



ISSN: 0975-833X

RESEARCH ARTICLE

A SURVEY ON MEDICAL DATA MINING - HEALTHCARE RELATED RESEARCH AND CHALLENGES

***Dr. Mohan Kumar, S. and Jaya, R.**

Department of Computer Science and Engineering, New Horizon College of Engineering,
Bangalore, Karnataka, India

ARTICLE INFO

Article History:

Received 09th October, 2015
Received in revised form
21st November, 2015
Accepted 15th December, 2015
Published online 31st January, 2016

Key words:

Data Mining,
DM Tools,
HealthCare Software,
DM Techniques,
Classification,
Clustering,
Decision Tree,
DM Challenges.

ABSTRACT

Data mining techniques have higher usefulness in medical data mining as there is voluminous data in this industry. General survey on Data Mining Techniques, Methods, Tools and Challenges of Data Mining in health care domain is at most important and so much in demand, applying machine learning concepts and techniques in medical field is also most essential in the present scenario. The challenges for Data Mining in the healthcare domain are also listed. This paper describes the healthcare related research work with diseases and data mining techniques applied- summary, healthcare related researchers and percentage of accuracy achieved through applied through DM. Through this paper it is possible to recognize the importance of Data Mining in general and especially in the healthcare industry.

Copyright © 2016 Mohan Kumar and Jaya. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Mohan Kumar, S. and Jaya, R., 2016. "A survey on medical data mining - healthcare related research and challenges", *International Journal of Current Research*, 8, (01), 25170-25173.

INTRODUCTION

The study of data mining, especially in the field of healthcare will be beneficial to the society. DM Techniques and how much the DM technique contributed in the healthcare field the presented in this chapter. The following table summarizes the healthcare related researchers, year of research, DM techniques applied, and Diseases considered to conduct the research (Shubpreet Kaur and Bawa, 2015; Monali Dey *et al.*, 2014; Dhanya *et al.*, 2015; Mohan Kumar *et al.*, 2015; Jaya *et al.*, 2015). The following table summarizes the healthcare related research works completed and also with accuracy achieved details through the effective utilization of DM Techniques (Beant Kaur and Williamjeet Singh, 2014; Rajalakshmi and Dhenakaran, 2015; Shinde Swati and Amrit Priyadarshi, 2015). Through the Table no 1 and 2 it is possible to realize the usefulness and competency of Data mining Techniques, especially in the field of healthcare, so many techniques and for different healthcare fields were considered for the research and the results achieved are measurable, to bring more

accuracy the combination of these DM techniques for all the mentioned fields can be applied and also it can be fine tuned further.

Challenges of data mining in healthcare

- DM can be limited by the accessibility to data that often is distributed in different settings.
- Data may be in complete, computed, noisy, or inconsistent.
- Ethical, legal and social issues
- DM of medical data requires specific medical knowledge as well as knowledge of DM technology.
- Without quality data there is no useful results.
- For successful Data Mining, complication in medical data. It is essential to maintain the quality and accuracy data for data mining to making effective decisions.
- Another difficulty with health care data is data sharing. Health care organizations are unwilling to share their data due to privacy concerns.
- Most of the patients do not want to disclose their health data.
- The health maintenance organization and health insurance organizations are not distributing their data for preserving the privacy of patient.

***Corresponding author: Dr. Mohan Kumar, S.**

Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India.

Table 1. Healthcare Related Research Work With Diseases and Data Mining Techniques

Authors	Year Of Publication	Techniques	Disease
Markus B rameier <i>et al</i>	2001	Genetic Algorithm	Diabetic Diseases
Asha Gowda <i>et al</i>	2009	Classification Algorithm	Cardio Vascular Diseases
Jenn -Lung <i>et al</i> & Weimin Xue <i>et al</i>	2001	Bayesian Network algorithm	Coronary Heart Disease
Yanwei Xing <i>et al</i>	2007	SVM	Diabetes
Dursun Delen <i>et al</i>	2005	ANN, DTREE, Logisitic regression	Breast Cancer
Asha Rajkumar <i>et al</i>	2010	Decision Tree, KNN, NB	Heart Disease
Akhil jabbar <i>et al</i>	2012	ANN, DTREE, NB	Heart Disease
K R Lakshmi <i>et al</i>	2014	ANN, DTREE, Logisitic regression	Kidney DIALYSIS
Syeda Farha Shazmeen <i>et al</i>	2013	ANN, DTREE, KNN,NB,SVM	Liver Disorder
Ms. Shinde Swati <i>et al</i>	2015	Navie Bayes, KNN	Heart Disease
Umair Shafique <i>et al</i>	2015	DTree, Navie Bayes	Heart Disease
Durairaj M, Revathi V	2015	MLP Back Propagation NN	Heart Disease
Dr. Mohan Kumar S , <i>et al.</i>	2010	SVM, KNN, SNE, SSNE, T-SNE	Breast Cancer
	2011		
	2012		
	2013		
	2014		
	2015		

Table 2. Healthcare Related Research work with Diseases and Data Mining Techniques and Accuracy achieved

Authors	Year	Technique	Accuracy
Yan, <i>et al.</i>	2003	Multilayer Perceptron	63.60%
Andreeva, P.	2006	Naive Bayes	78.56%
		Decision Tree	75.74%
		Neural network	82.77%
		Kernel density	84.44%
Palaniappan, <i>et al.</i>	2007	Naive Bayes	95.00%
		Decision Trees	94.93%
		Neural Network	93.54%
De Beule, <i>et al.</i>	2007	Artificial neural network	82.00%
T antimongcolwata, <i>et al.</i>	2008	Direct kernel selforganizing map	80.40%
		Decision Trees	60.40%
Rajkumar, <i>et al.</i>	2010	Naive Bayes	52.33%
		KNN	45.67%
		Decision list	52.00%
Abdi <i>et al.</i>	2013	SVM	94.56%
		AR_MLP	97.28%
		AR_PSO-SVM	98.91%
K R Lakshmi <i>et al</i>	2014	Artificial neural network	93.85%
		Decision Trees	78.44%
Abhishek Taneja	2013	Artificial neural network	93.83%
		Decision Trees	94.29%
		Naive Bayes	91.96%
Asha Rajkumar <i>et al</i>	2010	Decision Trees	52.00%
		Naive Bayes	52.33%
		Decision Tree	93.62%
		Artificial Neural Network	91.21%
Anbarasi, <i>et al.</i>	2010	Genetic with Decision Tree	99.20%
		Genetic with Naive Bayes	96.50%
		Genetic with Classification via Clustering	88.30%
Fan <i>et al.</i>	2010	CHAID	69.75%
		C & RT	69.73%
		QUEST	67.25%
		C 5.0	71.17%
Ms. Shinde Swati <i>et al</i>	2015	Navie Bayes	84.00%
		KNN	76.00%
Umair Shafique <i>et al</i>	2015	DECISION TREE	77.22%
		Navie Bayes	82.91%
Durairaj M, Revathi V	2015	MLP Back Propagation NN	96.30%
Dr. Mohan Kumar S <i>et al.</i>	2015	SVM & KNN	95.67%
		SVM , KNN, SNE	97.99%

- The startup cost of data warehouse is very high.
- Another most significant challenges of the data mining in health care is to obtain the quality and relevant medical data.
- It is difficult to acquire the precise and complete healthcare data.
- Health care data is complex and heterogeneous, because it is collected from various sources such as from the medical

reports of laboratory, from the discussion with patient or from the review of physician and so on.

All the mentioned challenges should be considered to conduct the research related to Healthcare data mining.

Conclusion

In this paper, we presented a survey of various data mining techniques that have been employed for medical data mining. Due to the rapid growth of medical data, it has become in expendable to use data mining technique to help decision support in the field of healthcare. The medical mining capitulate required business intelligence to support well-versed diagnosis and decisions. This paper has provided the summary of data mining techniques used for medical data mining in the field of healthcare and also the diseases they classified. It also expresses the importance and challenges of data mining.

REFERENCES

- Beant Kaur, Williamjeet Singh, 2014. "Review on Heart Disease Prediction System using Data Mining Techniques", *International Journal on Recent and Innovation Trends in Computing and Communication* ISSN: 2321-8169 IJRITCC | October.
- Dhanalakshmi, D, Dr. J. Komala Lakshmi, 2014. "A Survey on Data Mining Research Trends", *International Journal Of Engineering And Computer Science*, ISSN:2319-7242 Volume 3, Issue 10 October, Page No. 8911-8919
- Dhanya, P. Varghese, Tintu, P. B. 2015. "A Survey On Health Data Using Data Mining Techniques", *International Research Journal of Engineering and Technology (IRJET)* e-ISSN: 2395-0056 Volume: 02 Issue: 07 | Oct.
- Divya Tomar and Sonali Agarwal, 2013. "A survey on Data Mining approaches for Healthcare", *International Journal of Bio-Science and Bio-Technology*, Vol.5, No.5, pp. 241-266
- Durairaj, M., Revathi, V., 2015. "Prediction Of Heart Disease Using Back Propagation MLP Algorithm", *International Journal Of Scientific & Technology Research* Volume 4, Issue 08, August, Issn 2277-8616
- Feixiang Huang, Shengyong Wang, and Chien-Chung Chan, 2012. "Predicting Disease By Using Data Mining Based on Healthcare Information System", *IEEE International Conference on Granular Computing*.
- Gayathri, V. et al. 2014. "A Survey Of Data Mining Techniques On Medical Diagnosis And Research", *Singaporean Journal of Scientific Research (SJSR)* Vol. 6.,No.6, Pp. 301-310
- Georgeena. T., S. Thomas, Siddhesh. S. Budhkar et al, 2015. "Heart Disease Diagnosis System Using Apriori Algorithm", Volume 5, Issue 2, February, ISSN: 2277 128X *International Journal of Advanced Research in Computer Science and Software Engineering*.
- Jaya, R. et al. 2015. "Ayurveda Medicine Roles in Healthcare Medicine, and Ayurveda Towards Ayurinformatics", *International Journal of Computer Science and Mobile Computing*, Volume 4 Issue 12 publishes on 5th Dec to 30th Dec 2015 ISSN 2320-088X, Paper ID: V4I12,12
- Kittipol Wisaeng, 2013. "An Empirical Comparison of Data Mining Techniques in Medical Databases", *International Journal of Computer Applications (0975 – 8887)* Volume 77– No.7, September.
- Mohammed Abdul Khaleel Sateesh Kumar Pradham G.N. Dash, 2013. "A Survey of Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases", Volume 3, Issue 8, August, ISSN: 2277 128X *International Journal of Advanced Research in Computer Science and Software Engineering*.
- Mohan Kumar, S. et al. 2015. "Classification of Breast Mass Classification – CAD System and Performance Evaluation Using SSNE", *IJSET – International Journal of Innovative Science, Engineering & Technology*, 417-425, ISSN 2348 – 7968 Vol. 2 Issue 9, September.
- Mohan Kumar, S, et al. 2015. "Classification of Breast Mass classification – CAD System with Performance Evaluation , International Journal of Engineering And Computer Science", *International Journal of Engineering And Computer Science*, 14187-14193, ISSN 2319-7242 Volume 4 Issue 09 September.
- Mohan Kumar, S., et al. 2012. "Breast Cancer Diagnostic system based on Discrete Wavelet Transformation and stochastic neighbour Embedding", *European Journal of Scientific Research*, Volume 87, Issue 03, October, Page Numbers: 301-310,ISSN:1450-216X
- Mohan Kumar, S., et al. 2013. "Categorization of Benign And Malignant Digital Mammograms Using Mass Classification – SNE and DWT", *Karpagam Journal of Computer Science*, Volume-07, Issue-04, June-July, Numbers: 237-243. ISSN No: 0973-2926,
- Mohan Kumar, S., et al. 2013. "Classification of Microcalcification in digital mammogram using SNE and KNN classifier", *International Journal of Computer Applications*, IJCA 03, ISSN: 0975 - 8887.
- Mohan Kumar, S., et al. 2013. "The Performance Evaluation of the Breast Mass classification CAD System Based on DWT, SNE AND SVM", *International Journal of Emerging Technology and Advanced Engineering*, Volume 3, Issue 10, October 2013, Page Numbers: 581-587, ISSN 2250–2459,
- Mohan Kumar, S., et al. 2013. "The Performance Evaluation of the Breast Micro calcification CAD System Based on DWT, SNE AND SVM", *CiiT International Journal of Digital Image Processing*, November, DOI: DIP112013005, ISSN 0974 – 9691
- Mohan Kumar, S., et al. 2015. "Classification of Breast Microcalcification- CAD System and Performance Evaluation Using SSNE" , *International Journal of Advanced Research in Computer Science and Software Engineering*, 824-830, Volume 5, Issue 9, September, ISSN: 2277 128X
- Mohan Kumar, S., et al., 2012. "Statistical Features Based Classification of Micro calcification in Digital Mammogram using Stochastic Neighbour Embedding", *International Journal of Advanced Information Science and Technology*, Volume 07, Issue 07 , November, Page Numbers: 20-26, ISSN:2319-2682
- Mohan Kumar, S., et al., 2013. "Classification of Micro Calcification And Categorization Of Breast Abnormalities - Benign and Malignant In Digital Mammograms Using SNE

- And DWT", *Karpagam Journal of Computer Science*, Volume-07, Issue-05, July-Aug, 2013- Page Numbers: 253 to 259, ISSN No: 0973-2926
- Monali Dey, *et al.*, 2014. "Study and Analysis of Data mining Algorithms for Healthcare Decision Support System", (*IJCSIT International Journal of Computer Science and Information Technologies*, Vol. 5 (1), 2014, 470-477
- Nidhi Bhatla Kiran Jyoti, 2012. "An Analysis of Heart Disease Prediction using Different Data Mining Techniques", *International Journal of Engineering Research & Technology (IJERT)* Vol. 1 Issue 8, October.
- Rajalakshmi, K., Dr. S. S. Dhenakaran, 2015. "Analysis of Datamining Prediction Techniques in Healthcare Management System", *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 5, Issue 4, April.
- Reetu1, Narender Kumar, 2015. "Licensed Under Creative Commons Attribution CC BY Medical Diagnosis for Liver Cancer using Classification Techniques", *International Journal of Science and Research (IJSR)* ISSN (Online): 2319-7064 Volume 4 Issue 5, May.
- Sheetal, L. Patil, 2015. "Survey of Data Mining Techniques in Healthcare" *International Research Journal of Innovative Engineering*, Volume1, Issue 9 of September.
- Shelly Gupta *et al.*, 2011. "Data Mining Classification Techniques Applied For Breast Cancer Diagnosis And Prognosis", *Indian Journal of Computer Science and Engineering (IJCSE)* ISSN : 0976-5166 Vol. 2 No. 2 Apr-May.
- Shinde Swati, B. Prof. Amrit Priyadarshi, 2015. "Decision Support System on Prediction of Heart Disease Using Data Mining Techniques", *International Journal of Engineering Research and General Science*, Volume 3, Issue 2, March-April, ISSN 2091-2730.
- Shubpreet Kaur and Dr. R.K.Bawa, 2015. "Future Trends of Data Mining in Predicting the Various Diseases in Medical Healthcare System", *International Journal of Energy, Information and Communications*, Vol.6, Issue 4, pp.17-34
- Umair Shafique, Fiaz Majeed, Haseeb Qaiser, and Irfan Ul Mustafa, 2015. "Data Mining in Healthcare for Heart Diseases", *International Journal of Innovation and Applied Studies* ISSN 2028-9324 Vol. 10 No. 4 Mar., pp. 1312-1322
