Predicting Risk factor of Dengue using FP-Growth Mining Technique

S.Vasanthanageswari[1], M. Vanitha[2]
Research Scholar, Department of Computer Applications, Alagappa University, Karaikudi, Tamil Nadu, India.
Assistant Professor, Department of Computer Applications, Alagappa University, Karaikudi, Tamil Nadu, India.
Email : smilingsudhamathi@gmail.com, mvanitharavi@gmail.com

Abstract— Dengue is a debilitating malady which is caused by female mosquitoes (chomp of Aedes mosquitoes). It is regularly found in hot areas. The dengue maladies mostly caused in 4 serotypes (DENV-1, DENV-2, DENV-3 and DENV-4). A dengue malady grasp from gentle febrile malady to serious hemorrhagic fever. Anticipating the connection between serotypes of dengue and age of the people will help the biotechnologists and bioinformaticians to advance one stage to find solutions for dengue. Information Mining is a champion among the most completely and animating zones of research with the inspiration driving finding critical data from tremendous information accumulations . In Medical endeavors, Data Mining gives numerous purposes of enthusiasm, for instance, the area of the coercion in medicinal scope, sickness gauge, and availability of the helpful answer for the patients at bring down cost, acknowledgment of purposes behind disorders and recognizing verification of helpful treatment strategies. It is furthermore supportive to predict the risky diseases like- Dengue fever, Cancer, Diabetes et cetera. In this Research work to reduce the death rate, the risk factors of the dengue are predicted using Association rule Mining.

Keywords— Data Mining, Association Rule Mining,Heart Disease, Dengue disease.

I. INTRODUCTION

Data Mining is a striking of the most extrusive and convincing zones of research by the whole of the desire of finding moving information from massive first page new sets[20, 24]. In nature Medicinal Informatics is for the most part clinical or potentially organic, and information driven factual research has turned into a typical supplement. Foreseeing the result of an illness is a standout amongst the most intriguing and testing assignments where to create information mining applications. As the utilization of PCs fueled with computerized instruments, expansive volumes of therapeutic information are being gathered and influenced accessible to the restorative research to gathering. Therefore, Information Revelation in Databases (KDD), which incorporates information mining procedures, has turned into a prevalent research instrument for restorative scientists [18]. As knowledge discovery process comprises of an iterative grouping of the accompanying step Data cleaning, Data integration, Data selection, Data transformation, Data mining, Pattern evaluation, Knowledge presentation[19, 4]. Now-a-days, dengue fever(DF) has developed as one of the basic worldwide general wellbeing concern, specifically in the tropical countries. Currently, it is a standout amongst the most all around spread bug conceived infection, giving ascend to 50 to 100 million cases annually, incorporating in excess of 100 endemic nations on the planet [5]. Dengue has no particular medication or anti-microbial accessible to treat it . Dengue fever happens in type of cycles and this cycle is available inside the body of a tainted individual for two week or under two weeks. It causes stomach torment, drain (dying), and circulatory crumple and Dengue hemorrhagic fever[3]. The principle objective of research work is to predict the general population who are influenced by dengue based on the symptoms utilizing the association rule mining.[2].In this research work, authors predict the risk factors of dengue disease affecting majority of the patients with the aid of Weka data mining tool.

II. LITERATURE SURVEY

M.V.Jagannatha Reddy1 *et. al.,* proposed on Expert System to Predict the Type of Fever. They created in this undertaking another master framework to foresee the dengue fever in beginning times. This approach comprises of three critical advances: a) manual missing quality attribution strategy is connected that makes the information

steady. b) A specialist assessment is taken for choosing most compelling qualities for dengue fever additionally we done web review . c) A neural system demonstrates is utilized for precise forecast of dengue fever [1].

P.Manivannan *et. al.,* anticipated the general population who are influenced by dengue relying on order of age utilizing the K-medoid grouping calculation, which has been executed [2]. Mansaf Alam *et. al.,* have characterized the dengue informational collection and afterward analyzed the distinctive information mining strategies in knowledge flow , weka through Explorer and Experimenter interfaces in the research paper [3]. *Anurag Bhatt et. al.,* gives an audit of the part of data mining in healthcare with a specific end goal to give the better comprehension of the subject. Data Mining Algorithms and Techniques are being utilized to identify and anticipate Sudden Heart failure, Diabetes and so on and other medical problems utilizing Naïve Bayes classifiers , Decision Trees and so forth [4].

Zuiraini mustaffa *et. al.,* has investigated the normalization techniques usages incorporated with the prediction models Neural Network (NNM) and LS-SVM. Then, the results are compared by means of mean squared error(MSE) and prediction accuracy. According to the result it reveals that LS-SVM is a best prediction model when compared to the NNM [5]. A. I. Z. Abidin , *et. al.,* implemented about the present usage of dengue flare-up control in Malaysia and anticipate dengue fever cases utilizing information mining procedures. Distinctive information mining order procedures are connected onto these information with the execution of every method is estimated. The results highlight the best execution among systems utilized [6].

Gaurav Gupta, *et. al. ,*analysed the demographic profiles for dengue fever get examined through statistic profiles [7]. Marimuthu Thangam, *et. al.,* proposed an algorithm called Recurrence Finder (RECFIN) Mining Association Rules in Dengue Gene Sequence with Latent Periodicity in the paper that uses the addition tree for identifying the intermittent examples of dengue quality succession. We exhibit the adequacy of the proposed approach by looking at the test comes about performed on dengue infection serotypes dataset with NCBI-BLAST calculation [8]. M.Bhavani, *et. al.,* utilized the classification techniques as a part of this examination are REP Tree, J48, SMO, ZeroR and Random Tree. The execution of characterization strategies were looked at by plotting charts and table. Weka the data mining tool is utilized for the order [9]. Shilpa Serasiya, *et. al.,* foresee infections like hepatitis, Lung malignancy liver issue, bosom tumor or heart illnesses, diabetes and so forth, it helps in finding the examples to choose future patterns in restorative field. We can use in finding the learning and expectation, identification of Dengue too [10].

Dr. Stephen Kimani *et. al.,* displays a review and examination for existing strategies on both grouping and relapse models methods that have been connected for illnesses flare-up expectation in datasets in the work [11]. Thitiprayoonwongse, *et. al.,* proposed the initial two test comes about demonstrate the valuable information to order dengue disease from Srinagarindra Hospital's dataset and Songklanagarind Hospital's dataset, respectively. Each arrangement of information is tried by various dataset to ensure that the test information was a genuine inconspicuous information. The third exploratory outcomes demonstrate the valuable information when we coordinated 2 datasets. Another target of this examination is to distinguish the day of defervescence of fever which is called day0. The day0 date is the basic date of dengue patients that a few patients confront the deadly condition. Subsequently the doctors need to anticipate day0 inorder to treat the patients. They hope to have a savvy framework that can trigger the day0 date of every patient [12].

Nandini, *et. al.,*   investigated recurrence which corresponds the event of dengue and the sign of its indications over the months. The framework produces obvious precision and fills in as a profitable device for therapeutic specialists[13]. Dengue guidelines for diagnosis, Treatment, Prevention and Control This model for characterizing dengue has been recommended by a specialist gathering (Geneva, Switzerland, 2008) and is right now being tried in 18 nations by looking at its execution in handy settings to the current WHO case order. The procedure will be finished in 2010.For useful reasons this guide adjusts the refinement amongst dengue and serious dengue [14]. Nayyer Masood, et.al.,   predicted Dengue Fever using techniques J48 and SMO Naïve Bayesian, REP Tree, Random tree WEKA was utilized as Data digging device for order of information. Initially we will assess the execution of the considerable number of methods independently with the assistance of tables

and charts relying on dataset and also we will think about the execution of the considerable number of procedures [15]. Y. Lisnichuk ,*et. al.,* Accordingly they proposes finding an expectation model of the quantity of dengue cases with up to a month of foresight for areas of Paraguay finding most compelling climatic factors [16].

Ganesan.M, et.al., predicted the dengue disease using Decision tree and support machine by Experimenting collected dataset the *decision tree* Weka and Net Beans IDE by means of Support Vector machine , Fisher Filtering [17]. D.Manimegalai, *et.al.,* hasn't expected to give a far reaching review of therapeutic information mining but instead portrays a few territories which appear to be essential from our perspective for applying machine learning in restorative analysis for our genuine viral dataset [18]. Dr.A.Padmapriya, *et. al.,* proposed the fast algorithm to mine the large dataset This paper isn't expected to give a far reaching review of therapeutic information mining but instead portrays a few territories which appear to be essential from our perspective for applying machine learning in restorative analysis for our genuine viral dataset [19]. Vandana Rajput *et. al.,* has described the dengue fever, its contaminations, and signs. Further they characterized the *association rule* mining that has assorted unmistakable calculations which are furthermore useful in forecast [20].

III. DATASET

Simulated Dengue disease dataset containing 2087 patients' records were used for this analysis. These dataset contains 24 attributes as listed in Table 1. The data cleaning is required to handle the missing data. In this research work, replaces the value of zero (0) in the whole attribute which are having blank value in all the records of chosen dataset. Binarized dataset is used in this research work. The absence or presence of an attribute that causes Dengue disease as 0 or 1 respectively [16].

Table I: Attributes of Dengue Disease Dataset

Attribute Id	Attribute Name
1.	Sudden high fever
2.	Severe headache
3.	Pain behind the eyes
4.	Severe joint and muscle pain
5.	Fatigue/restlessness
6.	Nausea
7.	Persistent
8.	Skin rash, which appears two to five days after the onset of fever
9.	Mild bleeding (such a nose bleed, bleeding gums, or easy bruising)
10.	swollen lymph glands
11.	severe joint and muscle pains
12.	mild bleeding from the nose or gums
13.	mild bruising on the skin
14.	febrile convulsions
15.	Pain behind the eyes
16.	Swollen glands
17.	Rash
18.	Severe abdominal pain
19.	Bleeding from your gums or nose
20	Blood in your urine, stools or vomit
21	Bleeding under the skin, which might look like bruising
22.	Difficult or rapid breathing
23.	Cold or clammy skin (shock)
24.	Vomiting blood

IV. PROPOSED WORK

Each health care keeps up a gigantic volume of patients' information. It is exceptionally troublesome process for breaking down every one of these records physically. Data mining methods are utilized to extract the helpful data from the dataset contain tremendous measure of information. In the medicinal services field it serves to examination the patient's data for give a notice to the patient who have most extreme probability of opportunities to be influenced with the illness and bolster the specialists to give a conceivable treatment for diagnosing. In this work, Apriori association rule mining strategy is utilized to discover the most hazard components of dengue disease. FP-growth algorithm is the very essential calculation for mining continuous itemsets. By means of using minimum confidence value and minimum support value. The figure 1 demonstrates the well ordered procedure of the proposed work. This examination work predicts the hazard components of Dengue disease influencing larger part of the patients with the guide of Weka information mining apparatus.
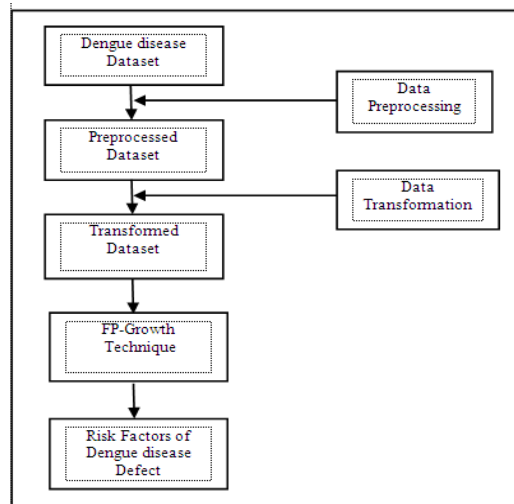

Fig. 1. Step by Step Process of the Proposed Work

A. FP-Growth Algorithm

Apriori algorithm is established to be better for mining association rules. There are different difficulties faced by Apriori algorithm [21]. This algorithm suffers from the following two limitations:
1. Costly in handling large number of candidate sets.
2. It is tedious to scan the database repeatedly and check a large set of candidates by pattern matching.

To overcome these two limitations, Han et al., proposed Frequent Pattern Growth (FP-Growth) algorithm, which avoids the generation of large numbers of candidate sets. FP-Growth uses divide and conquer strategy. FP-Growth requires the construction of FP-tree. The frequent itemsets are generated with only two passes over the database and without any candidate generation process. Two phases are involved in frequent patterns generation. In first phase, it constructs the FP-tree with respect to a given support. In the second phase, the algorithm uses the FP-tree and it does not use the transaction database any more [22]. The FP-Tree algorithm is an efficient rule mining algorithm because of the following reasons:

- The FP-Tree is a compressed representation of the original database because only those frequent items are used to construct the tree, other irrelevant information are pruned. Furthermore, by ordering the items according to their supports, the overlapping parts appear only once with different support count.
- This algorithm scans the database only twice.
- FP-Tree uses a divide and conquer method that considerably reduced the size of the subsequent conditional FP-Tree [23].

V. RESULTS AND DISCUSSION

Training dataset containing 2087 dengue disease affected patients' records is used in this research. This dataset contains 24 symptoms including sudden high fever, severe headache, pain behind the eyes, severe joint and muscle pain ,fatigue etc. The proposed method is implemented using FP-Growth technique with the assist of

WEKA data mining tool. The results obtained from this research work can be helpful to the medical practitioners to find out the very important symptoms by analyzing huge volume of existing records. Graph in Figure 2 shows the number of patients' affecting symptom-wise. Many number of the patients affected by Bleeding under the skin, which might look like bruising and minimum number of patients affected by Vomiting blood.
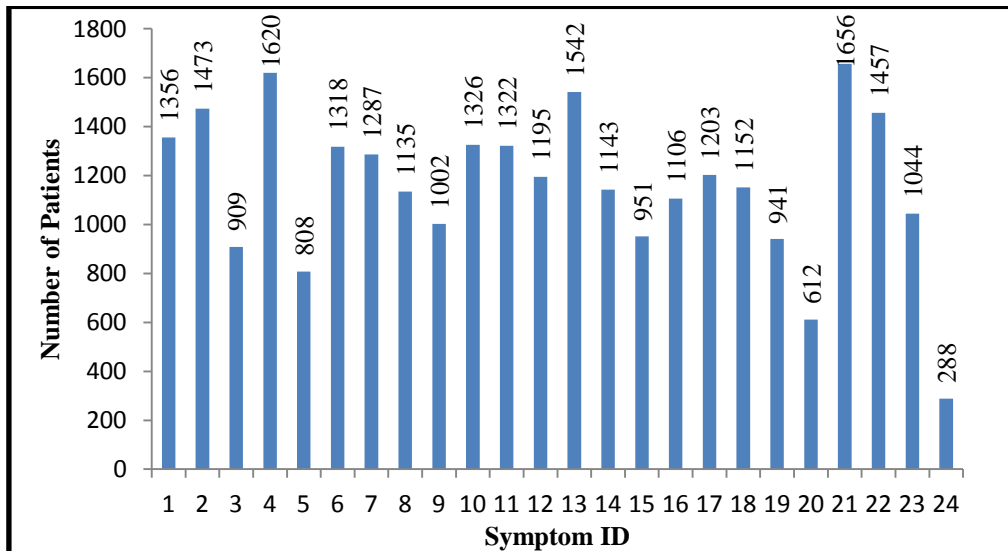

Fig. 2. Number of patients affected by various symptoms causing Dengue disease

The proposed work is experimented over Dengue disease dataset contains 2087 patients records. From the experimental result, the proposed work find outs the symptoms Skin rash, which appears two to five days after the onset of fever, mild bruising on the skin, febrile convulsions and Bleeding under the skin, which might look like bruising are most risk factors of Dengue disease affecting majority of the patients while given various support value ranging from 0.2 to 0.5. The following table 2 shows the symptom affected majority of the patients in for different support value over 2087 patients.

Table II: Symptoms affected majority of the patients for different support value

Support Values	Minimum Confidence	Symptoms ID	No. of Patients
0.2	0.8	10,11,14,813	434
0.3	0.8	14,8,13	651
0.4	0.8	23, 21	882
0.5	0.8	22, 21	1206

VI. CONCLUSION

Clinical database contains substantial amounts of patient records with their restorative conditions. It is hard assignment to recognize the concealed information from extensive volume of dataset by human. Association rule mining is a one of the significant method in data mining which is utilized as a part of human services for predicting, diagnosing and sickness forecast. Dengue dataset contains 2087 patients record in which FP-Growth association rule mining technique is applied in this research work. Experimental results demonstrates that, this exploration work anticipate the most hazard components of Dengue disease influencing dominant part of the patients' effectively.

References

[1]   M.V.Jagannatha Reddy and B.Kavitha ," Expert System to Predict the Type of Fever Using Data MiningTechniques on Medical Databases" , International Journal of Computer Sciences and Engineering Volume 03, Issue-09 ,2015.

[2]   P.Manivannan, Dr. P. Isakki @ Devi,"Dengue Fever Prediction using K-Medoid Clustering Algorithm","International Journal of Innovative Research in Computer and Communication Engineering",Vol. 5, Special Issue 1,2017.

[3]   Kashish Ara Shakil, Shadma Anis and Mansaf Alam, "Dengue Disease Prediction using WEKA Data Mining Tool",

[4]   M. Ilayaraja and T. Meyyappan, "Efficient Data Mining Method to Predict the Risk of Heart Diseases through Frequent Itemsets", Procedia Computer Science, Elsevier, Vol.70, pp.586-592, 2015.

[5]   Zuriani Mustaffa and Yuhanis Yusof "A Comparison of Normalization Techniques in PredictingDengue Outbreak,"International Conference on Business and Economics Reasearch , vol.1-2011.

[6]   N. F. Rahim,S. M. Taib,A. I. Z. Abidin,"Dengue Fatality Prediction using Data Mining", Journal of Fundamental and Applied Sciences,Special Issue-2017.

[7]   Ramandeep Kaur,Gaurav Gupta,Gurjit Singh Bhathal,"Demographic Analysis of Dengue Fever using Data Mining ",Volume 8, No. 7,2017.

[8]   Marimuthu Thangam and Balamurugan Vanniappan,"Mining Association Rules in Dengue Gene Sequence with Latent Periodicity",Hindawi Publishing Corporation Computational Biology Journal Volume 2015.

[9]   M.Bhavani and S.Vinod kumar,"A Data Mining Approach For Precise Diagnosis of Dengue Fever",International Journal of Latest Trends in Engineering and TechnologyVol.-7Issue-4.

[10] Dave Kaveri Atulbhai1, Shilpa Serasiya,"A Survey Predection & Detection of Dengue – Mining Methods &Techniques",IJARIIE-ISSN(O)-2395-4396,Vol-3 Issue-2, 2017.

[11] Hakizimana Leopord, Dr. Wilson Kipruto Cheruiyot, Dr. Stephen Kimani ,"The International Journal Of Engineering And Science (IJES) Volume -5 , Issue - 9,2016.

[12] Daranee Thitiprayoonwongse, Prapat Suriyaphol and Nuanwan Soonthornphisaj,"Data Mining of Dengue Infection Using Decision Tree",Latest Advances in Information Science and Applications",ISBN: 978-1-61804-092-3.

[13] Nandini. V and Sriranjitha. R and Yazhini. T. P, "Dengue Detection and Prediction System using Data Mining with Frequency Analysis", Computer Science & Information Technology (CS & IT).

[14] "Dengue Guidelines for Diagnosis, Treatment, Prevention and Control", A joint publication of the World Health Organization (WHO) and the Special Programme for Research and Training in Tropical Diseases (TDR),2009.

[15] Kamran Shaukat, Nayyer Masood, Sundas Mehreen and Ulya Azmeen,"Dengue Fever Prediction: A Data Mining Problem",Dar et al., J Data Mining Genomics Proteomics 2015.

[16] M. Ilayaraja and T. Meyyappan, "Medical Data Mining Method to Predict Risk Factors of Heart Attack and Raise Early Warning to Patients", International Journal of Applied Engineering Research, ISSN:0973-4562, Vol.10, No.55, 2015.

[17] Arun Kumar.P.M, Associate Professor, 1. Chitra Devi.B, Karthick.P, Ganesan.M and Madhan.A.S, "Dengue Disease Prediction Using Decision Tree and Support Vector Machine ",SSRG International Journal of Computer Science and Engineering- (ICET'17) - Special Issue - March 2017 ISSN: 2348.

[18] A.Shameem Fathima ,D.Manimegalai and Nisar Hundewale, "A Review of Data Mining Classification Techniques Applied for Diagnosis and Prognosis of the Arbovirus-Dengue"IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6, No 3, - 2011.

[19] N.Subitha M.phil, Research Scholar , Dr.A.Padmapriya M.C.A., M.phil., Ph.D,"Diagnosis for Dengue Fever Using Spatial Data Mining",International Journal of Computer Trends and Technology (IJCTT) – volume 4 Issue 8-2013.

[20] Vandana Rajput, Prof. Amit Manjhvar,"A Review Paper on Dengue Disease Forcasting Using Data Mining Techniques",IJSART - Volume 3 Issue 4 – 2017.

[21] Varsha Mashoria and Anju Singh. "Literature Survey on Various Frequent Pattern Mining Algorithm", IOSR Journal of Engineering (IOSRJEN), Vol.3, Issue.1, pp.58-64, 2013.

[22] Trupti A. Kumbhare and Santosh V. Chobe. "An Overview of Association Rule Mining Algorithms", International Journal of Computer Science and Information Technologies, Vol.5, No.1, 2014.

[23] Jyotsana Dixit and Abha Choubey. "A Survey of Various Association Rule Mining Approaches", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.4, Issue.3, pp.651-655, 2014.

[24] M. Ilayaraja and T. Meyyappan, "Mining Medical Data to Identify Frequent Diseases using Apriori Algorithm", Proceedings of the IEEE International Conference on Pattern Recognition, Informatics and Mobile Engineering, pp.194-199, 2013.

[25] Anurag Bhatt1 and Manish Joshi,"Analytical Study of Applied Data Mining in Healthcare",Special Issue NCETST-2017.

[26] V. Ughelli, Y. Lisnichuk, J. Paciello and J. Pane,"Prediction of Dengue Cases in Paraguay Using Artificial Neural Networks",Int'l Conf. Health Informatics and Medical Systems , HIMS'17.