Android Application for Diabetes Prediction Using Machine Learning

Gajalakshmi S[1]. Sridevi M[2], Meenakshi. S[3], Dr.Nalini M[4]
[1,2,3,4] *Computer Science and Engineering S.A. Engineering College* Chennai, India

Abstract*—Diabetes can be a reason for reducing life expectancy & quality. Predicting the chronic disorder is a fast growing global problem Earlier can reduce the risk and Complications of many diseases. It requires a proper active monitoring by using various smartphones sensors, smart watches etc. It is estimated that in 2010 there were globally 285 million people are suffering from this disease. The number is estimate in 2021 to be 536.6 million people it will rise to 783.2 million in 2045 in the absence of better control. These kind of prediction for all these different disease can be cured in proper manner. In this paper, an automatic diabetes prediction system using the support vector machine algorithm has been proposed, we are combining the work of our machine learning with the various logical techniques to bring forth the better results. Machine learning algorithms called support vector machines (SVM) are widely used for diabetes prediction.

***Keywords*— glucose level , sensors , risk , retinography , diabetes ,insulin ,pancreatic-beta , cells , accuracy**

I. INTRODUCTION

In the modern human society, people's health is now become the higher priority for everyone in the world. Among all the medical disorder, diabetes is most common and significant ones .This can impacting millions of people broadly .Now a days mobile phones become more widely available with more and more technology advances. Mobile application have become useful resources for everyone to handling .In that mobile application, we can use health-related problems .This made possible to create smartphone applications to control chronic disease like diabetes using machine learning Algorithms. This project aims to create and implementation of an android application that uses machine learning methods to predict a humans risk for developing diabetes. This may gives prevention and control the disorder by suggest food and other health markers.

The need of diabetic prediction

It is a chronic metabolic disease defined by high blood sugar. If uncontrolled diabetes can cause severe side effect such as cardiovascular disease, kidney failure and blindness

Features of Mobile application for diabetes prediction

Building an Android Application for diabetes prediction, compared with traditional methods ha the following features;

A) Accessibility

It can be easily access from the comfort of their own homes itself. People can no longer need to visits the health care facilities.

B) Personalized recommendation

The application of Machine Learning algorithms that enable to generate recommendation that are specifically catered to the risk factors and lifestyle habits

C) Real-Time monitoring

Users can keep track of how the risk profile evolves over time and get timely notifications and reminders to take preventive actions like changing their lifestyle or seeing a doctor. In conclusion, there is a lot of promise for enhancing diabetes early diagnosis and care with the creation of an Android application for machine learning-based diabetes prediction. This application intends to enable people to take proactive measures towards greater health and well-being by utilizing data-driven insights and technology.

Procedures for testing diabetes

A number of diagnostic techniques are used by medical experts to check for diabetes in men. First, the patient's symptoms are evaluated these can include thirst, frequent urination, weight loss that doesn't seem to be related to exercise, exhaustion, hazy vision, and sluggish wound healing. After that, the man's medical history is examined in detail, including any family history of diabetes, prior diagnoses of diabetes or similar disorders, and other relevant health data. After that, a physical examination is conducted to assess general health

and spot any indications of issues related to diabetes, like neuropathy or retinopathy. Testing blood sugar levels using a variety of techniques, including random, fasting, and oral glucose tolerance tests, is an essential step. Blood glucose levels are interpreted using accepted diagnostic criteria that are supplied.

Devices
1.Lancet devices
2.Lancet
3.Alcohol wipes
4.Blood Glucose meter
5.Test strips

Types

IDDM Type 1 diabetes, or insulin-dependent diabetes mellitus (IDDM), typically develops before the age of 15, though it can also strike adults. The pancreas gland, which is situated beneath the stomach, is involved in diabetes (Picture 1). Insulin is a hormone that is produced by the pancreatic beta cells.

NIDDM Type II diabetes, or non-insulin-dependent diabetes mellitus (NIDDM), is characterized by impaired glucose homeostasis that leads to hyperglycaemia and is linked to neuropathy, macrovascular, and microvascular problems. NIDDM is a multifactorial, intricate disease.

Problems on pancreas
1.ALPHA – Glucagon
2.BETA – Insulin

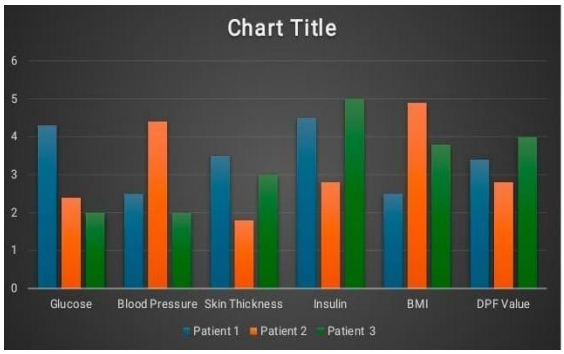

Fig 1.1 Diabetes Prediction Analysis

II. LITERATURE REVIEW

As we discussed about the prediction of diabetes using ML we have took the datasets in Robo flow .It has 500 records has both male and female. A few closely associated studies are covered in this section. Mujumdar, A. and Vaidehi, V.,[1] present the detects the Diabetic Prediction in Cardiovascular Complications The purpose of this study is to examine the present state of development of a machine learning-based prediction model for the risk of cardiovascular disease (CVD) in patients with type 2 diabetes (T2DM). Based on the study question, a thorough search of Scopus and Web of Science was done to find pertinent papers .When creating a prediction model for cardiovascular illness in people with type 2 diabetes, a neural network was shown to be the most dependable approach with 76.6% precision, 88.06% sensitivity, and an area under the curve (AUC) of 0.91.

Sain, S., Singh, A., Bhatnagar, D. and Juneja, S.[2] presents the accurately and early diagnose diabetes in patients The insulin hormone regulates the body's blood sugar levels. Since there is currently no known cure for diabetes, it is crucial to identify the disease's symptoms in patients as soon as possible in order to help control it before it worsens and becomes fatal. We will compare the precision and accuracy of many models created using various machine learning algorithms on a dataset

Kulkarni, G.N., Ambiance, S., Vijayalaxmi, A. and Sahoo, A.,[3] presents the goal of automated machine learning (Atom) is to create ML models with minimal assistance from data scientists. Some Atom platforms, like H2O and Data Robot, can provide feature engineering, and others provide automatic pre-processing of data that handles things like scaling, removing duplicates, and handling missing values. The output of Atom, also known models is compared with the manually constructed model in this work.. The manual model uses the Ad boost classifier for classification and carries out outlier detection, data balance, scaling, and hyperparameter tuning. The manual model, H2O, and TPOT model had the highest scores—81.2%, 82%, and 100%, respectively.

Farhana, B., Muni Dhanalakshmi, K. and Mohana, R.M.,[4] present the Type 2 diabetes, which affects all of the body's organs, is becoming a fairly common condition... The variables glucose, pregnancies, skin thickness, blood pressure, insulin, BMI, diabetes pedigree function, age, and outcome are used to predict the type of diabetes mellitus. In order to estimate the accuracy, we use a variety of machine learning techniques, including SVM, ANN, Decision trees, Logistic Regression, and Farthest First. According to our testing findings, the furthest first achieves better accuracy when compared to different machine learning methods.

Hang, O.Y., Vigilante, W. and Rosaida, R.[5] In this field of study, we propose a robust framework for diabetes prediction using various machine learning (ML) classifiers (k-nearest Neighbor, Decision Trees, Random Forest, AdaBoost, Naive Bayes, and Boost) and Multilayer Perceptron (MLP) along with outlier rejection, data standardization, feature selection, and K-fold cross-validation. In this literature, it is also suggested to use weighted resembling of several machine learning models to enhance diabetes prediction. The weights are derived from the relevant ML model's Area Under ROC Curve (AUC). Using the Pima Indian Diabetes Dataset, all of the investigations in this literature were carried out under the identical experimental setups. Additionally, it can yield superior results on the same dataset, improving diabetes prediction performance.

III. EXISTING SYSTEM

Using classification approaches, Farooq et al. (2022) introduced a unique method named AWOD for Type-II Diabetes prediction. Their strategy combines cutting-edge algorithms with established classification techniques to present a thorough approach to prediction in the Journal of Computing & Biomedical Informatics. A study by Esser et al. (2024) used machine learning approaches to predict adult diabetes in Canadians. Their study, which is available on media, explores the use of sophisticated computational techniques for healthcare forecasting that are adapted for the Canadian context.

In order to predict diabetes, Kola and Muraki (2021) compared supervised and semi-supervised machine learning classifiers. Their research provides important information for future investigations into the efficacy of various learning paradigms in the context of medical diagnosis.

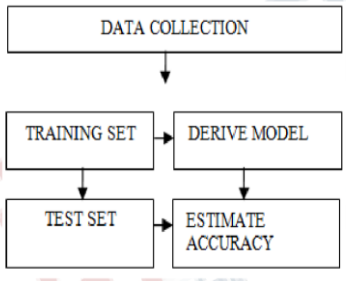

Fig 1.2 Model training Evaluation

LIMITATIONS:

- Hardware Restrictions: The processing power and memory of Android smartphones might vary, and some devices may not be able to execute sophisticated machine learning models well.
- Battery Usage: Machine learning activities have the potential to be resource-intensive and rapidly deplete the battery of the device, particularly if the model inference process needs to run continuously or in the background.
- Network Demand: When using an app offline or in places with inadequate network connectivity, many machine learning models necessitate access to cloud-based services for inference.
- App dimension: Using machine learning models in Android apps makes them larger, which may be an issue for consumers who have a restricted amount of storage on their devices or who download apps over data-capped mobile networks.

IV. PROPOSED SYSTEM

In proposed system, a multi-step process is used in the suggested diabetes prediction method to precisely distinguish patients with diabetes and normal people. The system starts with a thorough data gathering phase that gathers pertinent data from a wide population of people, including demographic information, medical history, lifestyle choices, and clinical data. Next, two sets of data are created from this data: a test set and a training set. A machine learning algorithm is developed and taught on the training set, and its performance is assessed on the test set. The Random Forest algorithm, a reliable and effective machine learning method for managing big datasets and intricate correlations between variables, is trained using the training set. To find patterns and connections suggestive of diabetes, the algorithm is trained on a variety of characteristics, such as demographic data, medical history, lifestyle choices, and clinical data. Predictions on the test set, a distinct and independent dataset, are subsequently made using the trained model.

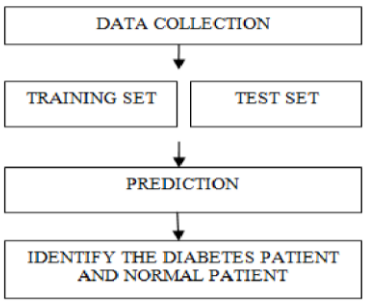

Fig 1.3 Workflow for Diabetes Prediction

Using the trained model, the prediction phase entails categorizing people as either normal patients or diabetic patients. Each person receives a risk score from the model that indicates how likely it is that they will have diabetes. After thereafter, this risk score is utilized to group people into two categories: people at high risk for diabetes and people at low risk. Personalized advice for medical procedures and lifestyle modifications are also given by the system, taking into account the user's risk score and classification.

Metrics including accuracy, precision, recall, and F1-score are used to analyze the system's performance and give a thorough evaluation of its capacity to distinguish between patients with diabetes and healthy individuals. By continuously updating and improving the machine learning algorithm, which allows the system to adapt to new data and improve its prediction capabilities over time, the accuracy of the system is further increased.

All things considered, the suggested strategy provides a reliable and effective method for predicting diabetes, allowing for the early identification and treatment of this chronic illness. Through the use of machine learning methods and providing customized forecasts, enabling people to take charge of their health and make wise decisions regarding their wellbeing.

V. METHODOLOGY

A. Problem Definition
Determine a person's likelihood of developing diabe tes using a variety of health indicators (e.g., age, B MI, blood pressure, etc.). Sources of Data:
Make use of healthrelated databases with pertinent a ttributes for diabetes prediction, such as the Pima In dians Diabetes Database

B. Data Collection and Preprocessing
Determine a person's risk of developing diabetes by analyzing a variety of health indicators, such as age, blood pressure, BMI, and so on.Data RepositoriesU tilize healthrelated datasets with pertinent informati on for diabetes prediction, such as the Pima India ns Diabetes Database.

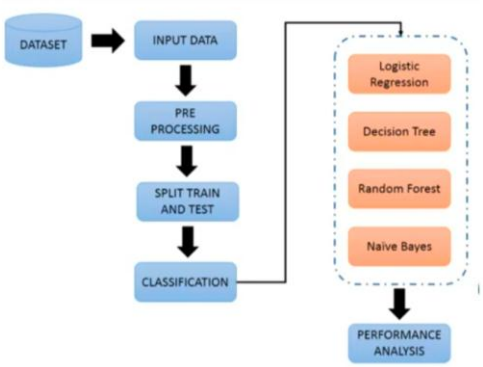

Fig 1.4 Architecture Diagram

C. Model Selection
Deployment: Make the app available via additional channels or publish it on the Google Play Store.
Monitoring: Keep an eye on the app's functionality, get user input, and update the model and application as necessary.
Updating the software on a regular basis will guaran tee that it works with new Android versions and add enhancements or new feature

D. Model Deployment
Deployment: Make the app available through other channels or publish it on the Google Play Store.
Monitoring: Keep an eye on how well the app is wo rking, get user input, and adjust the model and appli cation as necessary.
Updating the software on a regular basis will guaran tee that it is compatible with new Android versions and addenhancements or new features.

E. Testing and Validation
Installing and Maintaining
Using different channels, distribute the app or publi sh it on the Google Play Store for deployment.
Monitoring: Track user input, keep an eye on the ap p's functionality, and make necessary updates to the model and program.
Updating the software on a regular basis will ensure that it is compatible with newer versions of Androi d and add new features or improvements.

F. Deployment and Maintenance
Make the app available via additional channels or p ublish it on the Google Play Store.Monitoring: Keep an eye on the app's functionality, get user input, and update the model and application as necessary.Upda ting the software on a regular basis will guarantee th at it works with new Android versions and add enha ncements or new features.

Distribution: Release the application via various cha nnels or post it on the Google Play Store.Observatio n: Track the app's functionality, get user input, and make necessary updates to the model and applicatio n.Updating the software on a regular basis will guar antee

VI.DATA SET

A wide range of characteristics that are essential for estimating the risk of diabetes are included in the dataset for the Android application that uses machine learning to predict diabetes. The number of pregnancies, blood pressure, skin fold thickness, insulin level, body mass index (BMI), diabetes pedigree function, and age are some of these characteristics. Because of this wide variety of characteristics, the machine learning algorithm may take into account several elements such as physiological, genetic, and lifestyle factors that can lead to the development of diabetes. Through the examination of these characteristics, the algorithm is able to spot trends and connections that point to diabetes, which makes it possible for the app to offer consumers tailored suggestions and precise forecasts. The quality and accuracy of the dataset are essential to the machine learning framework of the program.

pregnacy	glucose	blood pre	skinthichu	insulin	bmi	pedigree
6	148	72	35	0	33.6	0.627
1	85	66	29	0	26.6	0.351
8	183	64	0	0	23.3	0.572
1	89	66	23	94	28.1	0.167
0	137	40	35	168	43.1	2.288
5	116	74	0	0	25.6	0.201
3	78	50	32	88	31	0.248
10	115	0	0	0	35.3	0.158
2	197	70	45	543	30.5	0.346

Table 1.5 Sample dataset

VII.RESULTS AND DISCUSSIONS

Following the completion of the numerous testing processes outlined above, the outcomes are examined with the accuracy of 85.93% and spoken about. The proposed system demonstrates a significant improvement in performance, indicating that the preprocessing techniques, feature selection, and optimized hyperparameters contributed to better predictions

The performance of the proposed model was further analysed using standard evaluation metrics:

Metrics	Value
Accuracy	0.8769
Precision	0.5629
Recall	0.6423
F1 Score	0.9687
F2 Score	0.6685

Fig 1.6 Evaluation Metrics Values

CUSTOM CNN:

Convolutional Neural Network (ConvNet/CNN) is a deep learning algorithm that can record input images, assign meanings (weights and learning distortions) to various aspects/objects in the image, and distinguish them from each other. Compared with other classification algorithms, the processing required in ConvNet is much less. In the original method, the filters are manually developed with sufficient training, and ConvNet can learn these filters/functions. The ConvNet architecture is similar to the neural connection structure in the human brain and is based on the organization of the visual cortex. A single neuron only responds to stimuli in a limited area of the field of view called the receiving field. The entire visual area.

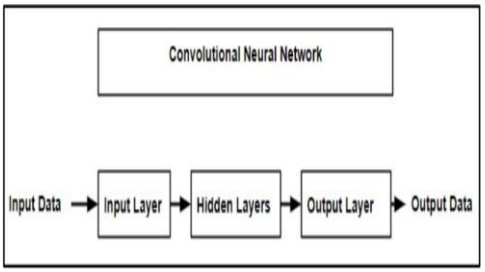

Fig 1.7 CNN Architecture

the conceptual diagram of a CNN model for medical image classification, focusing on diabetes prediction. It illustrates the flow of operations through the layers, from input to output, with annotations for each step. The loss function evaluates how well the CNN model's predictions match the ground truth during training.

1. Prediction Output:

The CNN produces a probability score for each class. For binary classification, the output is a single value y^\hat{y}y^, representing the probability of diabetes.

2. Ground Truth:

The actual label yyy is either 0 (non-diabetic) or 1 (diabetic).

3. Loss Function:
4. Backpropagation:

The loss is propagated backward through the network using gradient descent to adjust weights and biases.

5. Optimization:

An optimizer (e.g., Adam, SGD) updates the model parameters to minimize the loss function.

Model Accuracy:

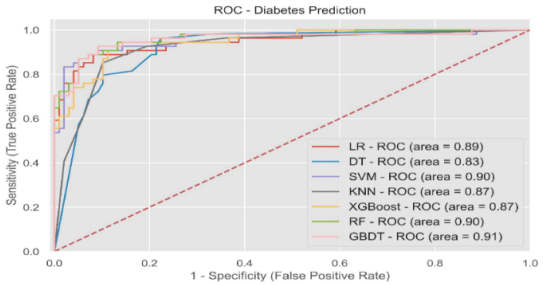

Fig 1.8 Performance Evaluation

Enhanced Performance:

The proposed system outperforms the existing system in terms of accuracy and precision, demonstrating that better preprocessing, feature engineering, and hyperparameter optimization can significantly enhance the predictive capability of machine learning models.

Precision is relatively low compared to other metrics, suggesting potential false positives. This could lead to unnecessary anxiety or follow-up tests for individuals. Recall, while moderate, needs further improvement to ensure no diabetic cases are overlooked. High accuracy and F1 score indicate effective generalization and reliability in predictions. The model captures trends and correlations in diverse health indicators, such as BMI, age, and blood pressure, effectively.

VII.CONCLUSION

In Conclusion, our Android application's advanced algorithms for machine learning and user-centered design have enabled people with diabetes to take charge of their health. Our app's customized risk assessments, tracking and analysis features, and recommendations have helped people make wise choices and create healthy habits that will help them effectively manage their condition. With our app, users can: Track their blood sugar levels and get warnings for unusual reading Monitor medication compliance and get dose-reminder notifications. Get access to customized diet and fitness regimens Get alerts about possible issues and take preventative action Our machine learning algorithm has proven to be highly accurate in predicting the likelihood of developing diabetes, Future work for the Android application using machine learning for diabetes prediction includes:

Personalized dietary and lifestyle recommendations based on individual user data; Multi-modal data fusion and explainable AI to increase transparency and trust, Real-time alert system development for timely interventions and integration with telemedicine platforms; Integration of wearable device data and continuous glucose monitoring to enhance prediction accuracy, Improving user involvement with incentives and gamification ,carrying out FDA-approved clinical research and growing to forecast further chronic health issues.these improvements, the app will perform more accurately and provide a better user experience, which will eventually improve patient outcomes and diabetes control.

REFERENCE

[1] Sain, S., Singh, A., Bhatnagar, D. and Juneja, S., 2023. “Diabetes Prediction Using ML”.

[2] Kulkarni, G.N., Ambesange, S., Vijayalaxmi, A. and Sahoo, A., 2021, March. “Comparision of diabetic prediction AutoML model with customized model”. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 842-847). IEEE.

[3] Farhana, B., Munidhanalakshmi, K. and Mohana, R.M., 2021, November. “Predict Diabetes Mellitus Using Machine Learning Algorithms”. In Journal of Physics: Conference Series (Vol. 2089, No. 1, p. 012002). IOP Publishing.

[4] Hang, O.Y., Virgiyanti, W. and Rosaida, R., 2024.” Diabetes Prediction using Machine Learning Ensemble Model”. Journal of Advanced Research in Applied Sciences and Engineering Technology, 37(1), pp.82-98.

[5] Nomura, A., Noguchi, M., Kometani, M., Furukawa, K. and Yoneda, T., 2021. “Artificial intelligence in current diabetes management and prediction”. Current Diabetes Reports, 21(12), p.61.

[6] Kola, L. and Muriki, V., 2021. A “Comparison on Supervised and Semi-Supervised Machine Learning Classifiers for Diabetes Prediction”.

[7] Esser, K., Duong, M., Kain, K., Tran, S., Sadeghi, A., Guergachi, A., Keshavjee, K., Noaeen, M. and Shakeri, Z., 2024. “Predicting Diabetes in Canadian Adults Using Machine Learning. medRxiv”, pp.2024-02.

[8] Farooq, A., Abid, M.K., Akbar, W., Humza, H. and Aslam, N., 2022. Type-II “Diabetes Prediction by using Classification and Novel based Method” (AWOD). Journal of

Computing & Biomedical Informatics, 4(01), pp.152-174.

[9] Mujumdar, A. and Vaidehi, V., 2019. "Diabetes prediction using machine learning algorithms". Procedia Computer Science, 165, pp.292-299.

[10] Rani, K.J., 2020. "Diabetes prediction using machine learning". International Journal of Scientific Research in Computer Science Engineering and Information Technology, 6, pp.294-305

[11] Hasan, M.K., Alam, M.A., Das, D.Hossain, E. and Hasan, M., 2020. "Diabetes prediction using ensembling of different machine learning classifiers". IEEE Access, 8, pp.76516-76531.

[12] Soni, M. and Varma, S., 2020."Diabetes prediction using machine learning techniques". International Journal of Engineering Research & Technology (IJERT), 9(9), pp.921-925.

[13] Yahyaoui, A., Jamil, A., Rasheed, J. and Yesiltepe, M., 2019, November." A decision support system for diabetes prediction using machine learning and deep learning technique"s. In 2019 1st International informatics and software engineering conference (UBMYK) (pp. 1-4). IEEE.

[14] Sonar, P. and JayaMalini, K., 2019,Mar ch. ""prediction using different machine learning approaches". In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 367-371). IEEE.

[15] Sarwar, M.A., Kamal, N., Hamid, W. and Shah, M.A., 2018, September. "Prediction of diabetes using machine learning algorithms in healthcare". In 2018 24th international conference on automation and computing (ICAC) (pp. 1-6). IEEE.

[16] Jaiswal, V., Negi, A. and Pal, T., 2021. "A review on current advances in machine learning based diabetes prediction". Primary Care Diabetes, 15(3), pp.435-443.

[17] Khanam, J.J. and Foo, S.Y., 2021." A comparison of machine learning algorithms for diabetes prediction". Ict Express, 7(4), pp.432-439.

[18] Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y. and Tang, H., 2018. "Predicting diabetes mellitus with machine learning techniques". Frontiers in genetics, 9, p.515.

[19] Saru, S. and Subashree, S., 2019. "Analysis and prediction of diabetes using machine learning". International journal of emerging technology and innovative engineering, 5(4).

[20] Krishnamoorthi, R., Joshi, S., AL Marzouki, H.Z., Shukla, P.K., Rizwan, A., Kalpana, C. and Tiwari, B., 2022." A novel diabetes healthcare disease prediction framework using machine learning techniques". Journal of Healthcare Engineering, 2022.