Life Science Informatics Publications

Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences

Journal Home page http://www.rjlbpcs.com/

RJLBPCS

ISSN 2454-6348

Original Research Article **DOI: 10.26479/2019.0501.32**

DENGUE SEROTYPE CLASSIFICATION USING DISCRETE WAVELET TRANSFORMATION AND MACHINE LEARNING TECHNIQUES

Pandiselvam Pandiyarajan[1*], Kathirvalavakumar Thangairulappan[2]

1. Department of Computer Science, Ayya Nadar Janaki Ammal College, Sivakasi, Tamilnadu, India.

2. Research centre in Computer Science, V.H.N.Senthikumara Nadar College, Virudhunagar, Tamilnadu, India

ABSTRACT: Dengue is a growing problem in tropical countries. It serves diseases especially in children. Different clinical and computerized methods like ELISA, Plateia, haematocytometer, RT-PCR, decision tree algorithms and artificial neural network are used to diagnose the dengue using blood specimen. These methods are not suitable for identifying the dengue serotypes. To overcome the problem, this paper proposes dengue diagnosis method based on nucleotides in the gene sequence, it needs only skin cells or hair or nail which can be collected easily from the patients. The proposed method not only diagnosing the dengue but also classifies serotypes using wavelet coefficients of EIIP indicator sequences. The experimental results demonstrate that proposed method identifies dengue and its serotypes correctly by EIIP indicator sequence, wavelet coefficients and neural network. Some clinical diagnostic methods take 30 – 60 minutes for identifying dengue. The proposed method takes a fraction of seconds for classifying the dengue serotypes. The results of this proposed system can be used for the drug designer for dengue.

KEYWORDS: Dengue diagnosis, classification, wavelet transformation, neural network**.**

Corresponding Author: Mr. Pandiselvam Pandiyarajan*

Department of Computer Science, Ayya Nadar Janaki Ammal College, Sivakasi, Tamilnadu, India

Email Address: pandiselvam.pps@gmail.com

1. INTRODUCTION

Dengue is a viral infection that causes severe joint pain, nausea, abdominal pain, fever lasting several days and in some cases bleeding from the gums, nose or ears and rashes. It can be identified with its distinct serotype namely DENV I (Strain Hawaii) or DENV II (Strain guineqe) or DENV III (Strain

H87) or DENV IV (Strain H241). In Tamilnadu, deaths have been reported due to the dengue rising to 87 included three girl children. Meanwhile, South Delhi Municipal Corporation (SDMA) report said, of the total 4,545 dengue cases, 2152 were residents of Delhi, while the rest were from other states. According to the SDMC, Aedes mosquito breeding has been reported from 1, 80, 687 households in Delhi [1]. Presently, there is no effective drug or antiviral drug against these viruses because it is difficult to differentiate and isolate the infection from other viral infections. In severe cases, people may cause death because of dengue by hemorrhagic fever and multi- organ failure. In 2017, [2] DENV III was causing infection in India, which leads to more hospitalizations. The global burden of dengue is to discover the affected serotypes in dengue patients. Among four serotypes, DENV II and DENV IV are more virulent and cause severe disease. Although infection with one serotype grants lifelong protection antigen against that serotype, it does not necessarily safeguard against a secondary infection with some other serotype. Some clinical diagnostic methods are used to diagnose the later stages of Dengue infections. These methods are based on the detection of IgG and IgM antibodies in the blood. After the infection, IgM becomes unrecognized between thirty and ninety days. Among these periods, the treatment of normal viral fever is given to dengue patients. This leads to severe of dengue infections. The recognition of IgG alone is not enough to confirm the dengue infection without the presence of IgM. Generally, serotypes refer to the subdivision of a virus that is divided based on their surface antigen. Each serotype has its own characteristics and there is no cross production. The burden of dengue in the world is to classify dengue serotypes hence this paper suggests a useful and stable method based on Discrete Wavelet Transformation (DWT) for classifying dengue serotypes of using dengue gene sequence. The paper is structured as follows: Section 2 describes the review of literature, Section 3 presents the classification process using DWT and feedforward neural network, Section 4 describes the experimental results.

2. MATERIALS AND METHODS

Numbers of different disease prediction models are used in medical diagnosis systems which are using data mining and machine learning techniques like Bayesian classification, decision tree, regression model, neural network, single best model and ensemble model. The different machine learning techniques (supervised, semi supervised, unsupervised, deep learning and reinforcement) and classification algorithms (Decision tree, kNN, Artificial Neural Network) are used in diagnosing different diseases such as Heart, dengue and cancer [3]. Raval et al., [4] have observed the current scenario of medical diagnosis system using different data mining techniques. Prediction model was proposed using feedforward neural network for obtaining the optimal result of swine flu diagnosis. A clustering method [5] was proposed for diagnosis of dengue in a human by using the factors like period of fever, fever temperature, rashes or red spots, nausea or vomiting, Low heart rate and fatigue collected from dengue patients, meanwhile many dengue patients are suffered by fever temperature, muscle and joint pain and rashes or red spots. An appropriate method [6] was proposed

for finding protein coding regions, which convert DNA into numerical sequences by EIIP (Electron Ion Interaction Pseudo potentials of nucleotides) values of nucleotides and are applied to discrete wavelet transformation for eliminating the high frequency noise and extracted the period-3 components from gene sequence as per the fixed threshold value of coding regions using Goertzel algorithm. From this method, the discrete wavelet transformation is also used for identifying protein coding and non coding regions of patient's gene sequences. A novel GUI tool [7] was developed for classification of dengue based on genotype and the information in sequences of patients. It classifies a patient's genotype using Naïve Bayes classification algorithm. The results obtained from this method are useful for drug designers. Rocha et al., [8] have studied that DENV I serotype was more prone to present with several clinical and laboratory features as compared with DENV IV patients together with spontaneous bleeding (DENV I: 33.0% and DENV IV: 20.0%), intense abdominal pain (DENV I: 29.7% and DENV IV: 14.1%), neurological symptoms (DENV I :6.7% and DENV IV: 2.2%) and thrombacy topenia (DENV I: 33.7% and DENV IV:18.2%). The immune status measurement of DENV I and DENV IV were same for 202 patients. They have shown that DENV I and DENV IV are more or less same in an antibody response patterns and severity of the diseases. Guzman et al., [9] have reviewed the various methods of identification of dengue. They have also discussed dengue virus pathogenesis, clinical signs and immunological response, laboratory diagnosis of dengue infection, virus isolation, serological testing, MAC ELISA, IgG ELISA, IgM: IgG ratio, neutralizations assays, nucleic acid amplification tests named as Reverse Transcriptase PCR (RT-PCR), Real Time RT-PCR, and antigen detection. They have also provided dengue control and prevention strategies. Saha et al., [10] have proposed three artificial neural network models for diagnosing and identifying the dengue-affected patient's data like high fever, chills, break bone aching, headache, sore throat, prostration and malaise from Jalpaiguri Sadar hospital, North Bengal, India. Ibrahim et al., [11] have developed the prediction system using multilayer feedforward neural networks based on clinical symptoms and signs of Dengue Fever (DF) and Dengue Hemorrhagic Fever (DHF). Artificial Neural Network (ANN) was trained for finding the regions and factors for spreading of dengue diseases by analyzing the factors mean temperature ranges from 23^{oC}– 31^{oC}, mean relative humidity and total rainfall ranges from 1,969 mm to 3,394 mm. The network is trained and tested with the real data obtained from Singaporean National Environment Agency (NEA) and the city of Iloilo Philippines [12]. The warning system of dengue made to predict the future outbreaks in Srilanka [13] and Jember [14] based on risk factors. ANN-based dengue diagnosing system [15] used for identifying the severity of dengue virus in microscopic images of blood cells. Existing methods are used to diagnose and classify the dengue infections based on various factors like high fever, chills, break bone aching, headache, sore throat, prostration, malaise, period of fever, fever temperature, rashes or red spots, nausea or vomiting, Low heart rate and fatigue obtained from dengue patients. The proposed system uses the gene sequence

of dengue patients as the gene sequence is responsible for making proteins and the functions of all living things. In the proposed method, the gene sequence is converted into EIIP indicator sequences based on EIIP values of nucleotides. The EIIP indicator sequence is composed into approximation and detailed coefficients using wavelet transformation. Each approximation coefficients are given to chi- squared based feature selection method for selecting high frequency genes in the EIIP indicator sequence. The selected coefficients are processed using feedforward neural network for classifying the dengue patients as DENV I, DENV II, DENV III or DENV IV. Proposed method has three segments i) preprocessing ii) feature selection and iii) classification of dengue serotypes using neural network. The procedure of the proposed method is depicted in Figure .1

Pre-processing

Numerical representation of biological sequences

The biological sequences as signals are to be encoded into a suitable format for data analysis and data mining tools. This is usually achieved by assigning a numeral to each symbol that forms the biological sequences. There are two fundamental kinds of biological sequences namely DNA nucleotide sequences and protein amino acids sequences relevant to dengue diagnosis. EIIP value of nucleic acid is a physical quantity that denoting the mean energy of valence electron [16] and also used to find the protein coding regions (hotspots of protein). In this proposed work, the DNA nucleotide sequence is used as the input and is converted into EIIP indicator sequences by replacing the nucleotides with its EIIP values as A=0.1260, G=0.0806, C=0.1340 and T=0.1335.

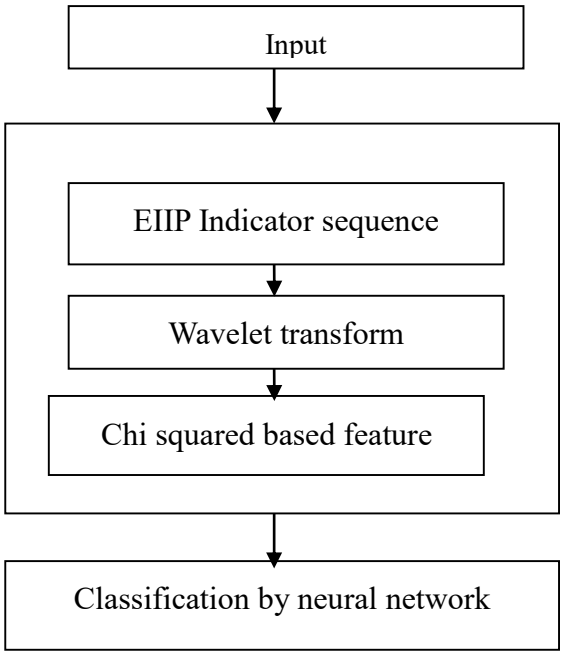

Figure 1. Proposed method for classifying dengue serotypes

Calculation of Wavelet Coefficients

Nowadays many digital signal processing (DSP) methods are used to identify the protein coding regions and also used to remove the background noise in genomic sequences [17,20]. In the proposed method, the discrete wavelet transformation (DWT) is used to find the highest frequency genomic signals. In this DWT, the approximate (LL) form in both horizontal and vertical directions, details in horizontal direction (HL) alone, details in vertical direction (LH) alone and details in both horizontal and vertical directions (diagonal edges) (HH) are extracted from the EIIP indicator sequences of dengue patients. This low and high pass filtering of genomic signals require the use of following filter functions through the multiplication of separable scaling and wavelet functions in h1 (horizontal) and v1 (vertical) directions [18,21,22,23,24].

$$\Phi(h1, v1) = \Phi(h1)\Phi(v1) \quad (1)$$

$$\varphi^h(h1, v1) = \varphi(h1)\Phi(v1) \quad (2)$$

$$\varphi^v(h1, v1) = \Phi(h1)\varphi(v1)$$

$$\varphi^d(h1, v1) = \varphi(h1)\varphi(v1) \quad (4)$$

where $\Phi(h1, v1), \varphi^h(h1, v1), \varphi^v(h1, v1)$ *and* $\varphi^d(h1, v1)$ denote the approximated genomic signals, genomic signal with horizontal details, genomic signals with vertical details and genomic signals with diagonal details respectively. The wavelet coefficients of decomposed genomic signals [5] are calculated using (5) and (6).

$$W_\Phi(j_0, k_1, k_2) = \frac{1}{\sqrt{h1, v1}} \sum_{h1=0}^{H1-1} \sum_{v1=0}^{V1-1} s(h1, v1)\Phi_{j_0, k_1, k_2}(h1, v1) \quad (5)$$

$$W_\varphi(j_0, k_1, k_2) = \frac{1}{\sqrt{h1, v1}} \sum_{h1=0}^{H1-1} \sum_{v1=0}^{V1-1} s(h1, v1)\ \varphi^i_{j_0, k_1, k_2}(h1, v1) \quad (6)$$

where i {h, v, d} denotes the detailed wavelet coefficients of horizontal, vertical and diagonal directions of genomic signals. $s(h1, v1)$ represents separable scaling function of wavelet [25, 26, 27, 28, 29, 30]. In this proposed work, Approximation wavelet coefficients are only considered for further processing.

Feature selection

Feature selection is the process of selecting the informative approximation wavelet coefficients

using chi-squared test value V from (7). Large value of V indicates there exists association between disease and exposure, small value of V indicates no association exists between disease and exposure [16, 20]. The existence of informative wavelet coefficient R is computed based on the *Mc* using (8). If chi-squared value is greater than Mc then set R with1 otherwise set R with 0. The wavelet coefficients are informative when R=1. These informative wavelet coefficients are used as features to identify dengue serotypes.

$$V = \frac{(observed\ \ value - expected\ \ value)^2}{expected\ \ value} \quad (7)$$

$$R = \begin{cases} 1 & if \quad V > Mc \\ 0 & otherwise \end{cases}$$

where $Mc = \sum_i^n v/n$ denotes mean of V, n represents number of approximation wavelet coefficients. R is the binary value for indicating the informative wavelet coefficients.

Classification of Dengue serotypes

Single hidden layer feedforward neural network is considered to classify the dengue serotypes into DENV I, DENV II, DENV III and DENV IV. The sigmoidal activation function is used in the hidden layer. The linear activation function is used in the output layer. X= [x_1 ..... x_n, 1] is the input pattern to the network. W is a matrix connecting input layer and hidden layer. Z is a matrix connecting hidden layer and an output layer. H and Y are vectors that represent the output of the hidden and output layer.

$$net_h = \sum XW, \qquad net_o = \sum HZ \quad (9)$$

$$H = f(net_h), \qquad Y = f(net_o) \quad (10)$$

$f(net)$

$$= \frac{1}{1 + e^{-net}}$$

Backpropagation algorithm is used to train the single hidden layer feedforward network. This algorithm is learned by samples and back propagates the error from the output layer. Weights are adjusted according to the deviation of errors. The procedure of backpropagation algorithm is shown below.

Input: set of patterns, each with input vector *x* and output vector *y,* single hidden layer feedforward network, weights W, V, activation function *f*

repeat

for each pattern p **do**

{

for each node j in the hidden layer **do**

{

$$net_j = \sum_{i=1}^{ni} x_i w_{i,j}$$

$$h_j = f(net_j)$$

$$= \frac{1}{1+e^{-net_j}}$$

}

for each node k in the output layer **do**

{

$$net_k = \sum_{j=1}^{nh} h_j w_{j,k}$$

$$y_k = f(net_k)$$

$$= \frac{1}{1+exp^{-net_k}}$$

$$err_k = expec_k - y_k$$

$$E = \frac{1}{2} \sum_{p=1}^{np} \sum_{k=1}^{no} err_k^2$$

}

$$new\ \ V = old\ \ V + \ \ \lambda\ \ \frac{\partial E}{\partial V}$$

$$new\ \ W = old\ \ W + \ \ \lambda\ \ \frac{\partial E}{\partial W}$$

}

Until (error of the network within required accuracy)

Note: *ni, nh, no, np* represent number of input neurons, number of hidden neurons, number of output neurons and number of patterns respectively. $expec_k$ represents expected value of neuron k and λ represents learning parameter. *W* is a weight matrix connecting input layer and hidden layer and *V* is a weight matrix connecting hidden layer and output layer.

Evaluation Metrics

TP is the True Positive, which are the serotypes of dengue patients correctly detected. FN is the False Negative, which are the serotypes of dengue patients wrongly detected. The overall prediction

accuracy *CC* is calculated using (12).

$$CC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (12)$$

Sensitivity (S_n) is the factor for calculating the correctly identified serotypes. Specificity (S_p) is a factor for calculating the wrongly identified serotypes. S_n and S_p are calculated using (13), (14).

$$S_n = \frac{TP}{TP + FN}$$

$$S_p = \frac{TN}{TN + FP}$$

3. RESULTS AND DISCUSSION

The proposed work has been implemented using Matlab 13. The gene sequences of 40 dengue patients are taken from NCBI (National Centre for Bioinformatics) [19]. Among these sequences, 9 patients were affected with DENV I, 20 patients were affected with DENV II, 17 patients were affected with DENV III and 5 patients were affected with DENV IV. Generally, single gene sequence dataset contains plenty of nucleotides (A, C, G, and T) at frequent intervals. Similarly, each gene sequence of dengue patients is having more than **10563** nucleotides. These sequences are converted into EIIP indicator sequences. For example, the sequence **{A, C, T, G, T, A, C, G}** is converted into **{0.1260, 0.1340, 0.1335, 0.0806, 0.1335, 0.1260, 0.1340, 0.0806}.** The obtained sequences are treated as genomic signal for processing and are represented in Figure 2. The genomic signals are applied on discrete wavelet transformation. In these transformations, approximation wavelet coefficients, detailed coefficients of horizontal, vertical and diagonal directions are obtained from (5) and (6). From these coefficients, approximation wavelet coefficients are used as the input for feature selection phase and are represented in Figure 3, 4, 5 and 6. For each patient, **10563** approximation wavelet coefficients ($\boldsymbol{W}_{\Phi}$) of each serotype are considered for the feature selection task. After removing missing and unsuitable wavelet coefficients, the size of genomic signals becomes **8166**. Chi squared value (V) is calculated for each and every approximation coefficients (**8166**) of the dengue patients (**40**). If an approximation wavelet coefficient has greater value of mean value of **V** then set R to **1**, otherwise set **R** value to **0**. The coefficient which has **R** value as **1** is selected for training the neural network. Among **8166** approximation wavelet coefficients, **553** approximations wavelet coefficients are selected for training the neural network. The same processes are performed for **40** dengue patients. After that, **553** approximation wavelet coefficients were selected for **40** patients by feature selection phase. The selected approximation wavelet

coefficients are used as the input for single hidden feedforward neural network. The neural network is trained with dynamic learning rate. The classification result is compared with the class labels as per the NCBI and is represented in Figure 7. According to the obtained result, patients affected with DENV I, DENV II and DENV III are correctly classified but among 5 DENV IV patients only 4 patients are correctly classified. The proposed method provides sensitivity as **97.8%** and specificity as **2.2%.** The learning curve of the proposed neural network is depicted in Figure 8**.** Some clinical diagnostic methods take 30 – 60 minutes for identifying dengue. The proposed method takes a fraction of seconds for classifying the dengue serotypes. The results of this proposed system can be used for the drug designer for dengue.

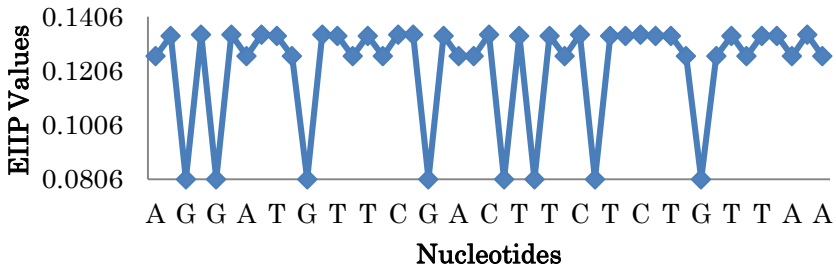

Figure 2. Original genomic signals

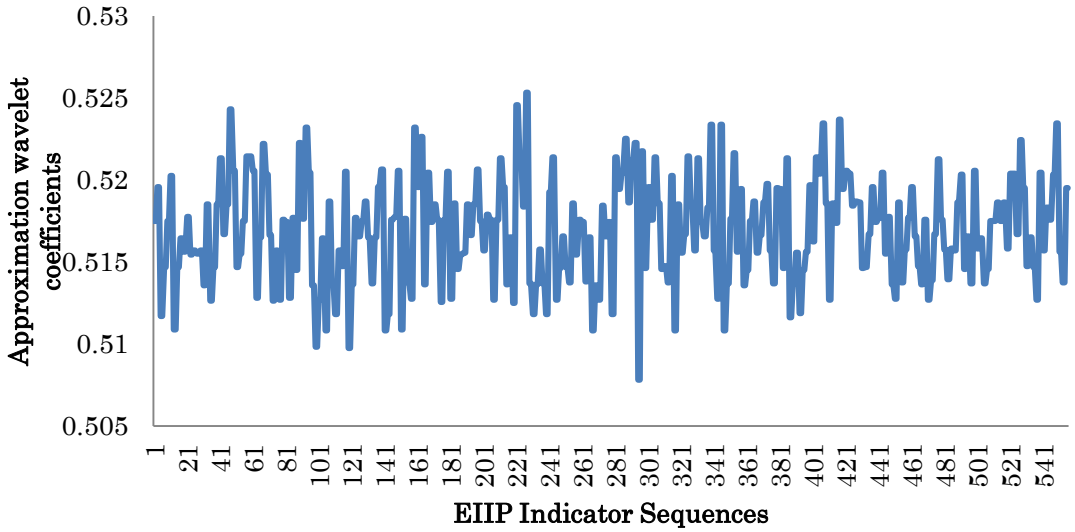

Figure 3. Approximation Wavelet Coefficients of DENV I patients

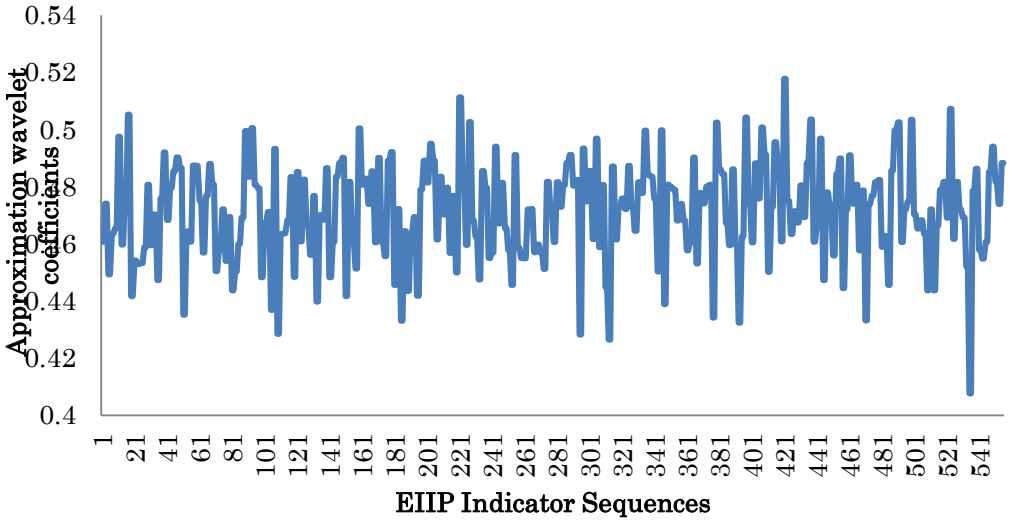

Figure 4. Approximation Wavelet Coefficients of DENV II patients

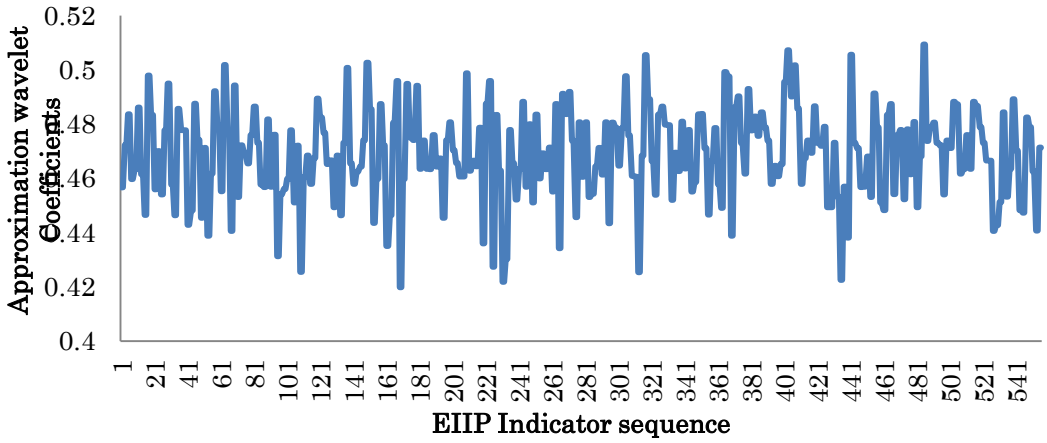

Figure 5. Approximation Wavelet Coefficients of DENV III patients

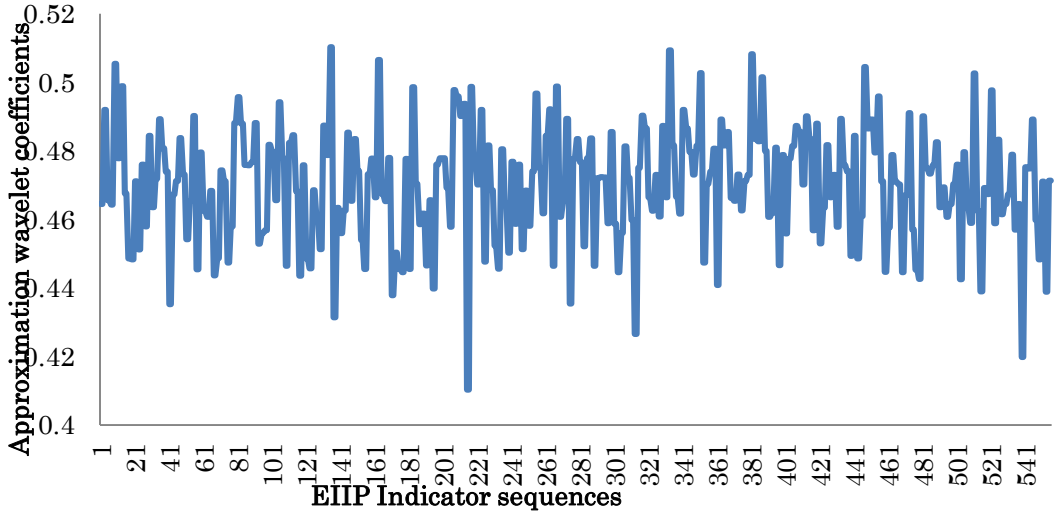

Figure 6. Approximation Wavelet Coefficients of DENV IV patients

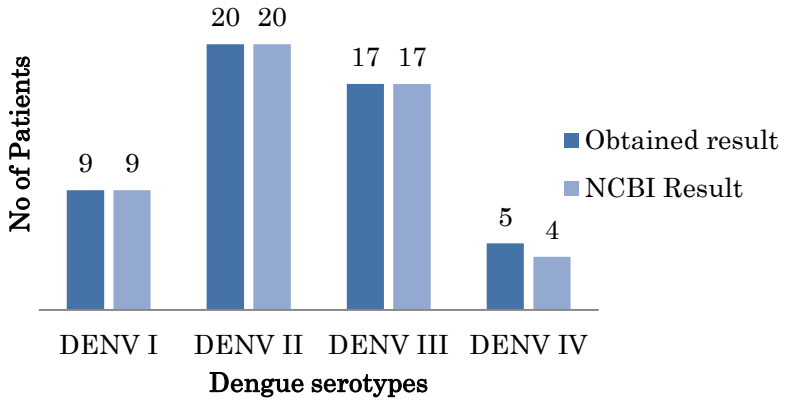

Figure 7. Obtained result and NCBI result

The benchmarks classification algorithms such as Naïve Bayes, Multilayer Perceptron, Iterative Dichotomiser (ID3), Decision tree algorithm (J48) and Random Forest algorithm are evaluated for the same dengue patient's gene sequences using Weka tool. After that, sensitivity and specificity for all benchmarks algorithms are calculated using (11), (12) and (13) and are shown in Figure 9. As per the error rates of all algorithms, the proposed method provides better performance for classifying dengue serotypes using dengue patient's gene sequence.

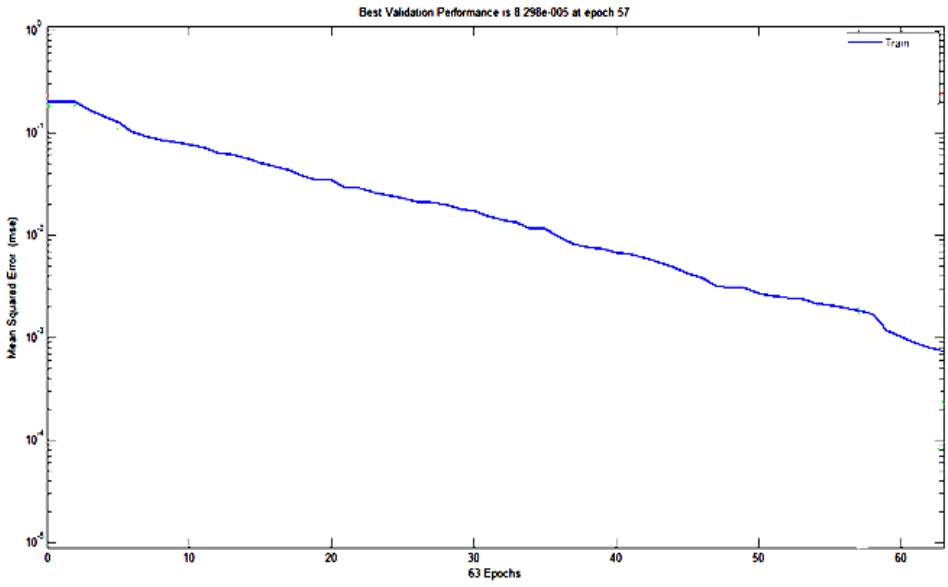

Figure 8. Learning curve

Figure 9. Comparative Chart

4. CONCLUSION

The Discrete wavelet transformation and neural network classification of gene sequence play an important role for classification of dengue serotypes as DENV I, DENV II, DENV III and DENV IV after converting nucleotide into EIIP values. The proposed method does not need to collect blood specimen but it needs only gene sequence which can be obtained from any patients at anytime from their hair or nail or skin. This method provides better result for classifying serotypes of the patients using gene sequence.

CONFLICT OF INTERST

Authors have no any conflict of interest.

REFERENCES

1. www.india.com/news/india/dengue-outtbreak.amp.
2. www.hindustantimes.com.
3. Razia S, Prathyusha PS, Krishna NV, Sumana NS. A review on disease diagnosis using machine learning techniques. Int. J. pure. Appl. Math. 2017; 117(16):79-85,.
4. Raval D, Bhatt D, Kumbar MK, Daiwat Vyal VP. Medical Diagnosis System using machine learning. IJCSC. 2016; 7(1):177-182.
5. Kaur RD, Gupta G, Bhathal GS. Hemographic analysis of dengue fever using data mining. Int. J. Adv. Res. Comp. Sci. 2017; 8(7):1233-1237.
6. Saberkari H, Shami M, Heravi H, Sedaagh H. A fast algorithm for exonic regions prediction in DNA sequences. J. Med. Sig. Sensor. 2013; 3(3):139-149.
7. Ali I, Humayur F, Azam S, Munir A, Riziwar M, Mahmood A, Adnan M. Computational tool for classification of dengue virus. J. Appl. Bioinformatics and Computational biology. 2017; 6(3):1-10.
8. Rocha BAM, Guilarde AO, Argolo AFLT, Tassara MP, Da Silveira LA, Junqueira IC, Turchi MD, Feres VCR, Martelli CMT. Dengue specific serotype related to clinical severity during the 2012/2013 epidemic in centre of Brazil. Inf. Dis. Poverty. 2013; 6(116): 2-11.

9. Guzman MG, Halstead SB, Artob H, Buchg P, Farrar J, Gubler DJ, Hunsperger E, Krooger A, Margolis HS, Martinez E, Nathan MB, Palegrino JL, Simmons C, Yoksan S, Peeling RW. Dengue: a continuing global threat. Nat. Rev. Microbiology. 2010; 8(12):1-10.
10. Saha, Mandal. Detection of dengue diseases using artificial neural networks. Int. J. Comp. Sci. Engg. 2017; 5(3): 65-68.
11. Ibrahim F, Taib MN, Baker wan Abas WA, Guan CC, Sulaiman S. A Novel dengue fever (DF) and dengue hemorrhagic fever (DHF) analysis using artificial neural network (ANN). Computer methods and program in biomedicine. 2005; 79: 273-281.
12. Gultenkin BC, Sari M, Aburas HM. Recognition of Dengue Disease Patterns using Artificial Neural Networks. paper Presented in 5th International Advanced Technologies Symposium (IATS), Turkey. 2009.
13. Munassinghe A, Premaratne HL, Fernando MGNAS. Towards an early warning system to combat dengue. Int. J. Comp. Sci. Elec. Engg. 2013; 1(2): 252-256.
14. Roziqin MC, Basuki A, Harsona T. Parameters data distribution analysis for dengue fever outbreak in Jember using monte carlo. Int. J. Comp. Sci.Soft. Engg. 2016; 5(4): 45-48.
15. Subitha, Padmapriya. Diagnosis for dengue fever using spatial data mining. Int. J. Comp. Trends. Tech. 2013; 4(8): 2646 – 2651,
16. www.ocw.jhsph.edu/courses/fund epiII/PDFs/Lecture17.pdf
17. Inbamalar TM, Sivakumar R. Filtering Approach to DNA Signal Processing. paper presented IACSIT Coimbatore Conferences IPCSIT 28, IACSIT Press, Singapore, 2012.
18. Stollnitz EJ, Derose TD, Salesin DM. Wavelets for computer graphics: A primer. Technical report 94-09-11. Department of computer science and engineering, University of Washington, Seatle, Washington, 1994.
19. National Centre for Bioinformatics www.ncbi.com.
20. Yin C, Yau SST. Prediction of protein coding regions by the 3-base periodicity analysis of a DNA sequence. J. Ther. Bio. 2007; 247: 687-694.
21. Haimovich AD, Byrne B, Ramasamy R, Welsh JW. Wavelet analysis of DNA walks. J. Comp. Bio. 2006; 13(7): 1289 – 1298.
22. Zhou X, Li Z, Dai Z, Zou X. Predicting promoters by pseudo –trinucleotide compositions based on discrete wavelets transform. J. Ther. Bio. 2013; 319:1-7.
23. Sharma R, Kaushik A. Exons identification using filter. Int. J. Engg. Res. Appl. 2012; 2(1):295-298.
24. Abbasi O, Rostami A, Karimian G. Identification of exonic regions in DNA sequences using cross correlation and noise suppression by discrete wavelet transform. BMC Bioinformatics. 2011; 12(430): 1-10.

25. Das J, Barman. DSP based entropy estimation for identification and classification of homo sapiens cancer genes. Springer-verlag. 2016.
26. Ahmad M, Jung LT, Bhuiyan AA. Biological inspired fuzzy adaptive window median filters (FAWMF) for enhancing DNA signal processing. Comp. Meth. Prog. Bio. Med. 2017; 149: 11-17.
27. George TP, Thomas T. Exon mapping in long noncoding RNAs using digital filters. Geno. Insights. 2017; 10:1-12.
28. Inbamalar TM, Sivakumar R. Improved algorithm for analysis of DNA sequences using multiresolution transformation. The. Sci. World. J. 2015.
29. Saberkari H, Farsani MS, Aminkar S. An efficient algorithm for small gene prediction in DNA sequences. Int. J. Comp. Vis. Sig. Proc. 2016; 6(1):1-10.
30. Zare F, Dow M, Monteleone N, Horny A, Nabavi S. An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. BMC. Bioinfo. 2017; 18(286): 1-13.