Data Mining Approaches Used in Healthcare – A Review

M. Sivasubramaniyan

Department of Computer Science, Government Arts and Science College, Kangeyam, Tamil Nadu, India.
sivasu31@gmail.com

N. Karthikeyani Visalakshi

Department of Computer Science, Government Arts and Science College, Kangeyam, Tamil Nadu, India.
karthichitru@gmail.com

Abstract – According to a report from the World Health Organization (WHO), in recent years, Non-Communicable Diseases (NCDs) have identified as the leading cause of death worldwide. Because these NCDs have considered highly controllable and preventable, the WHO has announced that the rapid and the early prognosis of these diseases will help to patients prevent from death and reduce treatment cost. Now, data mining techniques have the capability to handle large volumes of data in all sectors, including the healthcare sector and identify the problem across sectors and provide the optimal solution with the best accuracy. The main aim of this article is to present an overview of the current researches titled "Data Mining approaches used in Healthcare" and discussing the algorithms and techniques of data mining in the early detection of diseases. This paper presents a review on the use of data mining strategies for disease prognosis. An efficient review of the literature has conducted in the academic and scientific databases that taking into account the date of publication of the articles titled data mining techniques used in healthcare from 2015 to 2019. To achieve the review process, all selected articles have categorized by publication year, the objectives of the research, tools used to build a model, the type of diagnosis, the parameters used to improve the diagnosis, and the dataset used. Therefore, this article presents an overview of the literature on data mining used in the healthcare industry and the appropriate ideas for conducting research in early diagnosis using data mining strategies in the future.

Index Terms – Data Mining, Diagnosis, Healthcare, Non-Communicable Diseases (NCDs), Prediction.

1. INTRODUCTION

The World Health Organization (WHO) [71] had published a report comparing and analyzing the data related to top 10 diseases caused deaths from 2000 to 2016 in the world. In this report, the WHO stated that the rate of death caused by NCDs is higher than all other diseases. Figure.1 has been created by WHO which depicts the Top 10 diseases caused death. Citing the report of WHO, Hannah Ritchie et al. [70] have categorized the causes of death worldwide into three main categories. They are,

1. Death caused by accidents can be categorized as accidents on the road, murder caused by another human, affected by psychological disorder, conflict deaths, death caused by water, accident caused by fire, natural disasters and suicides.

2. Death caused by Non-Communicable Diseases (NCDs) like heart related diseases, stroke, different type of cancers, diabetes and breathing diseases.

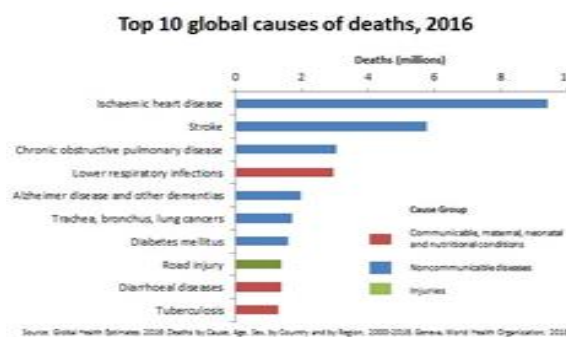3. Death caused by Communicable Diseases (CDs) like HIV/AIDS, malaria and tuberculosis.
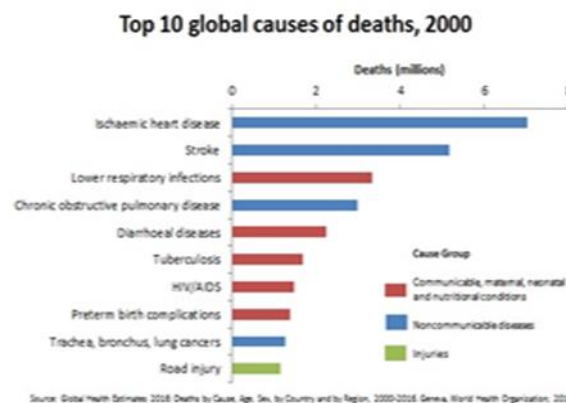

Figure.1 World Health Organization's Report from 2000 to 2016

Hannah Ritchie et al. [70] has generated a report by using the data collected from various sources like Institute of Health Metrics and Evaluation (IMHE), Global Burden of Disease (GBD), World Health Organization (WHO), Global Health Observatory (GHO), Global Terrorism Database (GTD), and Amnesty International. The Figure.2 illustrates the report of disease caused death. They are

- At each interval, the mortality rate caused by NCDs is very high compared to other causes, and the NCDs mortality rate increases rapidly.

- The mortality rate of CDs have controlled by some sources, such as awareness and newly introduced drugs among people, but the mortality rate of NCDs are not under the control.

Now, the main challenges for the healthcare industry is to control the growth of NCDs, save people from these types of diseases. Since the symptoms of NCDs [21 – 46] appears only in the human body at their last stage, no one medical procedure helps to control the severity of the disease and save the patient from death.
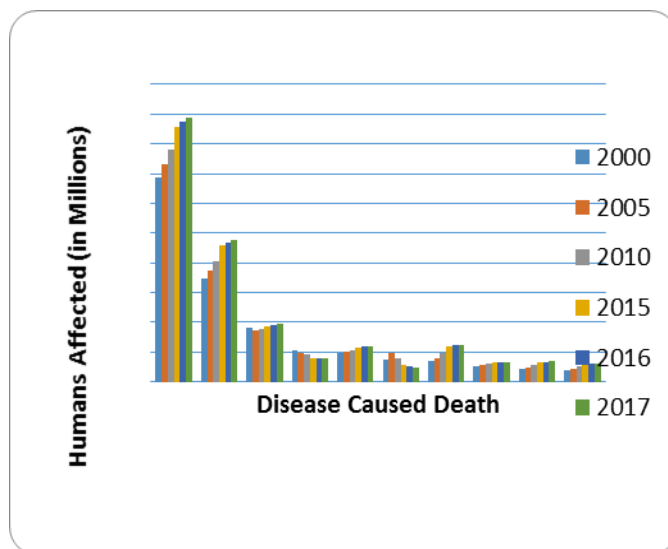

Figure. 2 Disease caused Death Rate (in Millions)

2. DATA MINING

The main solution for controlling the diseases is early diagnosis with high accuracy. Due to the high cost of diagnosis, early detection of the disease with high accuracy is impossible for all people. The early detection of non-communicable diseases [21-32] help in taking precautionary arrangements and effective treatment at an initial stage has always been found to be helpful for patients. Early detection of disease is fast and easy when the dataset containing data is exact, reliable and without noise. One of the data preprocessing techniques in data mining is feature selection [49] has the capability of identifying the most important risk factors in a dataset and also eliminating useless and unwanted attributes from the dataset. While applying the feature selection technique in the healthcare sector for early diagnosis, it provides quick and better results. Various data mining strategies [47] have been applied to disease datasets for the early diagnosis of non-communicable diseases and their results have been found to be useful in the prevention of diseases. Therefore, there is a major need to develop a new classification system to help diagnose NCDs faster and easier compare with traditional systems. In this digital era, large amounts of medical data are generated and constantly updated. They are called Electronic Health Record (HER). The disease diagnosis also depends upon the EHR [76, 84, 87] which is a collection of medical reports of patients, doctor's prescription, analytical test reports, and patient's health insurance. Currently, an effective data mining model capable of managing and analyzing large volume of medical data is the primary need in healthcare. This paper analyzed the recent researches held in healthcare industry related to disease prediction and offers ideas for developing a new data mining model for early detection of non-communicable diseases with high accuracy and fast. In this paper, the first part covers the introduction and the rest of the paper is organized in the following order. Section 2 presents the basic concepts of data mining and summary of data mining used in various fields. Section 3 addresses the methodology of our literature reviews. Section 4 presents a summary of data mining techniques used in the healthcare industry. Section 5 outlines of the discussion and Section 6 refers to the conclusion of this review.

All kinds of data such as images, videos and tables are stored separately in a large database known as data warehouse. Data Mining (DM) is the process of getting valuable set of information from the huge amount of raw data (like raw data present in data warehouse). It is a repeated process to validate the vast amount of data collected from various sources on a particular interval of time. The ultimate objective of data mining is Prediction and Description. The prediction used to forecast the suggestion related to present data and the description used to encapsulate the common properties of the data in the database. DM involves well-organized techniques known as algorithms or techniques used to manage enormous volume of data and it used to extract relevant information from the database.

2.1. Data Mining Process

The important step of Knowledge Discovery from the Database (KDD) is data mining. The following steps are involved in the data mining process (see Figure. 3)

Defining the problem: One of the most important steps used to identify the business goal and translate into specific data mining project goals.

Collect data: In this step, identifying the data for the particular business to achieve the goal and collected from the appropriate source.
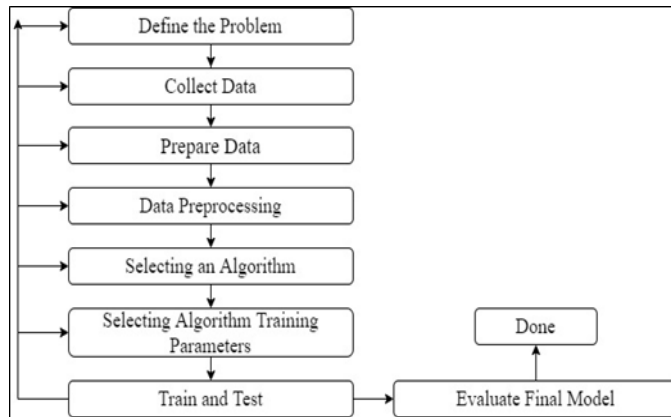
Figure .3   Data Mining Process

Preparing and pre-processing: prepare the final data in appropriate format for the business goal by using the data cleaning and formatting process.

Selecting an Algorithm: Select the suitable algorithms and required attributes for the business goal with the prior knowledge. Create a model and find the result.

Train and Test: Examine the built model in more number of iterations with different dataset and validate the result by using various suitable tools, tabulate the result.

Evaluate Final Model: One of the model produce the best optimal solution suitable for the business goal. Choose the best model with high optimum accuracy after the comparison of tabulated result produced the different models.

2.2.   Data Mining Algorithms Used in Various Disciplines

At present, all emerging sectors require a new systematic model to improve the goals of the business. DM strategies play an important role in the creation of a systematic model for financial services, telecommunications, natural sciences, healthcare, and education. DM strategies provide the best optimal solution for all business problems with great accuracy. Table.1 summarizes the various research findings related to data mining techniques used in different fields in the past.

Algorithm used	Author(s)	Objectives
ID3	B.V.Kiran mayee et al. [1]	Proposed the new model used to classify the medical images
C4.5	Felix Tamin et al. [2]	To regulate hospital readmission rate of diabetes patient in best accuracy.
Classification and Regression Trees (CART)	Shruti Dandge et al. [3]	To identify the required components are used to improve the outcome of Ultrasonic Machining process.

Chi-square Automatic Interaction Detection (CHAID)	Flora M Díaz-Pérez et al. [4]	To improve the quality of tourism market related information like cost reduction, customers satisfaction
Multivariate Adaptive Regression Splines	Xinglong Jua et al.[5]	Proposed to design the best wind farm layout with removal of redundant constraints in the traditional layouts with minimum cost.
Naive Bayes	Haviluddin et al.[6]	The model is used to evaluate student's academic performance and reduce the dropout.
Gaussian Naive Bayes	Hajer Kamel et al. [7]	The model is used to early detection all types of cancers with best accuracy
Multinomial Naive Bayes	Surya Praha PM et al. [8]	Proposed the model is used to classify the social media data for make decision related to that field with best accuracy.
Bayesian Belief Network	Golam Kabir et al. [9]	This model provides information related to natural hazard and create knowledge for Natural hazard triggering a technological disasters risk assessment.
Bayesian Network (BN)	Iraj Mohammadfam et al.[10]	This model provides the best prediction strategies for avoid risk factors in the workplace to employees
Feedforward Neural Network	Yongping Pan et al.[11]	The FNN based model used to simplify the production and investigation of the Neural-Network Learning Control (NNLC) while performing quite similar to the conventional NNLC.
Modular Neural Network	Manish Sharma et al. [12]	The Modular Feed forward Neural Network based model used to detect diabetic retinopathy by classifying retinal images as normal or abnormal with various parameters
Radial Basis Function Neural Network	Vandana Agarwal et al. [13]	The combination of Firefly algorithm and RBFNN used to improve the performance of traditional methods and efficiently handling variations in face images in real time.
Recurrent Neural Network	Liang Guoa et al. [14]	The Recurrent Neural Network with Health Indicator (RNN-HI) based model used to improve the Remaining Useful Life (RUL)

		prediction accuracy of the machineries.
Granular Support Vector Machine	A. Sowah et al. [15]	GSVM based model used to detect the variances and classification of health insurance claims into authentic and fraudulent claims
Proximal Support Vector Machine	Lerong Ma et al. [16]	PSVM with two-layer heuristic sampling algorithm used to classify and select preference-pair data from the data source produced the best output
Twin Support Vector Machine	Shifei Ding [17]	Proposed TWSVM has efficient problem solving capability, best accuracy and deal large data set after the comparative analysis performed on all SVM classifiers.
Multilevel association rule Mining	Golnar Assadat Afzali et.al[18]	Proposed model, can help to speed up the data mining process in big data mining
Multidimensional association rule Mining	Bhavsar. A. R et al. [19]	The model based on Multidimensional Association Rule used to diagnosis the cattle diseases with best accuracy.
Quantitative association rule	Yanjiao Li et al.[20]	The Quantitative Association Rule based Apriori algorithm model used to reduce the redundancy in data, time consumption of data and improve the efficiency of Blast furnace (BF) iron making process.

Table.1 The Various Research Findings Related to Data Mining Techniques

2.3. Data Mining Approaches Used in Healthcare

Data mining plays a precious role in the medical domain. DM techniques are efficient for identifying and predicting various diseases. By the use of DM techniques, several diseases namely, Breast Cancer, Cardiovascular Disease, Kidney Disorders, Heart Disease, Thyroid Diseases etc. can be predicted. This review has analyzed the performance of the classification algorithms for various disease predictions in the past research.

In the traditional methods of the healthcare sector, doctors have recorded patient data in paper-based forms and the accuracy were very low as data were analyzed using those forms. Traditional methods of diagnosing the disease are very complex, because the diagnosis depends on the accuracy of the patient's data and the processing speed is very low. Now, the

Healthcare sector has been maintaining Electronic Healthcare Records (EHR) [76, 84, 87], which is keeping the details about patient used for the diagnosis. Maintaining the unstoppable growth of EHR is challenging. But data mining has tremendous potential to maintain an EHR more efficiently and accurately. Now the data mining techniques used to maintain the EHR with high efficiency and diagnosis the disease early and this is where DM has proven to be extremely effective.

DM techniques are used in all dimensions of healthcare sector, such as predictive medicine, client relationship management, fraud and abuse detection [15], healthcare and measuring the effectiveness of certain treatments. The data mining have used to identify valuable and clear patterns by analyzing large sets of data. These data patterns help to predict the disease in earlier stage. Data mining techniques are used to improve the efficiency of diagnostics at lower cost, improve the life span of patients, and improve the characteristics of data processing used to save patient's life. Therefore, the healthcare industry is now getting the optimal solution to a large number of problems with the help of data mining service providers such as medical, financial and operational environments.

2.4. Usage of Data Mining Strategies in Various Healthcare Activities

The data mining strategies are used in various business and scientific domain for finding the different applications. Now, DM in the healthcare industry is mainly used to predict various diseases, assist in diagnosis, and to advise clinicians are making medical decisions. Because the efficiency of data processing is so great, it can provide services in anomaly-based findings, decision support, more informed decisions, probability measures and predictive modelling except the normal usage in the healthcare field. The following are the normal usage of data mining in the healthcare sector,

- Finding unnecessary usage of high cost diagnosis tests like X-rays, CT scan, MRI scan Radiographic tests, Ultrasound tests, Nuclear Medicine scans and the usage of emergency department.

- Analysis of patient flow at the medical center helps to reduce patients waiting time at the medical center.

- Investigation of drug related data from specialist's prescription of patients used to the prevention.

- Finding the ratio of diabetes patients from the given population used to analyse the growth of diabetes.

- Assessment of clinical outcomes is the important factor used to improve the quality of clinical practice. The patterns of clinical outcome data used to diagnosis and prediction of diseases. DM and Machine learning are used to identify the patterns of complex data sets related to healthcare.

- The NCDs and Thyroid related diseases (55-69) are diagnosed in early stage with high accuracy using data mining techniques.

3. METHODOLOGY

We have reviewed published research works titled by data mining used in the healthcare sector from 2015 to 2019. Scientific databases such as Google Scholar, IEEE, Xplore, PubMed, Science Direct, Scopus, Academia and Web of Science have been used to conduct the review. The databases includes an information about engineering, medicine and a variety of disciplines. They allow to find and access articles in academic, scientific journals or databases, documentations, and other scientific collections. Some search measures have established such as 'Data Mining used Healthcare' ,'how Data Mining used in Healthcare', 'Healthcare and Data Mining', 'Data Mining and their Applications', 'Thyroid diagnosis using Data Mining', 'Disease Diagnosis using Data Mining techniques', 'Data Mining + Healthcare' , 'Healthcare Prediction by using Data Mining techniques' for selecting the papers of greatest interest. The selection process (see Figure. 4) of the papers has carried out by reading the title abstracts, conclusion, observed result and year of publication. After reading the papers, were collected the best papers related to their abstract, view of analysis, publication along with the full article when essential. The following choice measures were used to classify the papers. They are (a) review in Data mining techniques applied to Healthcare. (b) Studies of Data mining techniques applied to Healthcare. (c) Studies in data mining algorithms used in healthcare. (d) Disease Diagnosis using Data Mining Algorithms. (e) Prediction and Diagnosis of diseases using various data mining techniques. We founded 99 articles are duplicated or with an irrelevant title for this research out of 302 publications. The remaining 203 articles were read and studied their abstracts to see were of interest, obtaining as a result, 91 documents which gave valuable information related to data mining used in the healthcare field. The findings of the searches are summarized and given in the following sections.
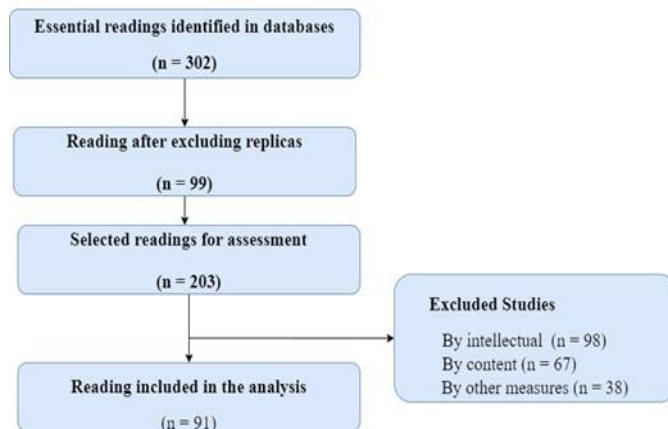

Figure .4 Review Process

4. RELATED WORK

Computerized clinical analysis helps clinicians to accurately predict disease in less time. It helps to patients to reduce the cost of diagnosis and protect the patient's life from NCDs such as cancer, liver and heart disease, or to increase their lifespan. The data mining models used to provide the disease diagnosis in the early stage with high accuracy. Table.2 has described the foremost objectives of the authors working in the medical field of predicting diseases using data mining methodology.

Author(s)	Year	Disease Considered	Objectives
V. Krishnaiah et al. [21], Sujata Joshi et al. [23]	2015	Heart Disease	Proposed a K-Nearest Neighbors algorithm based model used to diagnosis of Heart disease in earlier with high accuracy.
Hamed Majidi Zolbanin et al. [22]	2015	Comorbidity of Cancers	Prediction and diagnosis of cancer in early stage with better accuracy and minimum % of errors
Dheeraj Raju et al. [24]	2015	Pressure Ulcers	The Random Forest Based model have used to provides Prediction and diagnosis of Ulcer in the early stage with high accuracy
M.A.Jabbar et al. [25]	2015	Heart Disease	To designed one R + Naïve Bayes approach based predictive model is used to detect the heart disease with best accuracy.
Sajida Perveen et al. [27]	2016	Diabetes	The detection of disease in the early stage with best accuracy is possible only the Adaboost Ensemble method based new model
Moloud Abdar et al. [29]	2016	Liver Disease	Concluded the C4.5 algorithm performs more

			efficient for diagnosis of Liver disease in the early stage with high efficiency compared to other algorithms.
Han Wu et al.[32]	2017	Type2 Diabetes	Established a predictive model based on K-Means and Logistic regression used for diagnosis of Type2 Diabetes in earlier and avoids the risk.
Oliveira et al.[91]	2017	HIV-AIDS	Support Vector Machine based model produced an effective report without delay for diagnosing cases used to provide the necessary treatment in-time.
Hassoon et al.[35]	2017	Liver disease	Boosted C5.0 based model is a good and flexible algorithm for diagnosis of liver disease in earlier and also reduces cost of treatment.
Simone A. Ludwig et al. [19]	2018	Cancer	The Support Vector Machine (Sequential Minimal Optimization) algorithm is the best one compared to other classifiers for diagnosing cancer in the early stage with high accuracy.
Himansu Das et al.[38]	2018	Diabetes	Naive Bayesian algorithm performed more efficient than J48 for diagnosis and prediction of diabetes.
K. Mathan et al.[39]	2018	Heart Disease	Proposed the combination of decision tree and
			neural system classifiers used for diagnosis of heart disease in the early stage with high accuracy
Mohammad Shafenoor Amin et al.[41]	2018	Heart Disease	Proposed Naïve Bayes and Logistic Regression classifier is the best classifier compared to other classifiers to the prediction of heart disease in the early stage.
EI-Houssainy A.Rady et al. [42]	2019	Kidney Disease	Suggested the probabilistic Neural Networks technique provides the high accuracy and low execution time compare to other techniques used in diagnosis of chronic kidney disease in earlier stage
Senthilkumar Mohan et al.[44]	2019	Heart Disease	The Random forest and linear method (Support Vector Machine) based new hybrid model is used for the early detection of abnormal function of heart
Srabanti Maji et al. [45]	2019	Heart Disease	Proposed a new hybrid model (using decision tree and artificial neural network classifiers) for the prediction of heart disease with high accuracy.
Md. Faisal Faruque et al.[46]	2019	Diabetes	Suggested the C4.5 decision tree has produced better accuracy compare to other classifiers for prediction

			diabetes mellitus in earlier

Table. 2: Summary of Related Work Done by Authors on Diseases

4.1. Datasets Used in Healthcare

The dataset plays a vital role in data mining model performance evaluation process. The dataset is collected from various resources like online, hospitals, respective plants and data collection from the user by interview and questionnaire pattern. The information from our review about the dataset is that most authors have used online data and the remaining authors have also used the data collected from various sources.

Author(s)	Disease	Data Set / Database Used
V. Krishnaiah et al. [21], Sujata Joshi et al. [23] , M.A.Jabbar et al. [25] , Veenita Kunwar et al. [28] , Moloud Abdar et al. [29] , Narander Kumar et al.[33] , Hassoon et al.[35], Simone A. Ludwig et al. [19] , Chandraraj Singh et al.[37], K. Mathan et al.[39], Han Wu et al.[32] , Mohammad Shafenoor Amin et al.[41], EI-Houssainy A.Rady et al. [42] , Pavleen Kaur et al.[43], Senthilkumar Mohan et al.[44], Srabanti Maji et al. [45]	Heart Disease, Kidney Disease, Liver Disease, Brest Cancer	UCI Data depository
Hamed Majidi Zolbanin et al. [22]	Comorbidity of Cancers	Surveillance, Epidemiology, and End Results (SEER)
Dheeraj Raju et al. [24]	Pressure Ulcer	Military Nursing Outcomes Database (MilNOD)
Ahmed Alsayat et al. [26]	Heart Disease	BUPA Medical Research Ltd

Sajida Perveen et al. [27]	Diabetes	Canadian Primary Care Sentinel Surveillance Network (CPCSSN)
Azam Davari Dolatabadi et al. [31]	CAD	MIT-BIH data bank
Md. Faisal Faruque et al.[46]	Diabetes	Medical Centre Chittagong (MCC), Bangladesh

Table. 3: Summary of the Data Sources Used in Research by Different Authors

4.2. Data Mining Tools Used in Healthcare - Comparative Study

The Data Mining Tools are very essential component of designing new data mining models and validating their accuracy. More number of data mining tools are available and each carried out different data mining tasks. Each data mining tool has its own unique advantages and disadvantages. The functionalities of the data mining tools are:

- Characterization and classification of data.
- Patterns evaluation.
- Associations and correlations.
- Prediction over the data.

According from our review, the different authors have used various data mining tools in their research. The Table.4 presents the summarized data about the various tools such as WEKA (W),Rapid Miner (RM), R, CHAID (C) ROC and Python (P) have used by different authors in their research.

Author(s)	Year	Data Mining Tool Used					
		W	RM	R	C	ROC	P
Sujata Joshi et al. [**Error! Reference source not found.**]	2015	√	×	×	×	×	×
Dheeraj Raju et al. [24]	2015	×	×	√	×	×	×

M.A.Jabbar et al. [25]	2015	√	×	×	×	×	×
Ahmed Alsayat et al. [26]	2015	×	×	√	×	×	×
Veenita Kunwar et al. [28]	2016	×	√	×	×	×	×
Moloud Abdar et al. [29]	2016	×	×	×	√	×	×
Tapas Ranjan Baitharua et al. [30]	2016	√	×	×	×	×	×
Azam Davari Dolatabadi et al. [31]	2017	×	×	×	×	√	×
Han Wu et al.[32]	2017	√	×	×	×	×	×
Narander Kumar et al.[33]	2017	√	×	×	×	×	×
Oliveira et al. [91]	2017	√	×	×	×	×	×
Simone A. Ludwig et al. [19]	2018	√	×	×	×	×	×
Chandraraj Singh et al.[37]	2018	×	×	×	×	×	√
Himansu Das et al.[38]	2018	√	×	×	×	×	×

Mohammad Shafenoor Amin et al.[41]	2018	×	√	×	×	×	×
Han Wu et al. [32]	2018	√	×	×	×	×	×
Senthilkumar Mohan et al.[44]	2019	×	×	√	×	×	×
Srabanti Maji et al. [45]	2019	√	×	×	×	×	×
Pavleen Kaur et al.[43]	2019	√	×	×	×	×	×

Table. 4:   Assessment of Different Data Mining Tools Used

4.3. Study of Accuracies in Healthcare

The performance of the new data mining model is measured by using various data mining tools with various parameters and data set collected from different sources. Table-5 presents information about a review conducted in the title "Data Mining approaches used in the Healthcare" over the past 5 years from 2015 to 2019. It also summarizes the components like data mining algorithms, data mining tools, data sets, accuracy produced, year of article published, used in the research by various researchers.

Author	Type of Disease	Year & Publisher	Data Source	DM Techniques used	Accuracy %
V. Krishnaiah et al. [21]	Heart Disease	2015 ,Springer	UCI Data Repository	Fuzzy K-NN	98
Hamed Majidi Zolbani et.al [22]	Comorbidity of Cancers	2015 , Elsevier	SEER	Random Forest	77.80
Sujata Joshi et al. [23]	Heart Disease	2015 ,Springer	UCI Data Repository	K-Nearest Neighbors algorithm	100

Dheeraj Raju et al. [24]	Pressure Ulcer	2015 , Elsevier	MilNOD	Random Forest	83
M.A.Jabbar et al. [25]	Heart Disease	2015, IEEE	UCI Data Repository	One – R+NB	86.29
Ahmed Alsayat et al. [26]	Liver Disease Heart Disease	2016, IEEE	UCI, BUPA Medical Research Ltd	SOM Genetic K-Means	73.84 69.90
Sajida Perveen et al. [27]	Diabetes.	2016, Elsevier	CPCSSN	adaboost ensemble method	99.14
Veenita Kunwar et al. [28]	Chronic Kidney Disease	2016, IEEE	UCI Data Repository	Naïve Bayes	100
Moloud Abdar et al. [29]	Liver Disease	2016, Elsevier	(ILPD) of UCI	Decision Tree (C5.0)	93.75
Tapas Ranjan Baithar ua et al. [30]	Liver Disease	2016, Elsevier	Liver Function Tests (LFTs)	Multilayer Perceptron	71.59
Azam Davari Dolatabadi et al.[31]	CAD	2017, Elsevier	MIT-BIH data bank	SVM	99.2
Narander Kumar et al.[33]	Chronic Kidney Disease	2017, IEEE	UCI Data Repository	Random Forest	100
Oliveira et al.[91]	HIV-AIDS	2017, Springer	WHO, NHS, PHC	SVM	62
Hassoon et al.[35]	Liver Disease	2017, IEEE	UCI	boosted C5.0	81.87
Simone A. Ludwig et al. [19]	Cancer	2018, Springer	UCI Data Repository	SVM (SMO)	100
Chandraraj Singh et al.[37]	Brest Cancer, Mental Health	2018, IEEE	UCI OSMI ltd	K-Nearest Neighbors algorithm	73.4 85.3
Himansu Das et al.[38]	Diabetes	2018, Springer	Local MCH	Naïve Bayesian	72.5
K. Mathan et al.[39]	Heart disease	2018, Springer	UCI Data Repository	Neural Network	87.89
Han Wu et al.[32]	Type 2 Diabetes Mellitus	2018, Elsevier	Pima Indians Diabetes Dataset	improved K-means and logistic regression	93.5
Mohammad Shafenoor Amin et al.[41]	Heart Disease	2018, Elsevier	UCI Data Repository	Hybrid (NB +LR)	87.41
EI-Houssainy, A.Rady et al. [42]	Chronic Kidney Disease	2019, Elsevier	UCI Data Repository	PNN	96.7
Pavleen Kaur et al.[43]	General	2019, Springer	UCI Data Repository	Random Forest	97.26
Senthilkumar Mohan et al.[44]	Heart Disease	2019, IEEE	UCI Data Repository	Hybrid (RF +LM)	88.14
Srabanti Maji et al. [45]	Heart Disease	2019, Springer	UCI Data Repository	Hybrid (DT +ANN)	78.14
Md. Faisal Faruque et al.[46]	Diabetes	2019, IEEE	MCC Bangladesh	C4.5 Decision Tree	73.5

Table. 5: Information Related to Performance Accuracy of Different Data Mining Techniques

Using some of the contents of Table.5, we have given the graphical representation to illustrate the accuracy of the data mining models that various researchers have used in their research, using the heart disease and diabetes dataset collected from the UCI data repository.

Figure.5 shows the measurements of various data mining techniques used for the cardiovascular dataset collected from the UCI data repository, and we conclude that the K-Nearest Neighbours algorithm based algorithms which provides better accuracy for the detection of Heart disease.

The Figure. 6 displayed the measurements of the various Data Mining techniques applied to the Diabetes dataset collected from the UCI data repository, and we conclude that the SVM algorithms provides better accuracy for diagnosis of heart disease.
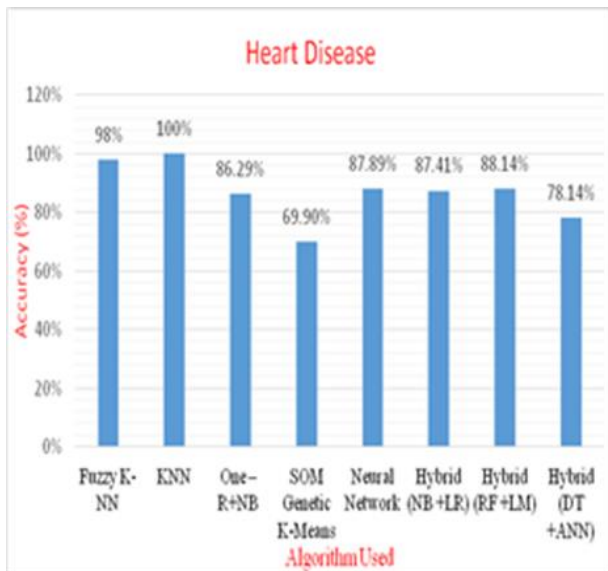

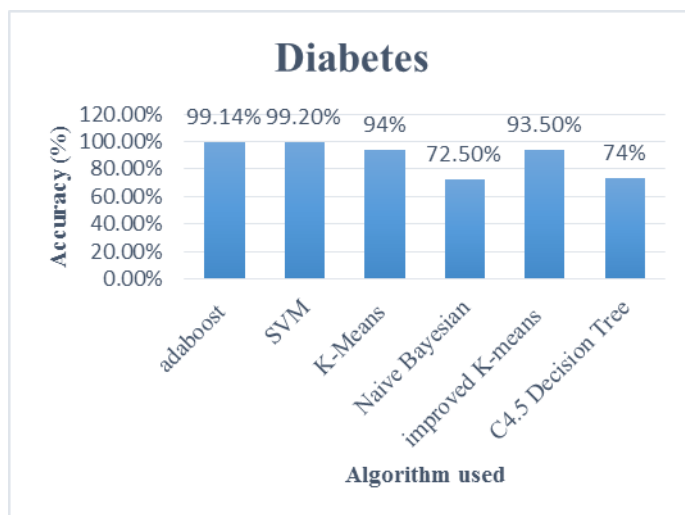Figure. 5 Different Data Mining Algorithms Used for Heart Disease

Figure. 6 Different Data Mining Algorithms Used for Heart Disease

5. DISCUSSION

The objective of this review is identifying the valuable information from the researches titled on 'data mining used in healthcare field' and reviewed of their emerging trends. Using this information and assessment, we have tabulated the information about the research conducted by various researchers using various data mining techniques on healthcare records from 2015 to 2019. According to that assessment, we have learnt some important information about data mining used in the healthcare industry. This study has designed to review the effectiveness of data mining strategies used in disease diagnosis. Earlier studies have explained the importance of data mining strategies in the healthcare sector. Another purpose of this study is to find which data mining methods have had the greatest impact on the early detection of diseases in the medical field, so we analyzed the past studies based on this objective. Researchers have used appropriate data mining techniques and dataset repositories in their research to solve their research problem. Another important point is that many researchers have used the WEKA tool for building the classification model to solve their research problem. This implies that the WEKA tool is particularly useful for researchers in the use of data mining strategies to improve disease diagnosis. This study further enhances our knowledge of data mining strategies in diagnosis. Therefore, this review offers valuable recommendations for researchers doing research on the topic of data mining strategies used in the healthcare industry.

6. CONCLUSION

World health is severely affected by diseases which are spreading and increasing every day. So, the healthcare sector faces many challenges related to their operations. The main challenge in the healthcare sector is the death rate caused by non-communicable and thyroid comparatively high to other factors. Lack of treatment and delayed diagnosis are key factors in the death of patients. Therefore, diagnosis and prognosis plays an important role in the medical field. This paper presents a study of different data mining algorithms and strategies which are very useful for diagnosis of disease in earlier stage. Over the last 5 years, the number of research has been done in the predictive analysis of diseases using classification strategies. This article, have reviewed the results and other details of the last 5 years of research on data mining strategies used in the healthcare industry. The accuracy of classifier, the performance speed of classifiers and dataset quality are important factors for improving the activities of the healthcare field like disease diagnosis, maintenance of healthcare record, etc. Therefore, this article presents a brief description of the various research conducted in the field of health, the classifiers used in the research, and the accuracy and effectiveness of their results. After completing our review of studies from 2015 to 2019 titled Data Mining used in the Healthcare Sector, we have

learnt that the hybrid model is better compared to all other traditional models used in the healthcare industry. Therefore, this article presents valuable information from the current research on data mining used in the healthcare industry and new ideas for developing a better predictive model for diagnosis. So, future research should focus on ways to improve the classification accuracy and their performance by designing new hybrid classification models. The main objective of this review is to provide a new set of guidelines and dimensions to the people involved in the healthcare field regarding data mining techniques used in the healthcare industry.

REFERENCES

[1] Kiranmayee, B. V., T. V. Rajinikanth, and S. Nagini. "A novel data mining approach for brain tumour detection." In 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), pp. 46-50, 2016.

[2] Tamin, Felix, and Ni Made Satvika Iswari. "Implementation of C4. 5 algorithm to determine hospital readmission rate of diabetes patient." In 2017 4th International Conference on New Media Studies (CONMEDIA), pp. 15-18. 2017.

[3] Dandge, Shruti, and Shankar Chakraborty. "Selection of Machining Parameters in Ultrasonic Machining Process Using CART Algorithm." In Advanced Engineering Optimization through Intelligent Techniques, pp. 599-607. 2020.

[4] Díaz-Pérez, Flora Ma, and Ma Bethencourt-Cejas. "CHAID algorithm as an appropriate analytical method for tourism market segmentation." Journal of Destination Marketing & Management 5, no. 3, pp. 275-282. 2016.

[5] Ju, Xinglong, and Feng Liu. "Wind farm layout optimization using self-informed genetic algorithm with information guided exploitation." Applied Energy 248, pp. 429-445, 2019.

[6] Dengen, Nataniel, Edy Budiman, Masna Wati, and Ummul Hairah. "Student Academic Evaluation using Naïve Bayes Classifier Algorithm." 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT), pp. 104-107, 2018.

[7] Kamel, Hajer, Dhahir Abdulah, and Jamal M. Al-Tuwaijari. "Cancer Classification Using Gaussian Naive Bayes Algorithm." International Engineering Conference (IEC), pp. 165-170, 2019.

[8] Surya, Prabha PM, Lakshmi V. Seetha, and B. Subbulakshmi. "Analysis of user emotions and opinion using Multinomial Naive Bayes Classifier." In 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 410-415, 2019.

[9] Kabir, Golam, Haruki Suda, Ana Maria Cruz, Felipe Munoz Giraldo, and Solomon Tesfamariam. "Earthquake-related Natech risk assessment using a Bayesian belief network model." Structure and Infrastructure Engineering 15, no. 6, pp. 725-739, 2019.

[10] Mohammadfam, Iraj, Fakhradin Ghasemi, Omid Kalatpour, and Abbas Moghimbeigi. "Constructing a Bayesian network model for improving safety behavior of employees at workplaces." Applied ergonomics 58, pp. 35-47, 2017.

[11] Pan, Yongping, and Haoyong Yu. "Biomimetic hybrid feedback feedforward neural-network learning control." IEEE transactions on neural networks and learning systems 28, no. 6. pp.1481-1487, 2016.

[12] Sharma, Manish, Praveen Sharma, Ashwini Saini, and Kirti Sharma. "Modular Neural Network for Detection of Diabetic Retinopathy in Retinal Images." In Proceedings of First International Conference on Smart System, Innovations and Computing, pp. 363-370, 2018.

[13] Agarwal, Vandana, and Surekha Bhanot. "Radial basis function neural network-based face recognition using firefly algorithm." Neural Computing and Applications 30, no. 8, pp. 2643-2660, 2018.

[14] Guo, Liang, Naipeng Li, Feng Jia, Yaguo Lei, and Jing Lin. "A recurrent neural network based health indicator for remaining useful life prediction of bearings." Neurocomputing 240, pp. 98-109, 2017.

[15] Sowah, Robert A., Marcellinus Kuuboore, Abdul Ofoli, Samuel Kwofie, Louis Asiedu, Koudjo M. Koumadi, and Kwaku O. Apeadu. "Decision support system (DSS) for fraud detection in health insurance claims using genetic support vector machines (GSVMs)." Journal of Engineering, 2019.

[16] Ma, Lerong, Dandan Song, Lejian Liao, and Jingang Wang. "PSVM: a preference-enhanced SVM model using preference data for classification." Science China Information Sciences 60, pp. 122103, 2017.

[17] Ding, Shifei, Nan Zhang, Xiekai Zhang, and Fulin Wu. "Twin support vector machine: theory, algorithm and applications." Neural Computing and Applications 28, pp. 3119-3130, 2017.

[18] Afzali, Golnar Assadat, and Shahriar Mohammadi. "Privacy preserving big data mining: association rule hiding using fuzzy logic approach." IET Information Security 12, pp.no. 15-24, 2017.

[19] Bhavsar, Ankit R., and Harshal A. Arolkar. "Multidimensional association rule based data mining technique for cattle health monitoring using wireless sensor network." In 2014 international conference on computing for sustainable global development (INDIACom), pp. 810-814, 2014.

[20] Li, Yanjiao, Sen Zhang, Yixin Yin, Jie Zhang, and Zhengguang Xu. "Quantitative association rule mining for blast furnace production data." In 2017 Chinese Automation Congress (CAC), pp. 2275-2280, 2017.

[21] Krishnaiah, V., G. Narsimha, and N. Subhash Chandra. "Heart disease prediction system using data mining technique by fuzzy K-NN approach." In Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1, pp. 371-384, 2015.

[22] Zolbanin, Hamed Majidi, Dursun Delen, and Amir Hassan Zadeh. "Predicting overall survivability in comorbidity of cancers: A data mining approach." Decision Support Systems 74, pp. 150-161, 2015

[23] Joshi, Sujata, and Mydhili K. Nair. "Prediction of heart disease using classification based data mining techniques." In Computational Intelligence in Data Mining-Volume 2, pp. 503-511, 2015.

[24] Raju, Dheeraj, Xiaogang Su, Patricia A. Patrician, Lori A. Loan, and Mary S. McCarthy. "Exploring factors associated with pressure ulcers: a data mining approach." International journal of nursing studies 52, pp. 102-111, 2015

[25] Jabbar, M. A., B. L. Deekshatulu, and Priti Chandra. "Computational intelligence technique for early diagnosis of heart disease." In 2015 IEEE International Conference on Engineering and Technology (ICETECH), pp. 1-6, 2015.

[26] Alsayat, Ahmed, and Hoda El-Sayed. "Efficient genetic K-means clustering for health care knowledge discovery." In 2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA), pp. 45-52, 2016.

[27] Perveen, Sajida, Muhammad Shahbaz, Aziz Guergachi, and Karim Keshavjee. "Performance analysis of data mining classification techniques to predict diabetes." Procedia Computer Science 82, pp. 115-121, 2015.

[28] Kunwar, Veenita, Khushboo Chandel, A. Sai Sabitha, and Abhay Bansal. "Chronic Kidney Disease analysis using data mining classification techniques." In 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence), pp. 300-305, 2016.

[29] Abdar, Moloud, Mariam Zomorodi-Moghadam, Resul Das, and I-Hsien Ting. "Performance analysis of classification algorithms on early detection of liver disease." Expert Systems with Applications 67, pp. 239-251, 2017.

[30] Baitharu, Tapas Ranjan, and Subhendu Kumar Pani. "Analysis of data mining techniques for healthcare decision support system using liver disorder dataset." Procedia Computer Science 85, pp. 862-870, 2016.

[31] Dolatabadi, Azam Davari, Siamak Esmael Zadeh Khadem, and Babak Mohammadzadeh Asl. "Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM." Computer methods and programs in biomedicine 138, pp. 117-126, 2017.

[32] Wu, Han, Shengqi Yang, Zhangqin Huang, Jian He, and Xiaoyi Wang. "Type 2 diabetes mellitus prediction model based on data mining." Informatics in Medicine Unlocked 10, pp. 100-107, 2018.

[33] Kumar, Narander, and Sabita Khatri. "Implementing WEKA for medical data classification and early disease prediction." In 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT), pp. 1-6, 2017.

[34] Domadiya, Nikunj, and Udai Pratap Rao. "Privacy preserving distributed association rule mining approach on vertically partitioned healthcare data." Procedia computer science 148, pp. 303-312, 2019.

[35] Hassoon, Mafazalyaqeen, Mikhak Samadi Kouhi, Mariam Zomorodi-Moghadam, and Moloud Abdar. "Rule optimization of boosted c5. 0 classification using genetic algorithm for liver disease prediction." In 2017 international conference on computer and applications (icca), pp. 299-305, 2017.

[36] Ludwig, Simone A., Stjepan Picek, and Domagoj Jakobovic. "Classification of cancer data: analyzing gene expression data using a fuzzy decision tree algorithm." In Operations Research Applications in Health Care Management, pp. 327-347, 2018.

[37] Singh, Chandraraj, Naveen Cheggoju, and Vishal R. Satpute. "Implementing Classification algorithms in Medical Report Analysis for helping Patient during unavailability of Medical expertise." In 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-5, 2018.

[38] Das, Himansu, Bighnaraj Naik, and H. S. Behera. "Classification of diabetes mellitus disease (DMD): a data mining (DM) approach." In Progress in computing, analytics and networking, pp. 539-549, 2018.

[39] Mathan, K., Priyan Malarvizhi Kumar, Parthasarathy Panchatcharam, Gunasekaran Manogaran, and R. Varadharajan. "A novel Gini index decision tree data mining method with neural network classifiers for prediction of heart disease." Design Automation for Embedded Systems 22, no. 3, pp. 225-242, 2018.

[40] Songthung, Phattharat, and Kunwadee Sripanidkulchai. "Improving type 2 diabetes mellitus risk prediction using classification." In 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 1-6, 2016.

[41] Amin, Mohammad Shafenoor, Yin Kia Chiam, and Kasturi Dewi Varathan. "Identification of significant features and data mining techniques in predicting heart disease." Telematics and Informatics 36, pp. 82-93, 2019.

[42] Rady, El-Houssainy A., and Ayman S. Anwar. "Prediction of kidney disease stages using data mining algorithms." Informatics in Medicine Unlocked 15, 100178, 2019.

[43] Kaur, Pavleen, Ravinder Kumar, and Munish Kumar. "A healthcare monitoring system using random forest and internet of things (IoT)" Multimedia Tools and Applications 78, no. 14, pp. 19905-19916, 2019.

[44] Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." IEEE Access 7, pp. 81542-81554, 2019.

[45] Maji, Srabanti, and Srishti Arora. "Decision tree algorithms for prediction of heart disease." In Information and Communication Technology for Competitive Strategies, pp. 447-454, 2019.

[46] Faruque, Md Faisal, and Iqbal H. Sarker. "Performance analysis of machine learning techniques to predict diabetes mellitus." In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1-4, 2019.

[47] Kaur, Shubpreet, and R. K. Bawa. "Future trends of data mining in predicting the various diseases in medical healthcare system." International Journal of Energy, Information and Communications 6, pp. 17-34, 2015.

[48] Alonso, Susel Góngora, Isabel de la Torre-Díez, Sofiane Hamrioui, Miguel López-Coronado, Diego Calvo Barreno, Lola Morón Nozaleda, and Manuel Franco. "Data mining algorithms and techniques in mental health: A systematic review." Journal of medical systems 42, pp. 161, 2018.

[49] Jain, Divya, and Vijendra Singh. "Feature selection and classification systems for chronic disease prediction: A review." Egyptian Informatics Journal 19, pp. 179-189, 2018.

[50] Rojas, Eric, Jorge Munoz-Gama, Marcos Sepúlveda, and Daniel Capurro. "Process mining in healthcare: A literature review." Journal of biomedical informatics 61, pp. 224-236, 2016.

[51] Sohail, Muhammad Noman, Ren Jiadong, Muhammad Musa Uba, and Muhammad Irshad. "A comprehensive looks at data mining techniques contributing to medical data growth: a survey of researcher reviews." In Recent Developments in Intelligent Computing, Communication and Devices, pp. 21-26, 2019.

[52] Krishna, Chittem Leela, and Poli Venkata Subba Reddy. "An Efficient Deep Neural Network Multilayer Perceptron Based Classifier in Healthcare System." In 2019 3rd International Conference on Computing and Communications Technologies (ICCCT), pp. 1-6, 2019.

[53] K. Sindhya , "A Study On Data Mining Classification Algorithms In Various Disease Predictions", 2019, published in EPRA International Journal.

[54] Vikas Chaurasia, Saurabh Pal, BB Tiwari "Prediction of benign and malignant breast cancer using data mining techniques" 2018, published in Journal of Algorithms & Computational Technology (JACT).

[55] Aswathi, A. K., and Anil Antony. "An intelligent system for thyroid disease classification and diagnosis." In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 1261-1264, 2018.

[56] Sumathi, A., G. Nithya, and S. Meganathan. "Classification of thyroid disease using data mining techniques." Int J Pure Appl Math 119, pp. 13881-13890, 2018.

[57] Qin YU, Tao Jiang, Aiyun Zhou, Lili Zhang, Cheng Zhang, Pan Xu "Computer Aided diagnosis of malignant or benign thyroid nodes based on ultrasound images", 2017, published in European Archives of Oto-Rhino-Laryngology

[58] Begum, Amina, and A. Parkavi. "Prediction of thyroid disease using data mining techniques." In 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), pp. 342-345, 2019.

[59] Umar Sidiq, Dr, Syed Mutahar Aaqib, and Rafi Ahmad Khan. "Diagnosis of various Thyroid ailments using data mining classification techniques." Int J Sci Res Coput Sci Inf Technol 5, pp. 131-136, 2019.

[60] Razia, Shaik, and MR Narasinga Rao. "Machine learning techniques for thyroid disease diagnosis-a review." Indian Journal Science Technology 9, pp. 1-9, 2016.

[61] Aswathi, A. K., and Anil Antony. "An intelligent system for thyroid disease classification and diagnosis." In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 1261-1264, 2018.

[62] Ioniţă, Irina, and Liviu Ioniţă. "Prediction of thyroid disease using data mining techniques." BRAIN. Broad Research in Artificial Intelligence and Neuroscience 7, pp. 115-124, 2016.

[63] Srinivasan, B., and K. Pavya. "Diagnosis of thyroid disease using data mining techniques: a study." International Research Journal of Engineering and Technology 3, pp.1191-1194, 2016

[64] Smayra, Tarek, Zahra Charara, Ghassan Sleilaty, Gaelle Boustany, Lina Menassa-Moussa, and Georges Halaby. "Classification and Regression Tree (CART) model of sonographic signs in predicting thyroid nodules malignancy." European Journal of Radiology, pp. 343-349, 2019.

[65] Priyal, S. Sathya, and D. Anitha. "Survey on Thyroid Diagnosis using Data Mining Techniques" International Journal Advanced Research Computer Communication Engineering, pp.161-4, 2017.

[66] Razia, Shaik, and MR Narasinga Rao. "Machine learning techniques for thyroid disease diagnosis-a review." Indian Journal Science Technology, pp. 1-9, 2016.

[67] Pal, Rekha, Tanvi Anand, and Sanjay Kumar Dubey. "Evaluation and performance analysis of classification techniques for thyroid detection." International Journal of Business Information Systems 28, pp.no. 163-177, 2018.

[68] Ioniţă, Irina, and Liviu Ioniţă. "Prediction of thyroid disease using data mining techniques." BRAIN. Broad Research in Artificial Intelligence and Neuroscience 7, pp. 115-124, 2016.

[69] Sindhuja, D., and R. Jemina Priyadarsini. "A survey on classification techniques in data mining for analyzing liver disease disorder." International Journal of Computer Science and Mobile Computing 5, pp. 483-488, 2016.

[70] Hannah Ritchie and Max Roser, "Causes of Death", Published by University of OXFORD. https://ourworldindata.org/causes-of-death, 2019.

[71] World Health Organization. https:// www.who.int/healthinfo / global_burden_disease / en

[72] Jothi, Neesha, and Wahidah Husain. "Data mining in healthcare–a review." Procedia computer science, pp. 306-313, 2015.

[73] Ahmad, Wan Muhamad Taufik Wan, Nur Laila Ab Ghani, and Sulfeeza Mohd Drus. "Data Mining Techniques for Disease Risk Prediction Model: A Systematic Literature Review." In International Conference of Reliable Information and Communication Technology, pp. 40-46, 2018.

[74] Das, Amit Kumar, Aman Kedia, Lisha Sinha, Saptarsi Goswami, Tamal Chakrabarti, and Amlan Chakrabarti. "Data mining techniques in Indian healthcare: A short review." In 2015 International Conference on Man and Machine Interfacing (MAMI), pp. 1-7, 2015.

[75] Mia, Md Robel, Syed Akhter Hossain, Amit Chakraborty Chhoton, and Narayan Ranjan Chakraborty. "A Comprehensive Study of Data Mining Techniques in Health-care, Medical, and Bioinformatics." In 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), pp. 1-4 , 2018.

[76] Pandey, Subhash Chandra. "Data mining techniques for medical data: a review." In 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), pp. 972-982, 2016.

[77] Batista, Edgar, and Agusti Solanas. "Process mining in healthcare: a systematic review." In 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA), pp. 1-6, 2018.

[78] Gorade, S. M., Ankit Deo, and Pritesh Purohit. "A study of some data mining classification techniques." International Research J. of Engineering and Technology (IRJET), 2017.

[79] Vougas, Konstantinos, Theodore Sakellaropoulos, Athanassios Kotsinas, George-Romanos P. Foukas, Andreas Ntargaras, Filippos Koinis, Alexander Polyzos et al. "Machine learning and data mining frameworks for predicting drug response in cancer: An overview and a novel in silico screening process based on association rule mining." Pharmacology & therapeutics , 2019.

[80] Domadiya, Nikunj, and Udai Pratap Rao. "Privacy preserving distributed association rule mining approach on vertically partitioned healthcare data." Procedia computer science 148, pp. 303-312, 2019.

[81] Altaf, Wasif, Muhammad Shahbaz, and Aziz Guergachi. "Applications of association rule mining in health informatics: a survey." Artificial Intelligence Review 47, pp. 313-340, 2017.

[82] Christy, A., G. Meera Gandhi, and S. Vaithyasubramanian. "Cluster based outlier detection algorithm for healthcare data." Procedia Computer Science 50, pp. 209-215, 2015.

[83] Deepika, M., and K. Kalaiselvi. "A Empirical study on disease diagnosis using data mining techniques." In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 615-620, 2018.

[84] Rahman, Nafisur, and Samar Wazir. "Frequent Itemset Mining Algorithms for Knowledge Discovery-A Compendious Review of Various Approaches." International Journal of Engineering and Management Research (IJEMR) 7, pp. 379-382, 2017.

[85] Tan, Junyan, Tianyu Xiong, Hongxia Miao, Rurong Sun, and Min Wu. "A case study of medical big data processing: Data mining for the hyperuricemia." In 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), pp. 196-201, 2018.

[86] Jothi, Neesha, and Wahidah Husain. "Data mining in healthcare–a review." Procedia computer science 72, pp. 306-313, 2015.

[87] Saastamoinen, Kalle, and Jaakko Ketola. "Medical data classification using logical similarity based measures." In 2006 IEEE Conference on Cybernetics and Intelligent Systems, pp. 1-5, 2006.

[88] Sharma, Richa, Shailendra Narayan Singh, and Sujata Khatri. "Data mining classification techniques-comparison for better accuracy in prediction of cardiovascular disease." International Journal of Data Analysis Techniques and Strategies 11, pp. 356-373, 2019.

[89] Svenstrup, Dan, Henrik L. Jørgensen, and Ole Winther. "Rare disease diagnosis: a review of web search, social media and large-scale data-mining approaches." Rare Diseases 3, no. 1, 2015.

[90] Kavakiotis, Ioannis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, and Ioanna Chouvarda. "Machine learning and data mining methods in diabetes research." Computational and structural biotechnology journal 15, pp.104-116, 2017.

[91] Oliveira, Alexandra, Brígida Mónica Faria, A. Rita Gaio, and Luís Paulo Reis. "Data mining in HIV-AIDS surveillance system." Journal of medical systems 41, pp. 51, 2017.

Authors

M. Sivasubramaniyan is Research Student (Part Time) in the Department of Computer Science, Government Arts and Science College, Kangeyam. He is working as a Assistant Professor in the Department of Computer Applications, Guru Nanak College (Autonomous), Chennai. He has 20 years of teaching experience. His areas of interests include Data Mining, Microprocessors and Microcontrollers.

Dr. N. Karthikeyani Visalakshi is working as an Assistant Professor in the Department of Computer Science, Government Arts and Science College, Kangeyam. She completed her Ph. D. degree in the area of Distributed Data Mining. She has got 26 years of teaching experience and 18 years of research experience. She has published around 40 research papers in various National and International Journals. She received Best Oral Presentation award in 4 th International Science Congress (ISC-2014) at Pacific University, Udaipur, Rajasthan, organized by International Science Congress Association. She is a reviewer for two International Journals. Her areas of interests include Distributed Data Mining, Clustering, Image Processing, Rough sets, Fuzzy logic and Intuitionistic Fuzzy sets.