ISSN: 2249-6661
Vol-44, No.-03 January-March(2021)

IMPLEMENTATION OF KNOWLEDGE BASED SYSTEM FOR PREDICTION OF VIRAL INFECTIOUS DISEASES OUTBREAKS

DR. ASHWINI BRAHME [1], DR. SHIVAJI MUNDHE [2], PROF. SAGAR KULKARNI [3], PROF. MANASI KULKARNI [4]

[1] Assistant Professor, Sinhgad Institute of Management and Computer Application (SIMCA), Pune, Maharashtra, India,
[2]Director, Yashaswi Education Society's, International Institute of Management Science, Pune, Maharashtra, India
[3,4] Asst. Prof., Department of Computer Engineering, Mumbai University PCE, New Panvel, India
Email id: ashwiniak47@gmail.com, ashwini_kulkarni21@gmail.com

ABSTRACT

The ample amount of data available all over the globe is in structured, semi structured and unstructured format; it may be in the forms like emails, full text documents, html files, web data, audio, video, etc. The paper intended towards the study of viral infectious diseases, its frequency, seasons and diseases outbreaks prediction. The paper aimed towards the analysis of outbreaks prediction through manually and through the proposed knowledge based system.

Keywords: Data Mining, Knowledge Based System, Viral Diseases, Outbreaks, Prediction

INTRODUCTION

Healthcare domain plays very significant role in our day to day life. The various viral and bacterial diseases are spread through various types of viruses and bacteria's. The some of the common viral diseases are dengue, Chikungunya, malaria, swine flu, H1N1, Influenza Flu, Cholera, Rabies, etc. The viral and bacterial diseases had great impact on public health. [4] Therefore the present research will focus on various infectious diseases controls and so that the remedial measures can be considered. The different applications of text mining are covers diverse areas namely bioinformatics, business intelligence, national security, telecommunication, banking, Information Technology, media, insurance, political analysis, pharmaceutical, health care, etc.

STATEMENT OF THE PROBLEM

Knowledge Based System is valuable for medical domain as there are numerous challenges to mine the data and get the specified results for decision making. The research aimed towards the generating the patterns from enormous and diverse data and apply these patterns with respect to various parameters. These patterns, knowledge discovery will be significant for healthcare organizations, medical experts, doctors, professional and government can take remedy measures. Therefore, the researcher desired to attempt "**Design and Development of Knowledge Based System for Infectious Diseases using Advanced Data Mining Techniques**" for the present study.

ISSN: 2249-6661
Vol-44, No.-03 January-March(2021)

OBJECTIVES AND HYPOTHESIS OF THE STUDY

Objectives:

1. To study the frequently occurring viral infectious diseases.
2. To understand the impact of viral diseases.
3. To design and implement the knowledge based system for infectious diseases.
4. To compare the diseases outbreak prediction manually and through knowledge based system.

Hypothesis:

H_1: There is significant discrepancy between the predictions of viral diseases outbreaks manually and by knowledge based system.

H_0: There is insignificant discrepancy between the predictions of viral diseases outbreaks manually and by knowledge based system.

REVIEW OF LITERATURE

A research article entitled "**A survey on Data Mining approaches for Healthcare"** by Tomar and D. Agarwal S. (2013) describes about Knowledge Discovery from Database (KDD). The knowledge discovery is step by step process and includes various steps specifically data selection, pre-processing, and transformation.

The author K.L.Sumathy and M. Chidambaram focused on information retrieval feature of text mining in their research paper entitled **"Text Mining: Concepts, Applications, Tools and Issues – An Overview.** Test mining includes two major phases one is text refining and another is knowledge distillation. The paper also elaborates on diverse implications of text mining namely data mining and natural language processing, information extraction, information retrieval, applications of text mining like telecommunication, bank, IT, media, insurance, political analysis, pharmaceutical, healthcare, bioinformatics, business intelligence, national security, etc. [2]

The Data Mining techniques and WEKA tool is used for prediction of Dengue disease by the researcher in "**Dengue disease prediction using Weka data mining tool.** This research compares the algorithms specifically Naïve Bayes, J48, SMO, REP Tree and Random tree and predict the results of each algorithm for Dengue diseases prediction. [3]

Milovic B. and Milovic M. (2012) elaborates knowledge discovery from database(KDD) and various data mining methods such as link analysis , classification, clustering, association rule, text mining in their research article entitled "**Prediction and Decision Making in Health Care using Data mining**". The paper described on the diversity and heterogeneous large volume of data is the biggest challenges of test mining and information retrieval. [5]

DATA COLLECTION

Primary Data: The primary data is gathered using questionnaire, formal and informal interviews and discussion with the doctors and medical experts.

Secondary Data: The secondary data is collected from reports, thesis, research journals, articles, newspaper, periodicals,

ISSN: 2249-6661
Vol-44, No.-03 January-March(2021)

published and unpublished reports, websites, etc.

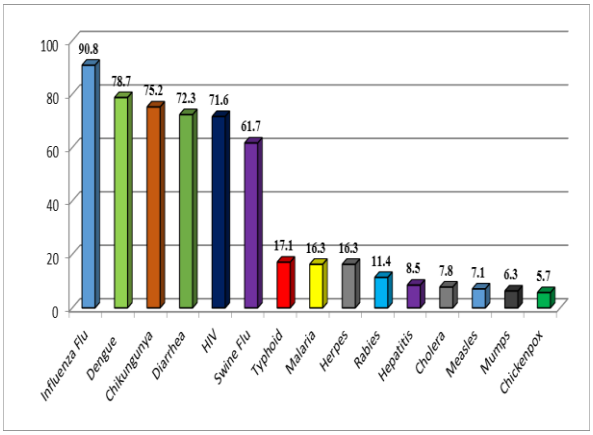
DATA ANALYSIS AND RESULT

This research is carried out on Filed survey in which the questionnaire is used for collecting the information from doctors and hospitals from Pune municipal corporation and Pimpri Chinchwad Municipal Corporation. Researcher has collected data from 140 respondents based on common viral diseases, their frequency, impact, locations, seasons, outbreaks of diseases, etc. The data is collected, formulated and analysed using SPSS statistical tools. Based on the analysis the result and Interpretation will be done.

- **Viral Diseases Frequency**

The following graph no.1 depicts the frequency of highly occurring diseases, it has resulted that, patients of Influenza Flu are 98.8 %, Dengue (78.7 %), Chikungunya (75.2 %), Diarrhea (72.3%), HIV (71.6 %), Swine Flu (61.7 %). The remaining disease specifically Typhoid, Malaria, Herpes, Rabies are in 10 to 20 % range and Hepatitis's, Cholera, Measles, Mumps and chickenpox are less than 10 %.

Hence it has been concluded that, there are maximum number of patients suffering from Influenza Flu, Dengue, Chikungunya, Diarrhea, HIV and Sine Flu.



Graph No.1 : Source: PMC Viral Infective Diseases details (2014-18)

- **Viral Infective Diseases Seasons, Outbreaks and Prediction**
  - It has resulted from the data analysis that maximum diseases (71 %) spread in rainy season as compare to 16% in winter and summer (13 %). Hence it has been concluded that in rainy season there are maximum patients suffers from common viral infections.
  - The study depicted that maximum respondents can't predict the outbreaks of viral diseases so there is need of system to for outbreak prediction.
  - Approximate 50 % of doctors predict the geographical location of highly spread viral infections.
  - 78 % of doctors suggested that there is need of system for diseases outbreak prediction so that the further decision making.

HYPOTHESIS TESTING

H_1: There is significant discrepancy between the predictions of viral diseases outbreaks manually and by knowledge based system.

H_0: There is insignificant discrepancy between the predictions of viral diseases outbreaks manually and by knowledge based system

To test the above hypothesis researcher had selected sample t-test (paired) to compare two means of prediction of diseases outbreaks manually and through the knowledge based system.

Table No. 1: Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Diseases outbreaks Prediction through Manually	1.6950	141	.86967	.07324
	Diseases outbreaks Prediction through Knowledge Based System	2.0496	141	.98797	.08320

(Source: Compiled by researcher using SPSS)

Table No.2 : Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Outbreak Prediction manually and through System	141	-.140	.097

(Source: Compiled by researcher using SPSS)

Table No.3: Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Diseases Outbreak Prediction Manually – Through knowledge based system	-.3546	1.40476	.11830	-.5885	-.1207	-2.997	140	.003

(Source: Compiled by researcher using SPSS)

The above table no. 1, 2, 3 signifies the paired sample statistics of difference between predictions of diseases outbreaks manually and using designed knowledge based system.

It has observed from table no.3 that the p-value = .003; this p-value is <0.05 at 95 % confidence level. If the p-value is < 0.05 then reject the null hypothesis and accept the alternate hypothesis. Therefore, the null hypothesis (H_0) is rejected and alternate hypothesis (H_1) is accepted; which means that significant discrepancy between the predictions of viral diseases outbreaks manually and by knowledge based system.

Therefore, it can be concluded that the designed and developed knowledge based system is significant for the prediction of viral infective diseases outbreaks.

CONCLUSION

The designed knowledge based system is useful for to get the required and necessary information of viral diseases from huge dataset of viral infective diseases and to succeed the better information retrieval. The healthcare experts, different stakeholders like doctors, practitioners, healthcare organizations take the help of this designed knowledge based system for diseases outbreak predictions. Thus, the investigator senses that the outcome of this study is going to be beneficial for hospitals, healthcare domain and related organizations, government, experts, doctors, social workers for enhanced protective actions, conclusions, educative arrangements to construct cognizance in the society.

References

[1] Tomar D., Agarwal S. (2013). A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology 5(5), 241-266. http://dx.doi.org/10.14257/ijbsbt.2013.5.5.25

[2] Sumathy K., Chidambaram M. (2013). Text Mining: Concepts, Applications, Tools and Issues – An Overview. International Journal of Computer Applications, 80(4)
[3] SHAKIL K., ANIS S., ALAM M. (2013). Dengue Disease Prediction Using Weka Data Mining Tool. Department of Computer Science, Jamia Millia Islamia New Delhi, India
[4] Kulkarni A., Mundhe S. (2018). Implementation of Effective Stemming Algorithm of Text Mining for Knowledge Discovery in Healthcare. Savitribai Phule Pune University
[5] Milovic B., Milovic M. (2012). Prediction and Decision Making in Health Care using Data Mining. International Journal of Public Health Science (IJPHS), 1(2), 69-78, ISSN: 2252-8806 _69. http://iaesjournal.com/online/index.php/IJPHS
[6] S. Vijayarani., J. Ilamathi, Nithya. Pre-processing Techniques for Text Mining - An Overview, International Journal of Computer Science & Communication Networks, 5(1), 7-16
[7] Bijan Raheemi. (2014). Data Mining and Knowledge Discovery in Healthcare and Medicine. Telfer School of Management and School of Electrical Engineering and Computer Science, University of Ottawa. Retrieved from : http://www.ieeeottawa.ca/aicn/data-mining-and-knowledge-discovery-in-healthcare-and-medicine/
[8] Mandal A. (2013). Human Diseases Caused by Viruses. http://www.news-medical.net/health/Human-Diseases-Caused-by-Viruses.aspx. (Retrieved on 22 Nov 2016)
[9] https://www.healthgrades.com/conditions/viral-diseases
[10] https://pmc.gov.in/en/health. (Retrieved on 6 April 2017)
[11] https://www.pcmcindia.gov.in/departments-details.php?Id=16. (Retrieved on 6 April 2017)
[12] Kulkarni A., Mundhe S. (2017). A study of Viral Diseases and their Impact on Public Health using Knowledge Based System. BIONANO FRONTIER, Print ISSN 0974-0678, Vol.10 , Issue -2, July – Dec 2017, Page No. 273
[13] Zakie M. (1998), Scalable Data Mining for Rules, University of Rochester, New York.https://pdfs.semanticscholar.org/9450/70ac0791162653219782cabbf0ba7aac8f88.pdf, Retrieved on 9 January 2018
[14] Grzymala-Busse J.W. (2009) Rule Induction. In: Maimon O., Rokach L. (eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA, https://doi.org/10.1007/978-0-387-09823-4_13
[15] Borgelt Christian (2012). Frequent item set mining. WIREs Data Mining Knowledge Discovery 2012, 2: 437-456, doi: 10.1002/widm.1074
[16] Christian Berget(2000). Data Mining with Graphical Models, http://www.borgelt.net/diss/dmwgm.pdf
[17] Jiawei Hon, Micheline Kamber, Jain Pei. (2012). Data Mining concepts and techniques, 3rd Edition
[18] http://www.zentut.com/data-mining/data-mining-techniques/ (Retrieved on 28 Nov 2016)