Data mining Techniques for Estimating the Type of Fever

[1]P.Hima Keerthi [2]M.Shashi

Department of Computer Science and Systems Engineering, Andhra University, Vishakapatnam, India.

Abstract---- By finding the most significant medical indications and research center data helps in structure an expert system to predict the dengue fever in the beginning periods. We created in this undertaking another expert system to predict the dengue fever in the beginning times. This approach comprises of three significant advances: a) manual missing worth ascription strategy is connected that makes the data predictable. b) An expert specialists assessment is taken for choosing the most persuasive characteristics for dengue fever additionally we completed a web study. c) A neural network model is utilized for precise prediction of dengue fever. The expert system is created utilizing MATLAB 2013. This procedure is by all accounts giving great predictive outcomes contrasted with different techniques.

Keywords: Dengue Fever, Expert system, Neural Network, Prediction

I. INTRODUCTION

Dengue fever is an ailment transmitted by mosquitoes[1] and causing abrupt high fever and torments in the joints. Otherwise called break bone fever. The principal case is recognized in the Philippines in 1953, the malady is distinguished as a standout amongst the most hazardous infections in humans[2]. Exact prediction of infection is conceivable simply after a few trial of lab and clinical side effects.

A multivariant model was developed for predicting hemoglobin utilizing predictors i.e., they have utilized different properties, for example, retching sensation, weight, sex, and different elements. These techniques are utilized simply following two to twelve days from the day of disease.

The world wellbeing association is made an arrangement for distinguishing influenced singular people dependent on the research facility and clinical manifestations. The models produced for the determination of dengue fever is influenced by missing qualities and compelling highlights. This might be because of wrong data passage or not gathered appropriately at the season of data accumulation.

So as to maintain a strategic distance from wrong prediction, we pursued the strategy as underneath

1) Missing qualities are filled physically with fitting ordinary qualities

2) In request to keep away from an excessive number of properties for examination, we took guidance from various expert doctors. With the goal that lone most powerful characteristics are gathered.

3) After data preprocessing, we utilized MATLAB 2013a for accurate prediction of dengue fever using neural networks.

II. RELATED WORK

2.1: - Using Decision Tree:

In this method[3] they said Dengue contamination is an infection commonly found in the hot and sticky district. The specialists need to comprehend the highlights of dengue disease so as to effectively classify the patients since these patients require diverse treatment. Their dataset comprises of clinical and research center data. The data was gathered from the principal visit of the patient to the medical clinic until the date of release. They got two wellsprings of datasets from various districts of Thailand, which are Srinagarindra Hospital and Songklanagarind Hospital. These datasets comprise of in excess of 400 qualities. They utilized a choice tree as a data mining device. They propose a lot of important characteristics from the worldly data. Their analyses are separated into four sections. In each of the four examinations, they use choice trees. The initial two test results show helpful learning to group dengue disease from Srinagarindra Hospital's dataset and Songklanagarind Hospital's dataset, separately. Each arrangement of information is tried by various dataset to ensure that the test data was a genuine inconspicuous data. The third trial results show helpful information when they incorporated two datasets. Another goal of this

examination is to recognize the day of defervescence of fever which is called day0. The day0 date is the basic date of dengue patients that a few patients face the deadly condition. Along these lines the doctors need to predict day0 so as to treat the patients. They hope to have a savvy system that can trigger the day0 date of every patient. They set up four trials. In the initial three examinations, they discover information so as to arrange the sort of dengue disease. For forward test, they attempted to predict the day of fizz with the data before the day0 date. They connected a choice tree way to deal with all trials. Note that they use affectability, explicitness, and precision as execution measures. Their rough exactness of every one of the four trials utilizing choice tree is around 96.5%. In another choice tree calculation technique specialists Tanner, et al and Tarig, et al [5]. They arranged 1,200 patients utilizing a choice tree approach. They discovered six significant highlights and they got 84.7 % accuracy.

2.2 Genetic Algorithm and ANN :

Predicting the dengue fever should be possible by various techniques one of the strategies depicted by[4] distinguishing the significant medical side effect and research center data without a medical expert. In their examination exertion, a savvy based system that distinguishes the analysis of dengue fever is proposed. They pursued three noteworthy advances they are: (1). A strategy to enter missing info esteems that contains differing dataset. (2).A hereditary calculation for trait determination for dealing with a subset of primarily significant side effect that can recognize the illness. (3).An fake neural network technique that utilizes a back-engendering strategy for expanding the precision of predicting dengue fever. The analysts state that this technique decreases the quantity of false prediction and builds the accuracy of predicting dengue fever contrasted with choice trees.

2.3: Self-arranging map and multilayer FFNN :

A mix of oneself arranging guide and multilayer feed-forward neural networks was utilized for the hazard prediction of dengue patients in the Tarig's examination. They bunched patients into two gatherings which are generally safe and high hazard utilizing tree criteria [6]. They utilized just models from Day0 until Day2.they got 70% of rightness. Fatimah Ibrahim et al. [7] predicted the day of defervescence of fever (day0) from 252 dengue patients (4 DF and 248 DHF). Analysts utilized Multi-Layer Perceptrons and got 90% rightness.

2.4: - Multiple Linear Regressions:

In this technique [8], they have done research to see if the climatic things can be connected to arrange the yearly dengue influenced people of Dhaka city of Bangladesh. They got the month to month dengue influenced cases and atmospheric data for 2000-2008 from DGHS(directorate general of wellbeing services).MDD(meteorological branch of Dhaka) of Bangladesh. Furthermore, data for the time of 2000-2007 is utilized to fabricate a model utilizing different direct relapses. Approval of the model is finished utilizing 2001,2003,2005 and 2008. Ordinary month to month moistness, precipitation, the base, and greatest temperature is utilized as self-ruling factors and measure of dengue things report month to month is utilized as the reliant variable. the rightness of the model for ordering ailment found through ROC (collector usable characteristics)curve. Air properties like precipitation, greatest temperature, and relative stickiness were significantly associated with month to month report dengue cases. The model containing environmental data of two slack month clarify 61 percent of the distinction in estimations of report dengue assets and this system was found to arrange dengue illness ($\geq$ 200 possessions) through typical exactness [area in ROC bend = 0.89, 95% CI = (0.89-0.98)]. The characterization system has some weakness in grouping the month to month instances of dengue assets.

2.5 – New Fuzzy Association Rule Mining:

In this strategy [9], they clarify a novel prediction system using Fuzzy Association Rule mining to get connections among climatic, meteorological, medical, and socio-political data from Peru. These affiliations are as principles. The best arrangement of guidelines is naturally select and frames a classifier. This classifier is then used to predict future dengue event as either HIGH (found) or LOW (not discovered) Results: Their programmed procedure manufactured three different fluffy affiliation rule systems. Utilizing the initial two week after week systems, they predicted dengue event three and a month ahead of time, correspondingly. The third prediction is a four-week span, particularly four to seven weeks from time of prediction. By methods for beforehand unused test data for the period 4–7 weeks from time of prediction gives a positive predictive estimation of 0.686, a negative predictive estimation of 0.976, an affectability of 0.615, and an explicitness of 0.982.

2.6-Wrapper-Based Attribute Selection and Decision Tree:

IN THIS METHOD [10], they build up a novel strategy that predicts the finding progressively, that limit the measure of wrong positive and wrong negative qualities. Their technique incorporates three primary advances:

(1) They utilized for contributing missed qualities new strategy that can be utilized on any data comprises of blended dataset (2) Wrapper based characteristic determination techniques are utilized to acquire powerful highlights. That could predict the sickness

(3) The choice Tree model is utilized to create rules. They said their predictive models created are to be more exact than different techniques connected in the discovering dengue fever.

2.7-Time Series Poisson Multi-Variate Regression Model:

In this technique [11], they created and verify a predictive strategy that can predict dengue effects and

offer early alert hint in Singapore. The time arrangement Poisson multivariate relapse technique is created by week after week normal temperature and aggregate precipitation for the length 2000–2010. Atmosphere data is likewise gathered utilizing piecewise straight spline capacities. They investigate different slack occasions among dengue and climate factors to predict the best dengue estimating period. Autoregression, pattern and seasons are estimated in the strategy. They tried the strategy by speculating the dengue cases for week one of 2011 up to week sixteen of 2012 utilizing climate data as it were. The best span for dengue conjecture was four months. Their technique assessed effectively with mistakes of 0.3 and 0.32 of the standard deviation of positive cases during the model structure and avocation time, correspondingly. In the end, they said separate among event and nonoccurrence to a 96% (CI = 93–98%) in 2004–2010 and 98% (CI = 95%–100%) in 2011. The technique gauge the event in 2011 precisely with under 3% probability of false alert.

They have developed a climate based dengue fore-throwing strategy that allows alert four months in right on time of dengue illness with more accuracy. They said that models utilizing temperature and precipitation could be straightforward and value effective measures for dengue determining.

2.8-Artificial Neural Network Model Using Humidity, Temperature, and Rainfall:

This exploration [12] is finished by creating the examples for dengue illness utilizing Artificial Neural Networks They gather the genuine data from the Singaporean National Environment Agency (NEA). This data was utilized to display the habits of dengue patients dependent on the physical traits of mean relative stickiness, mean relative temperature and all out precipitation. The data gathered week after week comprises of dengue positive cases for a time of six years, January 2001 to April 2007.

The Neural Networks model created produces the outcomes in 1000 ages in almost no time. The network is additionally used to predict the 2005 event. The relationship coefficient for the year 2005 period is 0.70. There was a drop in relationship to 0.70 from 0.76. The issue in prediction seen between weeks 34 and 38.

Strategy: The Artificial Neural Network model with Back Propagation calculation is utilized in this exploration [12]. The dataset comprises of 330 week after week estimations i.e., sets of mean relative moistness, mean temperature, complete precipitation and the all out number of dengue positive cases. At that point it is separated into two distinct data subsets called preparing dataset and test dataset, individually. The preparation data is utilized in building up the neural network model and test data is utilized to test the model with train data. The preparation dataset comprises of 104 weeks of data for a long time. For testing the model staying 226 weeks data is utilized. The fake neural network model utilized had three layers i.e., input layer, shrouded layer, and yield layer. They utilized three parameters, they are mean relative moistness, mean relative temperature and all out precipitation. The ANNs yield is the quantity of dengue affirmed cases. The fig beneath demonstrates the plan of ANN. They said that the network is done well in predicting the dengue cases aside from the year 2005.

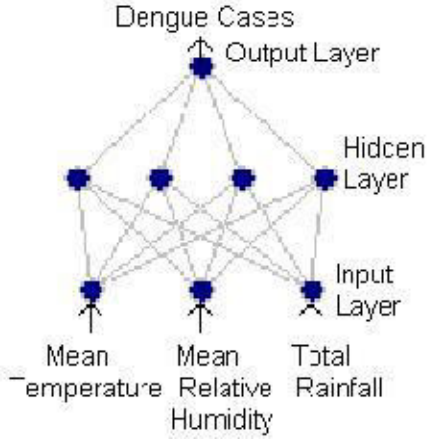


Fig: Artificial Neural Network Model

This ANN model gives the correlation coefficients of 0.84 for training dataset and 0.76 for the test dataset.

III. EXISTING SYSTEM

1) In the current strategies for missing qualities, they utilized mechanized data mining missing worth attribution techniques in every one of the techniques clarified previously. These techniques may fill rough or wrong qualities as a rule. This will influence the last outcomes.

2) In the current strategies for highlight determination, they utilized calculations. This procedure additionally may pick less significant qualities. So this makes the handling time increment. Likewise, it might influence the last outcomes.

3) By utilizing the above techniques they utilized Artificial neural networks(ANN) in view of Humidity, precipitation, and temperature.

IV. PROPOSED SYSTEM

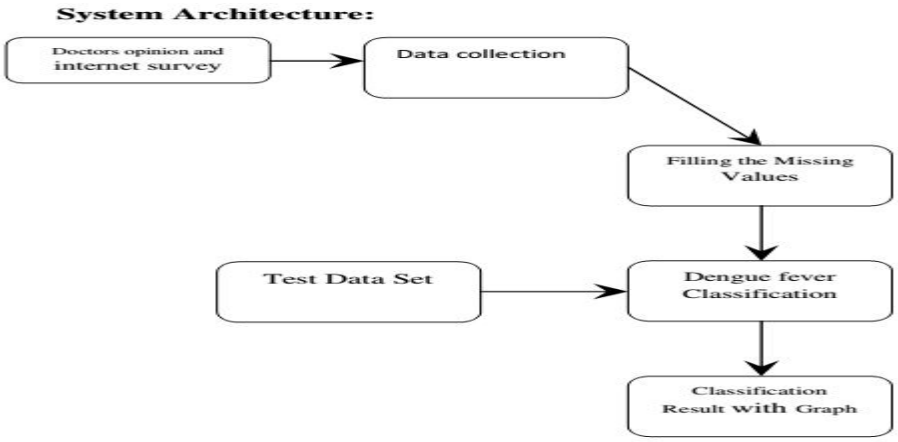
We propose another expert system for predicting dengue fever. Our system comprises of three noteworthy advances.

1) A manual missing quality ascription strategy is utilized. This decreases the bogus esteem passage. With the goal that our outcomes will improve imperceptibly.

2) For choosing the most powerful qualities that predict the dengue fever we took expert specialists assessment and a web review. This procedure decreases gathering pointless qualities during data accumulation. This aides in precise prediction of dengue fever.

3) After preprocessing the data we utilize neural networks for predicting dengue fever. This will be actualized by utilizing MATLAB 2013a.

So as we are normal this strategy gave exact outcomes as clarified in the execution area.



Execution:

1. Missing worth filling

2. Data Representation

3. Disease prediction

1. Missing Value filling:

We visited Hyderabad for gathering data in different medical clinics. During this period we connected with a portion of the specialists to gather assessment and to finish the most powerful ascribes to predict dengue cases. We gathered data through

physically and a portion of the reports from patients. During this period we associated with patients to gather the data identified with clinical data, for example, migraine, regurgitating body torments and so on.,

Data gathered is arranged and missing qualities are filled physically with fitting ordinary qualities. This makes to get the prediction esteems progressively precise.

2. Data Representation

The data gathered is of blended data. That is it incorporates straight out data and numerical data. So for the data to be utilized in MATLAB it should numerical data for neural networks. So we changed over the straight out data to numeric data set by supplanting if the characteristic esteem is YES at that point it is supplanted by 1. In the event that the trait esteem is NO, at that point the esteem is supplanted by 0. Staying numerical qualities kept immaculate.

The dengue fever dataset gathered can be utilized to predict the new dengue fever case. This task is executed by utilizing MATLAB 2013a.

Work process for Neural Network [13]

The Basic strides for neural network configuration are,

1. Collecting required data

2. Develop the network

3. Organize the network

4 Then instate the loads

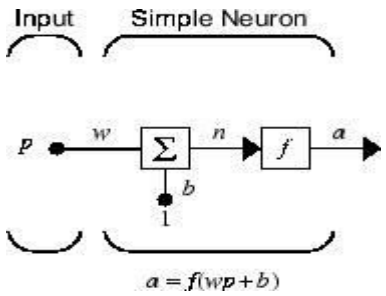
5 Using known dataset the network is prepared

6. Affirm the network utilizing test data

7 at long last we can utilize the network for another dataset

Neuron Model:

Basic neuron is as per the following



In the above model, there are three unique capacities happens. p is the primary capacity, it is scalar information work which is increased by scalar weight w which produces wp scalar item. The second capacity is wp is added to scalar inclination b(wp+b) to acquire net info n. Finally the net information is passed by means of exchange work f. This delivers the yield a.

3. Ailment prediction

ALGORITHAM:

Info: df_dataset

Yield: execution of the neural network

Stage 1: Perform data cleaning

Stage 2: Do data change

Stage 3: select dataset to the contribution of NN

Stage 4: select target data for network yield

Stage 5: Train the network

Stage 6: Validate and test the network

Stage 7: plot the ROC and Confusion bends

Approval and test data:

In this progression, our dataset comprises of 203 examples are separated into three examples preparing tests, approval tests, and testing tests. 143 examples (70%) are utilized for preparing the neural network and the staying 15% each i.e., 30 tests are utilized for approval and testing reason. Preparing: Training tests are exhibited to the network during preparing and the network is balanced by its mistakes.

Approval stage: In this stage tests are utilized to network speculation, and to end preparing when speculation quits improving.

Testing stage: In this stage tests have no impact on preparing network thus give an autonomous proportion of neural network execution during and in the wake of preparing.

Train Network:

Train the neural network to arrange the contributions as indicated by the objectives. Preparing consequently stops when speculation quits improving as shown by an expansion in the mean square mistake of the approval tests.

Results:

	Samples	**MSE**	**%E**
Training	143	$2.49422e^{-2}$	$2.79720e^{-0}$
Validation	30	$6.52483e^{-2}$	$6.66666e^{-0}$
Testing	30	$2.10199e^{-3}$	0

MSE: Mean Square Error is the normal squared contrast among yields and targets. In the event that the MSE esteem is less classifier precision is great.

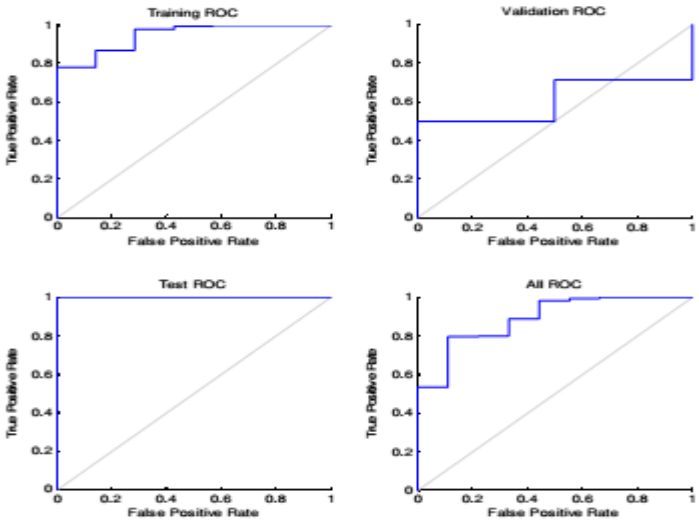
Rate Error: rate mistake shows the piece of datasets that are misclassified. An estimation of 0(zero) implies that no misclassification. 100 demonstrate the most astounding misclassification.

Test Network Results:

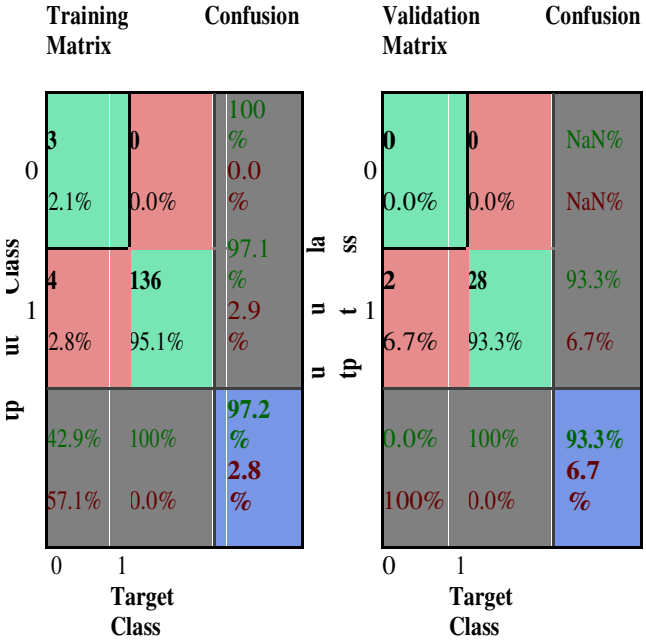
MSE:	$2.75234e^{-2}$
%E:	$2.95566e^{-0}$

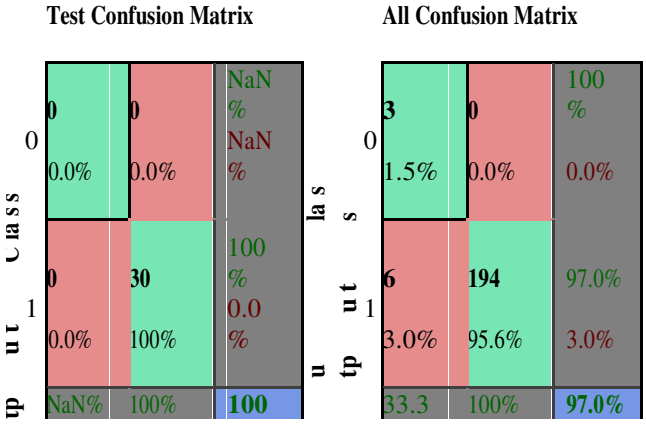
ROC Curves:

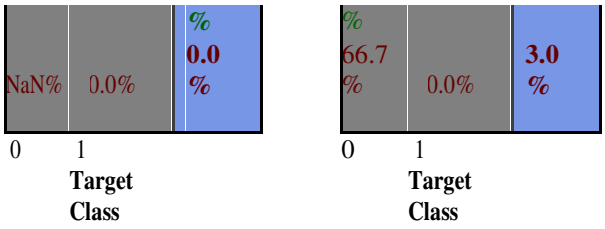
In the collector working trademark is a measure used to guarantee the nature of classifiers. For each class of a classifier, roc applies limit esteems over the interim [0,1] to yields. For every limit, two qualities are determined, the True Positive Ratio (the quantity of yields more noteworthy or equivalent to the edge, partitioned by the quantity of one target), and the False Positive Ratio (the quantity of yields not exactly the edge, isolated by the quantity of zero targets).ROC bends are additionally called relative working attributes [14] on the grounds that it is segregation between two qualities. i.e, True positive rate and false positive rate.



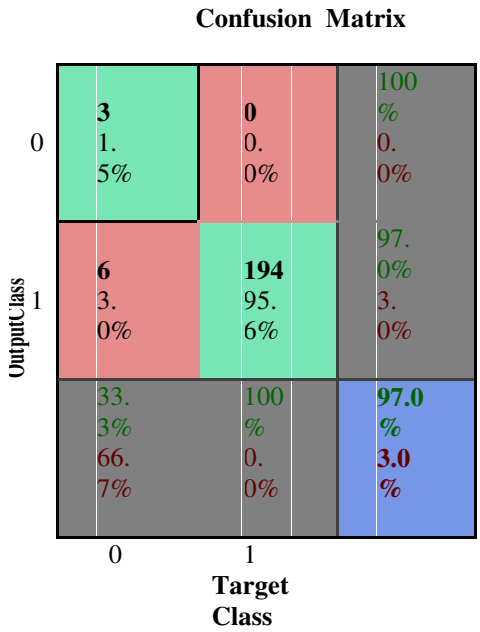
Perplexity grid: Confusion network or it is likewise called a mistake matrix[14]. This grid gives the presentation of the calculation or model. In this, every segment speaks to an occasion in a predicted class while each column speaks to an example in a genuine class and the other way around.







Test confusion:



Conclusion and Future Enhancement

A Neural network model created utilizing Matlab is tried to predict dengue fever cases. This model is utilized to produce execution bends, ROC bends, disarray bends for both preparing and test data this examination demonstrates that this approach is great than different techniques in diagnosing the dengue fever. This model gives an exactness of 100.0% in youngsters and grown-ups utilizing both clinical and research facility highlights. In light of the presentation of the model we finish up and prescribe that this neural network model can be utilized to construct an expert system to predict the new dengue cases in the beginning times.

Later on, this model can be reached out to predict any kind of fever like jungle fever, typhoid, viral fever etc.based on the clinical and research center reports

REFERENCES

[1] D.J. Gubler, "Dengue and dengue hemorrhagic fever," Clin. Microbiol. Rev., vol. 11, pp. 480– 496, 1998.
[2] T. P. Monath, "Dengue: The risk to developed and Developing countries," Proc. Nat . Acad. Sci. USA, vol. 91, no. 7, pp. 2395–2400, 1994.
[3] Thitiprayoonwongse, Prapat Suriaphol and Nuanwan Soonthornphisaj Data Mining of Dengue Infection Using Decision Tree Daranee Latest Advances in Information Science and Applications, ISBN: 978-1-61804-092-3
[4] Revathi N. Prof.S.J.K.Jagadeesh Kumar Genetic Algorithm Optimization And Neural Network For The Diagnosis of Disease. International Journal of Computer Applications & Information Technology Vol. II, Issue I,. January 2013 (ISSN: 2278-7720)
[5] L. Tanner, M. Schreiber, J.G. Low, A. Ong, T. Tolfvenstam, Y.L. Lai, L.C. Ng, Y.S. Leo, L. Thi uong, S.G. Vasudevan, C.P. Simmons, M.L. Hibberd and E.E. Ooi, Decision Tree Algorithms Predict the Diagnosis and Outcome of Dengue Fever in the Early Phase of Illness,PLoS Neglected Tropical Disease, Vol.2, 2008.
[6] T. Faisal, F. Ibrahim and M.N. Taib, A noninvasive intelligent approach for predicting the risk in dengue patients, Expert Systems with Application,Vol.37, No.3, 2010, pp. 2175- 2181.
[7] F. Ibrahim, M. N Taib, W. A. B. Wan Abas, C. G. Chan and S. Sulaiman, A novel dengue fever (DF) and dengue haemorrhagic fever (DHF) analysis using artificial neural network (ANN), Computer Methods and Programs in Biomedicine, No.79, 2005, pp. 273-281.

[8] Md. Nazmul Karim, Saif Ullah Munshi*, Nazneen Anwar & Md. Shah Alam**. "Climatic factors influencing dengue Cases in Dhaka city: a model for dengue prediction". Indian J Med Res 136, July 2012, pp 32-39.
[9] Anna L Buczak*, Phillip T Koshute, Steven M Babin, Brian H Feighner and Sheryl H Lewis Buczak et al"A data-Driven epidemiological prediction method for dengue outbreaks using local and remote sensing data." BMC Medical Informatics and Decision Making 2012, 12:124
[10] Vadrevu Sree Hari Rao, Senior Member,IEEE, and Mallenahalli Naresh Kumar "A New Intelligence- Based Approach for Computer-Aided Diagnosis of Dengue Fever".IEEE Transactions on Information Technology in biomedicine. Vol. 16, no. 1, January- 2012.
[11] Yien Ling Hii1*, Huaiping Zhu2, Nawi Ng1, Lee Ching Ng3, Joacim Rocklo¨ v1Francis Mutuku, DVBNTD/CWRU/Emory University, Kenya.Forecast of Dengue Incidence Using Temperature and rainfall. Received May 6, 2012; Accepted October 2 2012; Published November 29, 2012. PLOS Neglected Tropical Diseases www.plosntds.org. November 2012 | Volume 6 | Issue 11 e1908
[12] B. Gultekin Cetiner a, Murat Sari b and Hani M. Aburas c. Recognition of dengue disease patterns using artificial neural networks. 5th International Advanced Technologies Symposium (IATS'09), May 13-15, 2009, Karabuk, Turkey.
[13] http://www.mathworks.in/help/pdf_doc/nnet/
[14] https://en.wikipedia.org