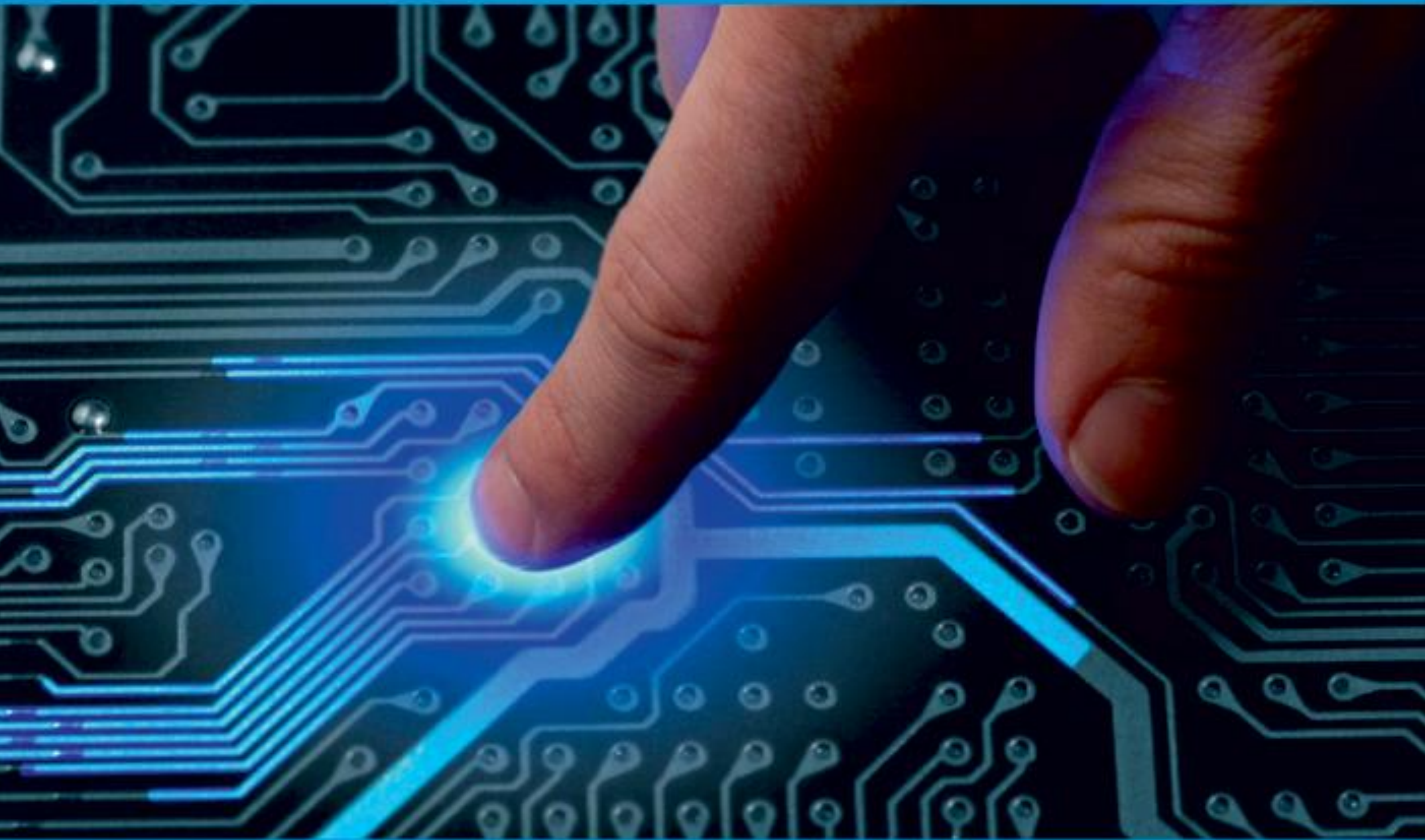
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205257 |

Heart Disease Prediction System Using a Machine Learning

Moin H. Khan, Ayaan S. Shakil, Samruddhi S. Kanade, Prof. Kapil Dere

UG Student, Department of C.S, Jaihind College of Engineering, Kuran, Pune, India

Assistant Professor, Department of C.S. Jaihind college of Engineering, Kuran, Pune, India

ABSTRACT: Heart disease prediction is very essential in today's environment, various researches has already done to predict heart disease from large dataset. IoT environment basically generate data from different sensors and predict the disease possibility accordingly. Various synthetic data sets content different body parameters which are extracted by specific sensor values, the major role played by machine learning algorithm. In this research we propose heart disease prediction with the combination of IoT and machine learning approach, the IoT environment has established to extract the data from real-time Body Sensor Network (BSN) with intermediate sensing System and store data in the cloud server adequately. Such audit data has considered synthetic information which is basically used to predict heart disease possibility. In this research, we illustrate various machine learning algorithms as well as some deep learning algorithms to achieve drastic supervision for diseas e prediction. The experimental analysis shows the effectiveness of proposed deep learning classification algorithms over the classical machine learning algorithms.

KEYWORDS: Disease prediction system, IoT, machine Learning, Supervised learning, NLP, Heart Disease.

I. INTRODUCTION

The field of "telemedicine" refers to the fast evolution of a services in recent years that enable wearable technologies to facilitate wireless communication between a physician and a patient .As of late, diabetes is the leading cause of death worldwide .In the year 2000, 171 million It was projected that there would be up to 642 million people on the planet by 2040. This rise in the number demands that this illness be taken seriously. Billionaires are spent on diabetes treatment by numerous hospitals worldwide. Type 1 diabetes, pre-diabetes, Type 2 diabetes, and gestational diabetes are the four categories of diabetes patients .Type 1 diabetes was brought on by adults' and children's insufficient insulin. Pre-diabetic refers to the stage that precedes Type 2 and gestational diabetes in pregnant women.

Furthermore, there is an infinite list of disorders that are directly linked to the heart; according to the International Cardiac Society, there are over 15 different kinds of heart-related ailments. These illnesses just need minimal historical data to be directly traced. However, conditions like diabetes, cancer, and TB, The term "indirectly related to heart diseases" refers to these. Careful historical monitoring and observational pattern analysis of the ECG waveforms are necessary for these disorders. Usually, the actions listed below are taken to complete this task:

There are also some existed models for diabetes prediction. El Jerjawi et al [2] established a neural network model for diabetes prediction. They used some attributes such as PG Concentration (Plasma glucose at 2 hours in an oral glucose tolerance test), Diastolic BP (Diastolic Blood Pressure (mm Hg)). Most of them need some professional medical test, so it is not accessible for every person. Also, the final accuracy of the prediction is around 87 percent . Peter W.F et al [3] also uses the regression models to make predictions for diabetes mellitus. However, in its samples, 99 percent of them are white and non-Hispanic which does not include other races. Hence, we believe the model is not representative.

We selected characteristics like abrupt weight loss and obesity, which are easier to comprehend and use, to build our prediction model in this study. The patients are not need to perform certain medical tests, which improves the readability and applicability of our approach. Building upon the six machine learning models outlined above, we develop diabetes prediction models. In terms of the testing error, we also compare how well they performed. As a result, we discover that neural networks and boosting have the best accuracy, at 95.5% and 96.1%, respectively.

II. LITERATURE SURVEY

According to Sunil S. Khatal & Dr. Yogesh Kumar Sharma in “Analyzing the role of Heart Disease Prediction System using IoT and Machine Learning.’’ Prediction of heart disease is very important in today's environment; various studies

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205257 |

have been done to predict heart disease from large database. The IoT environment basically generates data from various sensors and predicts the probability of disease accordingly. A variety of synthetic data, the main role of machine learning algorithms, is a variety of body parameters obtained by certain sensor values. In this study, we propose a combination of IoT and machine learning approaches to predict heart disease, an IoT environment designed to extract data from real-time body sensor networks (BSN) and remote sensing systems and store the data efficiently in a cloud server. Such studies are considered synthetic data used to predict the likelihood of heart disease. In this study, we present a variety of machine learning algorithms as well as some deep learning algorithms to achieve better control over the prediction of various diseases. Experimental analysis shows that deep learning classification algorithms are more effective than classical machine learning algorithms.

According to Chipara et. all in “Reliable clinical monitoring using wireless sensor networks: experiencesin a step-down hospital unit” presents the design, deployment, and empirical investigation of a wireless clinical monitoring system that collects pulse and oxygen saturation readings from patients. The main contribution of this paper is an indepth clinical trial evaluating the potential of a wireless sensor network for patient monitoring in a public hospital. We present a detailed analysis of system reliability in a seven-month long-term hospital setting involving 41 patients in a step-up cardiology unit. The network achieves high reliability (average 99.68%, range 95.21% - 100%). The overall reliability of the system was similar to pulse oximeters (mean 80.85%, range 0.46% to 97.69%). Sensing errors usually occur in short bursts, but there are longer durations due to sensor cutoff. We show that sensitivity can be significantly increased by implementing a connectivity alarm system that maximizes reliability and minimizes intervention costs. Retrospective data analysis showed that the system provided adequate temporal resolution to ensure transfer to the intensive care unit in three patients who experienced clinically significant events. These results show the potential and promise of using wireless sensor networks for continuous patient monitoring and detection of clinical deterioration in general hospital wards.

According to Khambete, N. D & A. Murray in “National efforts to improve healthcare technology management and medical device safety in India” In the practice of contemporary modern medicine, effective and safe use of healthcare technology is acknowledged worldwide as essential for any healthcare system. Achieving these goals can be particularly challenging in developing countries such as India, where an estimated 75% of medical technology is imported and studies have shownthat almost 30% of medical equipment is out of service. Furthermore, concerns regarding medical equipment safety have been raised in newspaper reports and also reported by a pilot study. However, recently, substantial efforts are being made to introduce changes in the health care system that will help in improving this situation. Discussions on these issues were initiated at two `International Clinical Engineering Workshops' (Trivandrum 2009 and Pune 2011) and two `Regional Clinical Engineering Workshops' (Latur and Mumbai 2011). A clear consensus emerged from these Workshops that urgent action was essential to initiate effective Healthcare Technology Management (HTM) practices in all health care sectors and actively promote medical device and equipment safety in India. Subsequently ,in February 2012, a round table meeting of experts was held, which focused on confronting medical and healthcare management staff with the problems to be solved, while at the same time helping to develop an action plan to bring about the necessary changes. These consultations identified existing gaps and underlying reasons, thus leading to development of an action plan. This paper reviews all these efforts and highlights the outcomes.

III. MODELS

In this section, we use logistic regression, support vector machine, decision tree, random forest, boosting and neural network to establish diabetes prediction models. We evaluate the model by calculation the train error and test error.

Logistic Regression:

Logistic regression models the likelihoods of potential outcomes using logistic functions. The dependent variable in a binary logistic regression model is divided into two groups. It determines the linear relationship by using the logit function to transform the probability, which is between 0 and 1, to any real integer .We must first ascertain the logistic coefficients. The table shows that there are positive correlations between diabetes and polyuria, polydipsia, sudden weight loss, weakness, polyphagia, genital thrush, visual blurring, irritability, and partial paresis; on the other hand, there are negative correlations between diabetes and age, itching, delayed healing, muscle stiffness, alopecia, and obesity. Table 4 provides an overview of the confusion matrix. There appears to be no over fitting because the train accuracy of 87.6% and the test accuracy of 87.2% are close to each other.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205257 |

Decision Tree:
A decision tree is a tree-like structure that is constructed by dividing the source set in order to develop a model that makes predictions depending on input variables. Examples are categorized by a collection of dividing rules according to the characteristics of classification. One of the most often used machine learning algorithms is decision trees, which have the advantages of simplicity and comprehensibility.

Random Forest:
Because a single decision tree is unstable, ensemble learning—which mixes several models to increase overall prediction accuracy and decrease the variation. A common ensemble learning technique is random forest. Since the complexity of the model is determined by the number of decision trees, we experiment with different numbers of trees and plot the resulting mistakes in Figure 4. In contrast to SVM and decision trees, random forests' testing errors are unlikely to increase as the number of trees increases, suggesting that they are less prone to overfit. It is unlikely to be an overfitting model, comparatively, even while the training error keeps decreasing and the testing error remains stable after declining. The confusion matrix is Compared with single decision tree, the testing accuracy of random forest, which is 94.9%, increased near 1.7 percent.

Boosting:
Another traditional ensemble learning technique is called "boosting," which lowers variance by reweighting the samples for each decision tree while it is being trained. similar to the arbitrary Moreover, we also utilize the quantity of decision trees to ascertain the model's and the outcome's complexity. Boosting is slightly more likely to be overfitting than random forest. The train error first decreases and then straightens out. Similar to other models, the test error first decreases and subsequently increases. The image indicates that we selected 700 trees for our boosting model, and Table 8 displays the confusion matrix. The fact that the train correctness is nearly 100% indicates that overfitting is more likely to be the cause of the boosting. Compared to the random forest,its accuracy of boosting is further improved by around 0.7 percent.

Support Vector Machine:
In this method, The regularization parameter C, which regulates the model's complexity, has to be changed. Greater C indicates a harsher punishment for the misclassification, which causes More likely, the model is overfitting. In order to compute the training and testing errors for each model with a different C, we employ several values of C, ranging from 0.1 to 4.6. The results are presented in Figure 2. The train error decreases as the model gets more complicated, while the test error initially decreases and subsequently increases. As a result, we determine that 2.1 is the ideal parameter for our model, and the matching confusion matrix that results is shown in Table 5. Consequently, the testing accuracy of the support vector machine is higher than that of the logistic regression (87.2%) at 90.4%.

Neural Network:
Neural network is made up of layers and neurons, and it learns by processing samples from first layer to the last layer. Once enough samples have been processed, the algorithm can construct the right model. Each of the three layers we built has 64, 16 or 8 neurons in it. Additionally, we select the logistic activation function and the 0.01 learning rate. Next, we obtain the Table 9 confusion matrix. As we can see, the neural network outperforms all other models with an accuracy of 96.2, owing to its larger parameter set that allows for better data fitting.

IV. CONCLUSION AND FUTURE WORK

Physicians can identify patients more accurately and treat them more quickly if they have access to an effective diabetes prediction model. We use descriptive statistics to assess the risk of diabetes.predictive dataset to look into the factors that affect diabetes. Table 10 lists the train and test errors of the six machine learning models—logistic regression, support vector machines, decision trees, random forests, boosting, and neural networks—on which we base our diabetes prediction models.Compared to the final three models, which are more sophisticated, the first three models—logistic regression, support vector machines, and decision trees—are simpler, intuitive, and have lesser accuracy. As we can see, test accuracy increases with model complexity. In the future, we can experiment with models that are more capable of learning and adapting, and we can employ a wider range of datasets to increase forecast accuracy.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205257 |

REFERENCES

1. Sunil S. Khatal, Dr. Yogesh Kumar Sharma “Analyzing the role of Heart DiseasePrediction System using IoT and Machine Learning” 2020
2. Chipara, Octav, Chenyang Lu, Thomas C. Bailey, and Gruia-Catalin Roman, “Reliable clinical monitoring using wireless sensor networks: experiences in a step-down hospital unit,” In Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems, ACM, pp. 155–168, 2010.
3. Khambete, N. D, and A. Murray, “National efforts to improve healthcare technology managementand medical device safety in India,” Appropriate Healthcare Technologies for Developing Countries, 7th International Conference on, IET, pp. 1–5, 2012.
4. Priyan Malarvizhi Kumar and Usha Devi Gandhi,” A novel three-tier internet of things architecture with machine learning algorithm for early detection of heart diseases”, J. Computers and Electrical Engineering, pp. 222-235, 2018.
5. Mingyu Park and Younghoon Song,JAewonLee,JeongyeupPaek,” Design and implementation of a smart chair system for IOT”, IEEE,2016.
6. P.M. Kumar, S. Lokesh, R. Varatharajan, C. Gokulnath, P. Parthasarathy,” Cloud and IoT baseddisease prediction and diagnosis system for healthcare using Fuzzy neural classifier”, Future Generation Computer Systems, 2018.
7. Koppar, Anant R, and Venu gopalachar Sridhar, “A workflow solution for electronic health records to improve healthcare delivery efficiency in rural India,” In eHealth, Telemedicine, and Social Medicine, 2009. eTELEMED’09. International Conference on, pp. 227–232, IEEE, 2009
8. Sharma YK, Khatal Sunil S. Health Care Patient Monitoring using IoT and Machine Learning,in IOSR Journal of Engineering (IOSR JEN) National Conference on “Recent Innovations in Engineering and Technology” MOMENTUM-19
9. Jha, R. K., Henge, S. K., & Sharma, A. (2020). Optimal machine learning classifiers for predictionof heart disease. International Journal of Control and Automation, 13(1 Special Issue), 31-37.J. Y.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details

www.ijircce.com