

# Data Mining In Healthcare Industry: A Systematic Review on Recent Studies

Divyanshu Keshari<sup>1</sup>, Prof. Durgesh Pandey<sup>2</sup>

<sup>1</sup>Dept of Computer Science and Engineering

<sup>2</sup>Guide, Assistant Professor, Dept of Computer Science and Engineering

<sup>1,2</sup>Saraswati higher education and technical College of engineering, babatpur Varanasi Uttar Pradesh.

**Abstract-** *The emergence of tools and approaches for using data has increased knowledge discovery in databases. One of the most crucial steps in data mining. Data mining is the process of discovering and extracting patterns from massive amounts of data. From clinical and diagnostic data, both the data mining and healthcare industries have developed some of the most accurate early detection systems and other healthcare-related tools. In relation to this emerging issue, we examined the many papers in this subject in terms of technique, algorithms, and findings. The works evaluated in this review paper have been consolidated in accordance with the disciplines, model, tasks, and methodologies. For chosen studies, the results and assessment techniques are examined, and a summary of the findings is offered to finish the study.*

**Keywords-** Data Mining Data Mining in Healthcare Health Informatics

## I. INTRODUCTION

It is incredibly expensive to keep data or information. However, thanks to advancements in information collecting methods and the WWW during the previous twenty-five years, a vast quantity of information or data is now accessible in electronic format. Database sizes are continually increasing in order to hold such a big quantity of data or information. Such databases include a wealth of important information. This knowledge might be extremely beneficial in any decision-making process. It is made feasible by the use of data mining or knowledge discovery in databases (KDD). Data mining is the process of collecting usable information from a big collection of previously unknown data. A variety of correlations are concealed amid such a big collection of data, such as one between patient data and the number of days spent in the hospital.

### 1.1 Five Stages Are Identified In Data Mining In Healthcare Industry

- **Selection:** In this step, the data is chosen based on a set of criteria. For example, if all of those folks

possess a bicycle, we may define subsets of data in this manner.

- **Preprocessing:** This stage eliminates information that is unnecessary, such as noting a patient's gender while performing a pregnancy test. It is often referred to as the data purification step.
- **Transformation :** This step altered just the data that is valuable in a certain study, such as data connected to a specific demographic in market research.
- **Data mining :** Data mining is a step in the process of discovering knowledge. This stage is important for identifying relevant patterns in data.
- **Interpretation and evaluation :** In this level, the system's detected relevant patterns are interpreted into knowledge. This information may then be put to use in making sound judgments.

### 1.2 The Importance and Uses of Data Mining in Medicine and Public Health

Despite variations and disputes in methods, the health industry now has a greater need for data mining. There are various reasons that might be presented to promote the use of data mining in the health sector, spanning both public and private health problems (which, in fact, as can be shown later, are also stakeholders in public health). Overload of information. Computerized health records may provide a plethora of information. However, the massive amount of data kept in these databases makes it incredibly difficult, if not impossible, for people to filter through and uncover information. Indeed, some experts feel that medical advancements have halted, blaming the prohibitive quantity and complexity of today's medical knowledge. Computers and data mining are ideal for this task. (Evidence-based medicine and hospital error avoidance.) When medical organisations utilise data mining on their current data, they may find fresh, valuable, and even life-saving information that would otherwise be dormant in their databases. For example, an ongoing hospital safety research discovered that around 87 percent of hospital fatalities in the United States might have been avoided if hospital workers (including physicians) had been more vigilant in avoiding mistakes (Health Grades

Hospitals Study 2007). Such safety hazards might be identified and remedied by hospital management and government authorities by mining hospital information. Policy formulation in public health. Larva et al. (2007) analysed commonalities between community health centres in Slovenia using GIS and data mining technologies such as Weka and J48 (free, open source, Java-based data mining tools). They were able to find commonalities across health facilities using data mining, which led to policy suggestions to their Institute of Public Health. They came to the conclusion that "data mining and decision support tools, including innovative visualisation methods, may lead to improved decision-making performance."



**Fig. 1 Healthcare Data Mining**

### 1.3 DATA MINING CHALLENGES IN HEALTHCARE

As we all know, numerous healthcare institutions create and maintain a large amount of healthcare data. However, there are a number of issues associated with healthcare data that may provide severe impediments to making sound judgments. The first problem with healthcare data is that the format in which it is kept varies across healthcare institutions. To far, no standard format for storing data has been established. In epidemic settings, the absence of a consistent format might exacerbate the crisis. Assume that an epidemic sickness spreads over a country's many geographical areas.

### 1.4 HEALTH CARE INDUSTRY

Companies in the Healthcare Industry provide clinical services, produce medications and medical equipment, and provide healthcare-related support services such as medical insurance. It is also known as the medical industry. These businesses are critical in the diagnosis, treatment, nursing, and management of sickness, disease, and accident. Patients are also provided with preventative, corrective, and therapeutic services by the healthcare business. To deliver these services, healthcare providers such as physicians, nurses,

medical administrators, government agencies, pharmaceutical businesses, medical equipment manufacturers, and medical insurance companies must work together.

## II. LITERATURE REVIEW

### 2.1 Anand Sharma et.al “Emerging Applications of Data Mining for Healthcare Management - A Critical Review” (2014)

In this study, we give a critical evaluation of the current research in data mining applications for healthcare management. The goal of this research is to look at new and developing areas of data mining methods utilised in healthcare management. Infection control surveillance, disease diagnosis and treatment, healthcare resource management, customer relationship management, fraud and anomaly detection, healthcare administration, hospital management, and public health are among the applications covered in this article. This study examines the data mining objectives attained, functions performed, and techniques employed in various applications.

### 2.2 Srinivas et.al “Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks” (2008)

The healthcare environment is often regarded as 'information rich' yet 'knowledge poor.' Within healthcare systems, there is a lot of data. However, there is a scarcity of good data analysis tools for uncovering underlying correlations and patterns. Knowledge discovery and data mining have a wide range of applications in business and science. The implementation of data mining methods in the healthcare system may provide valuable insights. In this paper, we look at how classification-based data mining approaches including rule-based data mining, decision trees, Naive Bayes, and artificial neural networks may be used to large amounts of healthcare data. The healthcare business gathers massive volumes of data, which are regrettably not "mined" to uncover hidden information. One Dependency Augmented Naive Bayes classifier (ODANB) and naïve credal classifier 2 (NCC2) are utilised for data preparation and effective decision making. This is an extension of naive Bayes to imprecise probabilities that tries to provide robust classifications even with tiny or partial data sets. The discovery of hidden patterns and correlations is often underutilised. It can forecast the chance of individuals developing heart disease based on medical characteristics such as age, gender, blood pressure, and blood sugar. It allows for the establishment of substantial information, such as patterns and correlations between medical aspects associated to heart disease.

### 2.3 Diva Tamar et.al “A survey on Data Mining approaches for Healthcare” (2013)

Data mining is one of the most enthralling areas of study that is gaining traction in health care organisations. Data mining is vital for identifying new trends in healthcare organisations, which is beneficial to all parties involved in this industry. This review investigates the value of several Data Mining methods in the health area, such as classification, clustering, association, and regression. In this work, we provide a quick overview of various strategies, as well as their benefits and drawbacks. This report also emphasises Data Mining uses, problems, and potential concerns in healthcare. This study also discusses recommendations for selecting the best accessible Data Mining approach.

### 2.4 Nisha Joti et.al “The Third Information Systems International Conference Data Mining in Healthcare – A Review” (2015)

Knowledge discovery in databases (KDD) is concerned with the creation of methodologies and procedures for using data. Data mining is a critical phase in the KDD process. Data mining is the process of discovering and extracting patterns from massive amounts of data. From clinical and diagnostic data, both the data mining and healthcare industries have developed some of the most accurate early detection systems and other healthcare-related tools. In relation to this emerging issue, we examined the many papers in this subject in terms of technique, algorithms, and findings. The works evaluated in this review paper have been consolidated in accordance with the disciplines, model, tasks, and methodologies. For chosen studies, the results and assessment techniques are examined, and a summary of the findings is offered to finish the study.

### 2.5 Umar Shaniqua et.al “Data Mining in Healthcare for Heart Diseases” (2015)

Data Mining is a field of study that involves the extraction of relevant information or knowledge from prior data. For data mining, several approaches are used. Data mining might be used in a variety of industries, including healthcare. Heart or cardiovascular illnesses are a major concern in the worldwide healthcare business. Many patients perished as a result of a lack of information. Because the healthcare business generates a large quantity of data, we may utilise data mining to uncover hidden patterns and fascinating information that can aid in effective and efficient decision making. Data mining in healthcare is a critical and complex process that must be completed correctly. It aims to tackle real-world health concerns in illness detection and treatment.

This endeavour is also an attempt to discover intriguing patterns in cardiac patient data. Three algorithms are utilised in two separate contexts. Decision Tree, Neural Network, and Nave Bayes are the algorithms that have been implemented.

### 2.6 Joti Sony et.al “Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction” (2011)

The successful use of data mining in high-profile disciplines such as e-commerce, marketing, and retail has led to its use in other businesses and sectors. Healthcare is one of these areas that is still in its infancy. The healthcare environment continues to be "information rich" yet "knowledge poor." Within healthcare systems, there is a lot of data. However, there is a scarcity of good data analysis tools for uncovering underlying correlations and patterns. This study work aims to present an overview of current strategies for knowledge discovery in databases employing data mining techniques that are used in today's medical research, specifically in Heart Disease Prediction. A number of experiments have been carried out to compare the performance of predictive data mining techniques on the same dataset, and the results show that Decision Tree outperforms, and sometimes Bayesian classification has similar accuracy as Decision Tree, but other predictive methods such as KNN, Neural Networks, and Classification based on clustering do not perform well. The second result is that the accuracy of the Decision Tree and Bayesian Classification increases even more after using a genetic algorithm to minimise the real data quantity in order to get the ideal subset of attributes required for heart disease prediction.

## III. AIM

Is it true that the goal of this initiative is to develop data mining and the healthcare industry? Several researchers have attempted to address this issue in a number of ways in order to identify which solution works and delivers best results.

## IV. RESEARCH GAP

Essentially, a tool for identifying and defining existing health-care challenges. A gap analysis highlights the disparity between existing health-care reality and desired or optimum health-care scenario and provides an opportunity that may be addressed in a CPD intervention.

## V. OBJECTIVE OF THE STUDY

1. To list current applications and emphasise the significance of data mining in medical and public policy.
2. Health, in order to identify data mining methods utilised in other disciplines that may be implemented in the health sector.
3. Identifying concerns and obstacles in data mining as they apply to medical practise.
4. Outline some guidelines for using data mining to uncover information in electronic databases.

## VI. SCOPE OF THE STUDY

Increasing computer-based data analysis understanding, online educational availability, and building an integrated learning method among medical practitioners would undoubtedly aid in correct diagnosis and an efficient treatment management plan in India. Innovative medical technologies are critical for patient care. This is also true for the prevention of numerous illnesses connected to cleanliness, communicable diseases, and addiction-associated disorders such as lung cancer, mouth cancer, liver cirrhosis, and so on. In the future, the extent of technological applications such as data mining techniques-based systems in India's healthcare system would really create tremendous changes at every level. Today, the internet serves as a portal to global information as well as a vast platform for national media and documentation, which will be very beneficial in the future deployment of data mining methods.

## VII. CONCLUSION

Future study goals include evaluating selected algorithms against various medical datasets. The trials would be carried out on a broader spectrum of medical information, making the assessment even more exact. It is a good idea to include different algorithms in the trials and compare their performance in the medical area. This would result in a new ranking and aid in the creation of Medical Decision Support Systems via the selection of the most appropriate algorithms. We may also compare other approaches that are not included in this survey to determine the best one by weighing the benefits and drawbacks of the present one.

Various data mining approaches must be utilised in conjunction to obtain greater accuracy in illness prediction, improve survival rate in critical death-related situations, and so on. To attain higher-quality medical data, all required actions must be made to establish better medical information systems that give accurate information about patients' medical histories rather than information about their billing bills.

Because high-quality healthcare data is helpful not only for delivering better medical services to patients, but also for healthcare organisations and other organisations participating in the healthcare business. Takes all necessary efforts to reduce the semantic gap in data exchange across remote healthcare database environments in order to find meaningful patterns. These patterns may be highly valuable in improving treatment efficacy services, detecting fraud and abuse, and improving customer relationship management all around the globe.

## REFERENCES

- [1] Sharma, A. (2014). Emerging Applications of Data Mining for Healthcare Management - A Critical Review. 377–382.
- [2] Govrdhan, A. (2010). Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks. October 2014.
- [3] Durairaj, M., & Ranjani, V. (n.d.). Data Mining Applications In Healthcare Sector: A Study. <http://www.mfta.ir>
- [4] ALI, S. I. M., & BUTI, R. H. (2021). Data Mining in Healthcare Sector. MINAR International Journal of Applied Sciences and Technology, 03(02), 87–91. <https://doi.org/10.47832/2717-8234.2-3.11>
- [5] De Vos, A., & Soens, N. (2008). The power of career counseling for enhanced talent and knowledge management. Smart Talent Management: Building Knowledge Assets for Competitive Advantage, 17(8), 119–138. <https://doi.org/10.4337/9781848442986.00014>
- [6] Tomar, D., & Agarwal, S. (2013). A survey on data mining approaches for healthcare. International Journal of Bio-Science and Bio-Technology, 5(5), 241–266. <https://doi.org/10.14257/ijbsbt.2013.5.5.25>
- [7] Jothi, N., Aini, N., Rashid, A., & Husain, W. (2015). Data Mining in Healthcare – A Review. Procedia - Procedia Computer Science, 72, 306–313. <https://doi.org/10.1016/j.procs.2015.12.145>
- [8] Canlas Jr, R. D. (2009). DATA MINING IN HEALTHCARE: Data Mining in Healthcare: Current Applications and Issues. Unpublished Master Thesis, August, 1–10.
- [9] Tayade, M. (2014). Role of Data Mining Techniques in Healthcare sector in India Scholars Journal of Applied Medical Sciences ( SJAMS ) Review Article Role of Data Mining Techniques in Healthcare sector in India. June 2013, 1–4.
- [10] De Vos, A., & Soens, N. (2008). The power of career counseling for enhanced talent and knowledge management. Smart Talent Management: Building

Knowledge Assets for Competitive Advantage, 17(8),  
119–138. <https://doi.org/10.4337/9781848442986.00014>