Dengue Disease Prediction Using SMO Classification

[1]M. Preetha, [2]P. Santhosh, [3]R. Saran kumaR, [4]Ms. K. Poongodi,

[1-3]UG Student, Department of Computer Science and Engineering,
K.S.Rangasamy College of Technology, Tiruchengode
[4]Assistant Professor, Department of Computer Science and Engineering,
K.S.Rangasamy College of Technology, Tiruchengode

Email id: [1]Preethamurugan91097@gmail.com, [2]santhoshsathasivam1997@gmail. Com,
[3]sarankumarmtp@gmail.com,

Abstract

The aim of this work is to examine the performance of different classification techniques. A dengue disease can cause severe damages to the society. Hence, it is critical to predict a dengue disease in advance to minimalize the damage and loss caused by the disease. The clinical documents maintained are a pool of information regarding the infected patients. By keeping this voluminous data we can predict the future occurrences of the disease earlier and safe guard the people. Dengue the global problem is common in more than 110 countries. Dengue infection has endangered 2.5 billion populations all around the world. Every year there are 50 million people who suffer from it globally. Dengue fever is a vector borne disease caused by the female Aedes Aegypti and Aedes Albopictus mosquitoes which adapt well to human environments. Data mining is a well-known technique used by health organizations for classification and prediction of diseases.

Index Terms: Data mining, Support vector machine, Weka library, IDE Net beans, UCI repository dataset.

1. INTRODUCTION

The dengue disease accounts to be the leading cause of death worldwide. It is difficult for medical practitioners to predict the dengue disease as it is a complex task that requires experience and knowledge. Data mining enable the health sector to predict patterns in the dataset. Analyze the dengue disease dataset available from the UCI machine learning with the aim of developing accurate prediction models for dengue disease using data mining techniques. The collected dataset was investigated with Weka library and Net Beans IDE and a decision tree is produced using Fisher Filtering and classification is done effectively with Support Vector Machine. Sequential minimal optimization applied in this research for predicting dengue disease. The research result of SMO shows better prediction accuracy than the existing system.

Several research works has been carried out for first outbreak highlights the geographic expansion of Dengue hemorrhagic fever [6][10] in Bangladesh, where classic dengue fever caused by multiple serotypes had been reported using enzyme-linked immunosorbent assay (ELISA) foranti-dengue antibodies confirmed the case as Dengue hemorrhagic fever.

In [3] intelligent outlier detection algorithm (IODA)was developed to treat a time series as an image and segments the image into clusters of interest, such as ''nominal data'' and ''failure mode'' clusters. The algorithm customs density clustering techniques to identify sequences of coincident

clusters in both the time domain and delay space, where the delay space representation of the time series consists of ordered pairs of consecutive data points taken from the time series. A best cluster is selected in delay space and used to construct a ''feature'' in the time domain from a subset of the optimal time-domain clusters. Segments of the time series and each datum in the time series are classified using decision trees which gives final quality score or quality index for each data point.

In [10], [14]an empirical mode decomposition (EMD) was developed to isolate a 3-yr periodic mode of variance. The travelling wave revealed in this periodic mode is obscured in the raw incidence data by the occurrence of many periodic and roughly periodic components.

Here this system examine the spatial–temporal dynamics of DHF incidence in a data set describing 850,000 infections occurring in 72 provinces of Thailand. It observe the wave in a three year periodic component of variance, which is thought to reflect host– pathogen population dynamics. This finding provides an important starting point for detecting and characterizing the key processes that contribute to the spatial–temporal dynamics of Dengue hemorrhagic fever.

In [10] the study carried out an algorithm to select boundary instances as training data to substantially reduce number of instance(n). The algorithm is motivated by the result of that, removing non-support vectors from the training set does not change SVM training results. The algorithm eliminates instances that are likely to be non-support vectors. In the concept independent preprocessing step of the algorithm, the proposed system prepare nearest-neighbor lists for training instances. In the concept-specific sampling step, then effectively select useful training data for each target concept. Empirical studies show the algorithm to be effective in reducing n, outperforming other challenging down sampling algorithms without significantly compromising testing accuracy. In [11] due to the remarkable increase in dengue fever and deaths has pave way for an effective surveillance system to detect the cases and make necessary control measures against dengue. Laboratory diagnosis, monitoring and investigation of outbreaks are documented, discussed to prevent the outbreaks of dengue fever.

In [10], [12] due to the increasing dengue incidence has been attributed to climate change, the contradicting reports displays the inconclusive relationships between climate factors and dengue. On viewing the climate factors, this report investigates temperature and rainfall as climate factors affecting dengue. Results displays dengue incidence varies with changing rainfall patterns. Monitoring and control mosquitoes during periods with high rainfall are recommended.

In [5] due to the remarkable increase in dengue fever and deaths has pave way for an effective surveillance system to detect the cases and make necessary control measures against dengue. Laboratory diagnosis, monitoring and investigation of outbreaks are documented, discussed to prevent the outbreaks of dengue fever.

II. EXISTING SYSTEM

Dengue is a threatening disease caused by female mosquitos. It is typically found in widespread hot regions. From long periods of time, experts are trying to find out some of features on dengue disease so that user can rightly categorize patients because different patients require different types of treatment. For properly categorizing the dataset, different classification techniques are used. These techniques are Naïve Bayesian, REP Tree, Random tree, J48.

The algorithm C4.5 has classified only undesirable effect of changing a dengue patient's existing test data groups, potentially undoing the patient's own manual efforts in organizing the history. It involves a high computational cost, have to repeat a large number of attribute test data group similarity computations for every new test data. C4.5 helps to extract dengue disease prediction suffer from scalability. It is imperative to address the scalability issue. Connections in dengue prediction are not homogeneous.

III. PROPOSED SYSTEM

Methods that can accurately predict dengue disease are greatly needed and good prediction techniques can help to predict dengue disease more accurately. In this system, it used two feature selection methods, forward selection and backward selection, to remove irrelevant features for improving the results of dengue disease prediction. The results

1806

Preetha M et al., Inter. J. Int. Adv. & Res. In Engg. Comp., Vol.–06(02) 2018 [1804-1808]

show that feature reduction is useful for improving the predictive accuracy and density is irrelevant feature in the dataset where the data had been identified on full field digital mammograms collected at the UCI Repository. In addition, decision tree, support vector machine—sequential minimal optimization (SVM-SMO) and their ensembles were applied to solve the dengue disease diagnostic problem in an attempt to predict results with better performance. The results establish that ensemble classifiers are more accurate than a single classifier.

The proposed framework SMO based on disease prediction is shown to be effective in addressing this prediction. The framework suggests a novel way of network classification: first, capture the latent affiliations of actors by extracting disease prediction based on network connectivity, and next, apply extant data mining techniques to classification based on the extracted prediction.

In the initial study, modularity maximization was employed to extract disease prediction. The superiority of this framework over other representative relational learning methods has been verified with dengue prediction dengue data.

IV.   SUPPORT VECTOR MACHINE

To find a feature of subset of size m which contains the most informative features. The two well performing feature selection algorithms on the dataset are briefly outlined below.

Feature reduction applies a mapping of the multidimensional space into a space of lower dimensions. Feature extraction includes features construction, space dimensionality reduction, sparse representations, and feature selection all these techniques are commonly used as preprocessing to machine learning and statistics tasks of prediction, including pattern recognition.

The feature space having reduced features truly contributes to classification that cuts preprocessing costs and minimizes the effects of the peaking phenomenon in classification. Thereby improving the overall performance of classifier based intrusion detection systems. SVM is a linear transformation with linear ortho normal basis vectors; it can be expressed by a translation and rotation.

V. FEATURE REDUCTION BY SEQUENTIAL MINIMAL OPTIMIZATION

Classification is the type of data mining, which deals with the problematic things by recognizing and detecting features of infection, among patients and forecast that which technique shows top performance.

Feature reduction applies a mapping of the multidimensional space into a space of lower dimensions. Feature extraction includes features construction, space dimensionality reduction, sparse representations, and feature selection all these techniques are commonly used as preprocessing to machine learning and statistics tasks of prediction, including pattern recognition. Although such problems have been tackled by researchers for many years, there has been recently a renewed interest in feature extraction. The feature space having reduced features truly Contributes to classification that cuts preprocessing costs and minimizes the effects of the 'peaking phenomenon' in classification. Thereby improving the overall performance of classifier based intrusion detection systems. The commonly used dimensionality reduction methods include supervised approaches such as Linear Discriminant Analysis (LDA), unsupervised ones such as SMO, and additional spectral and manifold learning methods. It converts a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. Consider the two dimensional cases then the basic principle of this transformation.

VI. RESULTS AND DISCUSSION

This prediction provides the selection of minimum attributes by providing prediction and classification of data set with the support of Sequential Minimal Optimization and thus this produces better accuracy in the prediction of dengue disease.

Performance Accuracy
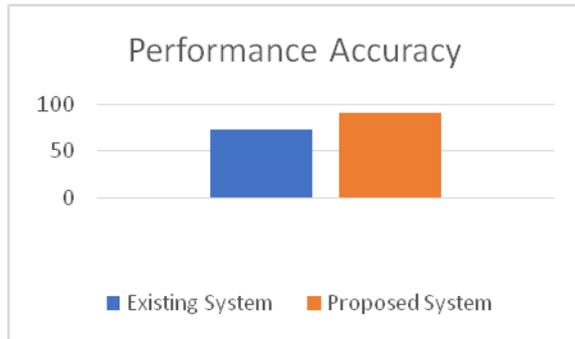
VII. FUTURE SCOPES

High effectiveness of the proposed algorithms in capturing test data relevance.

- Test data reformulation graph and the test data click graph into a single graph that it refer to as the test data fusion graph, and by expanding the test data set when classification relevance occur

- Also include other patient id

- High Relevance Measure

- Good classification and prediction accuracy

VIII. CONCLUSION

The infection rates of Aedes Aegypti mosquito's increases morbidity rate hence the decision tree is generated with the Aegypti rate as the root node and prevent further occurrences. The prediction of dengue infection carried out using Weka library and data mining techniques such as Sequential Minimal Optimization, decision tree and Support Vector Machine. This experiment can serve as an important tool for physicians to predict risky cases in the practice and advise accordingly. This model helps to predict the dengue disease by reducing the features in the data set and classify them with better accuracy. Thus the predictive accuracy determined by SMO classification algorithm suggests that parameters used are reliable indicators to predict the presence of dengue diseases.

REFERENCES

[1] Amin, MMM, Hussain, AMZ and Nahar Ket al (2000) , "Sero-diagnosis of Dengue Infections in Four Metropolitan Cities of Bangladesh", Dengue Bulletin , Vol 24.

[2] Aziz, KN Hasan and MA Hasanat et al (2002). "Predominance of the DEN-3 genotype during the recent dengue outbreak in Bangladesh". Department of Immunology, Bangladesh Institute of Research and Rehabilitation in Diabetes, Vol 33.

[3] Blackburn G L, Wang K A (2007),"Dietary fat reduction and Dengue Disease outcome: results from the Women's Intervention Nutrition Study (WINS)", The American journal of clinical nutrition, pp. 878-81. PMID 18265482.

[4] Dana a. Focks, Eric daniels and Dan G. Haile, "A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results", Society Tropical Hygiene, pp. 489-506.

[5] Delen D, Walker G, (2005), "Predicting Dengue Disease survivability: a comparison of three data mining methods", Artificial Intelligence in Medicine, vol. 34, pp. 113-127.

[6] Derek A. T. Cummings, Rafael A. Irizarry and Norden E. Huang et al (2004). "Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand", Nature PublishingGroup, |vol 427.

[7] Douglas M, and Watts,donald S, et al (1987) ," Effect of temperature on the vector efficiency of Aedes Aegypti for dengue 2 virus", The American Society of Tropical Medicine and Hygiene, pp. 143-152.

[8] Eli Schwartz, Leisa H. Weld and Annelies Wilder-Smith (2008), "Seasonality, Annual Trends, and Characteristics of Dengue among Ill Returned Travelers", Emerging Infectious Diseases, Vol. 14, No. 7.

[9] Glenn L. Sia Su (2008)," Correlation of Climatic Factors and Dengue Incidence in Metro Manila, Philippines", Royal Swedish Academy of Science, Vol. 37, No. 4.

[10] Kanchana Nakhapakorn and Nitin Kumar Tripathi (2005), "An information value based analysis of physical and climatic factors affecting dengue fever and dengue haemorrhagic fever incidence",International Journal of Health Geographics, pp. 1-13.

[11] Mahbubur Rahman and Khalilur Rahman et al (2002). "First outbreak of dengue hemorrhagic fever", Emerging Infectious Diseases, Vol. 8, No. 7.

[12] Michael A. Johansson, Francesca Dominici and Gregory E. Glass2 (2009), "Local and Global Effects of Climate on Dengue Transmission in Puerto Rico", vol. 3.

[13] Mohammad ali, Yukiko wagatsuma and Michael Emch et al (2003), "Use of a geographic information system for defining spatial risk for dengue transmission in Bangladesh: Role for Aedes Albopictus in an urban outbreak", The American Society of Tropical Medicine and Hygiene, pp. 634–640.

[14] Rico-Hessea R, and Harrisona LM et al, "Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas", Department of Epidemiology and Public Health, vol. 230, pp. 244–251.

[15] Rohani1, YC Wong1, I Zamre1, HL Lee1 and MN Zurainee, "The effect of Extrinsic incubation temperature on development of Dengue Serotype 2 And 4 Viruses In Aedes Aegypti" , Institute for Medical Research, Vol. 40 No. 5

[16] Supawan promprou, Mullica Jaroensutasinee and Krisanadej Jaroensutasinee(2005) , "Impact of Climatic Factors on Dengue Haemorrhagic Fever Incidence in Southern Thailand", Walailak J Sci & Tech ,pp. 59-70.

[17] Thomas W. Scott and Amy C. Morrison, "Aedes aegypti density and the risk of dengue-virus transmission", Department of Entomology, University of California, Davis, CA 95616 USA.