Big Data Analytics in Heart Attack Prediction

[1]Anmary Sunny, [2]Anagha Saju, [3]Sigma Sathyan

[1] Student, [2] Student, [3] Assistant Professor
[1]Department of Computer Science,
[1]Santhigiri College of Computer Sciences, Vazhithala, Idukki

Abstract: In today's world, heart attack is very common. Therefore, it is a big challenge to diagnose the patients timely and correctly. The treatment of heart disease is quite expensive and not affordable for most of the patients. The main goal is to develop a system to predict heart attacks using Big Data techniques. This system is able to discover and extract hidden knowledge about heart diseases from a historical database of heart diseases. It can answer complex queries about heart disease diagnosis, helping doctors make intelligent clinical decisions, which traditional decision support systems cannot. By providing effective treatments, it also helps reduce treatment costs. Big Data analytics, known in the corporate world for its valuable use in controlling, contrasting and managing large amounts of data, can be applied with great success to the prediction, prevention, management and treatment of cardiovascular disease. Therefore, a medical diagnostic system such as the Heart Attack Prediction System would likely be extremely useful.

Index Terms: Mobile computing, Healthcare, Communication, Pregnancy. (Key words)

I. INTRODUCTION

Acute Myocardial infarction (AMI), commonly referred to as a heart attack, is among one of the deadliest of cardiovascular diseases. AMI happens as circulation or blood flow to heart muscle is interrupted, causing the heart muscle to damage or die (become necrotic) . The primary reason for most heart attacks is a blockage which causes blood flow to one of the coronary arteries, vital channels through which blood travels to the heart muscle, to become reduced or obstructed. When blood flow is obstructed or reduced, the heart muscle is rapidly deprived of red blood cells which carry the necessary oxygen essential for sustaining life and consciousness in the human body. It takes as few as six to eight minutes without oxygen to cause the heart muscle to arrest, leading to the individual's death. The cause of most heart attacks is plaque, a hard substance which builds up over time in the coronary arteries. Plaque, a substance made up of numerous cells and cholesterol (fat), draws platelets, which increase over time, causing a blockage large enough to diminish or block blood flow to heart muscle. Some individuals have a build-up of plaque in the arteries over many years and this is known as atherosclerosis. Examining the cause and ethology of atherosclerosis, it can be described as a chronic inflammation. And when examining AMI, it also could be described as acute inflammation. White blood cell production in the bone marrow is increased due to signaling via the sympathetic nervous system aier the AMI as well as in the spleen. His increase in white blood cell production migrate to the heart and vessel wall and can be recruited into other atherosclerotic plaques, causing more inflammation and likely subsequent ischemic events such as reinfarction or stroke.

There are generally two phases of wound healing when it comes to monocytes and macrophages. Initially there is an early inflammatory phase and aierwards, a reparative phase begins. However, both phases are necessary for proper wound healing; but if either of these phases is stalled or if the inflammation continues too long, resolution of the Big Data..

II. WHAT IS BIG DATA

Big Data refers to data sets that are too large or too complex to be processed by conventional data processing software. Data with many fields (rows) offer greater statistical power, while data with higher complexity (more attributes or columns) can lead to a higher false discovery rate. Challenges in Big Data analysis include data collection, data storage, data analysis, search, sharing, transmission, visualization, query, updating, information protection and data source. Big Data was originally associated with three key terms: volume, variety and velocity. The analysis of Big Data poses a sampling challenge, so only observation and sampling have been possible so far. A fourth concept, veracity, therefore relates to the quality or insight value of the data. Without sufficient investment in Big Data veracity expertise, the volume and variety of data can create costs and risks that exceed a company's ability to create and leverage value from Big Data [1].

Current use of the term "Big Data" usually refers to the use of predictive analytics, user behavior analytics or certain other advanced data analytics methods that add value from Big Data, and rarely to a specific size of data set. "There is little doubt that the data sets now available are indeed big, but that is not the most important feature of this new data ecosystem. By analyzing datasets, new relationships can be found to "identify economic trends, prevent disease, fight crime and so on". Scientists, business leaders, medical professionals, advertising executives and governments regularly encounter difficulties with big data sets in areas such as internet search, fintech, health analytics, geographic information systems, urban informatics and business informatics. Scientists encounter limitations in their e-science work in areas such as meteorology, genomics, connect omics, complex physics simulations, biology and environmental research.

III. BIG DATA ANALYTICS IN HEALTH CARE

Healthcare is one of the largest and most complex industries, with patients constantly demanding better care management.. There are numerous new data sets every day and it is becoming increasingly difficult to effectively capture and analyse all the data. Electronic health records (EHRs) and all digital data relevant to healthcare are being amassed in vast quantities, making it necessary to use [2] technology and its tools to take advantage of the data available. This is where Big Data comes in.

Big Data refers to the vast amounts of information created by the digitisation of everything, which is consolidated and analysed using specialised technologies. The data comes from numerous sources, including electronic health records, medical imaging,

genomic sequencing, payer data, pharmaceutical research, wearables, medical devices, etc. The application of data analytics in healthcare has many positive and also life-saving implications. It uses specific health data of a population (or a specific person) and can potentially help prevent epidemics, cure diseases, reduce costs, etc.

The use of Big Data in healthcare can transform the industry, moving it away from a fee-for-service model and towards value-based care. In short, it can deliver on the promise of reducing healthcare costs while identifying ways to provide better patient experiences, treatments and outcomes

New challenges In Big Data Analytics in Medical Field

- Collecting the data
  Data insights are often retrieved from any source, but healthcare lacks habits for data management. Therefore, it is a daunting task to collect clean, accurate and complete data that can be used in multiple systems.
- Cleaning the data
  Many healthcare providers focus on keeping their clinics clean, but fail to keep their data clean as well. Unclean data can instantly derail a Big Data analytics report, especially when combining different data sources in slightly different formats. Data cleansing is called 'cleansing' or 'scrubbing' to ensure that data sets are accurate, consistent, relevant and not influenced by third parties.
- Data storage
  Doctors and health professionals give little thought to where their data is stored. However, this information is crucial for the IT department as it can pose security and performance risks. With the growing amount of data in healthcare, many providers no longer keep track of their costs. This has a significant impact on data center space.
- Securing patients' data
  Data security is a top priority for healthcare organizations, especially given the risk of a rapid series of high-profile cyberattacks and ransomware cases. The HIPAA Security Rule provides a list of technical safeguards for healthcare organizations to protect healthcare data, including transmission security, authentication protocols and controls for access, integrity and auditing.

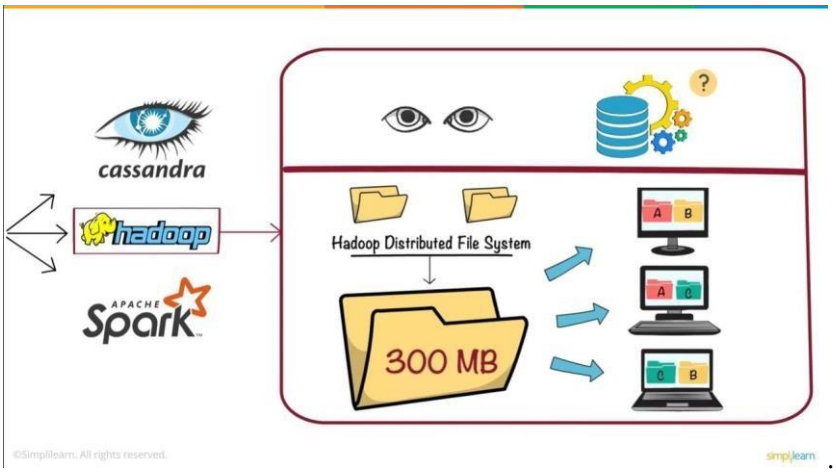
Advantages of Big Data Analytics in Medical Field

- Healthy Patients
  Keeping patients healthy and helping them avoid disease are high on the list of priorities. Controlling patients with high-risk problems and ensuring a more effective, tailored treatment approach can be made easier [2]. A lack of data makes it difficult to create patient-centered care programmers, so it is clear why the use of Big Data initiatives is important in the industry..
- Cost Reduction
  In various clinics, hospitals and medical facilities, there is often a high level of waste due to ineffective management of finances. The cause of losses in internal budgets is usually the under- or overbooking of staff. A proactive analysis can solve this very problem. It is much easier to get help for effective staff allocation together with a forecast of admission rates.
- Error Minimization and Precise
  Treatments Prescription errors are a serious problem in health facilities. Because people always make occasional mistakes, patients sometimes receive the wrong prescription, which can lead to harm or even death. Big Data can help drastically reduce these error rates by analyzing patient records with all prescribed treatments and flagging anything that seems inappropriate.
- Reduce Fraud and Enhance Security
  Personal data is extremely valuable. With this in mind, many companies have started to use analytics to prevent security threats by detecting changes in network traffic or other behaviors that indicate a cyber-attack.

IV. BIG DATA ANALYTICS IN HEART ATTACK PREDICTION

Smart healthcare, a current trend in healthcare that contributes to Big Data sources, has used several products such as home healthcare, wearable healthcare and bio transplant healthcare. For patients who need to be monitored at home, home health systems are sensors installed in the home that work with individual users and their smartphones to help manage the person's health. In wearable healthcare, sensors are worn on the human body and provide real-time personalized services by measuring, transmitting and analyzing the bio signals of the user's body [3]. Sensors provide invaluable real-time data to the provider.

Accurate analysis of Big Data can lead to further confident decision- timber. Apache Hadoop, an influential aspect of Big Data, was developed by Yahoo. It's an open- source software frame written in Java, primarily intended for distributed processing and distributed storehouse of huge quantities of data on computer clusters. Hadoop enables massive data. storehouse and briskly processing. The Hadoop Distributed train System(HDFS) creates multitudinous clones of each data block and distributes them across systems in a cluster for dependable access. HDFS supports pall computing with Hadoop, a distributed computing platform. A distributed column- acquainted database- HBase- is erected on top of HDFS. It can be used when we need arbitrary access to veritably large data sets. HDFS provides dependable and scalable data storehouse. The central core of Apache Hadoop consists of a storehouse part(HDFS) and a processing member(Map Reduce). Apache Mahout runs distributed or scalable machine literacy and data mining algorithms. The proposed system armature is shown in Figure 1



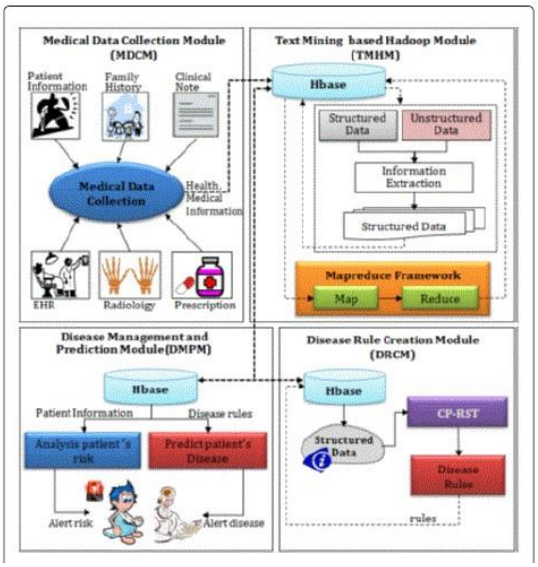
The Lullaby algorithm is separated into two main parts: the 12 second calibration window and the real-time processing which continues until the signal is finished. In the calibration phase, the first 12 seconds of the abdominal ECG are used to determine the features of the fetal peaks to selectively choose candidate peaks in the real-time processing phase. In the real-time processing phase, the abdominal ECG is segmented into 4 second windows and processed to determine the position of fetal peaks throughout the entire window. However, only the fetal peaks positions in the last 1 second of the 4 second window are outputted. The window then shifts forward in time 1 second and the process repeats until the end of the abdominal ECG is reached .The original algorithm was written in MATLAB and then converted to C code using the compiler provided through MathWorks Inc.

To store and process Big Data, we use different frameworks such as Cassandra, Hadoop and Spark. Let's take a look at Hadoop. Hadoop uses a distributed file system known as the Hadoop Distributed File System to store Big Data. In Hadoop's distributed file system, large files are split into smaller files and stored in different systems so that the data is not lost if one part of the system becomes corrupted

V. PATIENTS HEALTH CARE AND BIG DATA

Ciccone al conducted a feasibility study on the integration of care managers (specially trained nurses) into the health system to support general practitioners and specialists in the management of patients with cardiovascular disease, diabetes, heart failure or at risk of CVD. The care managers worked directly with individual patients, supporting them to make lifestyle changes, monitoring their health status and providing the necessary information and advice to encourage patient ownership and improve their selfmanagement skills. This led to a noticeable improvement in clinical care and patients had better control of their disease. Ultimately, the role of the care manager has a positive impact on patient health and self-management, and the results are due to the solid 'partnership' between the care manager and the patient, and the collaboration between the provider and the care manager. Specially trained nurses are at the forefront of the new data revolution in healthcare today. In the data-driven society we live in today, nurses need to support their patients to take more ownership, change their lifestyle and improve their overall health based on the data. Never before in history has this system of personalized medicine been on the cusp of revolutionary evidence-based patient care[7].

Recently, a patient-specific healthcare system based on Hadoop with Text Mining (PHSHT) has been developed as a method to efficiently manage diseases and healthcare. The PHSHT not only fills in the gaps in the existing decision-making algorithm that ignores the relationship between attributes, but also creates precise disease rules. It also communicates the status of the patient to the individual, thus avoiding unexpected accidents as the patient can then act immediately. The Hadoop platform, which is based on text mining, determines an individual's disease, predicts the course of disease and creates more precise information about diseases by converting the patient's unstructured generated data into structured data. PHSHT consists of four modules: (a) MDCM stores big data such as a patient's health information in Hbase, which is done within a hospital or mobile health system [6]. a further, the collected big data is divided into structured data such as patient information, family history, and medical prescriptions, and unstructured data such as clinical notes, EHR, and PACs data; (b) TMHM analyses the collected unstructured data using text miningbased Hadoop and converts it into structured data. TMHM also distributes and stores structured data in Hbase, then merges the stored structured data and stores it back in Hbase with a MapReduce framework; (c) DRCM generates disease rules using the disease information and CPST algorithm stored in Hbase and stores it back in Hbase; (d) DMPM notifies a patient or risk index provider as a result of disease prediction when analyzing patient risk with the patient's collected information or when predicting a disease by combining the DRCM generated disease rules with the collected information. The modules and flowchart of the PHSHT are shown in [8] Figure 2.



Jilani claims that since the team's initial study, the algorithm has already seen significant improvement. It is now both faster and more accurate than the previous version, and the team is trying to integrate it into a whole system.

This includes work on a mobile app that can be used on smart phones to support fetal heart monitoring [5].

VI. CONCLUSION

The analysis of big structured data, unstructured data and unordered data has led to significant insights. In the absence of crossborder alignment and technology integration, standards are needed to enable interoperability between elements of the Big Data value chain. Big Data offers tremendous opportunities to detect interactions and non-linear relationships between variables. Mobile devices such as smartphones and tablets, as well as sensors, will continue to be the most indispensable tools for providing heart attack prediction and telecardiology services via wireless networks to reduce cardiovascular disease morbidity and mortality. The use of cloud computing has facilitated the collaborative application of telecardiology between hospitals in a cost-effective manner and has expanded services from regional to global. However, the most important factor in the development and application of Big Data, telecardiology, sensor use, use of mobile phones or tablets and landlines is the protection of patient privacy and the preservation of the patient's ability to self-determine and ascertain the use of their health data. Care managers, specially trained nurses who are revolutionizing healthcare by empowering patients to change their lifestyle and habits based on research and data, are needed to support patients in this new data-driven healthcare scene. Nurses have always been at the forefront of revolutionary medicine, and in today's data-driven healthcare system, nurses are critical in helping patients help patients navigate the data mines and empower them to change unhealthy habits and achieve better health. Patients to navigate the data landmines and empower them to change unhealthy habits and reach a more improved health status.

VII. REFERENCES

1. What is Big Data? Introduction, Types, Characteristics, Examples (guru99.com)
2. https://www.drcatalyst.com/top-benefits-of-big-data-analytics-in-healthcare-industry
3. The use of Big Data Analytics in healthcare | Journal of Big Data | Full Text (springeropen.com)
4. BIG DATA ANALYTICS IN HEART DISEASES PREDICTION (researchgate.net)
5. https://illinois.touro.edu/news/applications-and-examples-of-big-data-in-healthcare.php
6. https://www.simplilearn.com/big-data-in-healthcare-sector-rar359-article
7. https://www.foreseemed.com/blog/big-data-analytics-in-healthcare