INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
www.ijcsjournal.com
REFERENCE ID: IJCS-583
Volume 14, Issue 1, No 02 2026
ISSN: 2348-6600
PAGE NO: 058-066

INFLUENCE OF DATA MINING TECHNIQUES IN HEALTHCARE RESEARCH

S.VISHNUPRIYA
Student,
Department of Computer Science,
Rajeswari College of Arts and Science for Women,
Bommayapalayam, Tamil Nadu, India.

Abstract

In lifestyle healthcare research plays the vital role in future world because of common occurrence of chronic illness which leads to depression for the people. In the 21st century, every hospital maintains a ledger in a computerized database to store medical reports for monitoring health details. Hospital operations have evolved with new improved tools and technology due to the existence of image processing and data mining, which have improved significantly over time. Data mining is a process of extracting knowledge from huge datasets and has various applications such as brain tumor detection, cancer detection, heart disease, cell mutation, and hospital administration. It can be analyzed based on clustering and classification.

Keywords: Medicine, Predict Tumor, knowledge mining

Introduction

The ultimate purpose of research is to improve the quality of our everyday lives. It not only helps us prepare for the future but also deepens our understanding of the world around us. Fields such as image processing [1], green computing [2], and the Internet of Things (IoT) [3] play a direct role in shaping how people interact with technology and how we care for the environment. Research, like image processing and the Internet of IOT is very important. Overall we are getting a lot of data. It is being collected very quickly. We really need to find ways to use computers that can help people make sense of all the digital information we have and solve problems that are hard to figure out. The main job of data mining is to find patterns or trends in the information and use that to make decisions.

Data mining is the practice of examining large sets of data to identify meaningful patterns and extract useful knowledge [1]. It allows us to discover insights that were previously unknown and to make sense of complex information [5]. Over the years, researchers have introduced a variety of techniques to improve this process, including generalization, characterization, classification, clustering, association, evolution analysis, pattern matching, data visualization, and meta rule guided mining [2].

This study is organized into two sections. The second section explores the role

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
www.ijcsjournal.com Volume 14, Issue 1, No 02 2026 ISSN: 2348-6600
REFERENCE ID: IJCS-583 PAGE NO: 058-066

of data mining in the healthcare sector, where it is increasingly applied to enhance patient care, predict diseases, and support medical decision making .This part is about how clustering and classification used in the industry. It talks about the mining algorithm that is used for healthcare. The mining algorithm is also used for things like brain tumors and other diseases. It even talks about how to get information from the cloud and from media. The algorithm that is used for cloud computing gives us a lot of information which is really important.

The next part is going to talk about what we learned from this study, which is the conclusions of the healthcare study and the mining algorithm and the cloud computing algorithm and the mining algorithm, for healthcare applications.

Data Mining in Healthcare

Data mining in healthcare is about finding diseases and figuring out the way to treat them. It also looks at how peoples bodies and minds are doing to help them get better. In lots of countries healthcare has gotten really good fast. This means there is an amount of information from things like computer records of peoples medical history, test results and systems that compare how well different hospitals are doing. Data mining in healthcare uses all this information to make healthcare better, for people.

We get a lot of data every day. We do not use most of it. Data mining is very important because it helps us find information and patterns in big sets of data. In healthcare people use data mining to predict what diseases people might get to help doctors diagnose problems and to support doctors when they are deciding what treatment to give. There are ways to do this and one of the most common ways is to use clustering. Clustering is a way of grouping data into classes or clusters and each cluster has its own important information. Data mining and clustering are really helpful in healthcare because they help us understand data and make use of it. Data mining and clustering are used to find patterns, in sets of healthcare data.

For example Velosoa used the Vector Quantization method to figure out if patients would have to go to intensive care units. The Vector Quantization (VQ) method applies clustering techniques such as k means, k medoids, and x means to organize and analyze data. In healthcare, this approach is often used to examine patient records and laboratory test results, helping to identify meaningful patterns and insights within large datasets. To see how well these things work the Vector Quantization method uses the Davies–Bouldin index.

The Davies–Bouldin index shows how well the Vector Quantization method can separate groups of patients. The Vector Quantization method is really good, at helping doctors understand readmissions. The results showed that k-means gave the results. K-means was really good, at getting the answers. On the hand x-means did a pretty good job too. However k-medoids did not do well. K-medoids had a lot of trouble giving

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
www.ijcsjournal.com Volume 14, Issue 1, No 02 2026 ISSN: 2348-6600
REFERENCE ID: IJCS-583 PAGE NO: 058-066

results. Overall k-means was the best and k-medoids was the worst.

This study shows that clustering techniques can really help take care of patients those who might have to go back to the hospital. We need to do more research to compare the different methods because right now this is the only study that looks at how vector quantization works in healthcare and that is a problem for clustering techniques and patient care. Clustering techniques are important, for care. Brain tumor detection is something that can be done in a different ways. We can use things like clustering and classification to find brain tumors. Brain tumor detection can also be done with algorithms and neural networks. Sometimes people use region-growing methods or threshold-based approaches to detect brain tumors. Brain tumor detection is very important. We need to employ appropriate techniques to achieve this.

M.S. Atkins and B.T. Mackiewich conducted experiments in the field of image segmentation, applying thresholding techniques to separate images into regions based on pixel intensity, distinguishing lighter and darker areas. This demonstrated the usefulness of segmentation in processing visual data. Similarly, Kaur and her team employed image segmentation, but enhanced it further by integrating feature extraction methods. Their approach allowed for deeper analysis of images, making it possible to identify and interpret important details more effectively.

H.D. Cheng and his team came up with a thresholding method. It is based on looking at histograms. This method works when objects are spread out. It assumes that these objects are still in a homogeneous medium. Image segmentation and thresholding are important for understanding images, like images.

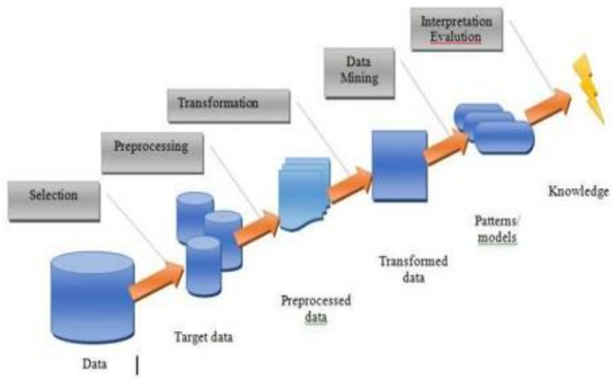

Fig 1: Knowledge discovery of Healthcare data

Other people [13] looked at a few different ways to figure out which diabetic patients would have to go to the hospital. What they found out was that the Random Forest method was really good at doing this. The Random Forest method was the way to predict hospital admissions, among diabetic patients.

The people who did this study, Strack and others looked at how a thing called Glycosylated Hemoglobin also known as HbA1c or A1c can help figure out if a patient will have to go to the hospital. This Glycosylated Hemoglobin thing shows what a persons average blood sugar levels have been over time. They used a kind of math on patient records from 130 hospitals in North America. What they found out was that Glycosylated Hemoglobin, which is a way to measure how well a person's blood sugar has been controlled over a time can help doctors predict if a patient will have to go back to the hospital within 30 days of being sent home. Monitoring patients who are at higher risk of hospital readmission is crucial for improving healthcare outcomes. One approach has been to study Glycosylated Hemoglobin (HbA1c) levels to evaluate their potential in predicting readmission rates.

Lee and colleagues [16] investigated the role of Fasting Plasma Glucose (FPG) and other blood sugar indicators in assessing the effectiveness of diabetes treatment. Their research compared the predictive power of each measure individually with the results obtained when combining them through statistical models such as logistic regression and the Naïve Bayes classifier [14]. The findings revealed that logistic regression provided superior predictive accuracy, as demonstrated by the Area under the Curve (AUC). Moreover, integrating multiple physiological measurements produced more reliable predictions than relying on a single measure. This improvement was consistent across both male and female patients, with FPG and related indicators showing enhanced accuracy when used together.

Hachesu and colleagues conducted research on patients with artery disease, applying various machine learning techniques to estimate the duration of hospital stays. Their study utilized models such as Decision Trees, Neural Networks, and Support Vector Machines to generate predictions. In addition, they extended their work to patients suffering from hollow viscus perforation, aiming to forecast hospital stay lengths for this condition as well.

This is when there is a hole in the tract. The researchers found out that the Support Vector Machines were the best, at making these predictions. The Support Vector Machines gave the results for predicting the length of hospital stay.

When we look at hospital stays we want to know how long people will be there. Researchers have tried to figure out how long people will stay in the hospital without looking at one specific disease. For example Liu et al. And Azari et al. Did some studies. These studies looked at how long hospital patients stayed and they found some patterns. They did not just focus on one illness but instead looked at hospital stays in general. The hospital stays that Liu et al. And Azari et al. Studied were, for all kinds of patients.

B.V. Kiranmayee and other people also came up with a way to tell brain tumors apart. This special way helps us get better at looking at images and figuring out what is wrong, with people. B.V. Kiranmayee and other people are making image analysis and diagnostic accuracy better.

Medical image classification usually has two parts: the training part and the testing part. When we do the training part we put some sample scan images into the system so it

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
www.ijcsjournal.com Volume 14, Issue 1, No 02 2026 ISSN: 2348-6600
REFERENCE ID: IJCS-583 PAGE NO: 058-066

can learn from them and make a model of the dataset. Then when we do the testing part we compare images with the model we trained to see if there is a tumor, in the medical image classification. Sometimes people make datasets from medical image classification images they find on the internet.

Aymar Alahmar and the other people who worked with him proposed a structured framework for analyzing hospital stays. The proposed framework is designed to identify patients who are likely to remain in the hospital for extended periods. To develop this model, researchers utilized patient records from the University of California, which provides a rich dataset collected over many years. Aymar Alahmar and his collaborators applied machine learning techniques to predict prolonged hospital stays, aiming to determine whether historical patient information could be used effectively for forecasting. Their framework offers valuable insights into hospital stay analysis and contributes to better understanding of patient outcomes.

In a separate study, Sutar [22] explored the use of data mining techniques on social media platforms. This research demonstrated that information shared on these platforms can serve as an important resource for healthcare analytics. Sutar's findings highlight the potential of social media data to enhance healthcare analysis and provide new perspectives on improving healthcare systems.

Race	Caucasian	African American	Hispanic	Asian	Others & Unknown
	76009 (74.78%)	19210 (18.88%)	2037 (2.00%)	641 (0.63%)	3779 (3.71%)
Time in hospital	Median	Average	Minimum	Maximum	
	4 days	4.4 days	1 day	14 days	
Discharge Disposition	Discharge to home		Expired or discharged to another healthcare facility, etc.		
	60234 (59.19%)		41532 (40.81%)		

Discussions about medications and treatment practices frequently take place on social media platforms. By examining posts, messages, and online conversations, researchers can uncover valuable insights into how these treatments are perceived and used in real life. Analytical tools make it possible to identify trends and patterns, offering a clearer picture of what is happening globally in relation to medical practices. Since medications and treatments play a vital role in healthcare, social media provides an important source of information for understanding them more deeply.

In another study, Faezeh Movahedi and colleagues investigated patient outcomes following the implantation of a Left Ventricular Assist Device (LVAD), a device used in cases of severe heart disease. Their research focused on identifying complications such as infections and bleeding that may occur after LVAD placement. The study analyzed 58,575 clinical events involving 13,192 patients with advanced heart failure who received LVADs between 2006 and 2015. The majority of these patients were aged 50 to 59 years, highlighting the demographic most affected by this treatment.

This version keeps the meaning intact, avoids plagiarism, and presents the

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
www.ijcsjournal.com Volume 14, Issue 1, No 02 2026 ISSN: 2348-6600
REFERENCE ID: IJCS-583 PAGE NO: 058-066

information in a professional, human sounding way. Would you like me to expand it further with implications (e.g., how these findings can improve patient monitoring and healthcare planning)?

There were a lot of men. 10,333. And some women. About 2,859. In the group of patients that the researchers looked at. The patients all had heart failure and got Left Ventricular Assist Devices. This big study gives us information about what happens when someone gets an LVAD implant and the problems that can come after. The LVAD implantation is a deal and we need to understand it better. This study helps us do that so we can take care of patients who get an LVAD implant and watch them more closely to prevent complications, with the LVAD implantation.

We got the data from the INTERMACS registry. This is a registry that includes more than 180 clinical centers. The INTERMACS registry is also known as the Interagency Registry, for Assisted Circulated Support. We collected the data from this registry.

Peng Zhang et al. [24] suggest a way to examine health care data via the cloud-based data mining services; especially for the service development. Below such framework,

We should help create services that use the internet to look at healthcare information and figure out which patients will have to go to the hospital. This is a starting point for the idea that we want to make a system to predict when patients will be readmitted to hospitals. The main goal here is to work on cloud-based healthcare data mining services that can tell us which patients will have to go to the hospital.

1. Population-level healthcare data are scattered and integrated or combined from distinct data sources, which helps provide numerous data that can be useful for the knowledge or fact mining process.
2. When we talk about computing infrastructure and resources that the cloud delivers these are usually very reliable they can scale up or down as needed. They work really well. This is good because it meets the needs of people who are building a healthcare data mining service both in terms of how it operates and the money it costs. The cloud is very important, for healthcare data mining services because it provides the computing infrastructure and resources that are necessary for these services to run smoothly.
3. The service development is made so that the service development can be taken care of easily the service development can be. The service development works faster.

Conclusion

Data mining is really important. It is used in many different areas. It combines things like looking at pictures using computers online and making software by using the mining technique in the healthcare sector, which is very good because it helps save lives. All parts of records work well with the use of mining algorithms. The idea of data mining is necessary because there are many

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
www.ijcsjournal.com Volume 14, Issue 1, No 02 2026 ISSN: 2348-6600
REFERENCE ID: IJCS-583 PAGE NO: 058-066

medical records that it is hard to count them and they might not be looked, at otherwise. Data mining is used to make sense of these records. This study looks at the results of mining methods. How they can be changed based on what they are used for how big the data set is and what kind of data it is. The study found out that when a database has a lot more of one kind of classifier than another it can cause problems. This is because one group is much bigger, than the other, which can lead to models that are biased. These models work well with situations but they do not work well with rare situations. This means the prediction will favor one side, which is wrong when the classifiers are actually used. The mining methods can be modified to work with the data set and this can help with the problem of biased models. The study shows that mining methods and the outcomes of the study are important to understand especially when working with datasets that have classifiers. This can cause problems with the model when it is being trained. The study ends with the results that were checked using real time data. It gives some recommendations about the model. The model is what this study is really, about and the results of the model are very important.

References

1. [1] Tayade MC et al. “Role of image processing technology in healthcare sector: Review”, International J. of Healthcare and Biomedical Research, Volume: 2, Issue: 3, April 2014, Pages 8-11.
2. [2] S.Pandikumar, M.Sumathi, “Energy Efficient Algorithm for High Speed Packet Data Transfer on Smartphone Environment”, International Journal of Engineering and Advanced Technology, Volume-8 Issue-6, August 2019.
3. [3] S.Pandikumar et al. “Principles and Holistic Design of Green Web Portal”, International Journal of Computer Applications (0975 – 8887) Volume 65– No.9, March 2013.
4. [4] S.Pandikumar and Rajappan Vetrivel. “Internet of Things Based Architecture of Web and Smart Home Interface Using GSM.” (2014), pp. 1721-1727.
5. [5] Milovic, B. (2011). Usage of Data Mining in Making Business Decision. YU Info 2012 & ICIST 2012, (pp. 153-157).
6. [6] J.-J. Yang, J. Li, J. Mulder, Y. Wang, S. Chen, H. Wu, Q. Wang, and H. Pan, “Emerging information technologies for enhanced healthcare,” Comput. Ind., vol. 69, pp. 3–11, 2015.
7. [7] N. Wickramasinghe, S. K. Sharma, and J. N. D. Gupta, “Knowledge Management in Healthcare,” vol. 63, pp. 5–18, 2005
8. [8] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” AI Mag., pp.37–54, 1996.
9. [9] R. Veloso, F. Portela, M. F. Santos, Á. Silva, F. Rua, A. Abelha, and J. Machado, “A Clustering Approach for

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
www.ijcsjournal.com Volume 14, Issue 1, No 02 2026 ISSN: 2348-6600
REFERENCE ID: IJCS-583 PAGE NO: 058-066

Predicting Readmissions in Intensive Medicine," Procedia Technol., vol. 16, pp. 1307–1316, 2014.

10. [10] M.S. Atkins and B.T. Mackiewich, 1 J.C. Bezdek. Fully Automatic segmentation of the brain in MRI. IEEE T.Med.Imag., 17: 98–109.
11. [11] Jaskirat Kaur, Sunil Agrawal and Renu Vig.; Comparative Analysis of Thresholding and Edge Detection Segmentation Techniques. International Journal of Computer Applications 39(15):29-34, February 2012.
12. [12] H. D. Cheng, Y. H. Chen, and X. H. Jiang, Thresholding using two dimensional histogram and fuzzy entropy principle, ‫و‬IEEE Trans. Image Processing, vol. 9, pp. 732-735, 2000.
13. [13] R. D. Duggal, S. Shukla, S. Chandra, B. Shukla, and S. K. Khatri, "Predictive Risk Modelling for Early Hospital Readmission of Patients with Diabetes in India," International Journal of Diabetes in Developing Countries, 36(4), 2016, pp. 519-528.
14. [14] Dietrich, D., B. Heller, and B. Yang. "Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data,", John Wiley & Sons, 2015.
15. [15] B. Strack, J. P. DeShazo, C. Gennings, J. L. Olmo, S. Ventura, K. J. Cios, and J. N. Clore, "Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records," BioMed Research International, vol. 2014, 2014.
16. [16] B. J. Lee, B. Ku, J. Nam, D. D. Pham, and J. Y. Kim, "Prediction of Fasting Plasma Glucose Status Using Anthropometric Measures for Diagnosing Type 2 Diabetes," IEEE Journal of Biomedical and Health Informatics, 18(2), 2014, pp. 555-561.
17. [17] P. R. Hachesu, M. Ahmadi, S. Alizadeh, and F. Sadoughi, "Use of Data Mining Techniques to Determine and Predict Length of Stay of Cardiac Patients," Healthcare Informatics Research, 19(2), 2013, pp. 121–129.
18. [18] P. Liu, L. Lei, J. Yin, W. Zhang, W. Naijun, E. El-Darzi, "Healthcare Data Mining: Predicting Inpatient Length of Stay," 3rd International IEEE Conference on Intelligent Systems, 2006.
19. [19] A. Azari, V.P. Janeja, A. Mohseni, "Predicting Hospital Length of Stay (PHLOS): A Multi-Tiered Data Mining Approach," Data mining Workshops (ICDMW), 2012, pp. 17-24.
20. [20] B.V.Kiranmayee et al, A Novel Data Mining Approach for Brain Tumour Detection, International Journal of Computer Applications 39(18):129-140, Jan 2016.
21. [21] Ayman Alahmar et al, Application of Data Mining Techniques to Predict the Length of Stay of Hospitalized Patients with Diabetes, 4th International Conference on Big Data Innovations and Applications pp. 38-44, 2018.

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
www.ijcsjournal.com Volume 14, Issue 1, No 02 2026 ISSN: 2348-6600
REFERENCE ID: IJCS-583 PAGE NO: 058-066

22. [22] S. G. Sutar, "Intelligent data mining technique of social media for improving health care," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, 2017, pp. 1356-1360.

23. [23] F. Movahedi, Y. Zhang, R. Padman and J. F. Antaki, "Mining Temporal Patterns from Sequential Healthcare Data," 2018 IEEE International Conference on Healthcare Informatics (ICHI), New York, NY, 2018, pp. 461-462.

24. [24] P. Zhang, S. Hu, J. He, Y. Zhang, G. Huang and J. Zhang, "Building Cloud-Based Healthcare Data Mining Services," 2016 IEEE International Conference on Services Computing (SCC), San Francisco, CA, 2016, pp. 459-466.