IJIRCCE

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

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 2, February 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

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

|| Volume 12, Issue 2, February 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1202009 |

Survey Paper on Machine Learning Algorithm for Smart Healthcare System

Jitendra Singh[1], Dr. Divya Jain[2]

M. Tech. Scholar, Department of Electronics and Communication Engineering, TIT-Advance, Bhopal, India [1]

Associate Professor, Department of Electronics and Communication Engineering, TIT-Advance, Bhopal, India [2]

ABSTRACT: The increasing demand in remote monitoring of health conditions of patients by their physicians and the members of the family has necessitated the development of healthcare system. Healthcare Monitoring system has been a growing technology that facilitates real-time monitoring of patient's essential body functions, especially pulse rate, temperature, respiration rate, and blood pressure during their daily life. It has been surveyed that the healthcare monitoring system has been designed to provide healthcare management for patients with cardiac diseases, hypotension, diabetes, hyper and hypothermia, hypertension etc. In this paper, we studied a diabetes prediction healthcare model for better classification of diabetes which includes few external factors responsible for diabetes along with regular factors like Glucose, BMI, Age, Insulin, etc. Classification accuracy is machine learning with dataset compared to existing model. Further with imposed a pipeline model for diabetes prediction intended towards improving the accuracy of classification.

KEYWORDS*:* Classification, Machine Learning, Healthcare System

I. INTRODUCTION

The increasing demand in remote monitoring of patients health conditions by their physicians and the members of the family necessitates the development of healthcare system. Healthcare Monitoring system facilitates real-time monitoring of patient's essential body functions, specifically pulse rate, temperature, respiration rate, and blood pressure during their daily life (at home, at work, at sport activities, etc.) or in a hospital environment, without affecting their day to day activities. It has been surveyed that the healthcare monitoring has been designed to provide healthcare management of the patients with cardiac diseases, hypotension, diabetes, hyper and hypothermia, hypertension etc.

Smart Healthcare is a patient-based healthcare system making use of SHS devices for remote monitoring of the patient. Networks are established Sensors, actuators, and personal medical devices (PMDs) that help to bridge the gap between the modern world and electronic records. These medical tracking systems will continually track a patient's health status and relay them to an expert or authorized healthcare provider [1, 2].

The entire structure of SHS is generally categorized into three separate phases or stages. The first stage consists of sensors that track a patient's vital signs and are light, low-power, and high-efficiency. These can be worn on, put into, or wrapped around the body. On-body sensors and stationary medical equipment are the two types of first-stage devices used to introduce smart healthcare. Biosensors that are connected to the human body for physiological monitoring are known as on-body sensors. In-vitro and in-vivo sensors are two different types of sensors. In-vitro electrodes are attached to the patient's psyche manually, removing the need for clinical or medical services in healthcare. In-vivo sensors are implantable devices that are positioned inside the body following surgery meeting all sterilization requirements [3, 4].

Pacemakers, motion detectors, and artificial retinas are all examples of first-stage devices. All of these sensors submit health data to a PMD. These PMDs binds to the Internet or a cloud server in the second level. In SHS, PMDs for patient monitoring and assessment are either implanted in the patient's body or externally attached to the patient's body to track their medical condition. These devices connect with a base station using a wireless interface, which is then employed to read system status, and diagnostic reports, adjust device parameters and update the status accordingly [5, 6].

The data is analyzed in the third level. The third tier of analysis assists healthcare professionals in assessing whether or not an intervention is needed to be based on sensed data SMS sensors and other autonomous instruments can classify data based on the patient and the examiner. A physician, for instance, needs historical knowledge about a patient's symptoms, while a chemist simply needs the patient's latest prescription log. For instance, an ambulance staff can need

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

|| Volume 12, Issue 2, February 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1202009 |

more information than a nurse or expert. Since sensors and apps interact wirelessly, SHS prioritizes guaranteeing that each health professional receives the information they require while maintaining the confidentiality and privacy of all collected data. Fig. 1 shows the general framework for SHS, the most vulnerable part for security and privacy is the transmission of data from a smart health device to a cloud storage server. The attackers can modify or change the data during transmission. In the transmission of processed data, attackers can still alter the information before getting to the organization that needs the information.

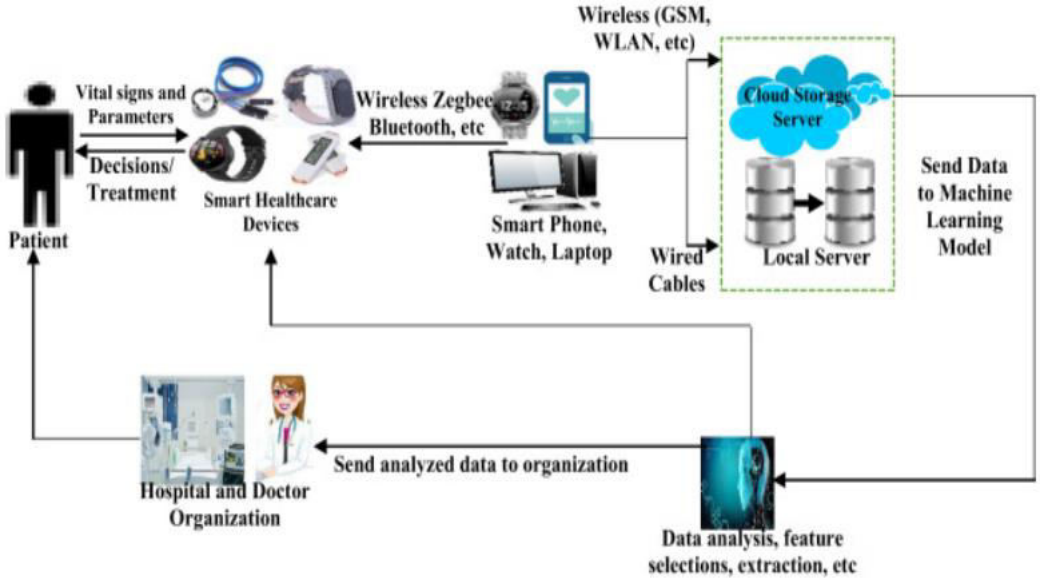

Figure 1: The general framework concepts for smart healthcare system

II. LITERATURE REVIEW

Isfafuzzaman Tasin et al. [1], in diagnosis or detection of temperature variations in the human skin are obtained from good resolution thermal images captured by the highly sensitive thermal imaging system. Historically, the first thermal image of human was acquired with a British prototype system; the "Pyroscan" were made in London for research in rheumatology, chronic fatigue, and pain management between 1959 and 1961. The thermal imaging system has been utilized for many clinical applications in hospitals since the late 20th century. The thermal imaging system has been used for mass fever detection localization of swelling in dentistry biometric authentication of facial signature and palm vein. Recently, it has also been utilized to record the presence and position of the fetus and other physiological factors associated with pregnancy..

Olisah et al. [2], is characterized in the form of intensity differences on a thermal image. The difference in the thermal pattern of pixel intensities results in texture variations in an image. These texture variations can be extracted in the form of statistical features and used for classification. As the temperature distribution of DM patients is different from healthy subjects, asymmetry amongst the features may be analyzed to gather significant information about the pathology. The human examination of plantar thermal images is susceptible to faults owing to carelessness, color blindness, fatigue, and repetitive tasks. The utility of the textural and higher-order statistical features in finding the temperature distribution changes in plantar thermal images has been determined. These features have been utilized to differentiate diabetic feet from healthy ones.

Deberneh et al. [3], in this work, many feature extraction approaches are explored to design and develop efficient machine learning frameworks intended for detecting diabetic feet in plantar thermal images. In extension to the broadly utilized asymmetry method using textural and temperature features, the subsequent feature extraction approaches are also analyzed, for plantar thermal images. When the performance of several extracted features could not be used seemingly, machine learning methods have been utilized for training and automated classification. Many CNN models are employed to automatically segment, classify diabetic foot ulcers from color (RGB) images. The most significant features are also extracted from different layers of the pre-trained CNN models using transfer learning.

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

|| Volume 12, Issue 2, February 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1202009 |

Nikos Fazakis et al. [4], the dataset contained nine variables, eight of which contained patient information. The ninth variable was the class that predicted whether patients would develop diabetes or not. Outliers and missing values make up the dataset. We have removed the outliers from the dataset using our proposed method. The mean filter method was used to impute missing values that were present in the dataset, preserving the dataset's consistency. Weka was used in each and every one of the experiments. It is used for "Data Mining" and classification purpose. Classification algorithms may be applied to a dataset. The other applications of weka are visualizing, regression, purpose of data mining methods. Full form of Weka is Waikato Environment for knowledge Analysis.

Naveen Kishore G et al. [5], one of the most fatal and long-lasting diseases that causes an increase in glucose is diabetes. If diabetes is left undiagnosed and untreated, many problems arise. The standard distinguishing factor results in a visit to a symptomatic focus and counseling specialist for the patient. However, the advancement of AI approaches addresses this primary problem. The goal of this study is to develop a model that can accurately predict a patient's likelihood of developing diabetes. As a result, this analysis employs Decision Tree, SVM, and Naive Bayes—three AI characterization calculations—to identify Diabetes in its early stages. P.I.D.D., which is sourced from the UCI AI storehouse, is the subject of analyses. Different estimations, such as Precision, Accuracy, F-Measure, and Recall, are used to evaluate the results of the multitude of three calculations. Estimates of exactness are made based on correctly and incorrectly ordered cases. With a precision of 76.30 percent, the obtained results demonstrate that Naive Bayes prevails over relatively different calculations. R.O.C. bends are used in a legitimate and methodical way to verify these results.

Chatrati et al. [6], there are a couple of simulated intelligence methodologies that are used to carry out judicious examination over colossal data in various fields. Perceptive assessment in clinical benefits is a troublesome endeavor regardless can assist experts with making gigantic data taught ideal decisions about open minded's prosperity and treatment. This paper discusses the perceptive examination in clinical consideration, six different man-made intelligence estimations are used in this assessment work. Six distinct AI calculations are applied to a dataset of the patient's clinical record for the purpose of analysis. The applied calculations' execution and precision are examined and analyzed. The best calculation for predicting diabetes is determined by correlating the various AI methods used in this study. Using AI strategies, the purpose of this paper is to assist specialists and experts in early diabetes prediction. One of the leading causes of disability and death is type II diabetes. These are incomplete and noisy data. A comprehensive estimation of missing data is presented in this paper. Correlation-based estimation models and k-NN were two of the five approaches to imputation of missing values that were compared.

Hasan et al. [7], In PC vision, the problem of extracting complex head and hand developments and their ever-shifting shapes for acknowledging communication through gestures is seen as challenging. This paper proposes the affirmation of Indian correspondence through marking movements using a mind blowing man-made awareness gadget, convolutional brain associations (CNN). CNN preparation is carried out using three distinct example estimates, each containing distinct subject and survey point arrangements. The prepared CNN is tested using the remaining two examples. With our selfie-based communication data, distinct CNN models were planned and tested for improved acknowledgment precision. In contrast to other classifier models written about the same dataset, our acknowledgment rate was 92.88%.

Naveen Kishore G et al. [10], Diabetes is considered one of the deadliest and most persistent illnesses which cause an increment in glucose. Numerous entanglements happen if Diabetes stays untreated and unidentified. The monotonous distinguishing measure brings about visiting a patient to a symptomatic focus and counselling specialist. The exhibitions of the multitude of three calculations are assessed on different estimates like Precision, Accuracy, F-Measure, and Recall. Exactness is estimated over effectively and erroneously ordered cases. Results got show Naive Bayes beats with the most elevated precision of 76.30% relatively different calculations. These outcomes are confirmed utilizing Receiver Operating Characteristic (R.O.C.) bends in a legitimate and methodical way.

III. METHODOLOGY

Learning

The main property of an ML is its capability to learn. Learning or preparing is a procedure by methods for which a neural system adjusts to a boost by making legitimate parameter modifications, bringing about the generation of wanted reaction. Learning in an ML is chiefly ordered into two classes as [9].

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

|| Volume 12, Issue 2, February 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1202009 |

- Supervised learning
- Unsupervised learning

Supervised Learning
Regulated learning is two stage forms, in the initial step: a model is fabricated depicting a foreordained arrangement of information classes or ideas. The model developed by investigating database tuples portrayed by traits. Each tuple is expected to have a place with a predefined class, as dictated by one of the qualities, called to have a place with a reclassified class, as controlled by one of the traits called the class name characteristic. The information tuple are dissected to fabricate the model all things considered from the preparation dataset.

Unsupervised learning
It is the kind of learning in which the class mark of each preparation test isn't knows, and the number or set of classes to be scholarly may not be known ahead of time. The prerequisite for having a named reaction variable in preparing information from the administered learning system may not be fulfilled in a few circumstances.
Data mining field is a highly efficient techniques like association rule learning. Data mining performs the interesting machine-learning algorithms like inductive-rule learning with the construction of decision trees to development of large databases process. Data mining techniques are employed in large interesting organizations and data investigations. Many data mining approaches use classification related methods for identification of useful information from continuous data streams.

Nearest Neighbors Algorithm
The Nearest Neighbor (NN) rule differentiates the classification of unknown data point because of closest neighbor whose class is known. The nearest neighbor is calculated based on estimation of k that represents how many nearest neighbors are taken to characterize the data point class. It utilizes more than one closest neighbor to find out the class where the given data point belong termed as KNN. The data samples are required in memory at run time called as memory-based technique. The training points are allocated weights based on their distances from the sample data point. However, the computational complexity and memory requirements remained key issue. For addressing the memory utilization problem, size of data gets minimized. The repeated patterns without additional data are removed from the training data set.

Naive Bayes Classifier
Naive Bayes Classifier technique is functioned based on Bayesian theorem. The designed technique is used when dimensionality of input is high. Bayesian Classifier is used for computing the possible output depending on the input. It is feasible to add new raw data at runtime. A Naive Bayes classifier represents presence (or absence) of a feature (attribute) of class that is unrelated to presence (or absence) of any other feature when class variable is known. Naïve Bayesian Classification Algorithm was introduced by Shinde S.B and Amrit Priyadarshi (2015) that denotes statistical method and supervised learning method for classification. Naive Bayesian Algorithm is used to predict the heart disease. Raw hospital dataset is employed. After that, the data gets preprocessed and transformed. Finally by using the designed data mining algorithm, heart disease was predicted and accuracy was computed.

Support Vector Machine
SVM are used in many applications like medical, military for classification purpose. SVM are employed for classification, regression or ranking function. SVM depends on statistical learning theory and structural risk minimization principal. SVM determines the location of decision boundaries called hyper plane for optimal separation of classes as described in figure 1.4. Margin maximization through creating largest distance between separating hyper plane and instances on either side are employed to minimize upper bound on expected generalization error. Classification accuracy of SVM not depends on dimension of classified entities. The data analysis in SVM is based on convex quadratic programming. It is expensive as quadratic programming methods need large matrix operations and time consuming numerical computations.

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

|| Volume 12, Issue 2, February 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1202009 |

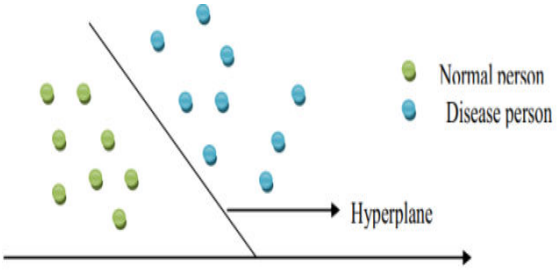

Fig. 1: Support Vector Classification

Particle Swarm Optimization

The notation is used in Particle swarm optimization are as follows:
Xid :Component in dimension d of the ith particle of swarm
Vid :The particle velocity of particle I in dimension d
PBi : the best position achieved so far by particle i
GB : The best global best position
C1,C2 : Constant weight factors
W: The inertia weight
r1,r2: Random factors in [0,1] interval
Vmin : The minimum velocity value of particle
Vmax: The maximum velocity value of particle
yi : The fitness value of particle i

PSO is a population base stochastic optimization technique inspired by the social behavior of swarm, such as bird flocking or fish schooling, to obtain a promising position to achieve certain objectives. The PSO algorithm works by having a population (called a swarm) of candidate solution (called particles).Each particle in a population has a fitness value computed from a fitness function and each particle has a position, and move based on an updated velocity according to few simple formula. The movements of the particles are guided by their own best known position in the search space as well as the entire swarm's best known position. The particle movements are directed by the position vector and velocity vector of each particle. In the n-dimensional space, the vector and velocity vector of the ith particle position are represented as Xi=[xi1,xi2,x3i,xi4,…….xin] and Vi=[vi1,vi2,……..,vin] respectively, where xid is a binary bit, i=1,2,…..,m (m is the number of particles). The record of the position of the previous best performance of the neighborhood is GBi=[gbi1, gbi2,…….gbin]. The particle velocity and position is updated and based on Eqs. (A1) and (A2), respectively.

Vidnew = w x vidold + c1r1(pbidold –xidold) + c2r2 (gbdold- xidold) , d=1,2 ely D

Xidnew = xidold + vidnew , d=1,2……elyN

Where c1 and c2 are the positive constant values between 0 and 4, indication the cognitive and the social learning factors, respectively. The inertia weight (w has a value between 0.4 and 0.9 and r1 and r2 are uniformly distributed with the numbers between 0 and 1. The values of the velocities are between vmin and vmax , N is the size of the swarms.

Feed Forward Back propagation Neural Network

Neural networks are predictive model that have ability to learn, analyses, organize the data and predict test results accordingly. Among several kinds of neural networks, feed forward neural network is usually employed in medical diagnosis applications and others. These networks are trained by a set of patterns called training set, whose outcome is already known. In our study Feed Forward NN consists of input, hidden and an output layer, and the data operates in forward direction, and the error is back propagated to update the weights at every epoch in order to reduce errors.

IV. CONCLUSION

The conclusion in smart healthcare is tremendous and there is a lot of room for improvisation. The emergence of the concept of Electronic Health Record (EHR) has also given rise to interesting ideas such as smart healthcard for patients, which can be implemented and can serve as invaluable information source at the time of need. The Next-Generation

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

|| Volume 12, Issue 2, February 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1202009 |

Genome Sequencing has brought about a revolutionizing change in the world of genomic research and researchers are attempting to translate it into routine clinical practice through the EHR. The vision for a robust EHR system can be thought of as a platform, which can skillfully manage health profiles of individuals right from childbirth and use it for faster and more effective medical decision-making processes. Along with the data collected from smart healthcare devices, which can track important parameters such as the vital signs of the patient, physiological changes associated with a particular disease, etc., EHR platform can help us analyze the disease from both the molecular standpoint and also from the symptoms that arise.

REFERENCES

[1] Isfafuzzaman Tasin, Tansin Ullah Nabil, Sanjida Islam, Riasat Khan, “Diabetes prediction using machine learning and explainable AI techniques”, Healthcare Technology Letters, pp. 01-10, Wiley 2022.

[2] Olisah, C.C., Smith, L., Smith, M., “Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective”, Comput. Methods Programs Biomed., Vol. 20, pp. 1–12, 2022.

[3] Deberneh, H.M., Kim, I., “Prediction of type 2 diabetes based on machine learning algorithm”, Int. J. Environ. Res. Public Health, Vol. 18, pp. 1–14, 2021.

[4] Nikos Fazakis, Otilia Kocsis, Elias Dritsas, Sotiris Alexiou, Nikos Fakotakis, and Konstantinos Moustakas, “Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction”, IEEE Access 2021.

[5] Naveen Kishore G, V.Rajesh, A.Vamsi Akki Reddy, K.Sumedh, T.Rajesh Sai Reddy, “Prediction of Diabetes using Machine Learning Classification Algorithms”, International Journal of Scientific & Technology Research, Vol. 9, No. 01, 2020.

[6] Chatrati, S.P., Hossain, G., Goyal, A., “Smart home health monitoring system for predicting type 2 diabetes and hypertension”, J. King Saud Univ. Comput. Inf. Sci., Vol. 34, No. 3, pp. 862–870, 2020.

[7] Hasan, M.K., Alam, M.A., Das, D., Hossain, E., Hasan, M., “Diabetes prediction using ensembling of different machine learning classifiers”, IEEE Access, Vol. 8, pp. 76516–76531, 2020.

[8] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease Prediction by Machine Learning over Big Data from Healthcare Communities," IEEE Access, vol. 5, no. c, pp. 8869–8879, 2017.

[9] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, "Machine learning on big data: Opportunities and challenges," Neuro computing, vol. 237, pp. 350–361, May 2017.

[10] J. B. Heaton, N. G. Polson, and J. H. Witte, "Deep learning for finance: deep portfolios," Appl. Stoch. Model. Bus. Ind., vol. 33, no. 1, pp. 3–12, Jan. 2017.

[11] Reddy S.S., Suman M., Prakash K.N. ., "Micro aneurysms detection using artificial neural networks", 2018, Lecture Notes in Electrical Engineering ,Vol: 434 ,Issue 3, pp: 409 to 417.

[12] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., Chouvarda, I., "Machine Learning and Data Mining Methods in Diabetes Research", Computational and Structural Biotechnology Journal 15, 104–116, 2017.

[13] Majid Ghonji Feshki and Omid Sojoodi Shijan, "Improving the Heart Disease Diagnosis by Evolutionary Algorithm of PSO and Feed Forward Neural Network", International paper on IEEE 2016.

[14] L. Hermawanti, "Combining of Backward Elimination and Naive Bayes Algorithm To Diagnose Breast Cancer", Momentum, vol. 11, no. 1, pp. 42-45, 2015.

[15] O.S. Soliman, E. Elhamd, "Classification of Diabetes Mellitus using Modified Particle Swarm Optimization and Least Squares Support Vector Machine", IEEE 2014.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details