

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

DIA

International Journal of Innovative Research in SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 3, March 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

 \bigcirc







|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

Volume 13, Issue 3, March 2024

| DOI:10.15680/IJIRSET.2024.1303008 |

Survey Paper on Prediction of Heart Disease using Supervised Machine Learning

Govind Singh¹, Prof. Suresh S. Gawande²

M. Tech. Scholar, Department of Electronics and Communication, Bhabha Engineering Research Institute,

Bhopal, India¹

Guide, Department of Electronics and Communication, Bhabha Engineering Research Institute, Bhopal, India²

ABSTRACT: Detection of heart disease through early-stage symptoms is a great challenge in the current world scenario. If not diagnosed timely, this may become the cause of death. In developing countries where heart specialist doctors are not available in remote, semi-urban, and rural areas, an accurate decision support system can play a vital role in the early-stage detection of heart disease. Invasive methods of diagnosing the disease are expensive and painful. Therefore, there is a need of a system that can diagnose heart disease in a noninvasive manner at less cost. **Classification model helps physicians to improve their prognosis, diagnosis or treatment planning procedures. Big Data Analytics plays an significant role in healthcare industries. Healthcare industries have large volume databases. Using big data analytics one can study huge datasets and find hidden information, hidden patterns to discover knowledge from the data and predict outcomes accordingly. In existing method, the classification and prediction accuracy is not so high.**

KEYWORDS: Classification, Machine Learning, Heart Disease

I. INTRODUCTION

Wellbeing doesn't exclusively portray the shortfall of infection or the arrangement of indicative, remedial, and preventive administrations. As indicated by WHO, Wellbeing likewise incorporates a condition of physical, mental, and social prosperity. Medical services is interdisciplinary and multi-sectorial. Clinical consideration is important for medical care and is requested at the hour of deviation from wellbeing, i.e., at the hour of sickness [1, 2]. With the progression in the area of innovation, an immense measure of information can be gotten to through high computational speed to upgrade information. It is consequently important for the data to be coordinated in a configuration that can be effectively enquired to decipher the outcomes. The most common way of putting together data includes the extraction of basic data from an immense volume of information [3]. Great many articles connected with illnesses are distributed day to day in diaries, thus losing all sense of direction in the subtleties of such a tremendous measure of information is simple. Established researchers is by and by dealing with the issue of extraction of applicable data and making it open to scientists, doctors, and non-experts [4]. The quick movement in the area of science and innovation particularly in genomics and data innovation has directed us to step into the rousing time of current science. The significant test for clinical science is the mix and extraction of pertinent data about the human sickness from the tremendous and speedily developing volumes of data into a comprehensive knowledge into the infections.

The expanded pace of information creation because of progression in the space of examination into the hereditary premise of illness, simultaneously with the rising DNA diagnostics and electronic medical care records, supports the clinical practices. To comprehend the effect of the hereditary premise on human wellbeing, reconciliation of the dispersed data is basic.

Consistently sicknesses are being found at a rising speed, further adding to the existing rundown. In any case, no such data set has been made far to the point that arranges significant data about human illnesses from all classifications to empower their valuable abuse [5]. Because of inability to carry out demonstrated methodologies because of diminished accessibility of precise illness data and inaccessibility of a data set that productively recovers all the data relating to a wide range of human sickness, there is a requirement for a cross-referred to semi dynamic non-benefit "Human Infection



|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

Volume 13, Issue 3, March 2024

DOI:10.15680/IJIRSET.2024.1303008

Information base (HDD)", which gives all suitable data on human infections. However sickness data sets are accessible that are more illness classification explicit, there is an absence of a complete openly accessible sickness data set grouping every known infection and giving data on all viewpoints of it.

Investigating the emerging nations, the irresistible illnesses are still enduring as significant medical conditions regardless of having public projects for the control of the majority of these infections for close to 50 years now. Bleakness (weakness) and mortality related with irresistible illnesses are high in emerging nations, especially babies

furthermore, youngsters are generally impacted. Around 26% of yearly passings overall are because of irresistible illnesses [6]. In India, a few transferable sicknesses, for example, smallpox and guinea worm have been effectively destroyed, while plague, yaws, and polio have been taken care of. While the quantity of non-transferable sicknesses like ischemic heart illnesses, malignant growth, stroke, diabetes, ongoing respiratory sickness, and wounds, is ascending, due to industrialization, urbanization, financial turn of events, changing age structure, and changing ways of life [7]. AI is an arising field in software engineering as it has given another aspect to the idea of learning. By gaining from past information, it imitates human knowledge in the framework. Because of its self-versatile ability and numerical premise, AI has tremendous applications in medical services, opinion examination, proposal frameworks, regular language handling, data recovery, ongoing interaction, market examination, text acknowledgment, and PC vision [8]. Wellbeing administrations can be improved with the utilization of AI. Choice emotionally supportive networks created utilizing AI can be utilized to analyze different sicknesses [9]. The examination work in this proposal centers around the conclusion of coronary illness utilizing an AI approach.

II. LITERATURE REVIEW

Romany Fouad Mansour et al. [1], the recent advancements in Internet of Things (IoT), cloud computing, and Artificial Intelligence (AI) transformed the conventional healthcare system into smart healthcare. By incorporating key technologies such as IoT and AI, medical services can be improved. The convergence of IoT and AI offers different opportunities in healthcare sector. In this view, the current research article presents a new AI and IoT convergence-based disease diagnosis model for smart healthcare system. The major goal of this article is to design a disease diagnosis model for heart disease and diabetes using AI and IoT convergence techniques. The presented model encompasses different stages namely, data acquisition, preprocessing, classification, and parameter tuning. IoT devices such as wearables and sensors permit seamless data collection while AI techniques utilize the data in disease diagnosis. The proposed method uses Crow Search Optimization algorithm-based Cascaded Long Short Term Memory (CSO-CLSTM) model for disease diagnosis. In order to achieve better classification of the medical data, CSO is applied to tune both `weights' and `bias' parameters of CLSTM model. Besides, isolation Forest (iForest) technique is employed in this research work to remove the outliers. The application of CSO helps in considerable improvement in the diagnostic outcomes of CLSTM model. The performance of CSO-LSTM model was validated using healthcare data. During the experimentation, the presented CSO-LSTM model accomplished the maximum accuracies of 96.16% and 97.26% in diagnosing heart disease and diabetes respectively. Therefore, the proposed CSO-LSTM model can be employed as an appropriate disease diagnosis tool for smart healthcare systems.

G. Muhammad et al. [2], an electroencephalogram (EEG)-based remote pathology detection system is proposed in this study. The system uses a deep convolutional network consisting of 1D and 2D convolutions. Features from different convolutional layers are fused using a fusion network. Various types of networks are investigated; the types include a multilayer perceptron (MLP) with a varying number of hidden layers, and an autoencoder. Experiments are done using a publicly available EEG signal database that contains two classes: normal and abnormal. The experimental results demonstrate that the proposed system achieves greater than 89% accuracy using the convolutional network followed by the MLP with two hidden layers. The proposed system is also evaluated in a cloud-based framework, and its performance is found to be comparable with the performance obtained using only a local server.

A. A. Mutlag et al. [3], in healthcare applications, numerous sensors and devices produce massive amounts of data which are the focus of critical tasks. Their management at the edge of the network can be done by Fog computing implementation. However, Fog Nodes suffer from lake of resources That could limit the time needed for final outcome/analytics. Fog Nodes could perform just a small number of tasks. A difficult decision concerns which tasks will perform locally by Fog Nodes. Each node should select such tasks carefully based on the current contextual information, for example, tasks' priority, resource load, and resource availability. We suggest in this paper a Multi-Agent Fog



e-ISSN: 2319-8753, p-ISSN: 2347-6710 www.ijirset.com | Impact Factor: 8.423 | A Monthly Peer Reviewed & Referred Journal |

Volume 13, Issue 3, March 2024

| DOI:10.15680/IJIRSET.2024.1303008 |

Computing model for healthcare critical tasks management. The main role of the multi-agent system is mapping between three decision tables to optimize scheduling the critical tasks by assigning tasks with their priority, load in the network, and network resource availability. The first step is to decide whether a critical task can be processed locally; otherwise, the second step involves the sophisticated selection of the most suitable neighbor Fog Node to allocate it. If no Fog Node is capable of processing the task throughout the network, it is then sent to the Cloud facing the highest latency. We test the proposed scheme thoroughly, demonstrating its applicability and optimality at the edge of the network using iFogSim simulator and UTeM clinic data.

M. S. Hossain et al. [4], new generation communication technologies and advanced deep learning models present a tremendous opportunity to develop fast, accurate, and seamless distributed systems in different sectors including the healthcare sector. in this article, we suggest a smart healthcare framework consisting of a pathology detection system, which is developed using deep learning. The pathology can be detected from electroencephalogram signals of a subject. in the framework, a smart EEG headset captures EEG signals and sends them to a mobile edge computing server. The server preprocesses the signals and transmits them to a cloud server. The cloud server does the main processing using deep learning and decides on whether the subject has pathology or not. Clients and stakeholders of the framework are connected via an authentication manager located in the cloud server. Experiment results on a publicly available database confirm the appropriateness of the proposed framework.

M. S. Hossain et al. [5], voice pathology disorders can be effectively detected using computer-aided voice pathology classification tools. These tools can diagnose voice pathologies at an early stage and offering appropriate treatment. This study aims to develop a powerful feature extraction voice pathology detection tool based on Deep Learning. In this paper, a pre-trained Convolutional Neural Network (CNN) was applied to a dataset of voice pathology to maximize the classification accuracy. This study also proposes a distinguished training method combined with various training strategies in order to generalize the application of the proposed system on a wide range of problems related to voice disorders. The proposed system has tested using a voice database, namely the Saarbrücken voice database (SVD). The experimental results show the proposed CNN method for speech pathology detection achieves accuracy up to 95.41%. It also obtains 94.22% and 96.13% for F1-Score and Recall. The proposed system shows a high capability of the real-clinical application that offering a fast-automatic diagnosis and treatment solutions within 3 s to achieve the classification accuracy.

V Krishnapraseeda et al. [6], diabetes is a major metabolic disorder which can affect entire body system adversely. Undiagnosed diabetes can increase the risk of cardiac stroke, diabetic nephropathy and other disorders. All over the world millions of people are affected by this disease. Early detection of diabetes is very important to maintain a healthy life. This disease is a reason of global concern as the cases of diabetes are rising rapidly. Machine learning (ML) is a computational method for automatic learning from experience and improves the performance to make more accurate predictions. In the current research we have utilized machine learning technique in Pima Indian diabetes dataset to develop trends and detect patterns with risk factors using R data manipulation tool. To classify the patients into diabetic and non-diabetic we have developed and analyzed five different predictive models using *R* data manipulation tool. For this purpose we used supervised machine learning algorithms namely linear kernel support vector machine (SVM-linear), radial basis function (RBF) kernel support vector machine, *k*-nearest neighbour (*k*-NN), artificial neural network (ANN) and multifactor dimensionality reduction (MDR).

Valasapalli Mounika et al. [7], the technological advancements in today's healthcare sector have given rise to many innovations for disease prediction. Diabetes mellitus is one of the diseases that has been growing rapidly among people of different age groups; there are various reasons and causes involved. All these reasons are considered as different attributes for this study. To predict type-2 diabetes mellitus disease, various machine learning algorithms can be used. The objective of using the algorithm is to construct a predictive model to critically predict whether a person is affected by diabetes. The classifiers taken are logistic regression, XGBoost, gradient boosting, decision trees, ExtraTrees, random forest, and light gradient boosting machine (LGBM). The dataset used is PIMA Indian Dataset sourced from UC Irvine Repository. The performance of these algorithms is compared in reference to the accuracy obtained. The results obtained from these classifiers show that the LGBM classifier has the highest accuracy of 95.20% in comparison with the other algorithms.

I.K. Mujawar et al. [8], an artificial neural network started to be used in medical processes and problems when it was claimed that these processes and problems might be complex in nature and that would be best handled by nonlinear approaches. An artificial neural network has been used and applied in great numbers for medical and pathobiological



[e-ISSN: 2319-8753, p-ISSN: 2347-6710] www.ijirset.com | Impact Factor: 8.423] A Monthly Peer Reviewed & Referred Journal |

Volume 13, Issue 3, March 2024

| DOI:10.15680/IJIRSET.2024.1303008 |

processes with huge success. It is used in disease processes to represent its every subspecialty. Disease diagnosis is one of the important and highly attentive biomedical areas where an artificial neural network has been used to a great extent. In the proposed work, patient's data was collected from hospital and used in the system design. The proposed work presents an architecture, design and development of the online system for a diabetes diagnosis. The proposed system is designed with an open source development environment which will be useful for people in the Diabetes Diagnosis process.

III. METHODOLOGY

Uproarious information is available in the heap of substance that will be identified through the anomaly strategies. The information can be spatial or can be a transient method spatial connected with the geological conditions and worldly connected with the time perspectives [14, 15]. The principle point of exception identification is to deal with the loud information that is introduced in the heap of text. Different methods for recognizing abnormalities in Text are specified in below:

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 [16].

- 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 [17].

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 [18].

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



e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

Volume 13, Issue 3, March 2024

| DOI:10.15680/IJIRSET.2024.1303008 |

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 3. 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.



Fig. 1: Support Vector Classification

IV. CONCLUSION

Heart disease is a major cause of death worldwide. About 31% of global deaths occur due to this disease. According to the World Health Organization report, approximately 17.9 million people worldwide die each year due to this disease. According to the American Heart Association reports, 121.5 million American adults were affected by this disease in 2016 [3]. Heart disease should be diagnosed at an early stage; otherwise, it can lead to life-threatening conditions. Heart disease can be diagnosed using invasive and non-invasive methods. Electrocardiogram, phonocardiogram, dynamic electrocardiogram, and echocardiography are some of the non-invasive methods. Non-invasive methods to diagnose the disease are less expensive and without any side effects. But, these methods are not completely accurate.

REFERENCES

- Romany Fouad Mansour, Adnen El Amraoui, Issam Nouaouri, Vicente García Díaz, Deepak Gupta, and Sachin Kumar, "Artificial Intelligence and Internet of Things Enabled Disease Diagnosis Model for Smart Healthcare Systems", IEEE Access 2021.
- [2] G. Muhammad, M. S. Hossain, and N. Kumar, ``EEG-based pathology detection for home health monitoring," IEEE J. Sel. Areas Commun., vol. 39, no. 2, pp. 603610, Feb. 2021,
- [3] A. A. Mutlag, M. K. A. Ghani, M. A. Mohammed, M. S. Maashi, O. Mohd, S. A. Mostafa, K. H. Abdulkareem, G. Marques, and I. de la Torre Díez, "MAFC: Multi-agent fog computing model for healthcare critical tasks management," *Sensors*, vol. 20, no. 7, p. 1853, Mar. 2020.
- [4] M. S. Hossain and G. Muhammad, "Deep learning based pathology detection for smart connected healthcare," *IEEE Netw.*, vol. 34, no. 6, pp. 120125, Nov. 2020.
- [5] M. S. Hossain G. Muhammad and A. Alamri "Smart Healthcare Monitoring: A Voice Pathology Detection Paradigm for Smart Cities" Multimedia Systems vol. 25 no. 5 pp. 565-75 Oct. 2019
- [6] V Krishnapraseeda, M S Geetha Devasena, V Venkatesh and A Kousalya, "Predictive Analytics on Diabetes Data using Machine Learning Techniques", 7th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 458-463, IEEE 2021
- [7] Valasapalli Mounika, Devi Sree Neeli, Gorla Suma Sree, Parimi Mourya and Modala Aravind Babu, "Prediction of Type-2 Diabetes using Machine Learning Algorithms", International Conference on Artificial Intelligence and Smart Systems (ICAIS), pp. 167-173, IEEE 2021



|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

Volume 13, Issue 3, March 2024

| DOI:10.15680/IJIRSET.2024.1303008 |

- [8] I.K. Mujawar, B.T. Jadhav, V.B. Waghmare and R.Y. Patil, "Development of Diabetes Diagnosis System with Artificial Neural Network and Open Source Environment", International Conference on Emerging Smart Computing and Informatics (ESCI), pp. 778-784, IEEE 2021
- [9] Cecilia Saint-Pierre;Florencia Prieto;Valeria Herskovic;Marcos Sepúlveda, "Team Collaboration Networks and Multidisciplinarity in Diabetes Care: Implications for Patient Outcomes", IEEE Journal of Biomedical and Health Informatics, Vol. 14(1), pp. 319-329, 2020.
- [10] M. S. Hossain and G. Muhammad, ``Emotion-aware connected healthcare big data towards 5G," IEEE Internet Things J., vol. 5, no. 4, pp. 23992406, Aug. 2018.
- [11] M. Pham, Y. Mengistu, H. Do, and W. Sheng, ``Delivering home healthcare through a cloud-based smart home environment (CoSHE)," Future Gener. Comput. Syst., vol. 81, pp. 129140, Apr. 2018.
- [12] A. Kaur and A. Jasuja, "Health monitoring based on IoT using raspberry PI," in Proc. Int. Conf. Comput., Commun. Autom. (ICCCA), Greater Noida, India, May 2017, pp. 13351340.
- [13] U. Satija, B. Ramkumar, and M. Sabarimalai Manikandan, "Realtime signal quality-aware ECG telemetry system for IoT-based health care monitoring," IEEE Internet Things J., vol. 4, no. 3, pp. 815823, Jun. 2017.
- [14] O. S. Alwan and K. Prahald Rao, "Dedicated real-time monitoring system for health care using ZigBee," Healthcare Technol. Lett., vol. 4, no. 4, pp. 142144, Aug. 2017.
- [15] P. Kakria N.K. Tripathi and P. Kitipawang "A Real-Time Health Monitoring System for Remote Cardiac Patients Using Smartphone and Wearable Sensors" t'l. J. Telemedicine and Applications vol. 2015 2015.









INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com