Data Mining in Healthcare: An Overview of Applications, Techniques, and Challenges

Naseer Ahmed[1], Rustum Ameer[2], Afrasiyab Khan[2], Nooruddin[2], Saba Gull[1]
[1]Department of Computer Science, Balochistan University of Engineering & Technology, Khuzdar, 89120, Pakistan
[2]Department of Software Engineering, Balochistan University of Engineering & Technology, Khuzdar, 89120, Pakistan

E-mails: naseerbajoi@gmail.com, ameerrustum@gmail.com, afrasiiyabkhan@gmail.com, nooruddin4201sn@gmail.com ,sabanaseerbajoi@gmail.com

Corresponding Author: Rustum Ameer: ameerrustum@gmail.com

Received: 01-02-2024; **Accepted:** 01-05-2024; **Published**: 01-07-2024

Abstract: Artificial intelligence (AI) and the widespread use of electronic health records (EHRs) are driving a data-driven revolution in the healthcare sector mining techniques are extremely useful for deriving significant insights from the large and intricate datasets shaped in healthcare environments. The state-of-the-art uses of data mining in healthcare are examined in this paper, with a focus on AI and machine learning for personalized medicine, risk assessment, disease detection, and therapy optimization. We explore the difficulties that come with data mining for healthcare, including the requirement for interpretable models, data quality, heterogeneity, and privacy data mining and artificial intelligence work together in a way that has the potential to completely transform healthcare delivery, improving patient outcomes, cutting costs, and speeding up medical research. But it's still critical to address ethical issues and make sure data usage is transparent while recognizing the challenges involved, this study emphasizes the revolutionary potential of data mining in healthcare and offers insightful advice for practitioners and scholars in this quickly developing sector.

Keywords: Electronic health records (EHRs), artificial intelligence (AI), machine learning, and data mining

1. Introduction

Healthcare data is expanding at an unprecedented rate as a result of the increasing use of electronic health records, or EHRs. Although scholars and physicians can get a lot of knowledge from this, advanced data mining techniques are necessary to extract relevant insights due to the volume and complexity of EHR data. Data mining has emerged as a powerful tool in all disciplines, data is composed of Things accumulated at a breakneck speed It is increasingly necessary to develop new computational theories and methods to help people extract useful knowledge derived from the vast volumes of digital data. The use of certain data mining techniques for pattern recognition and extraction forms the basis of the procedure[1] [2]. Generalization, characterization, classification, clustering, association, evolution, pattern matching, visualization of data, and meta-rule-guided mining are recent advancements in data mining technology[3] [4]. Healthcare is undergoing an evolution because of data mining and artificial intelligence (AI), which make it possible to extract insightful information from massive amounts of patient data [5]. Healthcare could be transformed by data mining techniques and artificial intelligence [6], [7]Patterns and correlations in the massive amount of healthcare data are sometimes obscured from traditional analysis[8], [9] AI algorithms, particularly machine learning models, excel at uncovering complex relationships within medical datasets [10], [11], [12] Predictive analytics for early disease identification, individualized treatment regimens, and efficient hospital operations can be powered by these findings. By obtaining, purifying, and processing the raw data, data mining offers the fundamental framework around which A models are constructed [13], [14], [15], [16], [17].

The healthcare sector is about to undergo a significant transformation because of data mining. Data mining techniques have the potential to transform illness diagnosis, treatment, and patient care by revealing hidden patterns and insights through the processing of large amounts of complex data. Healthcare data is expanding at an unprecedented rate as a result of the increasing use of electronic health records, or EHRs. Although academics and physicians can get a lot of knowledge from this, advanced data mining techniques are necessary to extract relevant insights due to the volume and complexity of EHR data. The synergistic collaboration of AI and data mining from drug discovery to patient risk evaluation boosts proactive

decision-making in healthcare behavior based on data. [18], [19], [20], [21], [22]. This convergence of technology can provide improved patients, lead to fewer healthcare costs, and quicken medical research [23], [24], [25], [26], [27]. As AI algorithms become increasingly sophisticated and widely accessible, their ethical connotations and the demand to make its data consumption more straightforward will be among the most critical considerations to consider, transforming it into an outstanding tool for the healthcare sector [28], [29], [30], [31], [32], [33].

This can result in PH tools such as forecasting models for early diagnosis of the disease, algorithms to develop patient-specific treatment plans and help hospitals optimize resource distribution. [34], [35], [36], [37]. This revolution's main engine is data mining, which gathers, purifies and organizes the unprocessed data that powers AI models.[38], [39], [40], [41]. This review delves into the state-of-the-art applications of data mining in healthcare, with a specific focus on Artificial Intelligence and Machine learning to help full for the researcher develop an AI tool related to healthcare. This data, which includes genetic information, electronic health records, and medical images, has enormous promise for enhancing diagnosis, estimating the risk of disease, enhancing treatment regimens, and customizing medicine. These complicated data can be analyzed by data mining technologies, such as machine learning and deep learning, to find hidden patterns, forecast patient outcomes, and develop intelligent decision support systems [42], [43], [44]. A data mining technique that is currently under investigation. Literature analysis and classification of articles published between 2019 to 2023 are used to analyse the progress of data mining techniques. The period is important because it coincides with a rise in the use of data mining techniques in the healthcare industry, where technology has been essential, particularly in the creation of systems for data collection from online databases. Using the keyword indices on the online databases Science Direct, Springer Link, Google Scholar, and IEEE Xplore, the search for this literature review began in 2019. There were 200 articles between 2019 to 2023. Search a complete article containing the phrases "data mining approaches, techniques and challenges in healthcare. The papers for this study were chosen from a pool of 90 publications use of data mining techniques and challenges in healthcare.

2. DATA MINING AN OVERVIEW

Data mining in healthcare is a rapidly developing discipline with enormous potential to revolutionize healthcare delivery, according to a study of the literature in this area. The first steps toward investigating the potential applications of data mining techniques to healthcare data were taken by Fayyad et al. and Koh and Tan [45]. indicating how well data mining works to forecast readmission rates in hospitals. Additionally, AI-powered data mining is proving valuable in disease detection[46] Singh et al. investigate its usage in cancer categorization, while [44] highlight its use in heart disease early detection. Data mining has the potential for tailored medicine that goes beyond diagnosis [47]. From day to day, the size of data is generally increasing. In all realms of technology, business, and research, the requirement to comprehend huge, complicated, information-rich data sets has grown. With such a big volume of data. In today's competitive climate, the ability to extract relevant knowledge from it and act on that knowledge is becoming increasingly important. The practice of extracting knowledge from data utilizing a computer-based information system (CBIS) and unique approaches is known as data mining [48].

2.1. FIELDS ADOPTING THE DATA MINING

Machine Learning Method [49].Applications of Artificial Intelligence [50].Predictions of Probability [51]. and Statistics [52] provide the foundation for data mining. Figure 1 shows above summarizes the fields recognized in the studies evaluated. The disciplines referenced in the papers evaluated are shown in Figure 1.

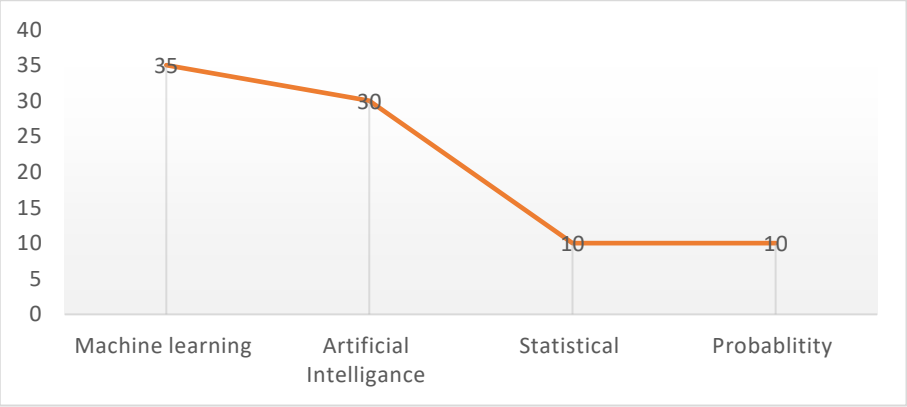

Figure 1 The different disciplines reviewed for data Mining in Health Care

Figure 1 also shows the number of documents reviewed for different disciplines relevant to data mining in healthcare. The line graph depicts how many documents were reviewed across various disciplines in data mining applied to healthcare. The x-axis represents these disciplines, while the y-axis indicates the document count. Machine learning appears to have the most reviewed documents, followed by artificial intelligence and statistical probability.

2.2. Models of Data Mining

There are two main types of data mining models such as descriptive predictive and Predictive models are particularly useful for forecasting future or unknown values of variables of interest. To glean insightful information from massive volumes of patient data, healthcare data mining primarily depends on a variety of models. Conventional models such as regression analysis [1], are used to find correlations between variables, which helps with jobs like forecasting readmissions to hospitals [2]. Classification models [3], play a vital part in the diagnosis of disease by classifying patients according to their attributes. Clustering techniques, detailed by Han and Kamber [4], assemble similar data points to allow for the identification of patient subgroups with different treatment outcomes or risk factors. The discipline has seen the introduction of strong new models due to the growth of artificial intelligence. Besides that, feed-forward architectures for deep learning [5], have excellent image analysis abilities, which benefit the field of cancer diagnosis [6]. On the other hand, due to their excellent skills in the management of sequential data RNNs [7], recurrent neural networks. this indicates that they are suited to tasks like the analysis of risk and prediction errors based on, say, patient medical records [8]. Finally, ensemble methods [9], can harness the strengths of multiple models, increasing robustness and accuracy. Naturally, the specific data properties and the task determine the best model to be used. However, issues like bias mitigation and model interpretability [10] still exist. All aforementioned models use supervised OL techniques to learn from previous data and make predictions on the new presented. For example, predictive models predict customer behavior, stock performance, and even the upcoming weather. On the contrary, descriptive models are used to determine certain patterns or correlations between various datasets but are not meant to forecast the value in the future. They are essential for better understanding the nature of the data and revealing unknown trends and relations. Descriptive models can be applied to the data grouping to find similarities or search for connections among the causality variables [8].

The quantity of research papers evaluated for different data analysis tasks is shown in Figure 2. The most reviewed studies were on classification tasks (18), then association (10), and clustering (8). This suggests that there was more of an emphasis on classification in the study that was examined, maybe as a result of things like more widely available data or a higher estimation of the significance of classification tasks.

2.3. Tasks of Data Mining

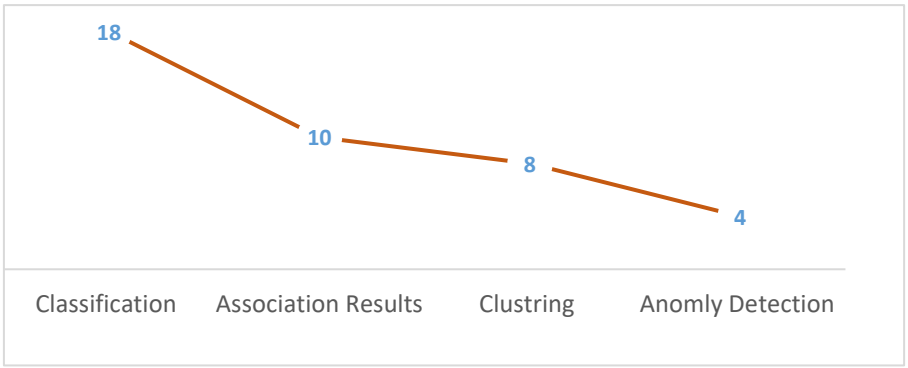

Figure 2 Papers reviewed for derived tasks

A model is typically put into action by carrying out a certain job. Examples include the frequent use of Clustering [9], Association Rules [53], and Correlation Analysis [54]techniques in Descriptive models. On the other hand, predictive models rely on methods like classification [55]. Regression [56]and categorization [57] is a list of the tasks obtained from the publications under examination in Table 1.

3. Literature Review

Once a data mining task and model have been selected, specific data mining methodologies are used to develop appropriate techniques, depending on the discipline under consideration. Various techniques are used for anomaly detection, including Gaussian mixture models, density-induced support vector data description, and standard support vector data description. The vector quantization technique is frequently utilized in clustering. There are several methods available for classification jobs, including statistical techniques, discriminant analysis, decision trees, Markov models, swarm intelligence, genetic classifiers, artificial neural networks, support vector machines, and association rules. These methods can be tailored to suit the specific data and desired outcomes and may involve the use of different algorithms, tools, and techniques. It's worth noting that data mining methodologies and techniques are constantly evolving, with new approaches and tools emerging all the time [88]-[89]. As such, it's important for the data miners to InTouch with the new techniques, by doing so, they can help ensure that they can effectively tackle complex data mining challenges and deliver meaningful insights [90].

Sr#	Description	Algorithm	Author/s	Reference
1	An approach for detecting the maximum impactful variations in a data set is anomaly detection. A balanced dataset has an average result of 93.59 percent.	Anomaly Detection	Bo Lie et al	[57], [58]
2	The process of classifying involves identifying a predictive learning function that divides a batch of data into some specified classes.	Classification	Clustering Rui Velosoa	[[59], [60], [61]
3	The class is predicted using discriminant analysis and measurements taken from brand-new, unlabelled observations.	Discriminant Analysis	Armañanzas et al. and Jen et al	[62], [63]
4	The decision tree cannot be used to suggest predicted decisions to address imbalanced problems since it recursively divides observations into branches to construct a tree.	Decision Tree	Sharma & Om Wang et al and Zolbanin et al	[64], [65], [66]
5	In large search spaces, the particle swarm optimization (PSO) method can quickly find optimal or nearly optimal solutions.	Swarm Intelligence	Yeh et al. and Abdi & Giveki	[67], [68]
6	The method assumes that all instances are points in the n-dimensional space RN and that the parameter units are the operation's samples.	K-NN	García-Laencina jen et al. rmañanzas et alBagui et al and aŞahan et al	[69], [70], [71]
7	It is uncomplicated to use and produces competitive results.	Logistic Regression	Mamiya et al. Su et alThompson et al and amanta et Zolbanin et al.	[72], [73], [74]
8	its intuitive and effective handling of missing data, as well as its computational effectiveness. Due to this advantage, both authors' models have demonstrated great forecast accuracy.	Bayesian Classifier	Bandyopadhyay et al. Armañanzas et al. Wang et al	[75], [76]
9	The method lowers the upper bound of the generalization error by using the structural risk reduction concept. In the same way that SVM training can be used to solve the original problem, it can also be used to solve a linear restricted quadratic programming problem.	Support Vector	Zheng et al. Kang et al and Su et al and García-Laencina et al	[77], [78], [79]

Table 1 shows different algorithms used in data mining healthcare

Anomaly Detection, Classification, Discriminant Analysis, Decision Tree, Swarm Intelligence, K-NN, Logistic Regression, Bayesian Classifier, and Support Vector Machines are among the data mining methods utilized in the healthcare industry that are listed in Table 1. It includes writers who have studied various methods as well as succinct explanations of the goals of each algorithm. The table also contains references, which most likely lead to particular research articles linked to the listed authors and algorithms. Algorithms belonging to Data Mining are used in healthcare. Healthcare encompasses the diagnosis, prevention, and treatment of illnesses, injuries, and other corporal and mental conditions in people [77]. The industry of healthcare is experiencing rapid change in many countries, with the advent of new technologies and practices. Given the vast amount of data generated by the healthcare sector, comprising administrative data, benchmarking statistics, and electronic medical records, it is often referred to as a "data-rich environment" Despite this wealth of data, it is often underutilized. Data mining has the potential to uncover valuable insights from these vast data sets, as described in section 2.0. By analysing this data, data mining can help doctors make more informed clinical decisions and improve patient care. Consequently, data mining is becoming increasingly prevalent in the healthcare industry [80].

4. Challenges in Healthcare Data Mining

Data mining holds immense potential to transform the healthcare sector, but its application still faces significant challenges. Data quality issues, such as missing values, inconsistencies, and noise, can hinder the reliability and accuracy of models [[81] [82] The heterogeneous and complex nature of healthcare data, encompassing structured records, unstructured notes, and medical images, adds to the complexity of data preprocessing and integration [83], [84]Privacy and security concerns remain paramount due to the sensitive nature of personal health information, and ethical considerations in data use must be meticulously addressed [85]. Moreover, the interpretability of data mining models for the healthcare provider is a vital factor. Complicated “black box” models are often poorly understood and thus do not foster trust on the part of clinicians [86], [87].

5. Discussion

This studied research reiterates many of the challenges that continue to persist when data mining is used in health and poor-quality data insecurity. Some of the features of poor data quality include noise, inconsistencies, and missing values, all of which hinder the dependability of generated models. The data in healthcare is complex, with unstructured notes, structured records, and various medical pictures all contributing to making data integration more difficult. Additionally, ethical and technological safeguards must be implemented in large-scale data undertakings to safeguard patient information. Lastly, the implementation of complicated data mining algorithms in clinical contexts may be hampered by their intrinsic lack of interpretability. Future studies should give special attention to data governance, sophisticated data integration methods, privacy-preserving technologies, and the creation of explainable AI models in the healthcare industry to fully realize the promise of data mining. According to the studies analyzed Data mining as previously noted, the accuracy of approaches varies between the training and testing sets, based on the properties of the size of the data sets. Highly unbalanced data sets are common in healthcare, and unbalanced majority and minority classifiers lead to erroneous predictions when the classifiers are run. Another element to examine is the presence of missing values in healthcare data sets. The sample size of the data is usually regarded as another feature because the data is frequently on a tiny scale. There is no single data mining method that can solve all of these issues.

6. Conclusion

Data mining is being used more and more in the healthcare industry, showcasing its ability to analyze illnesses, aid in medical diagnosis, and forecast health outcomes. Even if data mining has enormous benefits, difficulties still exist. To guarantee trustworthy and useful insights, data quality, complexity, privacy issues, and model interpretability must all be properly taken into consideration. Investing in explainable AI, advanced integration techniques, privacy-preserving technology, and data governance is crucial to achieving optimal accuracy and confidence in health diagnosis. The use of data mining in healthcare will probably require a careful balancing act between addressing its inherent challenges and maximizing its potential withstanding these difficulties, there is no denying data mining's revolutionary potential in the medical field. Through sustained investment in research, development, and ethical implementation, the potential to improve patient care can be fully realized.

References

[1] R. Agrawal and G. Psaila, "Active Data Mining," 1995. [Online]. Available: www.aaai.org

[2] R. Agrawal and G. Psaila, "Active Data Mining," 1995. [Online]. Available: www.aaai.org

[3] S. H. Liao, P. H. Chu, and P. Y. Hsiao, "Data mining techniques and applications - A decade review from 2000 to 2011," Expert Systems with Applications, vol. 39, no. 12. pp. 11303–11311, Sep. 15, 2012. doi: 10.1016/j.eswa.2012.02.063.

[4] S. H. Liao, P. H. Chu, and P. Y. Hsiao, "Data mining techniques and applications - A decade review from 2000 to 2011," Expert Systems with Applications, vol. 39, no. 12. pp. 11303–11311, Sep. 15, 2012. doi: 10.1016/j.eswa.2012.02.063.

[5] G. E. Vlahos, T. W. Ferratt, and G. Knoepfle, "The use of computer-based information systems by German managers to support decision making," Information and Management, vol. 41, no. 6, pp. 763–779, 2004, doi: 10.1016/j.im.2003.06.003.

[6] "12.2".

[7] N. P. Jayasri and R. Aruna, "Big data analytics in health care by data mining and classification techniques," ICT Express, vol. 8, no. 2, pp. 250–257, Jun. 2022, doi: 10.1016/j.icte.2021.07.001.

[8] E. Shirzad, G. Ataei, and H. Saadatfar, "Applications of data mining in healthcare area: A survey," Engineering and Applied Science Research, vol. 48, no. 3, pp. 314–323, 2021, doi: 10.14456/easr.2021.34.

[9] P. Mahindrakar and M. Hanumanthappa, "Data Mining In Healthcare: A Survey of Techniques and Algorithms with Its Limitations and Challenges." [Online]. Available: www.ijera.com

[10] D. Ananey-Obiri and E. Sarku, "Predicting the Presence of Heart Diseases using Comparative Data Mining and Machine Learning Algorithms," Int J Comput Appl, vol. 176, no. 11, pp. 17–21, Apr. 2020, doi: 10.5120/ijca2020920034.

[11] R. Ray, "Advances in Data Mining: Healthcare Applications," International Research Journal of Engineering and Technology, [Online]. Available: www.irjet.net

[12] F. Soleimanian Gharehchopogh, P. Mohammadi, and P. Hakimi, "Application of Decision Tree Algorithm for Data Mining in Healthcare Operations: A Case Study," 2012.

[13] N. P. Jayasri and R. Aruna, "Big data analytics in health care by data mining and classification techniques," ICT Express, vol. 8, no. 2, pp. 250–257, Jun. 2022, doi: 10.1016/j.icte.2021.07.001.

[14] I. Ioniță and L. Ioniță, "Applying data mining techniques in healthcare," Studies in Informatics and Control, vol. 25, no. 3, pp. 385–394, 2016, doi: 10.24846/v25i3y201612.

[15] I. O. Ogundele and P. Olusegun, "A REVIEW OF DATA MINING IN HEALTHCARE," 2018. [Online]. Available: https://www.researchgate.net/publication/370899263

[16] F. Soleimanian Gharehchopogh, P. Mohammadi, and P. Hakimi, "Application of Decision Tree Algorithm for Data Mining in Healthcare Operations: A Case Study," 2012.

[17] Ks. BKavihta Rani AGovrdhan Associate Professor, P. of CSE, and N. Karimnagar Jagtial, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks," 2010.

[18] I. Ioniță and L. Ioniță, "Applying data mining techniques in healthcare," Studies in Informatics and Control, vol. 25, no. 3, pp. 385–394, 2016, doi: 10.24846/v25i3y201612.

[19] D. Ananey-Obiri and E. Sarku, "Predicting the Presence of Heart Diseases using Comparative Data Mining and Machine Learning Algorithms," Int J Comput Appl, vol. 176, no. 11, pp. 17–21, Apr. 2020, doi: 10.5120/ijca2020920034.

[20] R. Ray, "Advances in Data Mining: Healthcare Applications," International Research Journal of Engineering and Technology, [Online]. Available: www.irjet.net

[21] F. Soleimanian Gharehchopogh, P. Mohammadi, and P. Hakimi, "Application of Decision Tree Algorithm for Data Mining in Healthcare Operations: A Case Study," 2012.

[22] Ks. BKavihta Rani AGovrdhan Associate Professor, P. of CSE, and N. Karimnagar Jagtial, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks," 2010.

[23] N. P. Jayasri and R. Aruna, "Big data analytics in health care by data mining and classification techniques," ICT Express, vol. 8, no. 2, pp. 250–257, Jun. 2022, doi: 10.1016/j.icte.2021.07.001.

[24] T. V, Vellore Institute of Technology, Institute of Electrical and Electronics Engineers. Madras Section, IEEE Communications Society., and Institute of Electrical and Electronics Engineers, Conference proceedings, International Conference on Vision Towards Emerging Trends In Communication and Networking (ViTECoN 2019) : 30-31, March 2019, Vellore, Tamilnadu, India.

[25] U. Shafique, F. Majeed, H. Qaiser, and I. U. Mustafa, "Data Mining in Healthcare for Heart Diseases," 2015. [Online]. Available: http://www.ijias.issr-journals.org/

[26] I. O. Ogundele and P. Olusegun, "A REVIEW OF DATA MINING IN HEALTHCARE," 2018. [Online]. Available: https://www.researchgate.net/publication/370899263

[27] B. Boukenze, H. Mousannif, and A. Haqiq, "Predictive Analytics in Healthcare System Using Data Mining Techniques," Academy and Industry Research Collaboration Center (AIRCC), Apr. 2016, pp. 01–09. doi: 10.5121/csit.2016.60501.

[28] A. Čaplinskas, G. 1958- Dzemyda, A. Lupeikienė, O. Vasilecas, International Baltic Conference on Databases and Information Systems 10 2012.07.08-11 Vilnius, and BalticDB & IS′2012 10 2012.07.08-11 Vilnius, Databases and information systems tenth international Baltic conference on databases and information systems ; local proceedings, materials of doctoral consortium. Žara, 2012.

[29] S. G. Jacob and G. Ramani, "ISSN : 2249-0868 Foundation of Computer Science FCS," 2012. [Online]. Available: www.ijais.org

[30] P. Ahmad, S. Qamar, S. Qasim, and A. Rizvi, "Techniques of Data Mining In Healthcare: A Review," 2015.

[31] "Data Mining Techniques to Find Out Heart Diseases: An Overview."

[32] M. Dey and S. S. Rautaray, "Study and Analysis of Data mining Algorithms for Healthcare Decision Support System." [Online]. Available: www.ijcsit.com

[33] W. : Www, N. Abirami, T. Kamalakannan, and A. Muthukumaravel, "International Journal of Emerging Technology and Advanced Engineering A Study on Analysis of Various Datamining Classification Techniques on Healthcare Data," 2008. [Online]. Available: www.ijetae.com

[34] S. A. Alowais et al., "Revolutionizing healthcare: the role of artificial intelligence in clinical practice," BMC Medical Education, vol. 23, no. 1. BioMed Central Ltd, Dec. 01, 2023. doi: 10.1186/s12909-023-04698-z.

[35] A. Kishor and C. Chakraborty, "Artificial Intelligence and Internet of Things Based Healthcare 4.0 Monitoring System," Wirel Pers Commun, vol. 127, no. 2, pp. 1615–1631, Nov. 2022, doi: 10.1007/s11277-021-08708-5.

[36] A. S. Albahri et al., "Role of biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19): A Systematic Review," Journal of Medical Systems, vol. 44, no. 7. Springer, Jul. 01, 2020. doi: 10.1007/s10916-020-01582-x.

[37] E. Shirzad, G. Ataei, and H. Saadatfar, "Applications of data mining in healthcare area: A survey," Engineering and Applied Science Research, vol. 48, no. 3, pp. 314–323, 2021, doi: 10.14456/easr.2021.34.

[38] Y. Miyashita, T. Hitsumoto, H. Fukuda, J. Kim, T. Washio, and M. Kitakaze, "Predicting heart failure onset in the general population using a novel data-mining artificial intelligence method," Sci Rep, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-31600-0.

[39] P. A. Bath, "Data Mining in Health and Medical Information," Annual Review of Information Science and Technology, vol. 38. Information Today, 2004. doi: 10.1002/aris.1440380108.

[40] C. and C. 2012. 03. 06-09 A. Japan-Egypt Conference on Electronics and JEC-ECC 2012.03.06-09 Alexandria, 2012 Japan-Egypt Conference on Electronics, Communications and Computers (JEC-ECC) 6-9 March 2012, Alexandria, Egypt.

[41] D. C. Tayade and P. M. Karandikar, "Review Article Role of Data Mining Techniques in Healthcare sector in India," Scholars Journal of Applied Medical Sciences (SJAMS, vol. 1, no. 3, pp. 158–160, 2013, doi: 10.36347/sjams.2013.v01i03.005.

[42] Z. S. Tawfik, A. H. Al-Hamami, and M. T. Abd, "Comparison of Data Mining Techniques in Healthcare Data," in 2022 International Conference for Natural and Applied Sciences, ICNAS 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 35–38. doi: 10.1109/ICNAS55512.2022.9944713.

[43] N. Ahmed, M. A. Khouro, A. Khan, M. Dawood, M. A. Dootio, and N. U. Jan, "Student textual feedback sentiment analysis using machine learning techniques to improve the quality of education," Pakistan Journal of Engineering, Technology & Science, vol. 11, no. 2, pp. 32–40, Dec. 2023, doi: 10.22555/pjets.v11i2.1039.

[44] N. Ahmed, "A review of existing Machine Translation Approaches, their Challenges and Evaluation Metrics," Pakistan Journal of Engineering, Technology & Science, vol. 11, no. 1, pp. 29–44, Dec. 2023, doi: 10.22555/pjets.v11i1.1002.

[45] P. Ahmad, S. Qamar, and S. Qasim Afser Rizvi, "Techniques of Data Mining In Healthcare: A Review," Int J Comput Appl, vol. 120, no. 15, pp. 38–50, Jun. 2015, doi: 10.5120/21307-4126.

[46] RVS Technical Campus, IEEE Electron Devices Society, and Institute of Electrical and Electronics Engineers, Proceedings of the International Conference on Electronics, Communication and Aerospace Technology (ICECA 2017) : date: 20,21,22, April 2017.

[47] C. S. Reddy, R. Singh, R. Bhavani, S. Dasgupta, Y. Singh, and S. Pratan Singh, "Using Machine Learning Techniques for Cancer Classification," in Proceedings of the 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems, ICSES 2022, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICSES55317.2022.9914370.

[48] G. E. Vlahos, T. W. Ferratt, and G. Knoepfle, "The use of computer-based information systems by German managers to support decision making," Information and Management, vol. 41, no. 6, pp. 763–779, 2004, doi: 10.1016/j.im.2003.06.003.

[49] G. E. Vlahos, T. W. Ferratt, and G. Knoepfle, "The use of computer-based information systems by German managers to support decision making," Information and Management, vol. 41, no. 6, pp. 763–779, 2004, doi: 10.1016/j.im.2003.06.003.

[50] I. H. Witten Eibe Frank Mark A Hall ELSEVIER AMSTERDAM, L. New York, S. Diego, and M. Kaufmann, "Data Mining Practical Machine Learning Tools and Techniques Third Edition M< Contents."

[51] M. Karegar, A. Isazadeh, F. Fartash, T. Saderi, and A. H. Navin, "Data-Mining by Probability-Based Patterns."

[52] N. Jothi, N. A. Rashid, and W. Husain, "Data Mining in Healthcare - A Review," in Procedia Computer Science, Elsevier B.V., 2015, pp. 306–313. doi: 10.1016/j.procs.2015.12.145.

[53] T.-P. Hong, K.-Y. Lin, and S.-L. Wang, "Fuzzy data mining for interesting generalized association rules," 2003. [Online]. Available: www.elsevier.com/locate/fss

[54] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Communicated by Erkki Oja Canonical Correlation Analysis: An Overview with Application to Learning Methods."

[55] M. Chau, R. Cheng, B. Kao, and J. Ng, "Uncertain Data Mining: An Example in Clustering Location Data," 2006.

[56] Z. Wu and C.-H. Li, "L0-Constrained Regression for Data Mining."

[57] A. Genkin, D. D. Lewis, and D. Madigan, "Large-scale bayesian logistic regression for text categorization," Technometrics, vol. 49, no. 3, pp. 291–304, Aug. 2007, doi: 10.1198/004017007000000245.

[58] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases) (© AAAI)," 1996. [Online]. Available: www.ffly.com/

[59] R. Veloso et al., "A Clustering Approach for Predicting Readmissions in Intensive Medicine," Procedia Technology, vol. 16, pp. 1307–1316, 2014, doi: 10.1016/j.protcy.2014.10.147.

[60] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases) (© AAAI)," 1996. [Online]. Available: www.ffly.com/

[61] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases) (© AAAI)," 1996. [Online]. Available: www.ffly.com/

[62] R. Armañanzas, C. Bielza, K. R. Chaudhuri, P. Martinez-Martin, and P. Larrañaga, "Unveiling relevant non-motor Parkinson's disease severity symptoms using a machine learning approach," Artif Intell Med, vol. 58, no. 3, pp. 195–202, Jul. 2013, doi: 10.1016/j.artmed.2013.04.002.

[63] C. H. Jen, C. C. Wang, B. C. Jiang, Y. H. Chu, and M. S. Chen, "Application of classification techniques on development an early-warning system for chronic illnesses," Expert Syst Appl, vol. 39, no. 10, pp. 8852–8858, Aug. 2012, doi: 10.1016/j.eswa.2012.02.004.

[64] H. M. Zolbanin, D. Delen, and A. Hassan Zadeh, "Predicting overall survivability in comorbidity of cancers: A data mining approach," Decis Support Syst, vol. 74, pp. 150–161, Apr. 2015, doi: 10.1016/j.dss.2015.04.003.

[65] K. J. Wang, B. Makond, and K. M. Wang, "An improved survivability prognosis of breast cancer by using sampling and feature selection technique to solve imbalanced patient classification data," BMC Med Inform Decis Mak, vol. 13, no. 1, 2013, doi: 10.1186/1472-6947-13-124.

[66] N. Sharma and H. Om, "Data mining models for predicting oral cancer survivability," Netw Model Anal Health Inform Bioinform, vol. 2, no. 4, pp. 285–295, 2013, doi: 10.1007/s13721-013-0045-7.

[67] S. W. Fei, "Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine," Expert Syst Appl, vol. 37, no. 10, pp. 6748–6752, 2010, doi: 10.1016/j.eswa.2010.02.126.

[68] W. C. Yeh, W. W. Chang, and Y. Y. Chung, "A new hybrid approach for mining breast cancer pattern using discrete particle swarm optimization and statistical method," Expert Syst Appl, vol. 36, no. 4, pp. 8204–8211, May 2009, doi: 10.1016/j.eswa.2008.10.004.

[69] P. J. García-Laencina, P. H. Abreu, M. H. Abreu, and N. Afonoso, "Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values," Comput Biol Med, vol. 59, pp. 125–133, Apr. 2015, doi: 10.1016/j.compbiomed.2015.02.006.

[70] S. Şahan, K. Polat, H. Kodaz, and S. Güneş, "A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis," Comput Biol Med, vol. 37, no. 3, pp. 415–423, Mar. 2007, doi: 10.1016/j.compbiomed.2006.05.003.

[71] S. C. Bagui, S. Bagui, K. Pal, and N. R. Pal, "Breast cancer detection using rank nearest neighbor classiÿcation rules," 2003. [Online]. Available: www.elsevier.com/locate/patcog

[72] H. Mamiya, K. Schwartzman, A. Verma, C. Jauvin, M. Behr, and D. Buckeridge, "Towards probabilistic decision support in public health practice: Predicting recent transmission of tuberculosis from patient attributes," J Biomed Inform, vol. 53, pp. 237–242, Feb. 2015, doi: 10.1016/j.jbi.2014.11.006.

[73] S. Şahan, K. Polat, H. Kodaz, and S. Güneş, "A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis," Comput Biol Med, vol. 37, no. 3, pp. 415–423, Mar. 2007, doi: 10.1016/j.compbiomed.2006.05.003.

[74] S. C. Bagui, S. Bagui, K. Pal, and N. R. Pal, "Breast cancer detection using rank nearest neighbor classiÿcation rules," 2003. [Online]. Available: www.elsevier.com/locate/patcog

[75] R. Armañanzas, C. Bielza, K. R. Chaudhuri, P. Martinez-Martin, and P. Larrañaga, "Unveiling relevant non-motor Parkinson's disease severity symptoms using a machine learning approach," Artif Intell Med, vol. 58, no. 3, pp. 195–202, Jul. 2013, doi: 10.1016/j.artmed.2013.04.002.

[76] S. Bandyopadhyay et al., "Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data," Data Min Knowl Discov, vol. 29, no. 4, pp. 1033–1069, Jul. 2015, doi: 10.1007/s10618-014-0386-6.

[77] P. J. García-Laencina, P. H. Abreu, M. H. Abreu, and N. Afonoso, "Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values," Comput Biol Med, vol. 59, pp. 125–133, Apr. 2015, doi: 10.1016/j.compbiomed.2015.02.006.

[78] S. Kang, P. Kang, T. Ko, S. Cho, S. J. Rhee, and K. S. Yu, "An efficient and effective ensemble of support vector machines for anti-diabetic drug failure prediction," Expert Syst Appl, vol. 42, no. 9, pp. 4265–4273, Jun. 2015, doi: 10.1016/j.eswa.2015.01.042.

[79] C. T. Su, P. C. Wang, Y. C. Chen, and L. F. Chen, "Data mining techniques for assisting the diagnosis of pressure ulcer development in surgical patients," J Med Syst, vol. 36, no. 4, pp. 2387–2399, 2012, doi: 10.1007/s10916-011-9706-1.

[80] “Creating Knowledge-Based Healthcare Organizations.”

[81] N. G. Weiskopf and C. Weng, “Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research,” Journal of the American Medical Informatics Association, vol. 20, no. 1, pp. 144–151, 2013, doi: 10.1136/amiajnl-2011-000681.

[82] A. F. Simpao, L. M. Ahumada, J. A. Gálvez, and M. A. Rehman, “A review of analytics and clinical informatics in health care,” Journal of Medical Systems, vol. 38, no. 4. Springer New York LLC, 2014. doi: 10.1007/s10916-014-0045-x.

[83] A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, “Big data analytics in healthcare,” Biomed Res Int, vol. 2015, 2015, doi: 10.1155/2015/370194.

[84] “4”.

[85] E. Vayena, A. Blasimme, and I. G. Cohen, “Machine learning in medicine: Addressing ethical challenges,” PLoS Med, vol. 15, no. 11, Nov. 2018, doi: 10.1371/journal.pmed.1002689.

[86] T. B. Murdoch and A. S. Detsky, “The Inevitable Application of Big Data to Health Care.” [Online]. Available: http://clarabridge.com/default.aspx?tabid=137&ModuleID=635

[87] A. S. Ahuja, “The impact of artificial intelligence in medicine on the future role of the physician,” PeerJ, vol. 2019, no. 10, 2019, doi: 10.7717/peerj.7702.

[88] N. Ahmed, M. G. Bizanjo, A. Khan, S. Gull, S. khaliq, and N. Uddin, “Analysis of Textual Feedback of Students for Course Evaluation in Universities Through Machine Learning Algorithms”, IJAIMS, vol. 2, no. 2, pp. 31-44, May 2024.

[89] N.Ahmed, “A review of existing Machine Translation Approaches, their Challenges and Evaluation Metrics ”, PJETS, vol. 11, no. 1, pp. 29-44, Dec. 2023..

[90] A.khan, N.Ahmed, “Student textual feedback sentiment analysis using machine learning techniques to improve the quality of education”, PJETS, vol. 11, no. 2, pp. 32-40, Dec. 2023..