DENGUE FEVER CLASSIFICATION TOOL USING MACHINE LEARNING

Khaing Thanda Swe[1], Phyo Thu Zar Tun[2]

Mandalay Technological University, Lecturer, +095, Myanmar

Abstract

Fever is generally not considered dangerous, but hyperthermia can cause dangerous rises in body temperature such as heat stroke, side effects of certain medications or illicit drugs. Another type of dangerous fever is dengue fever. Dengue fever is one of the most important mosquito-borne viral diseases in the world. Approximately 40% of the world population lives in the high-risk area of dengue fever, in which 400 million get infected and 22,000 die from severe dengue fever annually. In any case, it has been an awesome test utilizing blood and urine test for the doctors to distinguish the level of hazard in dengue patients. The main aim of proposed system is to reduce this unnecessary urine test and blood test for patients. Datasets are gathered from hospital of Lao PDR, which incorporates around 100 ground truth data containing numeric attributes. In the proposed fever classification tool, the user can choose the suffered symptoms. Then the system gives the result according to the symptoms using support vector machine that the fever is either dengue fever or severe fever or high temperature cold fever. Moreover the user can get the knowledge about the dengue fever and can share information about their cases via forum. The system is implemented with Java Programming Language. Finally the system analyses the performance for either how many cases increased or decreased in using this dengue fever classification tool for Champpassak region, Lao PDR.

Keyword: hyperthermia, dengue, PDR

1.INTRODUCTION

Dengue fever is an irresistible illness likewise called as a break borne fever or dandy fever or dengue haemorrhagic fever caused by a group of virus which are transmitted by mosquitoes. The family of mosquitoes which transmits dengue virus will also leads into diseases like Chikungunya, yellow fever and Zika virus [1].

Virus can influence anybody yet is more extreme in immunodeficiency individuals. This con-termination is caused through the chomp of Aedes, a female mosquito. Alternate mosquitoes get contaminated due to the admission of the blood from a man who is tainted with dengue infection. It is most regularly found in the areas of sub-tropical atmospheres around the world [2].

Diagnosing fever is challenging due to its nonspecific clinical presentation which varies with dengue stages, severity, patient's age, gender and pre-existing medical conditions. The management of fever is comparatively simple, inexpensive and highly effective in saving lives. Early detection and prompt access to proper medical care could lower mortality rates from 10% to below 1% [3].

The proposed system is applied only for the district area where difficulty in getting right clinical information. At these areas, there are little medical specialists. There may be healthcare volunteers and other medical students. In addition the people in these areas are poor medical knowledge background. So there occurs often unnecessary mortality rate increased related with dengue fever because of knowing later information which cases they suffer and how they should do the follow up treatment. If they know earlier the information about their fever, they can early take care how they do and what they need to overcome this fever. This inconvenience can be overwhelmed by utilizing machine learning calculations in medicinal and wellbeing spaces which can give the outcome in less time including less human work and money. There are many classification methods in machine learning such as Logistic Regression, Naïve Bayes Classifier, Nearest Neighbour, Support Vector Machine, Decision Trees and Random

Forest. Among them the proposed system is implemented with SVM according to the literature review that SVM can give the accurate results in low computation time. In addition to, SVM delivers a unique solution, since the optimality problem is convex.

2. LITERATURE REVIEWS

The latest research 2005, the author [5] proposed the "Dengue Anti-body Prevalence in German Travellers". He tested that his study was done on 2259 citizens of German after they returned from dengue-endemic countries. He used IgM, IgG, and ELISA test. The final result is that specific serotypes of dengue antibodies indicated acute infections in 51 travellers' i.e. 4.7% with recent fever and 13 travellers with no recent fever.

Then in the same year 2005, the author [6] submitted "A novel dengue fever (DF) and dengue hemorrhagic fever (DHF) analysis using Artificial Neural Network (ANN)". The author used the input for the ANN is that the clinical symptoms which uses multilayer feed-forward neural networks (MFNN). Initially, physical examinations were carried out and then SPSS software was used for data analysis. Finally, ANN was applied on the dataset using MATLAB tool. The final result is that the ANN presented in the study had successfully yielded promising results with the error of only 10%.

And then in 2012, the author in [8] implemented "A data-driven epidemiological prediction method for dengue out-breaks using local and remote sensing data". The author used the Fuzzy Association Rule Mining as the prediction method for extraction of relationships among clinical, meteorological, climatic etc. Data set includes dengue case data from 2001–2009 and case data is collected from the Peruvian Ministry of Health. Incidence of dengue is predicted three or four weeks in advance. Test data for a period 4–7 weeks from time of pre-diction yielded a positive, negative, sensitivity and specificity values as 0.686, 0.976, 0.615, and 0.982 respectively.

In 2014, the authors in [9] applied "Classification Rules Using Decision Tree for Dengue Disease". The objective is to create a prediction model, Decision tree. A total of 2424 records were gathered from various sources. Decision tree predicts the chances of occurrences of dengue diseases in a tribal area.

Also in 2014, the authors in Sri Lanka [10] designed "Prediction of Dengue Outbreaks in Sri Lanka using Artificial Neural Networks". ANN is designed for the forecast of Dengue out-breaks and it is a Multi-Layer Perceptron (MLP) network connects with three introductory layers. Correlation analysis is done on every individual variable with the dengue cases revealed. Variables with the highest correlation are selected as input to the ANN. ANN generates predictions with an accuracy of 68.5%, which implies that 68.5% of the produced prediction exists in 25% tolerance of the actual reported cases.

In 2015, the authors in [11] showed "Dengue disease prediction using Weka data mining tool". WEKA, a data mining tool is used in the Comparison of different data mining algorithms. Comparative study shows that Naïve Bayes is 100% accurate and J48 is 99.70% accurate.

In 2016, the authors in [12] submitted Machine Learning Models for Early Dengue Severity Prediction. Data was collected from the patients of Hospital which is used to validate several classification models. All the models were trained under 5-Fold Stratified Cross-Validation and these outcomes indicate good prediction capabilities and SVM offered the best results measured by receiver operating characteristic area.

In 2017, the authors in [14] carried on "A survey Prediction & Detection of Dengue – Mining Methods & Techniques". Classification techniques were used for ordering of dataset such as Naive Bayesian, REP Tree, Random tree, J48, SMO, SVM, Decision Tree Approach, and Spatial Data Analysis etc. Data set for dengue prediction is DNA microarray data which have information of gene's expression responsible for dengue virus. Comparison among the techniques concluded that Naive Bayes is prominent among all others as it delivers an accuracy of 92% with high probability and effectiveness.

The proposed system mainly references the paper [6] in 2005. It used the ANN for dengue fever and dengue hemorrhagic fever. But the system got the performance evaluation was below 89%. The error rate was 10%. Therefore, the proposed system uses the Support Vector Machine (SVM) to improve the accuracy rather than ANN. According to the literature review, when compared to ANN models, SVMs give better results.

3. METHODOLOGY

Machine learning involves predicting and classifying data and to do employing various machine learning algorithms according to the dataset.

SVM or Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple: The algorithm creates a line or a hyper plane which separates the data into classes.

Support Vector Machine (SVM) finds an optimal solution.

- Maximizes the distance between the hyper plane and the "difficult points" close to decision boundary
- One intuition: if there are no points near the decision surface, then there are no very uncertain classification decisions

SVM has been used successfully in many real-world problems

- Text (and hypertext) categorization
- Image classification
- Bioinformatics (Protein classification, Cancer classification)
- Hand-written character recognition [15]

3.1. Working Principles of SVM

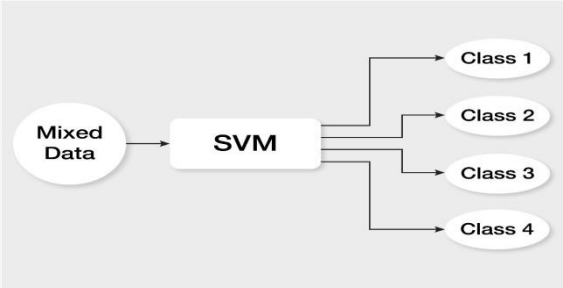
At first approximation what SVMs do is to find a separating line (or hyper plane) between data of two classes. SVM is an algorithm that takes the data as an input and outputs a line that separates those classes if possible.

- The classifier is a *separating hyper plane*.
- Most "important" training points are support vectors; they define the hyper plane.
- Quadratic optimization algorithms can identify which training point's x_i are support vectors with non-zero lagrange multipliers a_j.
- Both in the dual formulation of the problem and in the solution training points appear only inside dot products.

Datasets those are linearly separable with some noise work out great:

Support vectors are the data points nearest to the hyper plane, the points of a data set that, if removed, would alter the position of the dividing hyper plane. Because of this, they can be considered the critical elements of a data set.

As a simple example, for a classification task with only two features (like the image above), you can think of a hyper plane as a line that linearly separates and classifies a set of data. Intuitively, the further from the hyper plane our data points lie, the more confident we are that they have been correctly classified. We therefore want our data points to be as far away from the hyper plane as possible, while still being on the correct side of it [16].



3.2. Data Set

The proposed system uses the data set collected from various medical wards of Lao National hospitals as shown in Table 1. This data set considers only numeric attributes. This data set contains around 80 ground truth records and 20 attributes for each record.

Table 1: List of Attributes in Dengue Fever Dataset

Attributes	Attributes
Fever	Shock
Vomiting	Severe bleeding
Rash	Diagnosis
Tornique	Rising HCT
Leukopenia	Fluid Overload
Abdominal Pain	Mucosal
Irritable	Respiratory
Aches and Pain	Liver
Coma	Hemoglobin
Diarrhea	Hepatomegaly

4. PROPOSED SYSTEM DESIGN

The processing steps of dengue fever screening tool are as follow:

Step 1: Load test data and training dataset

Step 2: Preprocessing (Feature Selection)

Step 3: The classification algorithm (Support Vector Machine) is used to predict the dengue fever
Step 4: Based on the result, it will show the classification algorithm accuracy.
The future processing steps of surveillance system are as follow:
Step 1: Upload map from google earth pro
Step 2: Use GIS for mapping
Step 3: Report the number of dengue fever patients

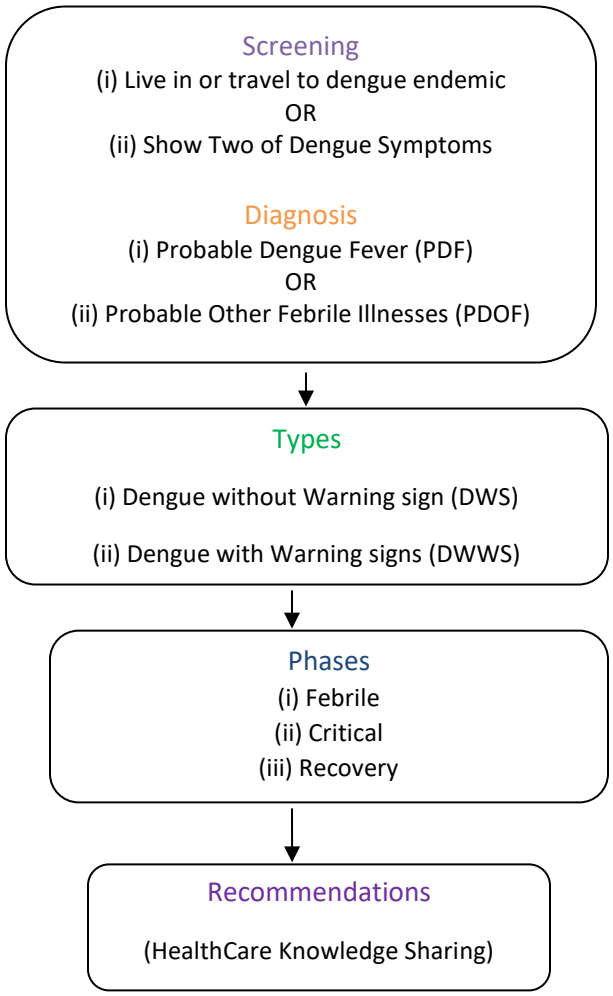


Figure 1. Proposed System Flow

The proposed system includes five sections: Screening, Diagnosis, Types, Phases and Recommendations.
In the screening phase, the system first checks the patient whether or not to travel or live in the dengue endemic area. And then how the symptom is that the patient suffers. The suffered symptoms are really dengue sign or not.

The second phase is Diagnosis, the system generates the illness is severe fever or dengue fever using SVM.
The third phase is detail types in dengue fever. There are three types such as Dengue without Warning Sign (DWS), Dengue with Warning Sign (DWWS) and Severe Dengue (SD).
The last portion of the system is level of the dengue fever. If the patient suffers the dengue fever, how should they do for the follow up instructions? There are three phases in the system. These are Febrile, Critical and Recovery.
The final part is recommendations. In proposed system the user can get the medical knowledge via Forum.

5.PERFORMANCE ANALYSIS

For accurate diagnosing, always accuracy performance metric is not sufficient to determine whether a patient is affected by dengue disease or not. And hence various performance metrics are used to measure the classifier's performance are Province A (Traditional Method) and Province B (Proposed Method).

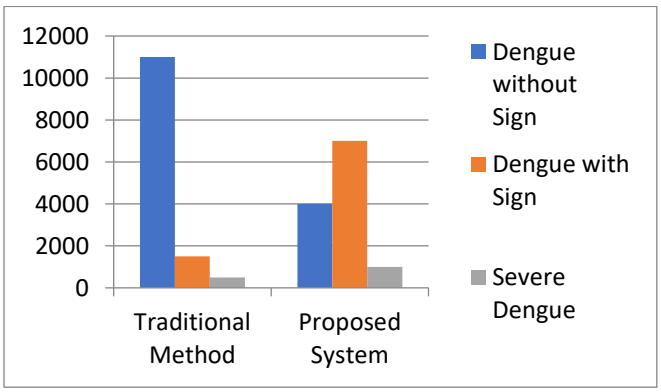


Figure 2. Performance Evaluation of Traditional and Proposed Method

There are two main objectives in implementing the proposed system. The first one is to reduce the amount of dengue fever patients in rural areas. The second one is to get high accuracy with SVM in the case of classification for dengue fever or severe fever instead of using ANN. Now the proposed system evaluates the performance analysis for rural areas especially Champpassak region depending on three factors such as DNWS, DWWS and SD. In this case there are two types: Province A (used Traditional Method) and

Province B (used Proposed Method). For traditional method, the patients and other family members used their traditional medicine without knowing the dengue fever earlier. Then the proposed system is applied to the same cases. The results of the proposed system over three factors are shown in Figure 2.

6.CONCLUSION

The proposed system mainly aims to reduce the unnecessary blood test and urine test for under-developing areas. Health is the most important in Life. So if they know early, which type of fever they suffer, they can do to get good health with the right ways. In this system, the data are gathered from the patients of the Lao National Hospitals. Then these data are applied into the system using Support Vector Machine. The final output is whether the patient suffers the severe fever or dengue fever. If the patient suffers the dengue fever, the system assigns the level of the dengue: (i) Febrile (ii) Critical (iii) Recovery. Otherwise, the system can also be used for primary healthcare workers, volunteers and healthcare students. The system includes the recommendation portion for sharing knowledge about dengue and if they suffer dengue fever, how should they do. In this way, the system gives the low rate of misunderstanding the signs of the fever and low infant mortality rate.

The system is implemented with the JAVA programming language with the help of XAMPP database.

REFERENCES

Article/ Research Paper

[1] Dave Kaveri Atulbhai and Shilpa Serasiya, "A Survey: Prediction & Detection of Dengue – Mining Methods & Techniques, IJAR-IIE–ISSN (O)-2395-4396, Vol-3, Issue-2, 2O17.

[3] Meena Kumari Bimal, Ms.Lakhbir Kaur2, Ms.Mandeep Kaur.Assessment of Knowledge and Practices of People Regarding Dengue Fever.International Journal of Nursing Education and Re-search 4(2):174-178; April-June 2016; DOI: 10.5958/2454-2660.2016.00035.1.

[4] NS Dighe, SR Pattan, SB Dighe, SB Bhawar, RB Saudagar,V M Gaware, MB Hole and VB Tambe.Chikungunya Disease: A Review.Research J. Pharmacology and Pharmacodynamics 2009; 1(1):07-12.

[5] Rohit Morlawar, V.A. Kothiwale, "A study of clinical profile in different serological diagnostic parameters of dengue fever", 2017, Volume: 10, Issue: 2, pp: 178-182

[7] M. Montes-y-Gomez et al. (Eds.): IBERAMIA 2016, LNAI 10022, pp. 247–258, 2016, DOI: 10.1007/978-3-319-47955-2 21

[8] M.Bhavani and S.Vinod kumar, "A Data Mining Approach For Precise Diagnosis of Dengue Fever", International Journal of Latest Trends in Engineering and Technology, Vol (7), Issue(4), pp.352-359, nov 2016, DOI: http://dx.doi.org/10.21172/1.74.048

[9] Purushottam Kumar, Rajendra T. Ankushe, Bina M. Kuril1, Mohan K. Doibale, Syed J. Hashmi, Sandeep B. Pund, "An Epidemiologi-cal Study of Fever Outbreak in Aurangabad, Maharashtra, India", International Journal of Community Medicine and Public Health Kumar P et al. Int J Community Med Public Health. 2016 May; 3(5), pp: 1107-1111

[10] Phong D. Tong, Vinh D. Le, Hieu N. Duong, Hien T. Nguyen, Vaclav Snasel, "Decision trees for diagnosis of dengue fever", In-ternational Conference on Information and Convergence Technolo-gy for Smart Society, Jan 2016, 19-21, in Ho Chi Minh, Vietnam

[14] P.H.M.Nishanthi, Herath, A.A.I., Perera, H.P.Wijekoon, "Predic-tion of Dengue Outbreaks in Sri Lanka using Artificial Neural Net-works", International Journal of Computer Applications (0975 – 8887) Volume 101– No.15, September 2014

Books

[1] Xavier-CarvalhoC, Cezar RDDS, Freire NM, Vasconcelos CMM, Solorzano VEF, de Toledo-Pinto TG, Fialho LG, do Carmo RF, Vasconcelos LRS, Cordeiro MT

[2] William Caicedo-Torres, Angel Paternina, and Hernando Pinzon, "Machine Learning Models for Early Dengue Severity Prediction", @Springer International Publishing AG 2016

[3]Danilo Bretas de Oliveira, Guilherme Machado, Gabriel Magno de Freitas Almeida, Paulo Cesar Peregrino Ferreira, Claudio Antonio Bonjardim, Giliane de Souza Trindade, Jonatas Santos Abrahao and Erna Geessien Kroon, "Infection of the central nervous system with dengue virus 3 genotype I causing neurological manifestations in Brazil", Revista da Sociedade Brasileira de Medicina Tropical 49(1):125-129, Jan-Feb, 2016, http://dx.doi.org/10.1590/0037-8682-0208-2015

[4] Kashish Ara Shakil, Shadma Anis and Mansaf Alam, "Dengue Dis-ease Prediction using Weka Data Mining Tool", 18 Feb 2015.
[5] 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, 6:3 http://dx.doi.org/10.4172/2153-0602.1000181

Online Sources

[1] http://www.cs.cmu.edu/~awm/tutorials **Mingyue Tan,** University of British Columbia Nov 26, 2004
[2] https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html