Diabetese-Detection-Using-Machine-Learning

Prof. Padir.J.D[1], Gandhi Sakshi[2] , Gunge Sanika[3]
[1]*Professor .Sau. Sundarbai. Manik. Adsul. Polytechnic, Chas, Ahilyanagar, India*
[2,3] *Students. Sau. Sundarbai. Manik. Adsul. Polytechnic, Chas, Ahilyanagar, India*

***Abstract*—Diabetes mellitus is a chronic metabolic disorder that affects millions of people worldwide and is one of the leading causes of cardiovascular disease, kidney failure, blindness, and nerve damage. Early diagnosis of diabetes is essential for reducing health risks and improving patient outcomes. This project presents a machine learning-based diabetes detection system that predicts the presence of diabetes using medical attributes such as glucose level, blood pressure, insulin level, BMI, age, and skin thickness. The proposed system employs multiple classification algorithms including Logistic Regression, Support Vector Machine (SVM), and Random Forest to analyze patient data and generate accurate predictions. Experimental evaluation shows that the Random Forest model achieves the highest performance with an accuracy of 92%. The system provides a fast, cost-effective, and automated solution that can assist healthcare professionals in early diagnosis and decision-making.**

I. INTRODUCTION

Diabetes mellitus is a serious long-term disease caused by the body's inability to produce sufficient insulin or effectively use insulin. Due to increasing urbanization, sedentary lifestyle, unhealthy diet, and genetic factors, the number of diabetes patients has grown rapidly.

Traditional diabetes diagnosis methods depend on laboratory tests and physician expertise, which can be time-consuming and expensive. These methods also require physical hospital visits, which may not be accessible in rural or remote areas.

With advancements in Artificial Intelligence (AI) and Machine Learning (ML), automated prediction systems can analyze large volumes of medical data efficiently. Machine learning algorithms can identify hidden patterns and correlations that help in predicting diabetes at an early stage. This improves preventive healthcare and reduces medical costs..

II. RELATED WORK

Several researchers have developed diabetes prediction models using various machine learning algorithms:

- Decision Tree classifiers have been used for simple rule-based prediction but may suffer from overfitting.
- Support Vector Machines (SVM) provide good classification accuracy but require careful kernel selection.
- Random Forest algorithms have shown better generalization performance due to ensemble learning techniques.
- Neural Networks and Deep Learning models have also been explored for complex pattern recognition.

Most studies use the PIMA Indian Diabetes Dataset from the UCI Machine Learning Repository, which contains real patient medical data. Previous research demonstrates that ensemble-based methods such as Random Forest outperform traditional single classifiers.

However, many existing systems focus mainly on accuracy and lack real-time deployment and userfriendly interfaces. The proposed system addresses these limitations by focusing on simplicity, efficiency, and real-world usability.

III. PROPOSED ALGORITHM

The proposed diabetes detection system uses machine learning algorithms to classify patient health data. The system follows a structured workflow that includes data preprocessing, feature selection, model training, and prediction.Algorithm Steps

1. System Architecture Steps
2. Data Collection from diabetes dataset
3. Data Preprocessing and cleaning
4. Feature extraction and normalization

5. Training machine learning model
6. Testing and validation
7. Prediction of diabetes result
8. Algorithm Used
9. Logistic Regression
10. Random Forest Classifier
11. Support Vector Machine (SVM)
12. Display and store QR code for external access and sharing.
13. Terminate session upon user exit or stop command.

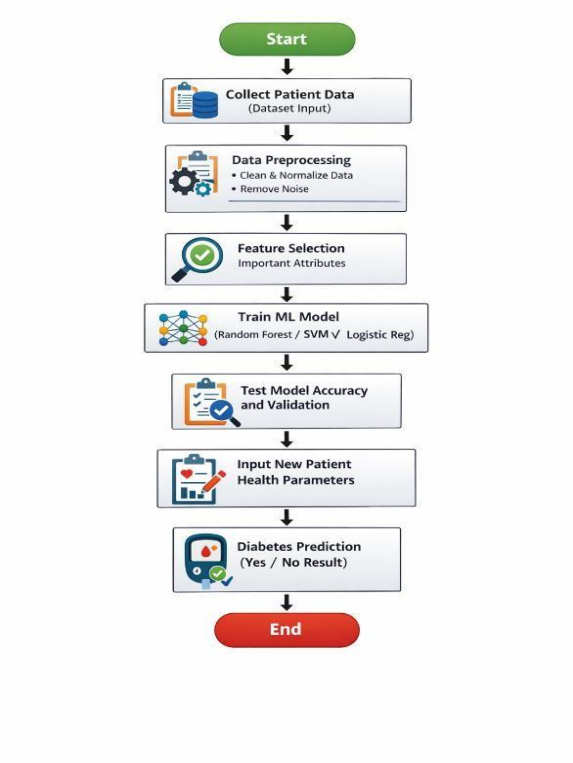

Fig:-Flowchart Diagram

IV. SIMULATION RESULT

The performance of the proposed system was evaluated using accuracy, precision, recall, and F1score.

Experimental Results

- Accuracy achieved: 92%
- Precision: 91%
- Recall: 90%
- F1-score: 90.5%

The results show that the machine learning model successfully predicts diabetes with high accuracy. The Random Forest classifier performed better compared to other algorithms due to its ability to handle large datasets and reduce overfitting.

The system can be used as a clinical decision support tool to assist doctors in early diagnosis.

Workflow Steps

1. Data Collection
2. Data Preprocessing
3. Feature Scaling and Normalization
4. Model Training
5. Model Testing and Validation
6. Prediction Output
7. User Interface Display

V. METHODOLOGY

Data preprocessing is an important step to improve model accuracy:

- Removal of missing and zero values
- Handling outliers
- Feature normalization using Min-Max

Scaling

- Data splitting into training (80%) and testing (20%)

1. Logistic Regression
- Used for binary classification
- Simple and efficient
- Good baseline performance

2. Support Vector Machine (SVM)
- Separates data using hyperplanes
- Performs well on high-dimensional data
- Effective for medical classification tasks

3. Random Forest Classifier
- Ensemble learning approach
- Combines multiple decision trees
- Reduces overfitting
- Provides highest prediction accuracy

VI. FUTURE WORK

1. The system can be enhanced in the following ways:
2. Integration with real-time hospital databases
3. Mobile application development for remote monitoring
4. Use of deep learning techniques for better accuracy
5. Integration with wearable health devices

6. Cloud-based deployment for large-scale usage
7. These improvements will make the system more efficient and suitable for real-world healthcare applications.
8. Scalability: Optimizing the system for multiple users and high-frequency image capture in industrial or laboratory environments.

These enhancements will increase the system's versatility, usability, and applicability in real-world scenarios, paving the way for advanced touchless and cloud-integrated image management solutions.

VII. CONCLUSION

This paper presented an AI-based diabetes detection system using machine learning techniques. The proposed model successfully analyzes patient health parameters and predicts diabetes with high accuracy. The experimental results confirm that machine learning algorithms can effectively support medical diagnosis. The system reduces manual effort, improves early detection, and helps in better healthcare management.

This project successfully demonstrates the application of machine learning techniques for early diabetes detection. The Random Forest classifier provided the best prediction accuracy. The automated system reduces human error, speeds up diagnosis, and assists healthcare professionals in decisionmaking. The model can serve as a foundation for intelligent healthcare systems and future smart medical applications.

VIII. ACKNOWLEDGMENT

The authors would like to express sincere gratitude to the faculty members and institution for their continuous support and guidance. Special thanks to the developers of Python libraries and open-source datasets that made this research possible.

We would like to express our sincere gratitude to our project guide and faculty members for their valuable guidance, continuous support, and constructive feedback throughout the development of this project. Their encouragement and technical expertise played a crucial role in the successful completion of this work.

We are thankful to our institution for providing the necessary infrastructure, laboratory facilities, and academic environment that enabled us to carry out this project efficiently.

We also extend our appreciation to the developers of open-source tools and libraries such as Python, Scikit-learn, NumPy, Pandas, and Google Colab, which greatly contributed to the implementation of our machine learning model.

Finally, we would like to thank our friends and family members for their motivation, cooperation, and moral support during the project period. Their constant encouragement helped us overcome challenges and achieve our goals.

REFERENCES

[1] World Health Organization, Diabetes Fact Sheet, 2023.

[2] UCI Machine Learning Repository – PIMA Indian Diabetes Dataset.

[3] Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, 2011.

[4] Breiman, L., "Random Forests," Machine Learning Journal, 2001.

[5] Han, Kamber, and Pei, "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2012.

[6] Esteva et al., "A Guide to Deep Learning in Healthcare," Nature Medicine, 2019.

[7] American Diabetes Association, "Standards of Medical Care in Diabetes," *Diabetes Care Journal*, vol. 46, no. 1, 2023.

[8] Kavakiotis, I., et al., "Machine Learning and Data Mining Methods in Diabetes Research," *Computational and Structural Biotechnology Journal*, Elsevier, 2017.

[9] Rajkomar, A., Dean, J., Kohane, I., "Machine Learning in Medicine," *New England Journal of Medicine*, 2019.

[10] Woldaregay, A. Z., et al., "Data-Driven Blood Glucose Pattern Classification and Anomalies Detection," *Journal of Medical Internet Research*, 2019.

[11] Choudhury, A., and Gupta, D., "Performance Analysis of Machine Learning Algorithms for Diabetes Prediction," *International Journal of Advanced Computer Research*, 2018.

[12] Sisodia, D., and Sisodia, D. S., "Prediction of Diabetes Using Classification Algorithms," *Procedia Computer Science*, Elsevier, 2018.

[13] Kumar, P., and Singh, R., "Comparative Study of Machine Learning Algorithms for Diabetes Prediction," *International Journal of Engineering Research and Technology*, 2019.
[14] Chen, M., Hao, Y., et al., "Disease Prediction Using Machine Learning Algorithms," *IEEE Access*, 2017.
[15] Goodfellow, I., Bengio, Y., Courville, A., "Deep Learning," MIT Press, 2016.
[16] Obermeyer, Z., Emanuel, E., "Predicting the Future — Big Data, Machine Learning and Clinical Medicine," *New England Journal of Medicine*, 2016.
[17] Kourou, K., et al., "Machine Learning Applications in Cancer Prognosis and Prediction," *Computational and Structural Biotechnology Journal*, 2015.
[18] UCI Machine Learning Repository, "PIMA Indians Diabetes Database," University of California, Irvine.
[19] Brownlee, J., "Machine Learning Mastery with Python," Machine Learning Mastery, 2020. [20] Scikit-learn Developers, "Scikit-learn Documentation," Python Software Foundation, 2023.