals managing diabetes.

Incorporating a diet chart feature into the chatbot service, with a focus on insulin control measures, can provide users with valuable dietary guidance for managing diabetes and regulating blood sugar levels. By analysing users' health data and preferences, the chatbot can generate customized diet plans and meal recommendations that align with their individual nutritional needs and insulin requirements.

Additionally, implementing a messaging alert feature to remind users to log in and engage with the chatbot at regular intervals can help promote consistent use and adherence to health management strategies, ultimately contributing to better long-term health outcomes. Overall, integrating these innovative features into diabetes management platforms can empower individuals to take proactive steps towards better health and well-being.

CONCLUSION :

This research endeavours to address a pressing medical concern: the early detection of diabetes. Recognizing the significance of this issue, a systematic approach is adopted to devise a predictive system capable of identifying diabetes in its nascent stages. Central to this endeavour is the evaluation of various machine learning classification algorithms, assessing their efficacy through a range of metrics.

To gauge the performance of these algorithms, extensive experiments are conducted employing the Kaggle raw dataset, a widely recognized resource in the domain of diabetes research. Through these experiments, invaluable insights emerge regarding the effectiveness of the designed system in accurately predicting diabetes.

The primary aim is to construct a system that can reliably predict the presence of diabetes with a high degree of accuracy. Following thorough experimentation and analysis, the developed system achieves an impressive accuracy rate of 96%. This signifies the system's efficacy in accurately identifying individuals at risk of diabetes, thereby emphasizing its potential for implementing early intervention and prevention strategies.

REFERENCES

[1] Chun-Yang Chou, Ding-Yang Hsu, Chun-Hung Chou, “Predicting the Onset of Diabetes with Machine Learning Methods”, Journal of Personalized Medicine 2023, Volume 13, Issue 3, 2023, https://doi.org/10.3390/jpm13030406

[2] Makroum, M.A., Adda, M., Bouzouane, A., Ibrahim, H., “Machine Learning and Smart Devices for Diabetes Management: Systematic Review”. *Sensors* **2022**, Volume 22, Issue 5, https://doi.org/10.3390/s22051843

[3] Kaur, H. and Kumari, V. (2022), "Predictive modelling and analytics for diabetes using a machine learning approach", Applied Computing and Informatics, Vol. 18 No. 1/2, pp. 90-100. https://doi.org/10.1016/j.aci.2018.12.004

[4] Fayroza Alaa Khaleel, Abbas M. Al-Bakry,” Diagnosis of diabetes using machine learning algorithms”, Materials today :Proceedings, Volume 80, Part 3, 2023, Pages 3200-3203, https://doi.org/10.1016/j.matpr.2021.07.196

[5] Chang, V., Bailey, J., Xu, Q.A. et al. Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. Neural Comput & Applic 35, 16157–16173 (2023). https://doi.org/10.1007/s00521-022-07049-z

[6] Abnoosian, K., Farnoosh, R. & Behzadi, M.H. Prediction of diabetes disease using an ensemble of machine learning multi-classifier models. BMC Bioinformatics 24, 337 (2023). https://doi.org/10.1186/s12859-023-05465-z

[7] G Ravi Kumar, Reddyvari Venkateswara Reddy, Jayarathna M, N Pughazendi, S Vidyullatha, Pundru Chandra Shaker Reddy, “Web application based Diabetes prediction using Machine Learning”, 2023 International Conference on Advances in Computing, Communication and Applied Informatics , https://doi.org/10.1109/ACCAI58221.2023.10200323

[8] Wee, B.F., Sivakumar, S., Lim, K.H. et al. Diabetes detection based on machine learning and deep learning approaches. Multimed Tools Appl 83, 24153–24185 (2024).https://doi.org/10.1007/s11042-023-16407-5

[9] Shubham Sain, Aayush Singh, Dharmender Bhatnagar, Asst. Prof. Shallu Juneja, “Diabetes Prediction Using ML”, International Research Journal of Engineering and Technology (IRJET), Volume: 10 Issue: 03, Mar 2023, Pages: 739-747, ISSN: 2395-0056,

[10] Isfafuzzaman Tasin, Tansin Ullah Nabil, Sanjida Islam, Riasat Khan, “Diabetes prediction using machine learning and explainable AI techniques”, The Institute of Engineering And Technology, 2022, https://doi.org/10.1049/htl2.12039