International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 04 | Apr 2021 www.irjet.net p-ISSN: 2395-0072

A Comprehensive Survey of Machine Learning Techniques for IoT Data Classifications of Heart Disease Prediction

Priti V. Shinde[1], Rupali R. Shewale[2], Priyanka P. Boraste[3]

[1]*Assistant Professor, Department of Computer Engineering, NDMVPS's KBTCOE, Nashik, Maharashtra, India*
[2]*Assistant Professor, Department of Computer Engineering, NDMVPS's KBTCOE, Nashik, Maharashtra, India*
[3]*Assistant Professor, Department of Computer Engineering, NDMVPS's KBTCOE, Nashik, Maharashtra, India*

Abstract - *The healthcare industries are in demand due to increase average age, population. The advancement in information and technology boosted the development of smart cities which is having various use cases. One of the use cases is Smart Health, which is used to improve healthcare industries by providing various services such as monitoring of patients, early disease diagnosis and so on. Nowadays due to various machine learning techniques that can assist s-Health services. In this paper a survey of machine learning techniques for IoT data classification of heart disease prediction is revived. Heart is the most important and critical organ of human body. It pumps blood around body, delivering oxygen and nutrients to body cells and removing waste product from blood. Many people die due to heart disease having symptoms such as chest pain, fast heartbeats. This boosts the healthcare industries for continuous data generation. Here we have given a survey of various machine learning algorithms that will found effective for heart disease prediction.*

Key Words: *Smart Health1, IoT2, Heart Disease3, Health Care4, Machine Learning5*

1. INTRODUCTION

In today's world at every single minute data is getting generated by the human population using smart devices. Researchers evaluate that over 3 million new devices are connected to the internet each month. Researchers also estimate that in the next four years, there are going to be over 30 billion connected devices worldwide. [1] Internet of Things (IoT) is becoming a highly anticipated fuel for big data revolution by interconnecting smart devices in cities i.e. connecting all the vehicles, appliances, industrial networks, health-care services, etc. to the internet [2]. To develop the quality of living style, the internet of things (IoT) is the solution and key technology for the current world having high abilities for the future. All types of data which are generated by the IoT devices are assumed to be uploaded and stored to the cloud viz. IOT/cloud integration. To e Xtract useful knowledge cloud-based data analysis is done, and results are generated for IoT data.

The machine learning algorithms classify the data in a particular area which helps to predict the future response or helps in the decision support system. Data Analysis understands the data to find patterns and try to obtain conclusion due to which the underlying patterns are observed. Machine Learning means when we train a system to learn those patterns and try to predict the upcoming pattern [11]. Machine Learning (ML) is a field that is developed from Artificial Intelligence (AI). It focuses on designing and developing algorithms that enable the computers to change their behaviors according to observed data. The ML approaches are growing faster as a result of the improvement of the ML algorithms, enhanced methods of capturing data, improved computer networks, new sensors/IO units, and the interest in self-customization to users' behavior [3]. ML plays a vital role in s mart health; it can improve the quality of healthcare services by providing correct medical diagnosis, predicting diseases in primary stages and disease analysis.

Heart disease damages our health badly. Every year many people are victims of heart disease. Due to the weakening of heart muscle, heart disease can occur. The heart disease can be defined as the low power of heart to pump the blood. The alternate term of heart disease is the Coronary artery disease or Coronary heart disease. Coronary artery disease can arise due to insufficient blood supply to arteries. Following are the common symptoms of heart disease.

- Chest pain.
- Shortness of breath.
- Sweating and Fatigue.
- Nausea, Indigestion, Heartburn, or Stomach pain.
- Pressure in the upper back pain that spreads to an arm.

- Types of heart disease are:

Coronary artery disease (CAD).
Angina pectoris.

Congestive heart failure.
Cardiomyopathy.
Congenital heart disease. [4]

1.1 Benefits of IoT in healthcare

There is a huge demand on the need for a digital healthcare system due to the rise of chronic illness that requires timely and quality care. Providing cost-effective and preventive care for patients is challenging without innovative and advanced technologies. There is a huge demand for an improved personal and tele healthcare system. The traditional method of offering treatment is non-effective and requires a modern way of handling the operation of the healthcare system [1]. The healthcare system has been modernized in a few ways by providing remote and continuous monitoring, quick diagnosis of disease, faster response to emergency situations, and timely assistance to patients, etc. The advancement in Internet networking technologies has given rise to the new technology called as Internet of Things. The introduction to IoT and cloud has enabled ubiquitous communication by providing connectivity across multiple objects for sharing data.

IoT offers a platform to connect different devices and provides processing, communication capabilities. It is deployed in several domains to develop smart homes, smart cities, smart healthcare systems, etc. IoT permits us to visualize objects that can sense, process, communicate and share info ration over the network of devices. These interconnected things and sensors collect the data which are analyzed, used for various purposes like classification, decision making, planning etc. IoT has gained popularity due to the sensing capabilities and in traction of the things i.e., devices via the wireless network. The smart device's communication is made possible through wireless technologies such as RFID, Bluetooth, NFC and Wi-Fi. This brings in flexibility and convenience in deploying in various environments for monitoring and communication purposes. The machines or protocols developed for IoT include embedded technology which is used for sharing the information with each another. The enhancement to the healthcare system with different IoT capabilities by facilitating remote monitoring and continuous tracking of patient's health conditions, long term review of patient's health records, decreasing medical expenses and prolonging the technology for providing patient - centric care instead of hospital-centric treatment. IoT system uses a wide variety of sensors, portable devices, wireless handheld devices, wearable devices that are connected either wired or wirelessly. These devices monitor and send the information to the nearby routers or gateways for next processing to take rapid responsive action on the patients. The development of wireless communication technologies and the Internet of Things gives us opportunities to assist the healthcare industry by providing advanced solutions.

The objective of this paper is to provide a review of the various data analytics techniques for heart disease prediction.

- Quick disease diagnosis: Patient's health data is checked real time which helps to diagnose the disease at an earlier stage or even before the symptoms develop.
- Proactive treatment: Enables to provide positive health care treatment of diseases through remote monitoring.
- Cost-effective: Expenditures of unwanted visits to clinics, an appointment with doctors and hospital stays can be avoided with the usage of wearable and assistive handheld electronic medical devices.
- Drugs and equipment management: the connected things are used in the management of equipment and drug management at a very low cost.
- Possibility of less error: IoT devices reduce human errors and help in eliminating. wrong decision making.
- Advanced treatment: Helps doctors to make correct diagnosis on patient's health reports and provide appropriate treatment.

1.2 The ML, Data Analytics and IoT

The union of machine learning and IoT covers the way for a prospective advancement in efficiency, accuracy, productivity, and overall cost-savings for resource-constrained IoT devices. When machine learning algorithms and IoT work together, we can achieve increased performance for communication and computation, better controllability, and improved decision making. Due to innovative monitoring from thousands to billions of ubiquitous sensing devices and improved communication capabilities, IoT has tremendous us potential to improve the quality of human life and capable applications for industrial growth (toward Industry 4.0). Due to IoT's capability with machine learning and artificial intelligence, Advanced machine intelligence techniques have made it possible to search the huge volume of IoT data to have better insights into a range of real-world problems [7]. Therefore, to solve real time complex problems and to meet the computation and communication requirements successfully, IoT, machine learning and data analytics must complement each other. In current years, IoT data analytics has gained significant importance and attention due to High volume of data generated from distributed IoT devices. This number will keep increasing over time due to the wide addition of IoT devices in a wide range of

complex applications. Intelligent data analytics will play vital role to identify and predict the future states of any process or system by mining this huge amount of data powerfully and intelligently. There exists variety of IoT devices, which include mobile phones, PC/ Laptop, tablets to short-range and wide area IoT devices. Due to the variety of the data, the features, formats, and attributes of the data are different. Also, based on different IoT applications domains, the data deluges also vary. For ex, the IoT devices used for medical applications will be different from a smart home IoT, smart agriculture. The factors such as quality, processing, and storage of data have also become a challenging because of its heterogeneity. Some papers find answers to critical questions such as: how to deal with the sampling procedure of the high-frequency streaming data, noise cancellation, filtering of the data, and gathering. Also, some papers find answers to critical questions such as merging of the data from different data sources, data interpretation, interoperability, reasoning, state awareness and knowledge creation from the data, gathering and storing data from heterogeneous data sources to meet application's constraints [8]. Uncertainty is also one problem. The problem may arise in the IoT data stream due to the failure of any IoT device or communication channel during data transfer. Gross errors and missing data are present in IoT data, which require advanced analytics to preprocess the data.

2. CLASSIFICATION OF DIFFERENT ANALYTICS TECHNIQUES FOR IOT

2.1.1 Descriptive Analytics

IoT systems can gather data from a few to more of smart devices and transmit them to cloud platforms. Based on past data, it is always possible to generate detailed insights into past events by using advanced machine learning techniques. These groups of machine learning-based algorithms that process and summarize the raw data and provide decisions which basically comprise the field of study called descriptive analytics. Data aggregation, data summarization, mathematical logical operations, data mining, clustering algorithms are some examples of descriptive analytics. Descriptive analytics requires a high-volume data. Current technological advances have confirmed that cloud storage can store large volumes of IoT data and cloud servers can process difficult tasks using high-performance computers, and by applying IoT cloud analytics.

2.1.2 Predictive Analytics for IoT

Predictive analytics depends on historical data and utilize advanced statistical or machine learning techniques to model the behavior or pattern so that it is possible to predict the similar possible future trends or patterns in data. It predicts what will happen in the future by learning the historical patterns and data relationship of existing data. Predictive analytics have been widely used for different applications including predictive maintenance, prediction of price, supply-demand trend or prediction any outcome. There are two types of data analytics models (i) classification-based models (ii) regression-based modeling techniques include statistical regression-based models, decision trees, and neural network or deep neural network-based models. Other widely used algorithms are based on Bayesian analysis, gradient boosting, and Ensemble model-based analysis. These predictive analytics techniques are dependent on data for decision making. The IoT paradigm helps to facilitate the data gathering process from smart IoT devices and provide an analytical framework using the cloud.

2.1.3 Prescriptive Analytics for IoT

Prescriptive analytics recommends how to respond to any future events based on data analysis. This class of analysis not only predicts the future states but also provides recommendations for the outcome. It combines the benefits of both descriptive and predictive analytics. While predictive analytics suggests what and when the event will occur based on future predictions, prescriptive analytics extends the capability by providing detailing of the future predictions with impact analyses. Prescriptive analytics is widely used to enhance the business outcome. Prescriptive analytics is suitable in an Industrial IoT (IoT) setup where business intelligence-based decisions are made by using the capabilities of cloud computing, big data analytics and machine learning. .

2.1.3 Adaptive Analytics for IoT

In actual implementation, the outcome obtained from the predictive analytics needs to adjust with real-time data. For this purpose, adaptive analytics are used to adjust or enhances the process outcome based on the recent history of the data and by looking at their correlations. This type of analysis helps to increase model performance and reduces errors. The advantage of adaptive analytics is that when a new set of input data is received, it can adjust the outcome of the solution. In an IoT environment, adaptive analytics are a good fit for real-time stream data processing.

3. IoT MODELS BASED ON ML ALGORITHMS

In this section based on different IoT applications, machine learning algorithms are surveyed.[12]

Table -1: IoT Applications

IoT Application	WSN IoT protocol	ML Algorithm
Smart Healthcare (Remote Patient Monitoring, Data analytics, appointment reservation)	Bluetooth , ZigBee	Unsupervised learning & supervised learning algorithms (SVM, ANN), Feed forward neural network
Smart Transportation (Intelligent VANET, highway monitoring, speed regulation, driverless vehicle)	Bluetooth , ZigBee	Unsupervised Clustering Approach (K means algorithm)
Smart Home (Home security, smart appliances, security system, assistance device)	Z-Wave, Insteon, Wavenis, Bluetooth , Wi-Fi, and ZigBee.	Supervised Learning algorithms (SVM , KNN, Naive Bayes) & Unsupervised Learning (k-means)
Smart Building (Energy efficiency, security, smart lightning, attendance monitoring)	Bluetooth , Wi-Fi, z-wave ,ZigBee	Supervised Learning algorithms (ANN, SVM, Bayesian Network) Unsupervised Learning(Hidden Markov)
Smart Education (Digital attendance, tracking, lab safety monitoring, s mart access)	Insteon, Wavenis, Bluetooth , Wi-Fi, and ZigBee.	Supervised Learning Algorithms (SVM , KNN, Naive Bayes)
Smart Infrastructure (Smart Lightning, automatic weighing, toll management, fault detection)	Bluetooth , Wi-Fi ,ZigBee	Unsupervised Learning Algorithms (K-means algorithm, Clustering and Anomaly detector)
Smart Grid (Energy efficiency, fault detection, monitoring)	ZigBee, 6LoWPAN, RPL, Z-wave, wavenis	Supervised Learning algorithms (SVM, Discriminant Analysis, Decision Trees)

3.1 Health Disease and its prediction

Over the last decade, heart disease or cardiovascular remains the primary basis of death over the world. On approximate by the World Health Organization, that over 17.9 million deaths occur every year because of cardiovascular disease, and 80% are because of coronary artery disease and cerebral stroke [9]. The vast number of deaths is common amongst low and middle-income countries [10]. Many factors such as personal and professional habits, genetic accounts for heart disease. Various bad habits such as smoking, overuse of alcohol, caffeine and stress, physical inactivity along with other physiological factors like obesity, hypertension, high blood cholesterol and pre-existing heart conditions are responsible factors for heart disease. The efficient, accurate and early medical diagnosis of heart disease plays an important role in taking preventive measures to prevent death. Heart disease becomes the major cause of death in the world. Medical diagnosis should be efficient, reliable and aided with computer techniques to reduce the effective cost for diagnostic tests. Data mining helps computers to build and classify various attributes. In this paper different classification techniques are surveyed to predict heart disease. Machine Learning - Machine learning is an emerging sub branch of artificial intelligence. Its main focus is to design systems, allow them to learn and make predictions based on the experience. It trains machine learning algorithms using a training dataset to test or create a model. The model uses the novel input data to predict heart disease. Using algorithms, it detects hidden patterns in the input dataset to build models. It makes correct predictions for new datasets. The dataset is cleaned first, and missing values are filled. Machine learning techniques are classified as: Supervised Learning: The model is trained on a labeled dataset. It has input data and its outcomes. Data is classified and split into training and test dataset. Training dataset trains model while testing dataset functions as new data to get accuracy of the model. The dataset exists with models and its output. The classification and regression are its example.

Unsupervised Learning: Data used to train are not classified or labeled dataset. Its main aim is to find hidden patterns in the data. The model is trained to develop patterns. It can predict hidden patterns for any new input dataset, but upon e exploring data, it finds conclusion from datasets to describe hidden patterns. The clustering method is an example of an unsupervised learning technique.

Reinforcement Learning: It does not use labeled dataset or the resultant data, thus model learns from the experience. In this technique, the model improves its presentation based on its association with environment and concludes how to discuss its faults and to get the right

outcome through assessment and testing various scenes. Classification algorithms are commonly used to define probability of heart disease occurrence.

3.2 ML Algorithms classification for Health disease prediction

The classification task is used for prediction which is dependent on past information. Many data mining techniques such as Naïve Bayes, neural network, decision tree have been applied by researchers to have a precision diagnosis in heart disease. The accuracy given by different techniques varies with number of attributes or input factors.

Support Vector Machine: The SVM is an A Technique which has been used for the most partite arrangement issues. Because of the high success of SVM in arrangement, different applications are generally applied it.

K-Nearest Neighbor: K-NN calculation predicts the class name of another data; K-NN utilizes the correspondence of new commitment to its origins of information tests in the readiness set.

Logistic Regression: Logistic regression is a supervised learning algorithm used to predict the binary form of a target variable. It is the easiest, easiest algorithm used in machine learning that can be used for various problems such as disease prediction, cancer detection and so on.

Naïve Bayes Classifier*:* Naive Bayes is a statistical classifier. It is based on Bayes' theorem. A naïve Bayesian classifier has good performance with decision tree and other selected classifiers. The computation cost can be reduced significantly. It is easy to implement.

Decision Tree Classifier:

This classifier belongs to supervised learning. It can be used for regression and classification problems. We can use this algorithm for problems where we have continuous but also categorical input and target features. It is the most effective machine learning algorithm used for describing the trees in a graphical manner.

Random Forest Classifier*:*

Random forest is a supervised learning algorithm. It can be used for classification and regression. It is simple and easy to implement. A forest is comprised of trees. This classifier generates decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting. The random forest composed of multiple decision trees.

By studying different classification techniques, here we have presented accuracy of those techniques. We have surveyed various papers and presented best accuracy for those classification techniques.

ML algorithm	Accuracy
K-Nearest Neighbor	87%
Logistic Regression	84%
Naïve Bayes	88.157%
Decision tree	80.263%
Random Forest Classifier	86.84%
Support Vector Machine	86%

4. CONCLUSIONS

IoT consists of many various devices that relate to each other and transfers huge amounts of data. The smart health is one of the most important applications of IoT and provides various services in domains such as disease prediction, tests. These services can be improved and boosted by analyzing the smart data collected from these areas. To e Xtract information from collected data, many data analytic algorithms can be applied. The overall aim is of the paper is to survey various data mining techniques useful in effective heart disease prediction. This paper shows that the machine learning algorithms can be used to predict the heart disease easily with various parameters. Machine learning is an effective way to solve the problems in different areas too**.**

REFERENCES

1. Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges. (2017). IEEE Access, 5, 5247-5261. doi: 10.1109/access.2017.2689040
2. Mahdavinejad, M., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., &Sheth, A. (2018). Machine learning for internetof things data analysis: a survey. Digital Communications AndNetworks, 4(3), 161-175. doi: 10.1016/j.dcan.2017.10.002
3. John D. Kelleher, Brian Mac Namee, Aoife D'Arcy, Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies, 1st Edition, The MIT Press, 2015
4. Heart Disease Prediction with Machine Learning Approaches, MeghaKambojDepartment of Computer Science & Engineering, IMS Engineering College, Ghaziabad, Uttar Pradesh, India, International Journal of Science and Research (IJSR) ISSN: 2319-7064 ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583
5. Heart Disease Prediction using Machine Learning Techniques, Devansh Shah1 · Samir Patel1 ·Santosh Kumar Bharti1,Received: 27 September 2020 / Accepted: 2 October 2020 / Published online: 16 October 2020© Springer Nature Singapore Pte Ltd

2020
6. Smart Health Monitoring System using IOT and Machine Learning TechniquesHoney PandeyDepartment of ECEHindustan Institute of Technology and Science ,S. PrabhaDepartment of ECEHindustan Institute of Technology and Science
7. Machine learning and data analytics for the IoT Erwin Adi1 • Adnan Anwar2 • Zubair Baig2 • Sheralieadally3,Received: 17 January 2020 / Accepted: 20 March 2020 / Published online: 11 May 2020 Springer-Verlag London Ltd., part of Springer Nature 2020
8. Sun Y, Song H, Jara AJ, Bie R (2016) Internet of things and big data analytics for smart and connected communities. IEEE Access 4:766–773. https://doi.org/10.1109/ACCESS.2016 2529723
9. Seckeler MD, Hoke TR. The worldwide epidemiology of acuterheumatic fever and rheumatic
10. Gaziano TA, Bitton A, Anand S, Abrahams-Gessel S, Murphy A.Growing epidemic of coronary heart disease in low-and middleincomecountries. CurrProblCardiol. 010;35(2):72–115.
11. https://datascience.stackexchange.com/
12. A survey on Machine Learning algorithms for Wireless Internet of Things (IoT) ,Ms. Priti V Shinde, Ms. Rupali R. Shewale,International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 05 | May20,www.irjet.net,p-ISSN:2395-0072
13. JithinJagannath , Nicholas Polosky , AnuJagannath , Francesco Restuccia , TommasoMelodia , Machine learning for wireless communications in the Internet of Things: A comprehensive survey, Ad Hoc Networks 93 (2019) 101913, Elsevier
14. K. Ashton , That 'Internet of Things' thing, RFID J. (2009)
15. Glen Martin (Forbes), How the internet A .Whitmore , A . Agarwal , L. Da Xu , The internet of things –a survey of topics and trends, Inf. Syst. Front.17 (2) (2015) 261–274 .
16. of things is more like the in- dustrial revolution than the
17. digital revolution(https://www.forbes.com/sites/ oreillymedia/2014/02/10/more- 1876- than-1995/#674c4e0b66d2
18. K means- https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1
19. Mohammad SaeidMahdavinejad, MohammadrezaRezvan, MohammadaminBarekatain,
20. PeymanAdibi, PayamBarnaghi, Amit P. Sheth, Machine Learning for Internet of Things Data Analysis: A Survey, Digital Communications and Networks, 10.1016/j.dcan.2017.10.002