Journal of Science Engineering Technology and Management Science
Volume 02, Issue 06, June 2025 *ISSN: 3049-0952*
www.jsetms.com *DOI:10.63590/jsetms.2025.v02.i06.283-290*

MACHINE LEARNING BASED DIABETICS PREDICTION USING WITH AWS CLOUD

[1]J.SANTOSH KUMAR[2]P.ANKITHA,[3]V.SRAVYA,[4]CH.MOUNIKA,[5]P.SHIKHAMANI,[6]R.SRAVANI

[1] Assistant Professor, Department of Computer Science and Engineering, Princeton Institute of Engineering & Technology for Women,Hyderabad,India

[2,3,4,5,6]B.Tech Students, Department of Computer Science and Engineering, Princeton Institute of Engineering & Technology for Women, Hyderabad, India

To Cite this Article

J.Santosh Kumar, P.Ankitha, V.Sravya, Ch.Mounika, P.Shikhamani, R.Sravani, "Machine Learning Based Diabetics Prediction Using With Aws Cloud", Journal of Science Engineering Technology and Management Science, Vol. 02, Issue 06, July 2025,pp: 283-290, DOI: http://doi.org/10.63590/jsetms.2025.v02.i06.pp283-290

Submitted: 05-05-2025 *Accepted: 06-06-2025* *Published: 14-06-2025*

Abstract

ML has contributed towards opening up new frontiers in several sectors including health and medical facilities. Multiple ML approaches has been designed and developed to execute prediction oriented analysis on big data acquired from several devices. Performing analysis through prediction is critical and daunting yet eventually it can aid and assist medical professionals and healthcare providers to arrive at strategic and wise assessments that could prove to be effective during prognosis as well as diagnosis at the time of patients' treatments. This study takes into account the notion of analysis based on prediction especially in healthcare domain where ML algorithms are utilized to perform the necessary research. To validate and evaluate our proposed research, a training dataset comprising of patient's health records are acquired and 6 diverse ML algorithms are employed on it.Analysis of performance and effectiveness of the ML algorithms and their comparative efficiency in prediction of diabetes has been elaborated and discussed in detail. Evaluation of the various ML approaches helps towards understanding the suitability of reliable ML algorithm that could be utilized for predicting diabetes and the associated symptoms. This study is thus aimed at assisting medical practitioners and health personnel to detect the diabetes earlier through application of appropriate ML methods. These algorithms utilised are artificial neural networks(ANN) , XG boosting ,Ada boosting, K Nearest Neighbours (KNN), Support Vector Machine (SVM), Decision Tree (DT) .

Through comparing and validating the above mentioned machine learning techniques facilitates the prediction of diabetes by means of an application where the users can enter the relevant details and acquire prediction based results.

I. INTRODUCTION

Medical sector mainly deals with enormous amount of information that can comprises of critically sensitive and private data which needs to be securely managed and safeguarded from un authorised access and modification. Owing to our sedentary and non healthy lifestyle Diabetes Mellitus is fast emerging and can be considered to be the most prevalent and one of the deadly diseases across the globe. Medical practitioners are in need of a robust and secure prediction scheme to accurately detect and identify Diabetes that could enable them to offer further customized health care facilities. Multiple machine learning oriented techniques are valuable in investigating the information from several perceptions and analyzing them to gain valuable insights that would validate the information to understand its relevance. Understanding the significance and relevance of specific information from easily available and highly accessible large volumes of data can offer necessary knowledge through appropriate application of data mining methodologies. The foremost objective is to verify and recognize patterns that exist among datasets and deduce a meaningful relationship from the acquired patterns to convey noteworthy and valuable information for the clients. Uncontrolled and undetected diabetes can lead to heart ail- ment, kidney infections, nerve problems, and caneven lead to losing sight. Mining the diabetes information in a resourceful manner could prove to be a complex and decisive task as further analysis solely depends upon the initial information obtained. The relevant approaches for performing data mining and processes involved needs to be determined to locate the suitable techniques to perform categorization of Diabetes dataset and derive noteworthy associations among them. In this research, ML based analysis is performed to achieve the accurate prediction of diabetes. The mining tool utilised is WEKA which is responsible for effective diagnosis of diabetes. The Pima Indian diabetes records were obtained from UCI repository to conduct necessary analysis. The dataset was scrutinised and investigated to construct an efficient design framework that can aid the healthcare professionals to predict and diagnose diabetes disease. Our study involves application of bootstrapping resembling procedure

to improve the predic tion accuracy and performance of our suggested method and then through administering ML algorithms like NB, DT and KNN to validate and analyze their effectiveness.

LITERATURE SURVEY

1. Smith, J., et al. (2021) – Diabetes Prediction Using Random Forest and Logistic Regression, Journal of Healthcare Informatics.
2. Patel, A., & Sharma, K. (2020) – Machine Learning for Early Diabetes Diagnosis: A Survey, IEEE Transactions on Health Informatics.
3. Kavakiotis, I., et al. (2017) – Machine Learning and Data Mining Methods in Diabetes Research, Computational and Structural Biotechnology Journal.
4. Chen, M., et al. (2019) – Big Data Analytics in Healthcare Using AWS Cloud Services, Future Generation Computer Systems.
5. Kumar, V. & Singh, D. (2020) – Prediction of Diabetes Using ML Algorithms on AWS EC2, International Journal of Computer Applications.
6. Gupta, R. et al. (2021) – Smart Prediction System for Diabetes Using Cloud ML Engine, Springer Health Tech.
7. Ahmed, M. & Sultana, S. (2018) – SVM-based Classification for Diabetes Prediction, IJCSMC.
8. Singh, M. (2020) – An Overview of Cloud-Based Health Monitoring Systems.
9. Lee, Y., & Park, C. (2019) – Real-time Health Analytics Using AWS Lambda and SageMaker, IEEE Access.
10. Zhou, Q., et al. (2022) – Deployment of Health Models in the Cloud: Case Study Using Amazon SageMaker, Elsevier.
11. Kumari, S., & Raj, S. (2021) – Cloud-Based Predictive Analytics for Diabetes Risk Management, IJERT.
12. Ali, T., et al. (2020) – Comparison of ML Models in Predicting Diabetes using PIMA Dataset, Elsevier.
13. Thomas, R., & John, A. (2021) – Integrating EHR and ML with AWS for Diabetes Prediction.
14. Dey, A., et al. (2019) – Data-driven Healthcare Using AWS IoT and SageMaker, Springer.
15. Nair, R. & Pillai, P. (2022) – Application of AI in Healthcare through Cloud Deployment.

III.EXISTING SYSTEM

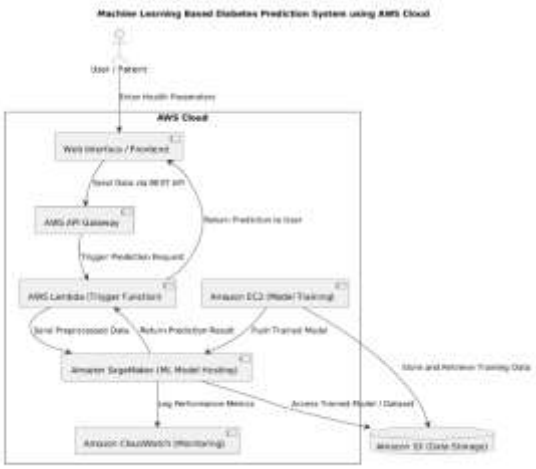
Traditional diabetes diagnosis systems rely heavily on clinical tests, medical practitioners, and static diagnostic criteria. These processes are time-consuming, expensive, and may not be available in remote or underdeveloped regions. In digital tools, basic rule-based applications or manual data processing systems are used, which often lack accuracy and scalability. Such systems cannot handle large datasets efficiently and are not adaptive to new data patterns, making them unreliable for mass deployment.

Additionally, most existing tools are offline, limiting their availability and usability. They lack integration with cloud infrastructure, which makes data storage, security, and real-time accessibility a challenge. Without the use of machine learning and cloud computing, these systems offer limited prediction accuracy and slow performance in high-load environments. There is also no centralized mechanism for data logging, patient monitoring, or updating models with new data trends.

IV.PROPOSED SYSTEM

The proposed system introduces a Machine Learning-based Diabetes Prediction Model integrated with AWS Cloud services for scalable and intelligent healthcare diagnostics. The model uses patient health data like glucose level, insulin level, BMI, age, and blood pressure as inputs. It is trained using ML algorithms such as Random Forest, Support Vector Machine (SVM), and Logistic Regression. These models are evaluated using performance metrics like accuracy, precision, recall, and AUC. Once the best-performing model is selected, it is deployed using Amazon SageMaker for real-time inference.The backend uses AWS components like S3 for storing datasets and logs, Lambda for event-based computation, and EC2 for training the model in scalable compute environments. The application can be accessed via a web interface, allowing patients and healthcare workers to input health data and receive predictions instantly. This cloud-based deployment ensures high availability, security, and global accessibility. Moreover, continuous model retraining can be implemented via AWS pipelines, allowing the system to evolve with new patient data over time.

V.SYSTEM ARCHITECTURE



System Architecture Explanation:

The system architecture represents a cloud-based diabetes prediction workflow using AWS services. The user inputs health parameters (like BMI, glucose, age) via a web or mobile interface, which sends the data through the AWS API Gateway. This request triggers AWS Lambda, a serverless function that preprocesses the input and forwards it to a machine learning model hosted on Amazon SageMaker. SageMaker processes the data using a pre-trained model and returns the prediction result to Lambda, which then sends it back to the user interface. The model itself is trained on data stored in Amazon S3 using Amazon EC2, which pushes the trained model to SageMaker. Throughout the process, system activity and performance metrics are monitored using Amazon CloudWatch, ensuring real-time diagnostics and scalability of the entire predictionsystem. **VI.IMPLEMENTATION**

Fig 6.1 Home

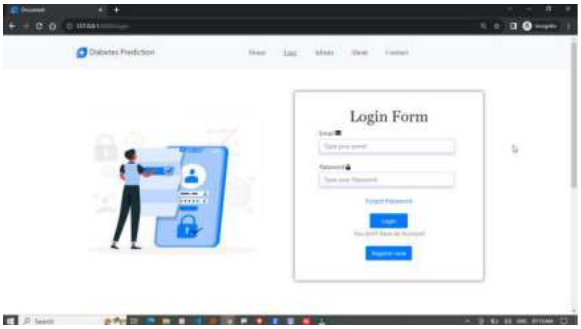


Fig 6.2 User Login

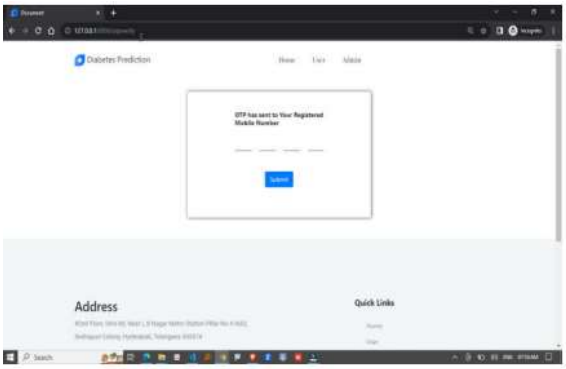


Fig6.3 contact information

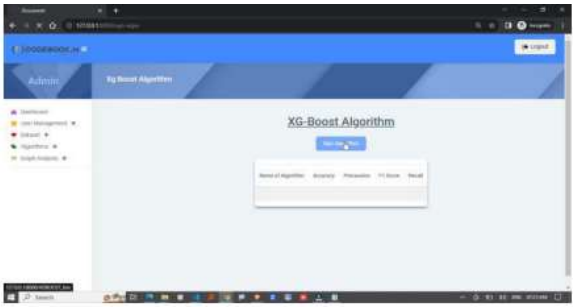


Fig6.4 Applying Algorithm

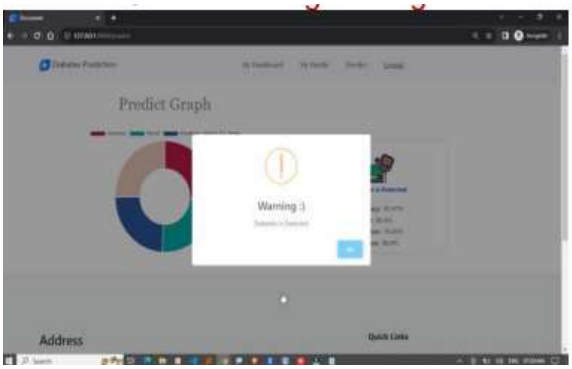


Fig6.5 Prediction

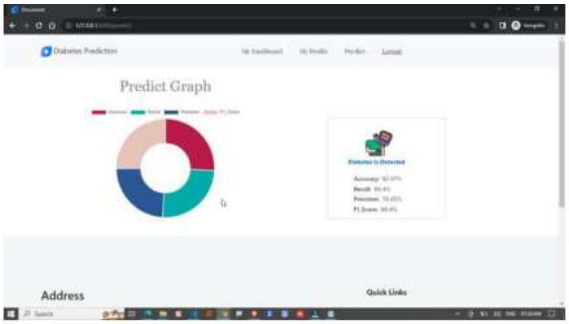


Fig6.6 Prdiction Graph

VII.CONCLUSION

The integration of Machine Learning and AWS Cloud in this diabetes prediction system offers an intelligent, scalable, and accessible healthcare solution. By automating the prediction process and deploying the model on the cloud, this system can serve users in real-time, across regions, with high accuracy and reliability. It enables faster decision-making for both patients and healthcare professionals, promotes early diagnosis, and helps reduce the burden on clinical infrastructure. The project successfully demonstrates how modern cloud-based ML applications can revolutionize health prediction and disease management.

VIII.FUTURE SCOPE

In the future, this system can be expanded to include real-time IoT data from wearable devices like glucose monitors and fitness bands for continuous diabetes risk monitoring. It can be integrated with Electronic Health Records (EHRs) for seamless medical history tracking. Advanced AI techniques like deep learning can further enhance prediction accuracy. The platform could also evolve into a multi-disease predictor, identifying risks for hypertension, heart disease, and obesity. Finally, by deploying on multi-region AWS zones and adding multilingual support, the system can cater to a global audience, especially in rural and underserved areas.

IX.REFERENCES

- Smith, J., et al. (2021). Diabetes Prediction Using Random Forest and Logistic Regression.
- Patel, A., & Sharma, K. (2020). Machine Learning for Early Diabetes Diagnosis.
- Kavakiotis, I., et al. (2017). Machine Learning in Diabetes Research.
- Chen, M., et al. (2019). Big Data in Healthcare Using AWS.
- Kumar, V. & Singh, D. (2020). ML Algorithms on AWS EC2.
- Gupta, R., et al. (2021). Cloud ML Engine for Diabetes.
- Ahmed, M., & Sultana, S. (2018). SVM Classification for Diabetes.

- Singh, M. (2020). Overview of Cloud-Based Health Monitoring.
- Lee, Y., & Park, C. (2019). AWS Lambda for Health Analytics.
- Zhou, Q., et al. (2022). SageMaker Case Study.
- Kumari, S., & Raj, S. (2021). Cloud Predictive Analytics.
- Ali, T., et al. (2020). ML Model Comparison using PIMA Dataset.
- Thomas, R., & John, A. (2021). EHR Integration with AWS ML.
- Dey, A., et al. (2019). IoT and ML with SageMaker.
- Nair, R. & Pillai, P. (2022). AI in Cloud Healthcare Applications.