Journal of Theoretical and Applied Information Technology
31st December 2020. Vol.98. No 24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DATA-DRIVEN NEURAL NETWORK MODEL FOR EARLY SELF-DIAGNOSIS OF DENGUE SYMPTOMS

[1]NOR AZURA HUSIN, [2] NURUL IMAN SAIFUL BAHARI, [1]HAZLINA HAMDAN, [3]NOOR AFIZA MOHD ARIFFIN, [1]TEH NORANIS MOHD ARIS

[1] Intelligent Computing Research Group, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia

[2] Institute of Tropical Agriculture and Food Security, Universiti Putra Malaysia, Malaysia

[3] Security Computing Research Group, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia

E-mail: [1]n_azura@upm.edu.my, [2]nuruliman.sf@gmail.com, [1]hazlina@upm.edu.my, [3]noorafiza@upm.edu.my,[1] nuranis@upm.edu.my

ABSTRACT

Dengue fever is one of the main public health concerns and endemic diseases in many countries, especially in tropical and subtropical regions. In severe cases, people infected with dengue may experience severe bleeding which may lead to death if the infection is not properly treated. It is a standard procedure when a person is referred to the hospital with a high fever for more than two days to be required to undergo a dengue fever screening at the triage before further clinical tests are done to confirm the patient's medical condition. Thus, an individual need to conduct an early self-diagnosis to identify the probability that he/she have been infected with dengue fever and further seek professional help from medical practitioners. There are many dengue fever symptoms outlined by the World Health Organization (WHO) such as sudden high fever, headache, abdominal pain, persistent vomiting, rapid breathing, bleeding gums, fatigue, restlessness, and vomiting blood, but the identification of the highly significant symptoms among the less significant symptoms are still scarce. Identification of significant symptoms from the dengue dataset may help patients and medical practitioners to acknowledge the alarming symptoms and endeavour for immediate action to prevent dengue outbreak and fatality. Hence, the identification of significant dengue symptoms may assist in system development to determine the weightage of each attribute in the system. This may result in better prediction of dengue. Therefore, the objective of this study is to develop an early self-diagnosis system using an artificial neural network with the ability to produce a reliable result based on the identification of significant symptoms. The model accuracy of 100% indicates the high reliability of the developed early self-diagnosis system. A mobile application was developed based on the prediction model for patients and medical practitioners as the target users. This study contributes to the field of public health by providing early detection for people who are at risk of being infected by dengue disease. The advancement in early detection technology brings such a huge positive impact in healthcare. Early identification of significant symptoms ensures that better focus can be given to the identified symptoms for a more reliable dengue fever assessment.

Keywords: *Dengue Fever, Diagnose, Early Self-Diagnosis, Machine Learning, Significant Symptoms.*

1. INTRODUCTION

Knowledge discovery from medical datasets is necessary in order to understand the hidden information about a particular disease. As for dengue fever (DF) cases, the dengue symptoms dataset can be obtained from the public health department. In recent decades, dengue cases have been an alarming disease as the dengue cases keep on increasing each year [1] DF is a mosquito-borne disease that is transmitted by the *Aedes* mosquitoes carrying a type of virus known as *Flaviviridae*. Once transferred on a human body, after 2 to 7 days of infection, the dengue virus can cause symptoms such as sudden high fever, headache, abdominal pain, persistent vomiting, rapid breathing, fatigue, and restlessness [2]. Severe dengue known as dengue haemorrhagic fever (DHF) is potentially life-threatening and there were fewer cases reported for DHF rather than DF. DHF symptoms such as inner mouth bleeding and consistent vomiting indicate that the patients are already in severe stages. World Health Organization indicated that early detection of DF/DHF and appropriate

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

treatment will decrease fatality rates from more than 20% to less than 1% [2]. Numerous research studies that can be referred for successfully and effectively implement early detection as an approach to mitigate risk especially for public health risk [3].

Considering the potentially fatal consequences of DF/DHF, the early symptoms of DF/DHF should not be neglected. There is a need to do an early self-diagnosis to prevent complications and reduce dengue mortality. The early symptoms should not be neglected before the occurrence of severe symptoms such as inner mouth bleeding and consistent vomiting. If the individual is already in severe cases, the chance of fatality is higher. There have been many models developed to assist the patients and medical practitioners in early diagnosis of dengue fever, but early self-diagnosis of dengue fever is still scarce. Most of the developed models make use of the patient's symptoms together with patients' medical reports as their attributes. In reality, early self-diagnosis is crucial for an individual to know if the fever that they have are potentially DF or not just by referring to their physical symptoms.

Moreover, identification of significant symptoms from the dengue dataset may help patients and medical practitioners to acknowledge the alarming symptoms and pay more attention towards these symptoms. A study done by [4], highlighted the importance of identification of important symptoms linked to persistent symptoms of dengue patients. The finding of the study shows that, hepatomegaly, abdominal pain and thrombocytopenia, and dermatological manifestations as the most important symptoms for dengue patients, one month after the onset of fever. While, age and health status were highly associated with the prolonged symptoms. The identification of these important symptoms enables medical practitioners to advice the patients on the possible persistent symptoms after the dengue fever episodes and pay more attention in order to monitor these symptoms among patients after they leave the treatment centre.

High accuracy of the early self-diagnosis model may help users to seek medical assistance when they are diagnosed with dengue. However, the identification of significant symptoms in order to build an accurate early self-diagnosed system is far from knowledge. A set of data has been obtained from a public health specialist in to gain knowledge about the significant DF symptoms. Data mining using an artificial neural network was also done onto the datasets to identify the weightage for each symptom. Identification of significant DF symptoms from expertise and data mining may help in developing the highly accurate DF early self-diagnosis system. Therefore, this study aims to improvise the DF early self-diagnosed system developed by [5] by adding the identified weightage for each attribute.

In the following section, the previous related work shall be briefly discussed, followed by an explanation about the materials and methods used in this study. Next, the model design section will elaborate on the data pre-processing and network design. The network design section will further explain the four general steps, i.e. initialisation, activation, weight training, and iteration. Then, the implementation of the rules will be discussed from developing the expert system framework to the early detection system. Finally, the findings will be evaluated and discussed.

2. RELATED WORK

Machine learning has been a great help in obtaining information from a set of data [6,7,8]. Various machine learning method has been implemented to achieve the accurate result in predicting disease diagnosis, especially on dengue. Table 1 represents the summaries of related works to the prediction of dengue fever based on symptoms. A review done by [9] on analysis of machine learning techniques for dengue disease detection reveals that there have been many applications of machine learning algorithms applied such as support vector machine, decision tree, naive Bayes, and artificial neural network. The developed prediction models have produced significant accuracy and are highly reliable.

[10] developed a fuzzy inference system for early diagnosis of dengue disease. It is proven that the system was able to produce high accuracy prediction with inputs of physical symptoms i.e. fever, nausea, diarrhoea, vomiting, headache, pains, skin rash, and medical test reports i.e. white blood count, platelet count, and more. [11] utilized WEKA data mining tools to classify dengue attributes data into five classification types i.e. Naïve Bayesian, REP Tree, Random tree, J48, and SMO. The attributes are fever, bleeding, flu, myalgia, and others. All the classifiers were able to produce high accuracy with Naïve Bayesian selected as the best classifier with an accuracy of 92%. [12] developed a prediction framework for early diagnosis of dengue disease using decision tree classifiers. The significant attributes were

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

identified in order to increase the prediction accuracy. The identified attributes are fever, headache, platelets, loss of appetite, blood pressure etc. The classifier was able to offer an accurate prediction with an accuracy level of 86.13%. [13] utilises three machine learning approaches in order to come up with an early detection technique of dengue disease. A total of 16 attributes were used in this study and further categorised into the non-clinical and clinical attributes. The non-clinical attributes are age, gender, vomit, abdomen pain, chills, body ache, headache, weakness, and fever. While, the clinical attributes are platelet count, temperature, heart rate, dengue antigen, IgM, IgG, and Dengue NSI Antigen. The three machine learning methodologies such as an artificial neural network (ANN), decision tree (DT), and naive Bayes (NB) are able to disclose an acceptable accuracy of 79.09%, 73.63%, and 76.36%, respectively. The study determined that the ANN-based diagnostic model is more suitable and reliable for early-stage dengue diagnosis.

Despite all the advantages of the produced model [10, 12, 13] the developed models are only suitable for the patient's early diagnosis with the assistance of the patient's symptoms together with the clinical results. While, early self-diagnosis potential has not yet been fully explored for their significant symptoms of the disease, accuracy, and reliability.

Identification of significant symptoms is one of the important steps to be considered to improve the model's accuracy. A study done by [14] highlighted the importance of significant symptoms identification for a more accurate early detection model. Initial physical symptoms for SARS-Cov-2 infected patients were collected comprising of gender, age, fever, cough, pneumonia, lung infection, runny nose, muscle soreness, diarrhoea, travel history and isolation. Five machine learning algorithms were used, and it is found out (in descending order) that lung infection, cough, pneumonia, runny nose, travel history, fever, isolation, age, muscle soreness, diarrhoea and gender are the important symptoms for this disease. The model can predict the disease at the early stage of infection with accuracy level of more than 85%.

In addition, the early self-diagnosis model is supposed to only utilise the physical symptoms of each patient before they went to seek medical assistance. Previously, [5] developed a Fuzzy Rules Base System for early self-diagnosis of dengue symptoms with 100% accuracy. The general rule generated was simplified as high fever symptom + more than 1 other symptoms + period of infection more than 2 days = positive, else = negative. The general rule was generated with the expert view from experienced doctors. Although the general rule generated results with 100% accuracy, it did not distinguish between the significant and the less significant symptoms. There is still a research gap in data exploration to identify the significant DF symptoms. Identifying the significant symptoms may help in increasing the accuracy of the developed model. Artificial neural network (ANN) is one of the most utilized machine learning techniques and produces a significantly high accuracy prediction model for disease diagnostic study [9,13,15]. Besides that, ANN is able to extract the information about the weight for each attribute in its model. This information can help the user to further understand the level of significance for each input.

3. EARLY SELF-DIAGNOSIS MODEL

The dengue symptoms data were collected from Aden Hospital in Yemen. There have been repeated large outbreak of dengue cases in Yemen over the last 10 years and the number of cases keeps increasing to an alarming number [16,17]. This fever has been declared as a pandemic and more public health awareness on early diagnosis of DF is crucial [18].

The dataset consists of 1054 cases with 15 attributes that were collected from patients attending the hospital with dengue symptoms in 2015. The datasets consisted of a combination of 15 attributes for DF and DHF symptoms as shown in Table 2. The dengue symptoms from patients were recorded as categorical data either 'yes' or 'no'. There is one target attribute for each patient case known as 'Status' categorized into positive dengue or negative dengue.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Summaries of the Literature Review on Dengue Fever Prediction.

Author	Study/ Finding	Method	Attributes		Performance	Limitation
			Physical symptoms	Medical test reports		
Shaukat *et al.* (2015)	Best classifier = Naïve accuracy 92%.	Naïve Bayesian, REP Tree, Random tree, J48 and SMO.	Fever, bleeding, flu, myalgia, and others.		The classifiers produced high accuracy range of 76% to 92%.	Potential to utilise ANN for better accuracy.
Saikia and Dutta (2016)	Fuzzy inference system for early DF diagnosis.	Fuzzy inference system.	Fever, nausea, diarrhoea, vomiting, headache, pains, and skin rash.	White blood count, platelet count, and more.	The system able to distinguish the result into No dengue/Probable dengue/Dengue confirm.	Not suitable for DF early self-diagnosis.
Anitha and Wise (2018)	Prediction framework for early diagnosis of dengue disease.	Decision tree classifiers.	Fever, headache, platelets, and loss of appetite.	Platelets, blood pressure and etc.	Accurate prediction with an accuracy of 86.13%.	Not suitable for DF early self-diagnosis.
Gambhir *et al.* (2018)	Early detection and precise diagnosis of dengue disease.	Artificial neural network (ANN), decision tree (DT), and naive Bayes (NB).	Age, gender, vomit, abdomen pain, chills, body ache, headache, weakness, and fever.	Platelet count, temperature, heart rate, dengue antigen, IgM, IgG, and Dengue NSI Antigen.	Accuracy for ANN = 79.09%, DT = 73.63%, NB = 76.36%.	
Husin *et al.* (2018)	Fuzzy Rules Base System for early self-diagnosis of dengue symptoms.	Fuzzy logic.	Sudden high fever, severe headaches, eyes pain, muscle pain, fatigue, nausea, vomiting, skin rash, loss appetite, abdominal pain, persistent vomiting, inner mouth bleeding, exhaustion, clinical fluid accumulation, and period by days.		Accuracy = 100%.	Need to explore the potential to distinguish between the significant and the less significant symptoms.

Journal of Theoretical and Applied Information Technology
31st December 2020. Vol.98. No 24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.1 Significant Symptoms

The confirmation of dengue status is completed after a clinical examination of the patients. However, there is no further detailed categorization into positive DF or DHF. By referring to Table 2, 'Period of days' representing a period of fever infection of more than two days. Then, after further examination, the patient will be diagnosed as having positive dengue or negative dengue attribute. The dataset consists of 1054 instances and further selected randomly into training for data validation and testing by 70: 15: 15 ratios. Frequency analysis of the dataset with information about the frequency of 'Yes' and 'No' in each attribute is presented in Table 3. The significant high 'yes' frequency of 93.18 % and 82.75% for sudden high fever and period by day attributes suggest that most of the patients attended hospital because of sudden high fever and period of infection for more than two days.

Table 2: DF and DHF Symptoms

Attributes	Symptoms
Sudden high fever	DF
Severe headaches	DF
Eyes pain	DF
Muscle pain	DF
Fatigue	DF
Nausea	DF
Vomiting	DF
Skin rash	DF
Loss of appetite	DF
Abdominal pain	DHF
Persistent vomiting	DHF
Inner mouth bleeding	DHF
Exhaustion	DHF
Clinical fluid accumulation	DHF
Period by days	DF DHF

Generally, a fever will be cured within a few days, but if the fever lasts more than two consecutive days with the occurrence of other symptoms, a blood sample will be taken for a dengue test. In the other hand, a high 'No' frequency of 89.76%, 93.18%, 86.16%, and 96.59% for fatigue, loss of appetite, inner mouth bleeding, and clinical fluid accumulation, respectively, suggest that these symptoms are not major symptoms in representing dengue fever. Overall, by referring to the 'Yes' and 'No' frequency it can be concluded that symptoms for DF are more common than for DHF. This is a positive observation as DF is less dangerous than DHF.

Table 3: Yes and No Frequency for each Attribute

Attributes	Yes (%)	No (%)
Sudden high fever	93.18	6.82
Severe headaches	61.90	38.10
Eyes pain	51.66	48.34
Muscle pain	55.17	44.83
Fatigue	10.24	89.76
Nausea	45.02	54.98
Vomiting	58.67	41.33
Skin rash	44.83	55.17
Loss of appetite	6.82	93.18
Abdominal pain	24.27	75.73
Persistent vomiting	31.18	68.82
Inner mouth bleeding	13.84	86.16
Exhaustion	20.85	79.15
Clinical fluid accumulation	3.41	96.59
Period by Days	82.75	17.25

3.2 Model Design

This study utilized Neurointelligence software (Alyuda Neurointelligence 2.2, Cupertino, CA) in order to produce the model. There are two main processes included i.e. data preprocessing and network design. Data preprocessing was done to enable the machine to understand the data better before further analysis, while network design involving the selection of the best network architecture and properties. These crucial steps enable the machine to work better and produce better accuracy.

Journal of Theoretical and Applied Information Technology
31st December 2020. Vol.98. No 24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A. Data Preprocessing

16 categorical columns with 15 columns were set as input and 1 column name 'Status' was set as target. The input consists of 'Yes' and 'No' category while target consist of 'Positive dengue' and 'Negative dengue' category. The data were randomly partitioned into 70% for the training set, 15% for the validation set, and the other 15% was as the testing set. During pre-processing, the input data were scaled into a scaling range of -1 to 1. The scaling formula is shown in Equation (1).

$$SF = \frac{SR_{max} - SR_{min}}{p_{max} - p_{min}} \quad (1)$$

$$p_s = SR_{min} + (p - p_{min})SF$$

Where:

- p - the actual value of a categorical column
- p_{min} - the minimum actual value of the column
- p_{max} - the maximum actual value of the column
- SR_{min} - lower scaling range limit
- SR_{max} - upper scaling range limit
- SF - scaling factor
- p_s -preprocessed value

This is a min-max normalization method for scaling technique. It is the most wide and simplest method to be used. The pre-processed values were between -1 and 1 enable the input data to be set at comparable range at numbers that are not very far from 0.

B. Network Design

The best architecture was searched automatically with logistic hidden layer activation function and cross-entropy error function. The classification model was set to a confidence limit with an acceptance level of 0.7. The parameters were set with a heuristic search method inverse test error value for fitness criteria. 1000 iteration was executed for each design with a hidden unit range between 2 to 38 and the best network was chosen. The best network design with 15-2-1 was selected. One hidden layer is sufficient in order to maintain the simplicity of this model. The architecture design of the ANN model is illustrated in Figure 1.

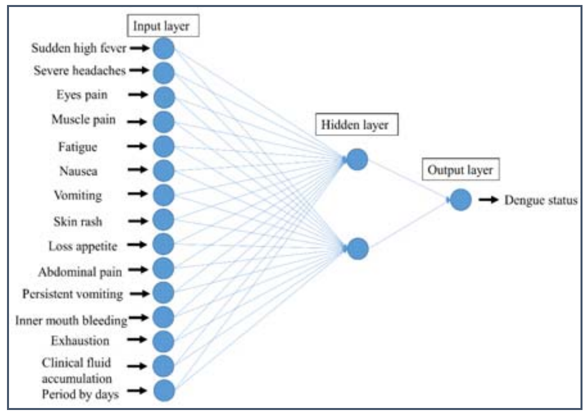

Figure 1: Architecture Design for the ANN Model

There are 3 main layers in a multilayer perceptron i.e. input layer, hidden layer, and output layer. The best network design of 15-2-1 indicates that there are 15 nodes for the input layer, 2 nodes for the hidden layer, and 1 node for the output layer. The input layer consists of the attributes data while the output layer consists of the target data.

The number of nodes in the hidden layer varies according to the requirement of the model and the nodes are interconnected through the varying weight of each node. The architecture design of 15-2-1 was chosen because the design satisfies the fitness criteria with having the least error value and highest accuracy. The four steps in structuring the ANN model are further described below:

I. Initialization

The network initialization was done to randomize the initial weights inside a small range on a neuron-by-neuron basis as shown in Equation (2).

$$-\frac{2.4}{N}, +\frac{2.4}{N} \quad (2)$$

Where:

- N – Total number of neuron inputs in the
- network.

II. Activation

The target data was scaled into a 0 to 1 scale according to the logistic hidden layer activation function. The function has a sigmoid curve as is presented in Equation (3).

$$y_j(p) = sigmoid\left[\sum_{i=1}^{n} x_i(p).w_{ij}(p) - \theta_j\right] \quad (3)$$

Where:

- n - Number of inputs of neuron j in the hidden layer
- $y_j(p)$ –Output from the output layer for p
- $sigmoid$ -The sigmoid activation functions
- $x_i(p)$ - Input for p
- $w_{ij}(p)$ - Weight for p

III. Weight Training

The network weight was updated according to the associated errors with the output neuron by backward propagation. The error gradient for the neurons in the output layer was calculated using the following in Equation (4a-f).

Calculate the error gradient for the neurons in the output layer:

$$\delta_k(x) = y_k(p)[1 - y_k(p)]e_k(p) \quad (4a)$$

$$e_k(x) = y_{d,k}(p) - y_k(p)$$

Where:

- $\delta_k(p)$ – Error gradient
- $y_{d,k}(p)$ – Desired output
- $e_k(p)$ - Error

Calculate the weight correction:

$$w_{jk}(p) = \alpha . y_j(p) . \delta_k(p) \quad (4b)$$

Where:

- $w_{jk}(p)$ – Weight corrector
- α – Alpha value

Update the weights at the output neurons:

$$w_{jk}(p+1) = w_{jk}(p) + \Delta w_{jk}(p) \quad (4c)$$

Calculate the error gradient for the neurons in the hidden layer:

$$\delta_j(p) = y_j(p).[1 - y_j(p)].\sum_{k=a}^{l} \delta_k(p) w_{jk}(p) \quad (4d)$$

Calculate the weight corrections:

$$\Delta w_{ij} = \alpha . x_i(p) . \delta_j(p) \quad (4e)$$

Update the weights at the hidden neurons:

$$\Delta w_{ij}(p+1) = w_{ij}(p) + \Delta w_{ij}(p) \quad (4f)$$

IV. Iteration

The iteration of *x* was increased by one and then returned to step 2 to repeat the process until the selected criterion is satisfied.

4. IMPLEMENTATION

After identification of the significant symptoms, the expert system was developed which consist of the knowledge base, rules, and convert rules. The next phase is to convert the expert system into an early detection system. The development of an application enables the finding of this study to reach target users such as patients and doctors.

4.1 Expert System

I. Knowledge Base

The knowledge for this study was obtained from the previously recorded dengue cases for 1054

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

patients. The expert domain has been interviewed for getting the correct information about the dengue disease and all the information from the interview session must be converted to rule-based and stored in the knowledge-based.

II. Rules

The identified significant rules are added to the general rules. Two experienced doctors have been interviewed to obtain the correct knowledge about dengue. From this study, we obtain that the diagnosis of DF should be considered in anyone who develops a fever and two others of the symptoms which is as elaborated in Table 1. Then, the general rule for dengue diagnosis was built as shown below and 20 rules about DF and DHF were extracted and considered as a testing dataset.

General rule:

High fever symptom + more than 1 other symptoms

+ period of infection more than 2 days = Positive,

Else = Negative

III. Convert Rules

The rules were converted according to the desired weight before added to the system. Each symptom was evaluated in percentage values. Other significant factors that increase the infection probability by dengue are the period of infection and the previous infection of dengue. When the period of infection is more than two days, the percentage of infection will increase by 10%. Besides, the patient who has a previous history of DF will increase the percentage of infection by 5%.

4.2 Early Detection System

To make this model more useful, the expert system was converted into a practical platform that will assist all users more effectively. This early self-diagnose system namely *DengueDiagnose* mobile application is a highly accessible platform that facilitates early detection and risk assessment for both the public and medical practitioners. This app helps patients to assess their condition and decide whether they should seek medical intervention from a doctor. Complete with separate access modules for patients and doctors, the application also provides accurate diagnostics and interactive functions to facilitate consultation and setting of appointment with doctors in real-time as shown in Figures 2 and 3.

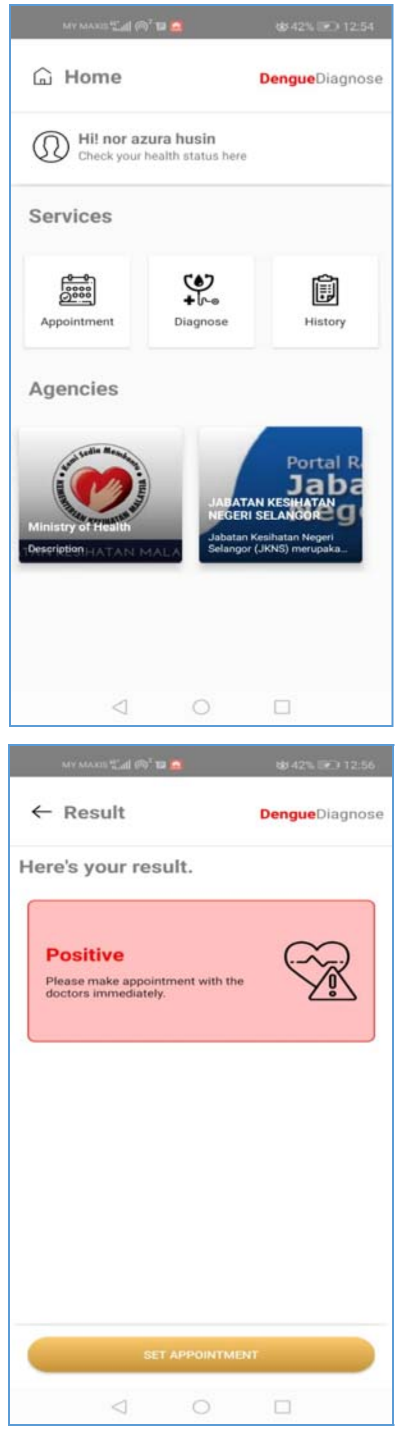

Figure 2: Early Self-diagnosis System

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

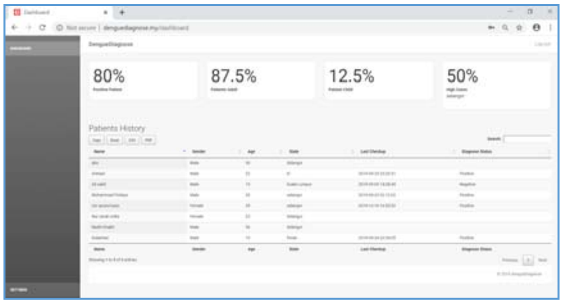

Figure 3: Dashboard for Doctor's Preference

5. RESULT AND DISCUSSION

In this study, the best ANN model has been developed with the characteristic in Table 4. The model is able to offer a correct classification rate of 100% and network error as low as 0.00056. Besides that, table 5 shows the confusion matrix of training, validation, and testing set. It is highlighted that there is no misclassification of any set data. This indicates that the developed model is highly dependable.

One of the advantages when using ANN is that it can allow the analyst to know the weight functions of different symptoms (Table 6). World Health Organization had issued a brief guideline regarding the management of dengue fever incidence.

Table 4: Network Parameters

	Training	Validation
Correct classification rate, %	100	100
Network error	0.00056	0
Error Improvement	0	
Training speed (iteration/second)	61.036622	
Architecture	15-2-1	
Training algorithms	Conjugate Gradient Descent	
Training stop reason	All iteration done	

Table 5: Confusion Matrix of Training, Validation and Test set

Target	**Output diagnosis**		**Total**
	Positive	Negative	
Training set			
Positive	348	0	348
Negative	0	371	371
Validation set			
Positive	93	0	93
Negative	0	75	75
Test set			
Positive	68	0	68
Negative	0	100	100
Total	509	546	1055

Table 6: Weight Functions of Different Symptoms

Attributes	**Importance (%)**
Sudden high fever	21.43
Severe headaches	0
Eye pain	21.43
Muscle pain	17.46
Fatigue	18.25
Nausea	0
Vomiting	0
Skin rash	0
Loss appetite	0
Abdominal pain	0
Exhaustion	0
Period by Days	21.43

It is stated that a patient should be suspected to have a dengue fever when the infection lasted for more than two days with high fever (40°C/ 104°F) and the other two symptoms such as severe headache, pain behind the eyes, muscle and joint pains, nausea, vomiting, swollen glands or rash [19]. As for this study, we want to learn from the data in order to identify the major symptoms of the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

minor symptoms. It is shown in Table 6 that sudden high fever, eye pain, muscle pain, fatigue together with a period of fever of more than two days have the importance of 21.43%, 21.43%, 17.46%, 18.25%, and 21.43%, respectively. This implies that these attributes do play an important contribution to the development of the accurate model.

This also suggests that the important attributes are the significant attribute besides the others. It is discovered that sudden high fever, eye pain, muscle pain, fatigue, and period by days are the major symptoms for dengue fever besides other attributes such as severe headaches, nausea, vomiting, loss of appetite, skin rash, abdominal pain, and exhaustion.

It is expected that by the exploitation of the significant attribute for the development of the model will enhance the accuracy and the performance of the prediction. As it is stated from [21], by removing less significant and redundant input variables, the produced model can work better in less training time and significantly improve the dengue prediction performance.

6. LIMITATION

Getting a published dataset, especially for medical data is very challenging. This is because the patient's medical data is considered as confidential.

7. CONCLUSION

The use of machine learning algorithms such as ANN has been a great benefit for data analysts in order to gain information from the dataset. As in this study, the application ANN for the dengue dataset has successfully produced a 100% accurate early self-diagnosis model and is able to differentiate the significant symptoms from the less significant symptoms. This model can be used as early after two days of fever. It is found out that sudden high fever, eye pain, and muscle pain, fatigue for more than 2 days are the main symptoms that a person may be infected by dengue fever and have higher weight functions in the model. Other symptoms like severe headaches, nausea, vomiting, skin rash, loss appetite, abdominal pain, exhaustion are the minor symptoms and therefore have lower weight function in the model. The developed early detection system from the ANN model is able to give 100% accuracy for the target users such as patients and medical practitioners. It is advised that a person suspected to have DF/DHF fever should reach out for immediate medical assistance.

ACKNOWLEDGEMENT

Recognition and gratitude are due to University Putra Malaysia for its funding under Geran Putra IPM vote number GP-IPM/2017/9522600 and other facilities provided to complete this study. Special thanks to Dr. Abdualkareem Alazee and Dr. Abalazeez Alrymy for their help in data collection and system validation.

REFERENCES

[1] World Health Organization. 2012. Global strategy for dengue prevention and control 2012-2020.

[2] World Health Organization, Dengue Control. 2017, http://www.who.int/denguecontrol/human/en/

[3] N.M. Noor, R. Abdullah, and M.H. Selamat, "A modul of Knowledge Management System and Early Warning System (KMS@EWS) for Clinical Diagnostic Environment," *2nd International Conference on Software Engineering and Computer Systems. ICSECS 2011. Communications in Computer and Information Science.* Vol. 179, Issues Part 1, 2011 pp. 78-91, 2011.

[4] D. C. Tiga-Loza, R. A. Martínez-Vega, E. A. Undurraga, C. A. Tschampl, D. S. Shepard, and J. Ramos-Castañeda, "Persistence of symptoms in dengue patients: a clinical cohort study," *Trans. R. Soc. Trop. Med. Hyg.*, vol. 114, no. 5, pp. 355–364, 2020.

[5] N. A. Husin, A. Alharogi, N. Mustapha, H. Hamdan, and U. A. Husin, "Early self-diagnosis of dengue symptoms using fuzzy and data mining approach," *AIP Conference Proceedings,* Vol. 2016, pp. 020048, September 2018.

[6] N. M. Sharef, N. A. Husin, K. A. Kasmiran, and M. I. Ninggal, "Temporal Trends Analysis for Dengue Outbreak and Network Threats Severity Prediction Accuracy Improvement," *Journal of Digital Information Management*, 17(3), pp. 122-132, 2019.

[7] A. B. Rashed, H. Hamdan, M. N. Sulaiman, N. M. Sharef, and R. Yaakob, "Multi-Objective PSO-fuzzy Optimization Approach to Improve Interpretability and Accuracy in Medical Data," *International Journal of Engineering & Technology*, 7 (4.31), pp. 316-321, 2018.

[8] K. Al-Rababah, S. Doraisamy, M. Rina, and F. Khalid, "A Color-Based High Temperature Extraction Method in Breast Thermogram to Classify Cancerous and Health Cases Using SVM," *2nd International Conference on*

ISSN: **1992-8645** www.jatit.org E-ISSN: **1817-3195**

Imaging. Signal Processing and Communication. pp. 70-73, 2018.

[9] M. Fatima, and M. Pasha, "Survey of machine learning algorithms for disease diagnostic," *Journal of Intelligent Learning Systems and Applications*, Vol. 9, pp. 1- 16, January 2017.

[10] D. Saikia, and J. C. Dutta, "Early diagnosis of dengue disease using fuzzy inference system," *2016 International Conference on Microelectronics, Computing and Communications (MicroCom),* Vol. IEEE2016, pp. 1-6, January 2016.

[11] K. Shaukat, N. Masood, S. Mehreen, and U. Azmeen, "Dengue fever prediction: A data mining problem," *Journal of Data Mining in Genomics & Proteomics*, Vol 6, pp. 1-5, 2015.

[12] A. Anitha, and D. J. W. Wise. "Forecasting Dengue Fever using Classification Techniques in Data Mining," *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Vol. IEEE2018, pp. 398-401, December 2018.

[13] S. Gambhir, S. K. Malik, and Y. Kumar, "The Diagnosis of Dengue Disease: An Evaluation of Three Machine Learning Approaches," *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, Vol 13, pp. 1-19, 2018.

[14] M. Ahamad *et al.*, "Expert Systems with Applications A machine learning model to identify early stage symptoms of SARS-Cov-2 infected patients," *Expert Syst. Appl.*, vol. 160, p. 113661, 2020.

[15] H. Hamdan, and J. M. Garibaldi, "Adaptive neuro-fuzzy inference system (ANFIS) in modeling breast cancer survival," *International Conference on Fuzzy Systems,* Vol. 2010 IEEE, pp. 1-8, July 2010.

[16] A. Barwad, P. Dey, and S. Susheilia, "Artificial neural network in diagnosis of metastatic carcinoma in effusion cytology," *Cytometry Part B: Clinical Cytometry*, Vol. 82, pp. 107-111, 2012.

[17] M. A. Al-garadi, "Epidemiological review of dengue fever in Yemen," 1, Int. J. Adv. Res., vol. 3, no. 7, pp. 1578–1584, 2015.

[18] A. S. Bin Ghouth, A. Amarasinghe, and G. W. Letson, "Dengue outbreak in Hadramout, Yemen, 2010: An epidemiological perspective," Am. J. Trop. Med. Hyg., vol. 86, no. 6, pp. 1072–1076, 2012.

[19] A. Jawfi, A. M. Al Jawfi, H. D. Fahmy, and A. K. Hassan, "Assessment of the Awareness and Practice of Health Team Toward Dengue Fever at Al-Hodeidah City –Yemen," Assiut Sci. Nurs. J., vol. 5, no. 12, pp. 33–42, 2017.

[20] World Health Organization, Dengue and Severe Dengue, 2014, http://www.who.int/mediacentre/factsheets/fs117.

[21] S. A. Balamurugan, M. S. M. Mallick, and G. Chinthana, "Improved prediction of dengue outbreak using combinatorial feature selector and classifier based on entropy weighted score based optimal ranking," *Informatics Med. Unlocked*, vol. 20, no. March, p. 100400, 2020.