5G-Smart Diabetes and Healthcare Big Data Clouds for Individualized Diabetes Diagnosis

E. Soumya, Asst. Professor, Department of IT, Hyderabad, India, esoumyait@smec.ac.in

P.Devasudha, Asst. Professor, Department of CSE, Hyderabad, India, pdevasudhacse@smec.ac.in

P. Sabitha Reddy, Asst. Professor, Department of CSE. Hyderabad, India, psabithacse@smec.ac.in

ABSTRACT

Recent advancements in wireless networking and big data technologies, such as 5G networks, medical big data analytics, and the Internet of Things, as well as recent developments in wearable computing and artificial intelligence, have made it possible to develop and implement innovative diabetes monitoring systems and applications. Diabetes patients experience long-term and systematic suffering as a result of their disease. It is vital to develop better diabetes diagnosis and treatment strategies for patients. Based on our thorough analysis, this paper divides those techniques into Diabetes 1.0 and Diabetes 2.0, both of which have networking and intelligence flaws. As a result, our goal is to provide a long-term, cost-effective, and intelligent diabetes diagnosis and treatment solution. In this article, we introduce the 5G-Smart Diabetes system, which combines cutting-edge technologies as wearable 2.0, machine learning, and big data to provide comprehensive sensing and analysis for diabetic patients. The data sharing mechanism and customized data analysis approach for 5G-Smart Diabetes are then presented.

KEYWORDS: Big data, Cloud, Diabetes, Machine Learning

INTRODUCTION

Diabetes is a condition that results in insulin shortage due to a lack of insulin in the blood. Frequent urination, thirst, and increased hunger are all warning signs of high blood sugar. It will cause a lot of problems if it isn't treated. This adversity resulted in death. Cardiovascular illness, foot sores, and eye blurriness are all symptoms of severe difficulty. Prior diabetes is characterized by an increase in blood sugar levels. As a result, the previous diabetes is not as valuable as the usual value. Diabetes is caused by the exocrine gland not producing enough hypoglycaemic agents or failing to respond appropriately to the hypoglycaemic agent produced. Different information mining algorithms present various decision support methods for supporting health professionals. The accuracy of the decision support system is a measure of its effectiveness. As a result, the goal is to create a decision support system that can accurately forecast and diagnose a certain condition. ML is a subfield of AI that handles real-world problems by "offering learning capability to workstations without the need for additional programme programming."

Types of Diabetes

1) The failure of the pancreas to supply adequate hypoglycemic agent results in type one diabetes. "Insulin-dependent polygenic disease mellitus" (IDDM) or "juvenile diabetes" were terms used to describe this kind. The cause has yet to be determined. The type one polygenic disease that affects youngsters under the age of twenty. People with type one diabetes suffer for the rest of their lives and rely on insulin vaccinations. Diabetic individuals

must frequently adhere to doctor-recommended workouts and fitness regimens.

2) Hypoglycemic agent resistance, a condition in which cells fail to respond to hypoglycemic medicines effectively, is the beginning of type 2 diabetes. The illness arises as a result of a lack of hypoglycemic agent, which has also built up. "Non-insulin-dependent polygenic disease mellitus" was the term used to describe this kind of diabetes. Excessive weight is the most common cause. By 2025, the number of people affected by type two diabetes will have increased. When compared to the urban zone, the prevalence of diabetes mellitus is reduced by 3% in rural areas. Bulkiness, fatness, and diabetes mellitus are all linked to pre hypertension. According to the research, each UN agency has its own traditional vital sign.

3) When a woman is pregnant and has high blood sugar levels without a history of diabetes, she is diagnosed with type 3 gestational diabetes. As a result, it has been discovered that 18% of pregnant women have diabetes. As a result, as people get older, they are more likely to develop gestational diabetes during pregnancy. Obesity is one of the most common causes of type 2 diabetes. Type-2 polygenic illness can be managed with proper exercise and adherence to a regimen. When higher measures fail to lower aldohexose levels, medicines are frequently prescribed. According to the polygenic disease static report, diabetes affects 29.1 million persons in the United States.

Diabetic patients' conditions are monitored using 5G technology at a reasonable cost. Many people are suffering from diabetic disease these days as a result of work stress or unhealthy lifestyles, and people will not be aware of their current health situation until symptoms appear or a diagnosis is made through a medical check-up, by which time the disease will be severe and there will be no way to know about it beforehand.

Diabetes will be of two type's diabetes 1 and diabetes 2. Diabetes 2 require hospitalization and in diabetes 1 condition we can monitor patient and alert him or doctors about his current condition using below techniques:

1) cost
2) Comfortable
3) Sustainability
4) Personalization
5) Smartness

Here we design two applications to implement above technique

Cloud Application: Diabetic patients' status are monitored at a low costutilising 5G technology. Many people are developing diabetic disease as a result of work stress or unhealthy lifestyles these days, and they will not be aware of their current health situation until symptoms appear or a diagnosis is made through a medical check-up, by which time the disease will be severe and there will be no way to know about it beforehand.

User Application: We will upload some test data to this application, which will be considered user sense data, and this data will be sent to a cloud server, which will use decision and SVM and ANN models on the test data to predict patient status and provide the results to this application. We regard uploaded test data to be sense data because we don't have sensors to sense data. Because we don't have user information to share data, I've left all forecasted data accessible for all users to see and share.

LITERATURE SURVEY

Veena Vijayan V. and Anjali C has discussed, the diabetes disease produced by rise of sugar level in the plasma. Various computerized information systems were outlined utilizing classifiers for anticipating and diagnosing diabetes using decision tree, SVM, Naive Bayes and ANN algorithms.

P. Suresh Kumar and V. Umatejaswi has presented the algorithms like Decision Tree, SVM, Naive Bayes for identifying diabetes using data mining techniques.

RidamPal ,Dr.JayantaPoray and Mainak Sen has presented the Diabatic Retinopathy (DR) which is one of the leading cause of sight inefficiency for diabetic patients. In which they reviewed the performance of a set of machine learning algorithms and verify their performance for a particular data set.

Dr. M. Renuka Devi and J. Maria Shyla has discussed about the analysis of various skills of mining to guess diabetes using Naive Bayes, Random forest, Decision Tree and J48 algorithms.

Rahul Joshi and MinyechilAlehegn has discussed the ML techniques which are used to guess the datasets at an initial phase to save the life. Using KNN and Naive Bayes algorithm

ZhilbertTafa and NerxhivanePervetica has discussed the result of algorithms that are implemented in order to progress the diagnosis reliability.

Prof. Dhomse Kanchan B. and Mr. Mahale Kishor M. has discussed the study of Machine Learning Algorithms such as Support Vector Machine, Naïve Bayes, Decision Tree, PCA for Special Disease Prediction using Principal of Component Analysis.

PROPOSED METHODOLOGY

In the proposed system we use different algorithms and compare them to find the best algorithm that gives a very personalized and accurate diagnosis.
The block diagram of the diabetes prediction system is given below.

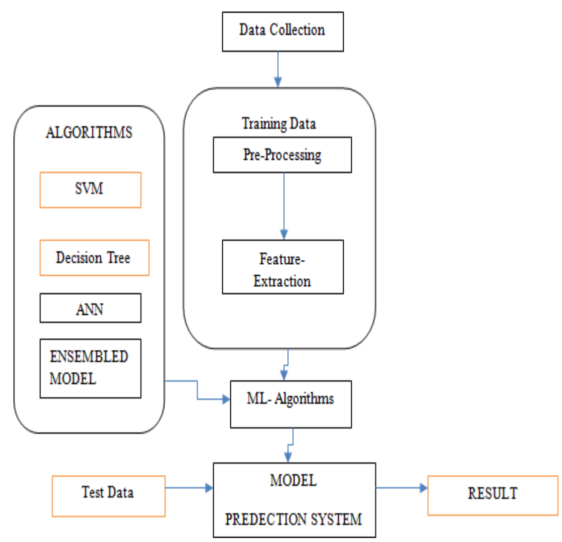

Fig 1: BLOCK DIAGRAM

Dataset collection:

The dataset contains seven hundred sixty-eight instances and nine features.

The dataset features are:

- Total number of times pregnant
- Glucose/sugar level
- Diastolic Blood Pressure
- Body Mass Index (BMI)

- Skin fold thickness in mm
- Insulin value in 2 hours
- Hereditary factor- Pedigree function
- Age of patient in years

Percentage split option is provided for training and testing. Out of 768 instances 75 % is used for training and 25% is used for testing.

Training and test data:

We divide the entire dataset in two parts: training data and testing data.

To develop a model, the training dataset is utilised to train the model. The algorithms learn and extract patterns from a training dataset. The model will be able to learn something from the features if the training set is labelled appropriately.

The testing dataset is used to put the model to the test and see how accurate it is. To see how accurate the algorithm's prediction is, we use a testing dataset.

Pre-Processing:

The adjustments we do to our data before passing it to the algorithm are referred to as pre-processing.

The raw data is converted into a comprehensible data set using the data pre-processing technique.

Steps in pre-processing are:

1. Data quality assessment
    - Mismatching in data types
    - Different dimensions of data arrays
    - Mixed of data values
    - Outliers in dataset
    - Missing data
2. Data cleaning
    - Missing data
    - Noisy data
3. Data transformation
    - Aggregation
    - Normalization
    - Feature selection
    - Discretization
4. Data reduction
    - Attribute feature selection
    - Dimensionality reduction

Feature Extraction:

Feature Extraction is a technique for transforming input data into features. Attribute square measurements are a feature of input designs that helps distinguish between distinct classes of input designs. If the input data is too large for processing, the algorithm will suspect it of being redundant, as evidenced by the appearance of images represented as pixels that are converted into a condensed collection of attributes. The specified task can be completed by using the extracted feature rather than the entire initial data.

Algorithms Used:

1. **Decision tree algorithm:**
   Decision Tree is a supervised learning technique that may be used to solve both classification and regression problems, however it is most commonly employed to solve classification issues.
2. **Support Vector Machine Algorithm (SVM Algorithm):**
   The SVM algorithm's purpose is to find the optimum line or decision boundary for categorising n-dimensional space into classes so that additional data points can be

readily placed in the correct category in the future.

3. **Artificial Neural Network Algorithm (ANN Algorithm):** An artificial neural network (ANN) is a component of a computer system that simulates how the human brain analyses and processes data.
4. **Ensembled model:** To improve the overall Analysis, they mix the decisions from multiple models.

RESULTS AND DISCUSSIONS

1. When we execute the code we get this window on our screen and we need to click on upload files, we need to upload the data set which we need to train.

Fig 2: Result1

2. After uploading the dataset we need to train the datasets by preprocessing it and find the accuracy for each algorithm.

Fig 3: Result2

3. After finding the accuracy when click on accuracy graph we get a graph showing which algorithm is best used for the given dataset, in this case ensembled is best as its accuracy is 88.76%

   In the graph the X-axis shows the algorithms used and Y-axis shows the percentage of accuracy.

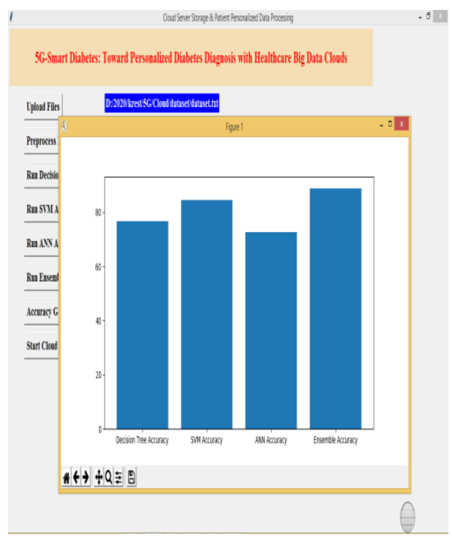

Fig 4: Result3

4. When we click on upload files, we need to upload our medical files and based on the trained data we get the results as for each users data we predicted 0 and 1 values and

also indicates patient values as normal or abnormal, and we also get the type of diabetes the user have.

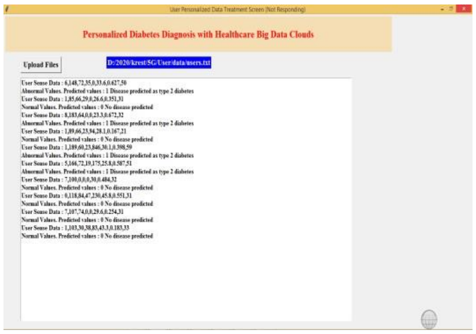

Fig 5: Result 4

CONCLUSION

In this paper, we first propose a 5G-Smart Diabetes system that includes a sensing layer, a personalized diagnosis layer, and a data sharing layer. Compared to Diabetes 1.0 and Diabetes 2.0, this system can achieve sustainable, cost-effective, and intelligence diabetes diagnosis. Then we propose a highly cost-efficient data sharing mechanism in social space and data space. In addition, using machine learning methods, we present a personalized data analysis model for 5G Smart Diabetes.

REFERENCES

[1] S. Mendis, "Global Status Report on Noncommunicable Diseases 2014," WHO, tech. rep.; http://www.who.int/ nmh/publications/ncd-status-report-2014/en/, accessed Jan. 2015.

[2] F. Florencia et al.,IDF Diabetes Atlas, 6th ed., Int'l. Diabetes Federation, tech. rep.; http://www.diabetesatlas.org/, accessed Jan. 2016.

[3] M. Chen et al., "Disease Prediction by Machine Learning over Big Healthcare Data," IEEE Access, vol. 5, June 2017, pp. 8869--79.

[4] O.Geman, I. Chiuchisan, and R. Toderean, "Application of Adaptive Neuro-Fuzzy Inference System for Diabetes Classification and prediction}," Proc. 6th IEEE Int'l. Conf. E-Health and Bioengineering, Sinaia, Romania, July 2017, pp. 639--642.

[5] S. Fong, et al. "Real-Time Decision Rules for Diabetes Therapy Management by Data Stream Mining," IT Professional, vol. 26, no. 99, June 2017, pp. 1--8.

[6] B. Lee, J. Kim, "Identification of Type 2 Diabetes Risk Factors Using Phenotypes Consisting of Anthropometry and Triglycerides Based on Machine Learning," IEEE J. Biomed. Health Info., vol. 20, no. 1, Jan. 2016, pp. 39--46.

[7] M. Hossain, et al., "Big Data-Driven Service Composition Using Parallel Clustered Particle Swarm Optimization in Mobile Environment," IEEE Trans. Serv. Comp., vol. 9, no. 5, Aug. 2016, pp. 806--17.

[8] M. Hossain, "Cloud-Supported Cyber-Physical Localization Framework for Patients Monitoring," IEEE Sys. J., vol. 11, no. 1, Sept. 2017, pp. 118--27.

[9] P. Pesl, et al., "An Advanced Bolus Calculator for Type 1 Diabetes: System Architecture and Usability Results," IEEE J. Biomed. Health Info., vol. 20, no. 1, Jan. 2016, pp. 11--17.

[10] M. Chen et al., “Wearable 2.0: Enable Human-Cloud Integration in Next Generation Healthcare System,” IEEE Commun. Mag., vol. 55, no. 1, Jan. 2017, pp. 54--61.

[11] E. Marie et al., “Diabetes 2.0: Next-Generation Approach to Diagnosis and Treatment,” Brigham Health Hub, tech. rep.; https://brighamhealthhub.org/diabetes-2-0-next-generation-approach- to-diagnosis-and-treatment, 2017, accessed Feb. 2017.

[12] M. Chen et al., “Green and Mobility-Aware Caching in 5G Networks,” IEEE Trans. Wireless Commun., vol. 16, no. 12, 2017, pp. 8347–61.

[13] C. Yao et al., “A Convolutional Neural Network Model for Online Medical Guidance,” IEEE Access, vol. 4, Aug. 2016, pp. 4094--4103.

[14] M. Anthimopoulos et al., “Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network,” IEEE Trans. Med. Imaging, vol. 35, no. 5, May 2016, pp. 1207--16.

[15] K. Hwang and M. Chen, “Big Data Analytics for Cloud/ IoT and Cognitive Computing,” Wiley, 2017. ISBN: 9781119247029.