

# Chapter 2

## A Comprehensive Survey of Data Mining Techniques in Disease Prediction

**Durgadevi Mullaivanan**

*Pondicherry Engineering College, India*

**Kalpana R.**

*Pondicherry Engineering College, India*

### **ABSTRACT**

*In recent days, data mining has become very popular, and numerous research works have been carried out of using data mining techniques in the healthcare sector. The healthcare transactions generate a massive amount of data which are very voluminous and complex to be processed. Therefore, data mining techniques have been employed, which provides a practical methodology for transforming the massive amount of data into efficient knowledge for the process of decision making. Prediction and classification are the two forms of data analysis methods. However, it is still difficult to explore the complete literature in the healthcare domain. This chapter reviews the research overview that is done in the healthcare sector utilizing different data mining methodologies for prediction and classification of diverse diseases. Also, a detailed comparison of reviewed methods takes place for better understanding of the existing models. An extensive experimental study is also performed to analyze the performance of data mining algorithms.*

### **1. INTRODUCTION**

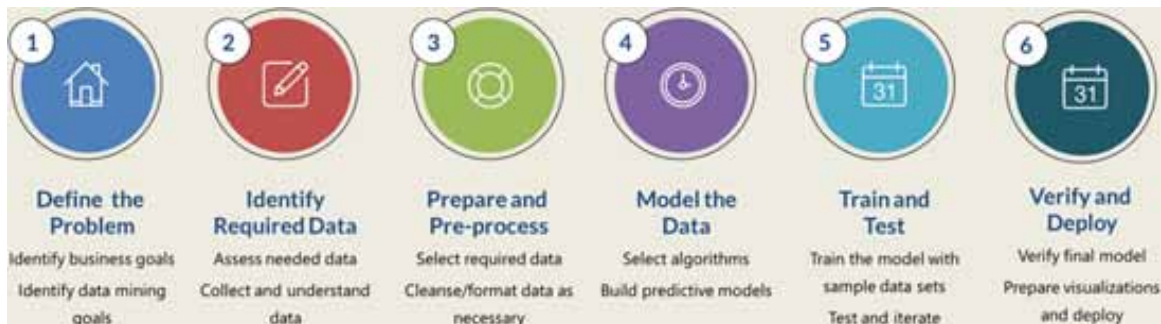
Data mining is a procedure of discovering knowledge from massive databases to uncover trends and patterns exist in data. The vast amount of data present in the information industry has to be converted to extract useful information from it (Neesha Jothi, et.al, 2015). Data mining also perform several other processes like cleaning, integration, transformation, mining, evaluation, and presentation. The different stages included in data mining are demonstrated in Fig. 1. There are two categories exist in the data mining patterns such as Descriptive, Classification and Prediction (G. E. Vlahos, et.al, 2004). The descriptive

DOI: 10.4018/978-1-7998-2566-1.ch002

function deals with the general properties of data whereas the classification and prediction deal with the concepts of data. The descriptive function involves summarization and mapping of data which is commonly known as data characterization and data discrimination. Predictive analytics is a combination of historical data, machine learning, and artificial intelligence. Predictive analytics helps to analyze the state of the current data to determine the future outcome. It becomes much more prevalent in areas like finance, marketing, healthcare, social networking, and other areas. The implementation of predictive analytics is a complex process as it comes with many challenges. The ultimate aim of predictive analytics in the digital world is to get better revenue and profit at reduced cost and risk. The problem of every prediction must come with a “remedy.” Predictive analytics can be applied to any type of mysterious data in the past, present or future (Chandamona and Ponperisamy, 2016).

In the healthcare sector, data mining becomes more familiar. Numerous factors have been inspired by the usage of data mining in healthcare (Salim A. Dewani and Zaipuna O. Yonah, 2017). The subsistence of medical insurance abuse and fraud makes numerous healthcare insurers to utilize data mining approaches to decrease their losses by identifying and tracking offenders. Fraud identification utilizes the applications of data mining in the profitable globe, for instance, the identification of false transactions in credit card. The massive sum of data generated through the healthcare sector is very complicated and huge to be processed and investigated through the conventional approaches.

*Figure 1. Data mining stages*



Data mining enhances the procedure of decision-making through exploring patterns and tendency in a massive quantity of composite data. The investigation has become very important since economical pressure has enhanced the requirement of healthcare sectors to create decisions using the investigation of medical as well as financial data. The healthcare organizations which make use of data mining techniques that are highly positioned to meet its long-term requirements (I. Witten, et.al, 2011). Number of applications related to healthcare is existed by the use of data mining techniques. In general, they can be integrated as the examination of treatment efficiency, healthcare management, management of customer relationship, and discovery of abuse and fraud. Applications of data mining are employed for validating the efficiency of medical treatments. Using the symptoms, causes, and treatments courses, data mining techniques offer an investigation of the patient status. Numerous dedicated medical data mining like predictive analysis and medicine of DNA micro-arrays are also developed. The primary use of predictive analytics in medical decision aiding systems is to predict the percentage of risk of developing certain diseases like diabetes, heart disease or other complications.

25 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/a-comprehensive-survey-of-data-mining-techniques-in-disease-prediction/267238](http://www.igi-global.com/chapter/a-comprehensive-survey-of-data-mining-techniques-in-disease-prediction/267238)

## Related Content

---

### E-Healthcare System for Disease Detection Based on Medical Image Classification Using CNN

Himansu Das, Mahendra Kumar Gourisaria, Badal Kumar Sah, Saurabh Bilgaiyan, J Chandrakanta Badajenaand Radha Mohan Pattanayak (2022). *Empirical Research for Futuristic E-Commerce Systems: Foundations and Applications* (pp. 213-230).

[www.irma-international.org/chapter/e-healthcare-system-for-disease-detection-based-on-medical-image-classification-using-cnn/309676](http://www.irma-international.org/chapter/e-healthcare-system-for-disease-detection-based-on-medical-image-classification-using-cnn/309676)

### A Review on Time Series Motif Discovery Techniques an Application to ECG Signal Classification: ECG Signal Classification Using Time Series Motif Discovery Techniques

Ramanujam Elangovanand Padmavathi S. (2019). *International Journal of Artificial Intelligence and Machine Learning* (pp. 39-56).

[www.irma-international.org/article/a-review-on-time-series-motif-discovery-techniques-an-application-to-ecg-signal-classification/238127](http://www.irma-international.org/article/a-review-on-time-series-motif-discovery-techniques-an-application-to-ecg-signal-classification/238127)

### Integrated Regression Approach for Prediction of Solar Irradiance Based on Multiple Weather Factors

Megha Kambleand Sudeshna Ghosh (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-12).

[www.irma-international.org/article/integrated-regression-approach-for-prediction-of-solar-irradiance-based-on-multiple-weather-factors/294105](http://www.irma-international.org/article/integrated-regression-approach-for-prediction-of-solar-irradiance-based-on-multiple-weather-factors/294105)

### Data Science for Industry 4.0

Indraneel Dabhade (2023). *Encyclopedia of Data Science and Machine Learning* (pp. 58-71).

[www.irma-international.org/chapter/data-science-for-industry-40/317438](http://www.irma-international.org/chapter/data-science-for-industry-40/317438)

### Autoencoder Based Anomaly Detection for SCADA Networks

Sajid Nazir, Shushma Patel and Dilip Patel (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 83-99).

[www.irma-international.org/article/autoencoder-based-anomaly-detection-for-scada-networks/277436](http://www.irma-international.org/article/autoencoder-based-anomaly-detection-for-scada-networks/277436)