**Heart Disease Prediction**

**A PROJECT REPORT**

**Submitted By**

**Neha Bansal**

**(2000290140075)**

**Aditya Awasthi**

**(2000290140010)**

**Mohd Bilal**

**(2000290140070)**

**Submitted in partial fulfilment of the**

**Requirements for the Degree of**

**MASTER OF COMPUTER APPLICATIONS**

**Under the Supervision of**

**Dr. Arun Kumar Tripathi**



**Submitted to**

**Department Of Computer Applications**

**KIET Group of Institutions, Ghaziabad**

**Uttar Pradesh-201206**

**(Jan 2022)**

ABSTRACT-

In addition to information, communications, and technology, there are new areas such as machine learning, the Internet of Things, and cloud computing. These technologies can save millions of lives in the healthcare world and can be used in healthcare systems that lack medical expertise. In the healthcare field, these technologies have the potential to save millions of lives and can be employed in healthcare systems that lack medical competence. Fast food consumption has risen in recent decades, contributing to high cholesterol, diabetes, and other health issues affecting the heart and other organs. Another factor that contributes to health issues, such as cardiovascular disease, is lifestyle changes. Cardiovascular is also a top cause of the deaths, according to the World Health Organization. The purpose of this study was to analyse existing cardiovascular disease data to predict heart disease early and prevent its onset. Records of heart disease patients are extracted from India and stored via the cloud. The proposed task predicts the likelihood of heart disease and ranks patient risk levels by performing various data mining techniques such as naive arrays, decision trees, logistic regression, and random forest. In this study we analysing the performance of different machine learning algorithm by comparison.

Introduction-

Heart attacks, arrhythmias, pericardial illnesses, and fatalities have all increased in the last several decades over the world (CVDs). In the India, CVD kills one person every minute. Many researchers have experimented with machine learning classification algorithms in order to diagnose cardiovascular illness and assist medical practitioners throughout the world in improving local health systems. According to the World Health Organization (WHO), cardiovascular disease causes about a quarter of all fatalities worldwide, with more than three-quarters of those deaths happening in low- and middle-income nations. In India, 25 percent of the population aged 2569 dies from cardiovascular disease [1]. The Internet of Things (IoT) is also known as the physical Internet since it is linked to the Internet but has limited storage and processing capabilities. We're still dealing with challenges like performance, interoperability, security, and privacy, and there's a lot of room for progress [2]. [3] [4].

Smart sensors are utilized for continuous monitoring of specific patient ailments, and IoT has shown to be the gold standard for medical systems. Biomedical sensors are examples of smart sensors that collect health-related data and send it to doctors via the cloud/edge for further diagnosis. As a result, regardless of region, IoT can assist in bridging the gap between patients and doctors [5]. Cloud computing provides nearly limitless processing and storage capacity. Data mining is a clever technology used to analyse large numbers of documents and extract new information, despite being a more advanced solution for solving IoT technological challenges. [6]. Various machine learning algorithms can be utilized to produce specific judgements, estimations, and predictions. The majority of medical data is now collected via computer systems, yet it is not used for analysis everywhere in the world. It accumulates as an old handwritten trace in the database and is no longer useful. Cancer, cardiovascular disease, diabetes, and dengue fever can all be predicted using this information [7]. As a result, we suggest a novel information technology (IT) model. The IoT ML Cloud model combines machine learning with IoT and cloud computing, resulting in three integrated technologies that work together to eliminate hurdles in existing and future global systems. future. Medical attention. Medical assistive technology and services are inextricably linked to public health and improved medical facilities. In the prediction of chronic diseases, the integration of cloud computing and IoT for healthcare-oriented modern technology applications is critical. High security, enhanced efficiency, virtualization, dependability, scalability, resource sharing, cost reduction, medical monitoring, management, and administration systems are all advantages of the growth of public cloud (cloud computing) in hospitals. The process is highly efficient and precise.

Related Work-

Many studies are being conducted on disease prediction using artificial intelligence, machine learning algorithms, the Internet of Things, and other technologies.

This paper suggested an IoT-based health monitoring system based on a random forest algorithm [8]. Using a random forest method, many diseases such as heart disease, diabetes, and breast cancer were predicted and attained with maximum accuracy in the dermatology dataset. [9] The combined cardiovascular illness recommendation system [10] [11] employs IoT in a cloud setting to solve a multi-class classification issue that can predict eight different types of cardiovascular disorders. To increase accuracy, the model used a feature selection strategy. Lung Cancer developed an Io-based prediction system employing segmentation and cluster-based fuzzy classification in one study. The proposed system, which was created in the MATLAB environment, is designed to classify radiographic images of the lungs. He has created an online platform for clinical decision support systems that is web-based (CDSS). It is based on the most effective deep neural network (DNN). They created a cloud-based CDSS system to forecast the severity of chronic kidney disease (CKD). They used Internet of Things sensors to collect clinical data from relevant patients, anticipate normal and abnormal states, and obtain the best DNN accuracy. [12] Researchers have developed an Android-based monitoring device that can track the heart rate of heart disease patients. We built a model that can trigger an alarm if a patient's heart rate is irregular using a decision tree method. It delivers an Io-based system that uses machine learning algorithms to detect cardiac problems early.

[13] Researchers looked at IoT security and privacy elements such as security criteria, utilization, and categories of healthcare attacks. To deal with noisy missing values, they employ a decision tree-based classifier of how people maintain social and economic bases for sustainability. Developed a mobile healthcare environment that is capable of driving.

Objective-

The objective

of this research is to analyze the available data pertaining to cardiovascular diseases for prediction of heart

diseases at an earlier stage to prevent it from occurring. The dataset of heart disease patients was taken from

Jammu and Kashmir, India and stored over the cloud. Stored data is then pre-processed and further analyzed

using machine learning techniques for the prediction of heart diseases.

The objective

of this research is to analyze the available data pertaining to cardiovascular diseases for prediction of heart

diseases at an earlier stage to prevent it from occurring. The dataset of heart disease patients was taken from

Jammu and Kashmir, India and stored over the cloud. Stored data is then pre-processed and further analyzed

using machine learning techniques for the prediction of heart diseases.

The objective

of this research is to analyze the available data pertaining to cardiovascular diseases for prediction of heart

diseases at an earlier stage to prevent it from occurring. The dataset of heart disease patients was taken from

Jammu and Kashmir, India and stored over the cloud. Stored data is then pre-processed and further analyzed

using machine learning techniques for the prediction of heart diseases.

The objective

of this research is to analyze the available data pertaining to cardiovascular diseases for prediction of heart

diseases at an earlier stage to prevent it from occurring. The dataset of heart disease patients was taken from

Jammu and Kashmir, India and stored over the cloud. Stored data is then pre-processed and further analyzed

using machine learning techniques for the prediction of heart diseases.

The objective

of this research is to analyze the available data pertaining to cardiovascular diseases for prediction of heart

diseases at an earlier stage to prevent it from occurring. The dataset of heart disease patients was taken from

Jammu and Kashmir, India and stored over the cloud. Stored data is then pre-processed and further analyzed

using machine learning techniques for the prediction of heart diseases.

The objective

of this research is to analyze the available data pertaining to cardiovascular diseases for prediction of heart

diseases at an earlier stage to prevent it from occurring. The dataset of heart disease patients was taken from

Jammu and Kashmir, India and stored over the cloud. Stored data is then pre-processed and further analyzed

using machine learning techniques for the prediction of heart diseases.

The goal of this study is to examine the current data on cardiovascular illnesses in order to predict heart disease at an earlier stage and so prevent it from occurring. The heart disease patient data was imported from India and stored in the cloud. For the prediction of cardiac illnesses, the stored data is pre-processed and further analysed using machine learning techniques.

The benefits of developing this study article include increased security, enhanced efficiency, virtualization, dependability, and scalability, which can promote resource sharing, cost savings, and a highly efficient and accurate medical monitoring, management, and administration system.

Additionally, this study aims to identify patients with cardiac disease.

Dataset description-

Among the 11 features in this dataset, there is a target variable. It has 6 nominal variables and 5 numeric variables. The detailed description of all the features are as follows:

1. Age: Age of patients in years (numerical)

2. Sex: Gender of patient (Male - 1, Female - 0) (Nominal)

3. Chest Pain Type: A patient's chest pain is classified into four categories: typical, angina, non-anginal pain, and asymptomatic (Nominal).

4. Resting Bps: Number of mm/HG of blood pressure at resting (Numerical)

5. Cholesterol: Number of milligrams of cholesterol in a liter of blood

6. Fasting blood sugar: A value of 1 indicates a true value and 0 indicates a false value (Nominal) for fasting blood sugar levels >120 mg/dl

7. Resting ECG: In the electrocardiogram while at rest, three values are presented: 0: Normal 1: ST-T wave abnormality 2: Left ventricular hypertrophy (nominal)

8. Max Heart Rate: Result of maximum heart rate (Numeric)

9. Exercise Angina: Angina induced by exercise 0 portraying NO, 1 depicting Yes (Nominal)

10. Oldpeak: Compared to rest, exercise caused ST-depression (Numeric)

**Target variable**

11. HeartDisease : It is the target variable which we have to predict 1 means patient is suffering from heart risk and 0 means patient is normal.

Machine Learning Algorithms-

Five machine learning techniques are employed to develop an illness prediction model in this research.K Nearest Neighbor (KNN), Decision Tree Classifier (DTC), Support Vector Machine (SVM), Random Forest (RF), and Nave Bayes are all used in this model (NB).As the dataset has output class labels, supervised algorithms are suitable for handling class label problems.

Big data in healthcare information processing-

In the healthcare system, big data is utilized to anticipate high-risk diseases that can save a person's life, predict the patient's state from two chronic diseases at the same time, and save money on medical services[18]. The management and treatment of diseases information is maintained in a big data database sequentially in many nations. The Conventional Electrocardiogram (ECG) collects a large amount of data in a short amount of time. The healthcare gadget uses cloud computing and an IoT structure. The more expensive home use devices are ECG based. For this reason, an ECG tracking device is designed, which works in conjunction with the IoT cloud.

Diagram

Description automatically generated

The five representations that differentiate the enormous facts are volume, velocity, diversity, truth, and value. The Size of the dataset is represented by the Volume in huge facts, and its length ranges from (10) 12 bytes to (10) 21 bytes[19]. Velocity represents the movement of the facts as well as the arrival of fresh facts. The numerous sorts of facts and the issues that come from them are represented by variety. The obligation of the facts is indicated by Veracity due to the lack of confidence in fact versions. Figure 1 depicts the data processing within the healthcare device. Value refers to the knowledge derived from the significance of a fact. However, there are several limitations to traditional data analytics within the large data's fitness informatics [20]. As a result, advancements in big data in healthcare devices increased quickly in the form of smart phones, implantable and wearable devices, and real-time sensors.

Cloud computing with IoT architecture in healthcare-

In a fitness care gadget, a large amount of data from the patient's frame sensor needs to be saved and regulated properly. Information from sensors is currently saved and controlled for processing via cloud computing with IoT via wearable or implanted sensors [23]. Previously, a completely wearable sensor based on the Wireless Body Area Network (WBAN) was employed, which included the data gathering, communication networking, and provider layer.

The patient's body temperature and blood pressure, as well as other physiological markers, are detected by this wearable sensor. Furthermore, IoT with cloud computing necessitates energy consumption in sensing devices for low-energy communication methods. In Zigbee, IEEE 802.15. four is commonly used in LR WPANs (low price WPANs) to improve the records replacing between Personal Operating Space (POS) at low energy environments of around 10m. IPv6 with Less Power Wireless Network (6LoWPAN) was developed to connect electricity-limited WPAN devices to the internet, further complicating the IoT concept [21]. It does, however, assist fragmentation systems in repairing IPv6 datagrams.

Diagram

Description automatically generated

The Conventional Electrocardiogram (ECG) collects a large amount of data in a short amount of time. The healthcare gadget uses cloud computing and an IoT structure. The more expensive home use devices are ECG based. For this reason, an ECG tracking device is designed, which works in conjunction with the IoT cloud.

Predictive analytics system for Cardiovascular data-

According to World Health Organization (WHO) data, cardiovascular illnesses are one of the worst infectious diseases in the world, affecting roughly 30% of the global population [23]. The goal of the pattern or data mining technique is to categorize Cardiovascular connected to determine illness severity. The severity is calculated using a parameter value that is depending on the diagnostic stage. Because a large amount of data is available in the network, some of it is unbalanced, filtering the dataset is critical [24]. For data categorization approaches, modern standard schemes such as the Naive Bayes (NB) classifier, j48, CART, plain logistic, decision tree, and multilayer perception are employed [25]. Furthermore, all of these classifiers were created by researchers for various body diseases.

Reference-

[1] V. Fuster, B. B. Kelly, eds., “Promoting cardiovascular health in the developing world: a critical challenge to achieve global health,” National Academies Press, 2010, pp 465.

[2] P. Kaur, R. Kumar, M. Kumar, “A healthcare monitoring system using random forest and internet of things (IoT)”, 2019, pp. 19905-1916.

[3] I. Cvitić, D. Peraković, M. Periša, M. Botica, “Novel approach for detection of IoT generated DDoS traffic,” Wireless Networks, 2019, pp. 1-14, doi: 10.1007/s11276-019-02043-1.

[4] A. Tewari, B.B. Gupta, “Security, privacy and trust of different layers in Internet-of-Things (IoTs) framework,” Future generation computer systems”, 2002, pp 909-920.

[5] I. Cvitic, D. Perakovic, M. Perisa,M. Botica, “Definition of the IoT device classes based on network traffic low features”, in 4th EAI International Conference on Management of Manufacturing Systems. Springer, Cham, 2020, pp 1-17.

[6] P.K. Senyo, E. Addae, R. Boateng, “Cloud computing research: “A review of research themes, frameworks, methods and future research directions”, International Journal of Information Management, vol. 38, no. 1, pp 128-139, 2018.

[7] Y. Deepthi, Kalyan K.P., Vyas M., Radhika K., Babu D.K., Krishna Rao N.V. “Disease Prediction Based on Symptoms Using Machine Learning”. In: Sikander A., Acharjee D., Chanda C., Mondal P., Verma P. (eds) Energy Systems, Drives and Automations. Lecture Notes in Electrical Engineering, vol. 664, Springer, 2020

[8] F. Jabeen, M. Maqsood, M. A. Ghazanfar, F. Aadil, S. Khan, K. Kim, “An IoT based efficient hybrid recommender system for cardiovascular disease,” Peer-to-Peer Networking and Applications, vol.12, no. 5, pp 1263-1276, 2019.

[9] D. Palani, K. Venkatalakshmi, “An IoT Based Predictive Modelling for Predicting Lung Cancer Using Fuzzy Cluster Based Segmentation and Classification,” Journal of medical systems,vol. 43, no. 2, pp. 1-12.

[10] A. Benjemmaa, H. Lti, M. Ben Ayed, “Design of Remote Heart Monitoring System for Cardiac Patients”, In International Conference on Advanced Information Networking and Applications. Springer, Cham, 2019. pp.963-976.

[11] Hashi, E. K., Zaman, M. S. U., Hasan, M. R. “An expert clinical decision support system to predict disease using classification techniques”. International Conference on Electrical, Computer and Communication Engineering (ECCE),2017. pp. 396-400

[12] K. Saravananathan, T. Velmurugan, “Analyzing Diabetic Data using Classification Algorithms in Data Mining,” Indian Journal of Science and Technology, vol. 9, no. 43, pp 1-6, 2016.

[13] D. Dziak, B. Jachimczyk, W. Kulesza, “IoT-Based Information System for Healthcare Application: Design Methodology Approach,” Applied Sciences, vol. 7, no. 6, pp. 596, 2017.

[14] A.M. Koli, M. Ahmed, “Machine Learning Based Parametric Estimation Approach for Poll Prediction”, Recent Advances in Computer Science and Communications, vol. 14, no. 4, pp. 1287 - 1299, doi: 10.2174/2666255813666191204112601.

[15] P. P. Chavda, “Early Detection of Cardiac Disease Using Machine Learning,” In 2nd International Conference on Advances in Science & Technology (ICAST) 2019.http://dx.doi.org/10.2139/ssrn.3370813.

[16] Thanh Noi, P., Kappas, M.” Comparison of random forest, k-nearest neighbour, and support vector machine classifiers for land cover classification using Sentinel-2 imagery.Sensors”, vol. 18, no. 1, pp. 18, 2018.

[17] Podgorelec V, Kokol P, Stiglic B, Rozman I. “Decision trees: an overview and their use in medicine”, Journal of medical systems, vol. 26, pp 445-463, 2002.

[18] Avati, Anand, et al. "Improving palliative care with deep learning." BMC medical informatics and decision making 18.4 (2018): 122.

[19] Gandomi, Amir, and Murtaza Haider. "Beyond the hype: Big data concepts, methods, and analytics." International journal of information management 35.2 (2015): 137-144.

[20] Wang, Yichuan, LeeAnn Kung, and Terry Anthony Byrd. "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations." Technological Forecasting and Social Change 126 (2018): 3-13.

[21] Liu, Ruiqing, et al. "Profile of Consecutive Fecal Calprotectin Levels in the Perioperative Period and Its Predictive Capacity for Early Endoscopic Recurrence in Crohn’s Disease." Diseases of the Colon & Rectum 62.3 (2019): 318-326.

[22] Manogaran, Gunasekaran, et al. "A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system." Future Generation Computer Systems 82 (2018): 375-387.

[23] World Health Organization. "Compendium of good practices in the implementation of the Tuberculosis Action Plan for the WHO European Region 2016–2020." (2019).

[24] Alonso-Betanzos, Amparo, et al. "Preprocessing in High Dimensional Datasets." Advances in Biomedical Informatics. Springer, Cham, 2018. 247-271.

[25] Chen, Wei, et al. "A Comparative Study of Functional Data Analysis and Generalized Linear Model Data-Mining Techniques for Landslide Spatial Modeling." Spatial Modeling in GIS and R for Earth and Environmental Sciences. Elsevier, 2019. 467-484.

Conclusion-

Heart disease is becoming more widespread in people everywhere, including in our own country (India). As a result, anticipating disease before infection lowers the risk of death. Lot of press has been gotten from this forecast.. Our findings are part of a bigger study on the identification and prediction of heart disease. With SVM and KNN, a genuine Algerian dataset with excellent results, a neural network obtained 93 percent accuracy. Our research is strong because we examined the algorithm's stability on datasets of varied sizes. In the end, neural networks were the most successful. To determine the connections between the attributes, we conducted feature selection experiments or employed correlation matrices. There are certain methods to magnify these approaches. B. Using deep learning techniques Increasing the size of the dataset, employing various approaches for attribute selection, and applying algorithms.