A HYBRID NEURAL KNOWLEDGE EXPERT SYSTEM WITH PARALLEL COORDINATES VISUALIZATION IN DENGUE DIAGNOSIS PREDICTION

J. Joshua Thomas[1], Jodene Ooi Yen Ling[2], Bahari Belaton[3]

[1,2]Department of Computing, School of Engineering, Computing and Built Environment
KDU Penang University College, 10400, Penang, Malaysia

[3]School of Computer Sciences, University Sains Malaysia

Email: joshopever@yahoo.com[1]; jodeneoyl@gmail.com[2]; bahari@usm.my[3]

DOI: https://doi.org/10.22452/mjcs.sp2019no1.3

ABSTRACT

This research presents the extension of DengueViz, a hybrid neural-knowledge-based expert system integrated with parallel coordinates as its visualization technique to assist in the diagnosis and severity assessment of dengue. This implementation involves the expert system with 140 rules for the classification of dengue, along with the multilayer perceptron with the stochastic gradient descent algorithm as the artificial neural network to learn data representations, support vector machine to systematically verify errors, and the Theil-Sen estimator to increase its robustness against outliers. The integration of parallel coordinates visually presents the large amount of dengue information into a single visualization space, where data interactions such as the selection of axes, filtering and highlighting reduces the clutter for it to be more comprehensible and enhances the correlation between the attributes of the information. The experiments of this system are conducted with the technology acceptance model, where the usefulness and ease of use of the system is analyzed.

Keywords: Dengue, Knowledge-based expert system, Information visualization, Parallel coordinates, Visual data interactions, Dengue classification, Probability diagnosis, Multilayer Perceptron, Support Vector Machine, Theil-Sen Estimator

1.0 INTRODUCTION

Transmitted by the female Aedes mosquitoes, Dengue is a viral disease that has rapidly and widely spread in tropical and sub-tropical climates worldwide. According to the World Health Organization (WHO) report in April of 2019, the global incidence of dengue has increased dramatically, putting half of the world's population at risk [1]. In 2015, the last highest number of cases was recorded with 120, 836 cases with 336 deaths, double in figures compared to the previous year for the same period [2]. Despite government efforts, dengue cases continue to rise year after year, hitting the all-time high of nearly 80,000 cases and 113 deaths reported from January to the beginning of August in 2019 with the number of cases expected to hit up to 150,000 by the end of this year [2].

Dengue fever starts with severe flu-like illness, which in its febrile phase may often be misdiagnosed as the common influenza, Chikungunya or Malaria [3]. The clinical diagnosis is often difficult, requiring extensive and time-consuming tests, with more details on the diagnosis in [3]. Dengue would have most likely evolved into more severe forms by the time the diagnosis is made, leading to serious illness and deaths, especially in young children [1]. In addition, the delayed diagnosis hinders the proper dengue patient management process, increasing the risk of dengue widespread transmission in the neighboring areas, that is, Aedes mosquitoes may be infected by the dengue patient and becomes the vector to transmit the disease to another person.

This research proposes an extension from DengueViz [4], a system to assist physicians and health care personnel in dengue diagnosis and assessing the severity of dengue by integrating machine learning methods. Further details will be discussed in the following chapters of this research. The outlines from the remaining

contents are as follows: Sect. 2 Literature Review, Sect. 3 Methodology, Sect. 4 Experiments and Discussions, and Sect. 5 Conclusion.

2.0 LITERATURE REVIEW

2.1 Dengue Diagnosis

A guideline from the WHO states dengue as flu-like symptoms which should be suspected when there is a sudden high fever of 40°C accompanied by any two from these symptoms, such as severe headache, pain behind the eyes, muscle or joint pain, rash, nausea and vomiting, or swollen glands, although the symptoms may vary according to the age of the patient [5]. The more severe form of dengue is characterized with the drop in temperature to below 38°C accompanied by the symptoms such as severe abdominal pain, persistent vomiting or vomit with the presence of flood, rapid breathing, bleeding in gums, fatigue or restlessness, which may occur within the first week of the dengue infection [5]. While there is no specific treatment for dengue fever so far, proper medical care must be administered to avoid further complications and risk of death [5].

The severe and non-severe dengue classification criteria revised by WHO has always been considered by many to be too broad for its diagnosis and management purposes [6]. Hence, additionally, the symptoms of dengue are studied in other dengue-related researches as well, noting the common clinical characteristics as symptoms which appear in more than one researches and affect the majority of the sample size. The common characteristics from the studies are summarized in Table 1 below.

Table 1: A summary of common clinical characteristics for dengue in adults and children respectively. AST- aspartate aminotransferase count, ALT- alanine transaminase count.

Research	Location	Description	Characteristics
Watt et al., 2003 [7]	Thailand	Dengue in adults	• Headache • Muscle joint pain • Rash • Bleeding • Decreased platelet count • Increased AST and ALT
Chang et al., 2009 [8]	Taiwan		
Mitra et al., 2017 [9]	India		
Sa-ngamuang et al., 2018 [10]	Thailand		
Xuan et al., 2004 [11]	Vietnam	Dengue in children	• Headache • Vomiting • Anorexia • Abdominal pain • Rash • Skin bruising • Bleeding • Decreased platelet count • Increased AST and ALT
Pongpan et al., 2013 [12]	Thailand		
Pone et al., 2016 [13]	Brazil		
Mishra et al., 2016 [14]	India		
Phakhounthong et al., 2018 [15]	Cambodia		

2.2 Knowledge-Based Expert Systems

In artificial intelligence (AI), a knowledge-based expert system is a system designed to solve complex problems by emulating the reasoning and decision-making behaviors of experts using its domain-specific expert knowledge base [16]. The advantages of expert systems include high performance, responsiveness, reliability, and understandability, provided an adequate and accurate knowledge base [16]. However, the capabilities of the expert system is limited to the extent of its knowledge base; an inadequate or inaccurate knowledge base will

still produce poor, inefficient solutions. The knowledge-based expert system is made up of two main components: the knowledge base and the inference engine.

The knowledge base of the expert system plays a key role in providing the intelligence and insights resembling a human expert through its vast collection of both factual and heuristics knowledge [16]. The knowledge is organized and stored meaningfully into multiple IF-THEN rules by the knowledge engineer- the person who had acquired all the relevant information in the problem domain [16]. For real-world applications, knowledge is typically derived from a combination of multiple verified sources, including human experts to provide sufficient and high integrity knowledge for the reasoning of the inference engine.

The inference engine holds the reasoning and decision-making ability of the system, based on the IF-THEN rules from the knowledge base, deploying either the forward or backward chaining strategies. An inference engine deploys the forward chaining strategy if the goal of the system is not known, where rules from the knowledge base are applied to produce possible outcomes as its goal [17]. Meanwhile, the backward chaining strategy is deployed to discover the root cause or prove an outcome, avoiding any superfluous reasoning paths [17].

One of the applications of the knowledge-based expert systems is in the field of medicine. MYCIN, a knowledge-based expert system was developed in 1974 by Stanford University for the diagnosis of bacterial infection [18]. The success of MYCIN was demonstrated with a series of experiments, proving its capabilities in generating accurate diagnosis, besides prescribing accurate drug dosage adjusted according to the health conditions of the patient [18]. This reveals the potential of expert systems in improving the quality of healthcare globally, as the skills and knowledge of human experts can now be replicated and transferred to be made available to anyone who needs them.

2.3 Multilayer Perceptron Neural Networks

Another branch of the AI is the neural networks (NN) inspired by the biological neurons in the human brain. NN is able to learn the representation and recognizing the underlying relationships in a set of data, adapting to changes as it learns to generate the best possible solution [19]. The basic building blocks of NN include the neurons, weights and activation function, in which each block forms a layer in the network [19]. A Multilayered Perceptron (MLP) is one of the simplest and most useful NN architectures with three or more fully connected layers: an input layer to receive data, a number of hidden layers to provide computational learning functions, and an output layer to return the final solution [20].

Typically, a weight and bias are assigned to each neuron, both are often initialized to small random values, although more complex initialization schemes can be used to solve more complex problems [19]. The weighted inputs are then passed through the activation function, which may be linear or non-linear functions to determine the activation strength of the output [19]. One of the optimization algorithms in NN that may be deployed in the MLP is the stochastic gradient descent (SGD), which calculates the relative error by comparing the output to the expected output [19]. By backpropagation, the error is propagated back through the layers of the NN, updating the weights of the neurons accordingly [19].

MLPs are often used in supervised learning problems, and they are able to perform both linear and non-linear classifications, depending on the activation functions chosen [20]. In the fields of medicine, MLP has been widely used in the diagnosis of diseases [21], disease risk classifications [22], and even disease outbreak predictions [23]. An example of application of MLP in the diagnosis of disease is its use in the research for predicting dengue diagnosis conducted in Paraguay in 2018, where the MLP classifier had managed to achieve the best average accuracy in prediction of 96%, an average sensitivity of 96%, and an average specificity of 97% - compared with other NN topologies and SVM with their various kernel functions [21]. In another example, the researchers had employed two MLPs in their diagnostic system for the risk classification in dengue patients [22]. The two models trained with Levenberg-Marquardt and Scaled Conjugated Gradient algorithms

managed to achieve a prediction accuracy of 75% and 70.7% respectively [22]. The results may still be improved with further fine-tuning or optimization techniques, but already satisfactory for assisting physicians in assessing the risk level in patients [22].

2.4 Support Vector Machine

A Support Vector Machine (SVM) is a machine learning algorithm defined by a separating hyperplane to distinctly classify data points in an N-dimensional space according to the N number of features of the set of data provided [24]. The hyperplane acts as a decision boundary, aiming to raise the confidence of the classification by maximizing the margin distance between the hyperplane and the data from different classes, where points on either side of the hyperplane are considered distinct classes [24]. To prevent the models from underfitting or overfitting the training data set, it is common for researchers to tune the regularization parameter for the balance between the margin maximization and the loss, and choose suitable kernel functions according to the data type and the desired linear or non-linear separation [24].

As a supervised learning MLP, the SGD optimizer is used as the basis of the backpropagation algorithm. The SGD aims to minimize the loss function of the MLP, that is, to minimize the difference between the actual and predicted value by iteratively updating the weights, W of the neurons as required, mathematically expressed as:

$$W_{new} = W_{current} - \eta \frac{\partial \mathrm{L}}{\partial \mathrm{W}_{current}} \quad \text{Eqn. (1)}$$

In early works, the linear activation function is used to solve binary classification problems, with only two possible output values of -1 or +1, such as in Eqn. 2. On the other hand, the sigmoid activation function is a non-linear function, described as in Eqn. 3.

$$\theta(x) = \begin{cases} -1 \; if \; x < 0 \\ +1 \; if \; x \geq 0 \end{cases} \quad \text{Eqn. (2)}$$

$$\theta(x) = \begin{cases} 0 \; if \; correctly \; classified \\ -\eta xj \; if \; -1 \; classified \; as \; +1 \\ +\eta xj \; if \; +1 \; classified \; as \; -1 \end{cases} \quad \text{Eqn. (3)}$$

Just as the NN or MLP, the SVM is also a commonly used machine learning classifier in the fields of medicine for various purposes. In [21], the research in predicting the diagnosis of dengue also employed the use of the SVM machine learning classifier, with the use of three different kernels- the linear kernel, Gaussian kernel, and the polynomial kernel. Although the results were not as good as the MLP classifier, the polynomial kernel managed to achieve results close to that, with an average accuracy, sensitivity and specificity of 92%, 93%, and 92% respectively. The SVM with the Gaussian kernel only achieved an average accuracy, sensitivity and specificity of 86%, 84%, and 89% respectively, with the linear kernel producing the worst results of average accuracy, sensitivity and specificity of 64%, 56%, and 71% respectively [21].

2.5 Theil-Sen Estimator

Similar to the SVM, the Theil-Sen estimator (TSE) discriminates between classes by fitting a line through the data points of the dataset, known as the Sen's slope estimator by choosing the median of the slopes through pairwise points as estimators [25]. The TSE is a robust simple linear regression model named after Henry Theil and Pranab K. Sen [25]. Despite the many properties and geometric interpretation, TSE is much under-developed and under-used due to its simple linear model [25].

2.6 Parallel Coordinates Information Visualization

Information visualization aim is to visually present data intuitively and comprehensively without information overloading, in a format that could be easily recognizable, navigated, and managed [26]. Complex problems and real-world applications are often multivariate with high-dimensional parameters, making it difficult to visualize the affiliations between the relationships of the parameters. One of the approaches to visualize many parameters in the same spatial position is the parallel coordinates, where the axes are parallel to each other and the scalability of this approach allows up to hundreds of axes to be plotted in the same visualization [27].

This visualization approach makes the correlation within the set of data more noticeable, such that, if the line segments of neighboring axes are parallel, the two parameters have positive correlation, else if the line segments cross over at a single point between the axes, the parameters are negatively correlated [27]. However, the visualization of large high-dimensional multivariate data can look intimidating, especially to new users. To ease the visualization, commonly interactions are allowed in the parallel coordinates, such as deletion or addition of axes to temporarily remove any unnecessary axes or add them back whenever required, highlighting to bring focused attention to a particular data value, and filtering to view only a range of data values [28].

3.0 METHODOLOGY

The Fig. 1 below illustrates the methodology of DengueViz. DengueViz accepts two different types of inputs upon the initialization. First is the interactions through the user interface of the system, which allows the user to select and input dengue symptoms as required for a prediction of dengue diagnosis and severity assessment. Second, the system also allows CSV formatted datasets to be imported as the inputs, directing the users directly to the parallel coordinate visualization for further interactions with visual cues. The knowledge-based expert system, MLP, SVM, Theil-Sen and visualization with parallel coordinates will be elaborated in the following subtopics in this chapter.

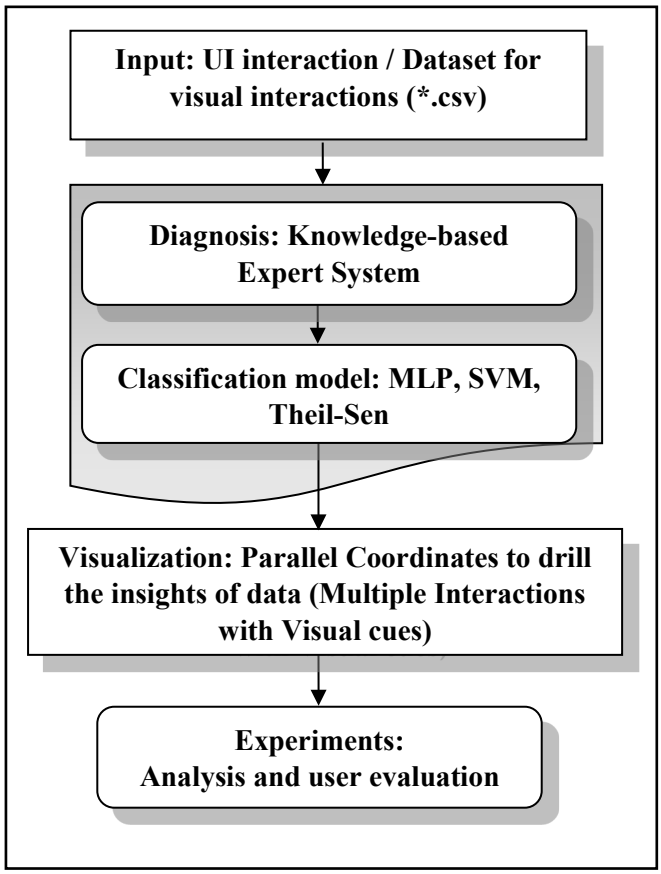

Fig. 1: Hybrid Neural Knowledge based Architecture

3.1 Knowledge-Based Expert System

The knowledge base of DengueViz contains the characteristics for the classification of dengue according to the three levels of severity, dengue fever, dengue hemorrhagic fever (DHF), and dengue shock syndrome (DSS). From the common clinical characteristics of dengue discussed in 2.1 above, in this system, each clinical characteristic is given certainty points to represent the severity and importance of each characteristics in each classification, as in Table 2 below. The default value for the certainty point of each characteristic is 1; the value is increased with the increase in severity and importance of the characteristic.

Table 2: Clinical characteristics for dengue severity classification with their certainty points

Classification	Clinical characteristics	Certainty points	Remarks
Dengue fever	Sudden high fever	2	Common dengue characteristics
	Severe headache	2	
	Pain behind eyes	1	Default points
	Muscle or joint pain	2	Common dengue characteristics
	Rash	2	
	Abdominal pain	2	
	Anorexia	2	
	Fatigue or lethargy	1	Default points
	Positive tourniquet test	6	Preliminary dengue test
Dengue hemorrhagic fever	Skin bruising	10	Blood below skin
	Bleeding from gums, nose, or mouth	12	Hemorrhagic indicator
	Black stools	10	Blood in stools
Dengue shock syndrome	Lymph node enlargement	15	Internal organs affected
	Liver enlargement	15	
	Shock	25	Shock indicator

The above knowledge of the clinical characteristics along with their certainty points are stored in the form of IF-THEN rules, where this knowledge base contains a total of 140 rules from the different combinations of clinical and blood test characteristics for the respective dengue severity classification. The C Language Integrated Production System (CLIPS) expert system shell [29] inference engine of the expert system checks the rules for conditions matching the user inputs, and returns the probability of the dengue classification. As it is common for the inference engine to return more than one matching rule at the same time, the probability of the classification is calculated as mutually exclusive events, represented mathematically as such:

$$Probability\ (DF) = \frac{100(cer1+cer2)-(\text{cer1}\times\text{cer2})}{100} \quad \text{Eqn. (4)}$$

; where cer1 and cer2 are probabilities from rules 1 and 2 respectively.

Example. An example scenario is given below based on the clinical characteristics and certainty points of the characteristics are referred from Table 2 above. The sum of certainty points from each characteristic is used to provide a probability value in classifications for each rule respectively.

Rule 1: If sudden high fever, severe headache and rash is present,

then dengue diagnosis is dengue fever with a probability value of 6.

Rule 2: If sudden high fever, severe headache, rash, and skin bruising is present,

then dengue diagnosis is dengue fever with a probability value of 6

and dengue hemorrhagic fever with a probability value of 12.

Rule 3: If sudden high fever, severe headache, rash, skin bruising, and shock is present,

then dengue diagnosis is dengue fever with a probability value of 6

and dengue hemorrhagic fever with a probability value of 12

and dengue shock syndrome with a probability value of 25.

From the rules above, with sudden high fever, severe headache, and rash, the conditions of Rule 1 are fulfilled. Therefore the diagnosis is a 6% probability of Dengue Fever.

Later, with an additional symptom of skin bruising, the conditions of both Rules 1 and 2 are matched. From the equation (1) above, the dengue diagnosis is now an 11.64% probability of Dengue Fever from the probability calculated in (2), with a 12% probability of Dengue Hemorrhagic Fever.

$$Probability\ (DF) = \frac{100(6+6)-(6\times6)}{100} = 11.64\% \quad \text{Eqn. (5)}$$

If the condition worsens with the presence of shock, the conditions of Rules 1, 2, and 3 are matched. The new Dengue Fever probability would be 16.94% as calculated in (3), and 22.56% for the probability of Dengue Hemorrhagic Fever as in (4), with a 25% probability of Dengue Shock Syndrome.

$$Probability\ (DF) = \frac{100(11.64+6)-(11.64\times6)}{100} = 16.94\% \quad \text{Eqn. (6)}$$

$$Probability\ (DHF) = \frac{100(12+12)-(12\times12)}{100} = 22.56\% \quad \text{Eqn. (7)}$$

The certainty points of the rules are then doubled if the patient belongs to a more vulnerable age group, such as a child or elderly, or has blood test results that deviate from the healthy range. For example, if the patient above is a child with sudden high fever, severe headache, and rash, the probability of Dengue Fever would be calculated as in (5), resulting in a 21.93% probability.

$$Probability\ (DHF) = \frac{100(11.64+11.64)-(11.64\times11.64)}{100} = 21.93\% \quad \text{Eqn. (8)}$$

3.2 MLP

The 3-layer MLP NN in this system is developed with the Python Scikit-Learn machine learning modules, along with Pandas, and Numpy as helper libraries. The MLP here is made up of an input layer, a hidden layer, and an output layer, as in Fig 2. The input layer of the system accepts clinical or blood test characteristics of the dengue patient, by importing patient dataset from DengueVIz into the MLP classification model. The hidden layer, along with its weights and bias attachments learns and calculates the representation of the data through an activation function. The output layer will return the prediction of the dengue classification.

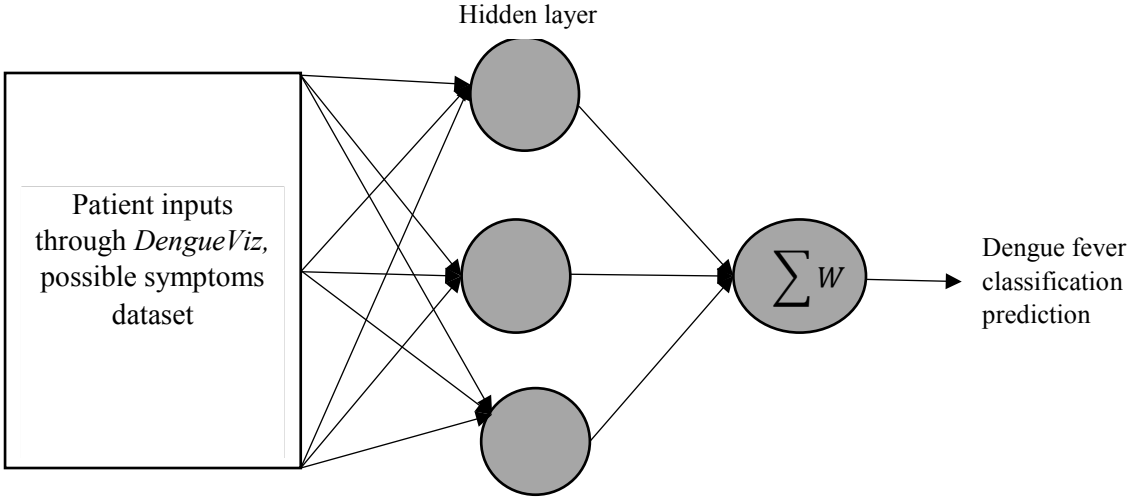

Fig. 2: MLP NN topology for dengue prediction

By plotting a small portion of the dataset, it can be seen, such as in Fig. 3 (a), the data points are linearly separable binary classifications- in which the classifications are only dengue or non-dengue. On the other hand, the learning rates from the user input is as in Fig. 3 (b).

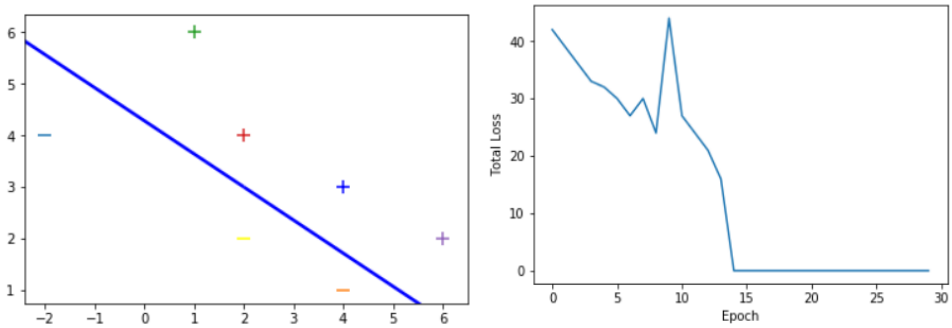

Fig. 3 (a): Linearly separable binary classification (b) Learning rates from user input

In this MLP, the weights are initially set to 0, with a learning rate of 1. The backpropagation technique was deployed along with the SGD in calculating the loss of the training. The training runs for 30 epochs and a sample size of 5. Fig. 4 (a) shows the characteristics of dengue as 1, 0, or -1 for symptom present, unknown, and absent respectively; Fig. 4 (b) shows that the MLP stops learning after the 4th epoch; Fig. 4 (c) shows the number of misclassifications according to the number of epochs.

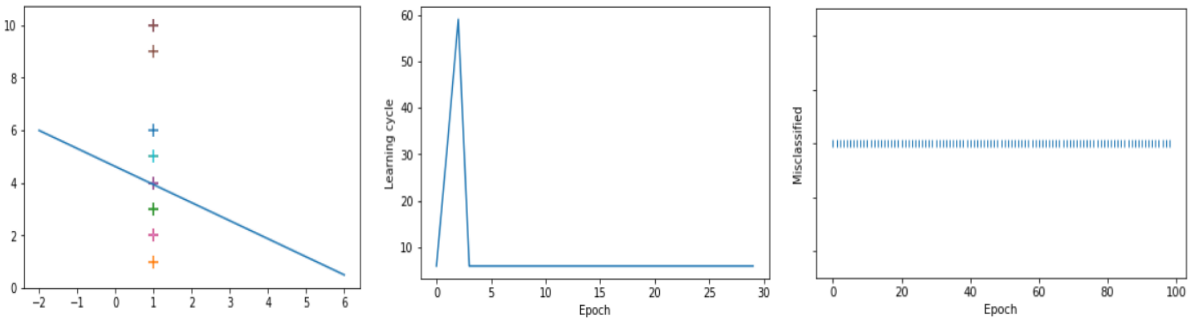

Fig. 4 (a): Dengue characteristics as 1, 0, or -1 (b) Learning rates of MLP in each epoch
(c) Misclassification by epochs

3.3 SVM

Like the MLP classifier, the SVM classifier in this research is also developed with the Python Scikit-Learn machine learning library, along with Pandas, Numpy and Matplotlib as helper libraries for data manipulation over the dataset imported from DengueViz. A few SVM kernel functions are developed, from the examples mentioned in 2.4, but only the model with the best performance according to its loss function is remained. The root mean squared error (RMSE) is chosen as the loss function, increasing the sensitivity of the model towards even small changes. The learning rate is set to 1 and the regularization parameter is set to 1/epochs. The Fig. 4 below shows the misclassifications when the SVM classifier is trained for 100 epochs.

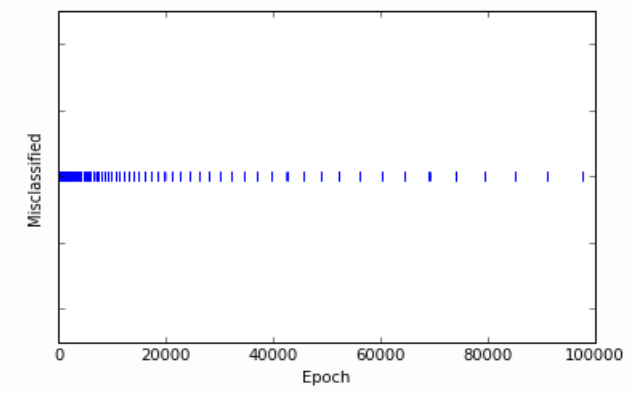

Fig. 4: SVM misclassification moving along the epochs

3.4 TSE

The TSE, along with its other flavors of the generalized linear models of Ordinary Least Squares (OLS) and Random Sampling Consensus (RANSAC) are developed with the Python Linear Model library, also along with other helper libraries as mentioned above for the data manipulation processes. Fitting on 100 rows of dengue characteristics dataset, the linear models are compared as in the Fig. 6 (a) and (b) below.

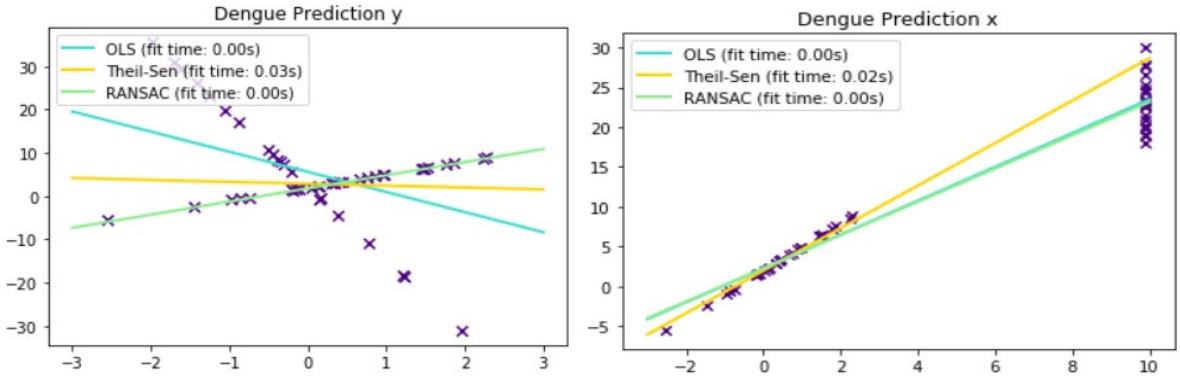

Fig. 5 (a): Dengue prediction y with linear models (b) Dengue prediction x with linear models

3.5 Parallel Coordinates

The visualization of parallel coordinates in DengueViz is developed with the use of Java Graphics 2D library. This native graphics library is chosen as although each element in the visualization had to be plotted individually, it allows flexibility in customizing the visual elements and the interactions between the elements.
The dataset contains 28 attributes, a combination of clinical and blood test characteristics, including some information on the demography of the patient. In the parallel coordinates of DengueViz, the axes represent the attributes and are labeled according to the names of the attributes in the dataset. The values within the axes are labeled with data tics to represent the data values in the attributes respectively. Each data row is plotted as data lines along the axes according to the attribute values, connecting one axis to another until the last axis in a continuous manner.

The parallel coordinates in DengueViz allows three basic user interactions within the visualization using mouse-related events registered by the Java MouseEvent library. The allowed interactions are filtering, highlighting and axes selection, as described below:

- Filtering. The filter isolates the range of data value within the upper and lower boundaries of the selected axes. The color of data values outside the filter boundaries is lightened to allow better visibility of the isolated data values within the filter range. The filter boundaries can be changed by mouse dragging motions over the upper or lower boundaries of the axes.

- Highlighting. The highlight brings focused attention to only one or more specific data values in the parallel coordinates. Mouse-hovering motions over the data lines temporarily changes the lines to a different color, while clicking on the line sets the color to the data lines until the line is clicked again to revert it to its original color. This frees the mouse for other interactions within the visualization.

- Axes selection. The list of attributes from the dataset is displayed in the left-most panel of the system view as checkboxes for each individual attribute. Only the selected checkboxes are plotted as axes in the parallel coordinates, where any attribute can be unchecked to prevent it from being plotted in the visualization.

4.0 EXPERIMENTS AND DISCUSSION

4.1 DengueViz

The Fig. 7 below demonstrates the user interface of DengueViz. All the attributes in this interface are the same as the attributes from the dataset for the visualization of parallel coordinates. By default, the form is left empty for users to input the clinical and blood test characteristics of the dengue patient. The probability of the dengue classification changes as the attributes are selected or filled in accordingly. Such as in 3.1, the probability of dengue fever shown in Fig. 7 is 6% when sudden high fever, severe headache, and muscle or joint pain is present.

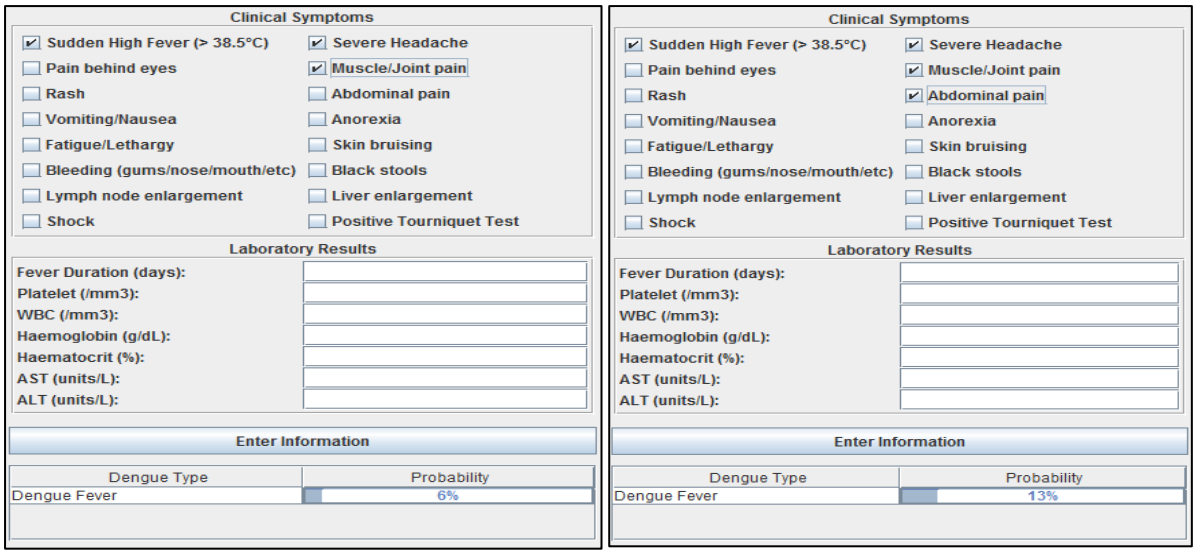

Fig. 6: DengueViz user interface

When DengueViz is first initialized, the default view for the parallel coordinates is as Fig. 8 below. The attribute selection panel is shown on the utmost left, with a majority of space for the parallel coordinates. Each axes is

labeled according to the attributes, and an indication box is provided at the bottom of the visualization space to provide the legends of the parallel coordinates.

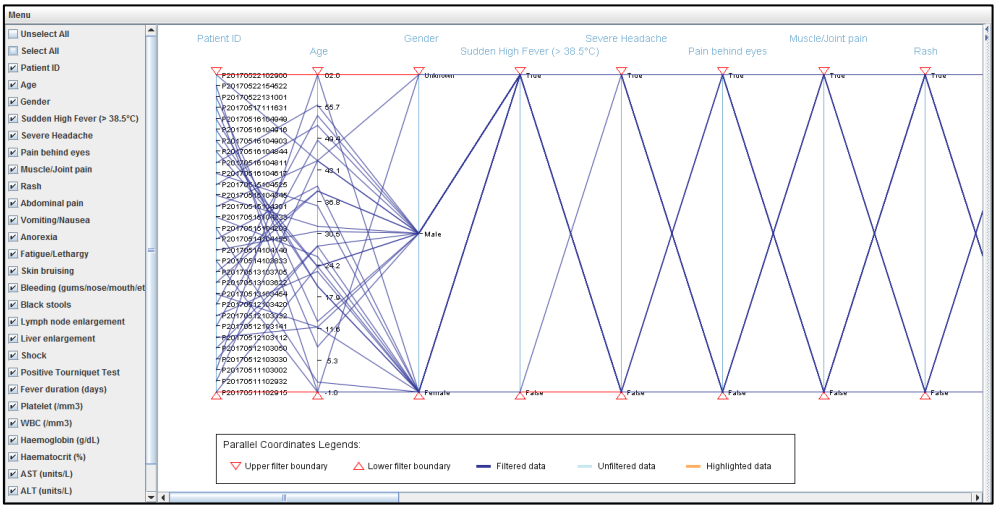

Fig. 7: Default view of the parallel coordinates

Meanwhile, the effects of user interactions on the parallel coordinates are illustrated as in Fig. 9 below. In this case, as only several attributes are selected from the left panel, only several axes plotted. The positions of the filter boundaries have been moved for the age axes to filter data of patients from age 0 to age 12, with red lines to indicate the upper and lower bounds of the filter, and light blue lines to represent the data lines outside the range of the filtered data, while the dark blue lines are data lines still within the boundaries of the filtered data. The orange line shows the highlighted data line, which in this case shows a female child at risk of DSS with the probability of 43%.

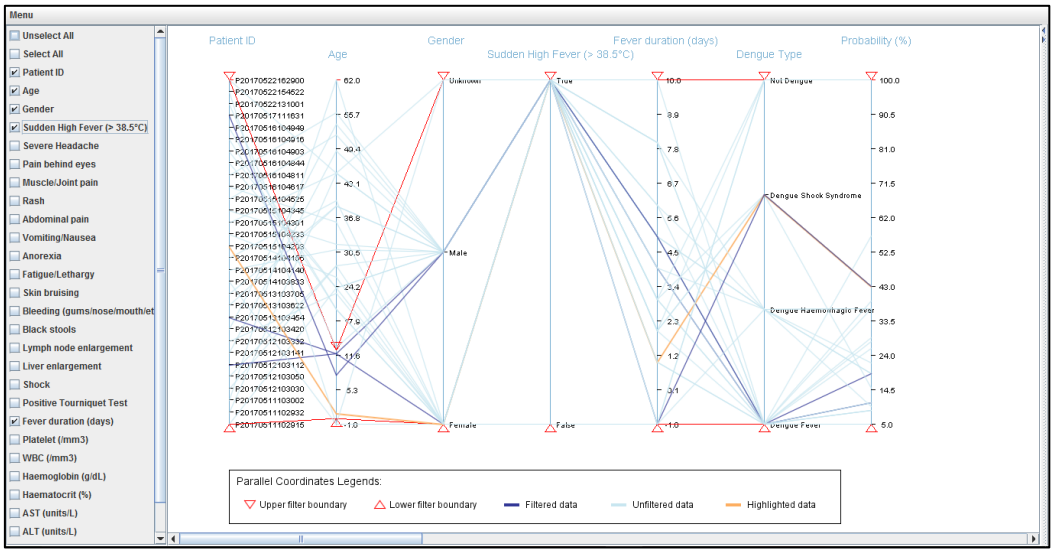

Fig. 8: The effects of user interactions on the parallel coordinates

4.2 Technology Acceptance Model

The experimentations done on DengueViz is based on the quantitative results and analysis by studying the usefulness and ease of use of the system through the Technology Acceptance Model (TAM) [30]. In this case, 20 final-year nursing students are gathered for a survey on their everyday routines, and how they would accept DengueViz as part of a dengue diagnosis assistance tool. The pointers are used in 6 different measures: (1) Extremely Disagree, (2) Moderately Disagree, (3) Slightly Disagree, (4) Slightly Agree, (5) Moderately Agree, and (6) Extremely Agree. The TAM inspection survey is conducted based on 2 parts, with survey items as listed below:

Part 1: TAM inspection on usefulness

1. Selecting and filling in symptoms individually on the input interface of DengueViz
   (S1) DengueViz helps me to be more productive.
   (S2) DengueViz helps me to be more efficient.
   (S3) DengueViz helps me to complete my tasks.
   (S4) DengueViz helps make it easier to accomplish my tasks.
   (S5) DengueViz saves my time.
   (S6) I find DengueViz useful.

1. Importing dataset for interactive parallel coordinates of DengueViz
   (S1) Filtering with upper and lower boundaries makes it easier to view only the desired range of data values.
   (S2) Highlighting with color-coded lines makes it easier to focus attention on the characteristics of a particular patient.
   (S3) Highlighting upon mouse hover makes it easier to preview the characteristics of the patient within the visualization
   (S4) Axial information makes it easier to differentiate the strength of the cluster of symptom characteristics from the dataset.
   (S5) Parallel coordinates show the overall visualization of the multiple characteristics for dengue within the same spatial view.
   (S6) I find DengueViz easy to use.

Part 2: TAM inspection on ease of use

2. Selecting and filling in symptoms individually on the input interface of DengueViz
   (S1) DengueViz is simple to use.
   (S2) DengueViz is user-friendly.
   (S3) DengueViz can recover my mistakes from the interactions easily.
   (S4) DengueViz completes my tasks in fewer steps.
   (S5) DengueViz can be used without written instructions.
   (S6) I find DengueViz easy to use.

3. Importing dataset for interactive parallel coordinates of DengueViz
   (S1) Filtering data values are easy with mouse dragging motions.
   (S2) Highlighting data values temporarily are easy by mouse hovering or clicking motions.
   (S3) Highlighted data values can easily be reverted by clicking on the mouse again.
   (S4) Axes selection is easy from the checkbox of dengue characteristics.
   (S5) The underlying relationship of the dengue characteristics are easy to see with parallel coordinates.
   (S6) I find DengueViz easy to use.

Table 3 and Table 4 below shows the results of the TAM inspection in terms of the usefulness and ease of use of DengueViz respectively. The Table 3 shows a slightly better mean score of 31.3 for the interactions within the visualization panel, compared with the 29.25 for the interactions on the input panel, while in Table 4, the mean scores were 30.65 for the interactions within the visualization panel, also slightly better than the means score of 28.15 for the interactions within the input panel. As in the results, the participants in the TAM inspection slightly prefers the use of DengueViz by visualization interactions described in 3.5 on the parallel coordinates with a dataset, rather than to select the symptoms individually for the dengue diagnosis through the user interface of DengueViz.

Table 3: TAM inspection results for usefulness

Understanding the usefulness														
	Original Version DengueViz Mockups							Visual Cues Interaction Steering Mockups						
Participant	S1	S2	S3	S4	S5	S6	Total	S1	S2	S3	S4	S5	S6	Total
P01	5	5	5	5	5	5	30	5	5	5	5	5	5	30
P02	4	4	2	4	4	4	22	4	4	5	5	5	4	27
P03	4	5	4	6	5	5	29	5	6	6	6	5	4	32
P04	5	6	6	5	5	4	31	5	6	6	5	6	6	34
P05	4	4	4	4	4	4	24	5	5	5	5	6	6	32
P06	3	4	4	3	3	3	20	3	3	3	3	3	3	18
P07	5	4	5	4	5	6	29	5	5	6	5	6	6	33
P08	6	4	5	5	4	6	30	4	5	5	5	6	5	30
P09	5	5	5	5	5	6	31	4	5	6	5	5	5	30
P10	6	5	6	5	6	5	33	4	4	5	5	6	5	29
P11	5	5	5	5	5	6	31	6	5	5	5	5	6	32
P12	6	5	4	4	6	5	30	6	5	6	5	6	6	34
P13	4	4	5	5	5	5	28	6	6	5	6	6	6	35
P14	4	4	4	5	6	6	29	5	6	6	6	6	6	35
P15	5	6	5	6	5	6	33	5	6	5	6	6	5	33
P16	5	6	6	5	5	4	31	5	5	6	4	5	6	31
P17	4	6	6	5	5	5	31	5	5	4	6	6	6	32
P18	4	5	5	5	5	5	29	6	5	5	5	5	5	31
P19	6	6	4	6	6	4	32	6	5	5	5	5	6	32
P20	6	5	5	6	6	4	32	6	6	6	6	6	6	36
Mean							**29.25**							**31.3**
Median							**30**							**32**
STDev							**3.46**							**3.83**

Table 4: TAM inspection result for ease of use

Understanding the Ease to use														
	Original Version DengueViz Mockups							Parallel Coordinate style visual represetation mockups						
Participant	S1	S2	S3	S4	S5	S6	Total	S1	S2	S3	S4	S5	S6	Total
P01	5	5	5	5	5	5	30	5	5	5	5	5	5	30
P02	4	4	2	3	4	4	21	3	4	5	5	5	4	26
P03	4	5	4	6	5	5	29	5	6	6	6	5	4	32
P04	4	6	6	5	5	4	30	5	6	6	5	6	6	34
P05	3	4	4	4	4	4	23	5	4	5	3	6	6	29
P06	3	4	4	3	3	3	20	4	6	3	4	4	3	24
P07	3	4	5	4	5	6	27	5	5	6	5	6	6	33
P08	3	3	5	5	4	6	26	4	5	5	5	6	5	30
P09	5	5	5	5	5	6	31	4	5	3	4	5	5	26
P10	6	3	6	5	6	5	31	4	4	5	5	6	5	29
P11	5	5	5	5	5	6	31	6	5	5	5	5	3	29
P12	6	5	4	4	6	5	30	6	6	6	5	6	6	35
P13	4	4	5	3	5	5	26	6	6	5	6	6	3	32
P14	4	4	4	5	6	6	29	5	6	6	6	6	6	35
P15	5	6	5	3	5	6	30	5	6	5	6	6	5	33
P16	5	6	3	5	3	4	26	3	5	6	4	5	6	29
P17	4	6	6	5	5	5	31	5	6	4	3	4	6	28
P18	4	5	5	5	5	5	29	6	5	5	5	5	5	31
P19	6	6	4	6	6	3	31	6	5	5	5	5	6	32
P20	6	5	5	6	6	4	32	6	6	6	6	6	6	36
Mean							**28.15**							**30.65**
Median							**29.5**							**30.5**
STDev							**3.48**							**3.25**

The TAM inspection is conducted based on the import of a specific dataset. Alongside the survey, there are no unexpected or mistakes done by the participants interacting with DengueViz. Table 5 below shows the overall results of the TAM survey on the usefulness and ease of use of DengueViz, in both the input interactions and the visualization interactions. It should also be noted that all the results of the TAM inspection are included in this research, no parts of the data have been cut off to provide a comprehensive and honest review on the DengueViz. From the overall results in Table 5, this shows DengueViz can be learned with only a slight learning curve in both interaction aspects of the system. However, the learning curve and acceptance might be different depending on the size of the dataset imported.

Table 5: TAM overall results

	Results per subject			
	Total Score understanding the usefulness		Total Score understanding the Ease of use	
Participants	DengueViz	Visual Cues	DengueViz	Parallel style
	Interface	Steering	Interface	Visual interaction
P01	30	30	30	30
P02	22	27	21	26
P03	29	32	29	32
P04	31	34	30	34
P05	24	32	23	29
P06	20	18	20	24
P07	29	33	27	33
P08	30	30	26	30
P09	31	30	31	26
P10	33	29	31	29
P11	31	32	31	29
P12	30	34	30	35
P13	28	35	26	32
P14	29	35	29	35
P15	33	33	30	33
P16	31	31	26	29
P17	31	32	31	28
P18	29	31	29	31
P19	32	32	31	32
P20	32	36	32	36
Summary of Stastical Results				
Mean	29.25	31.3	28.15	30.65
Median	30	32	29.5	30.5
Std Dev	3.483	3.249	3.483	3.249
Mode	31	32	31	29

To provide a better visual representation of the TAM inspection results, Fig. 10 shows the box plot of the overall results of usefulness and ease of use according to the pointers in Table 5.

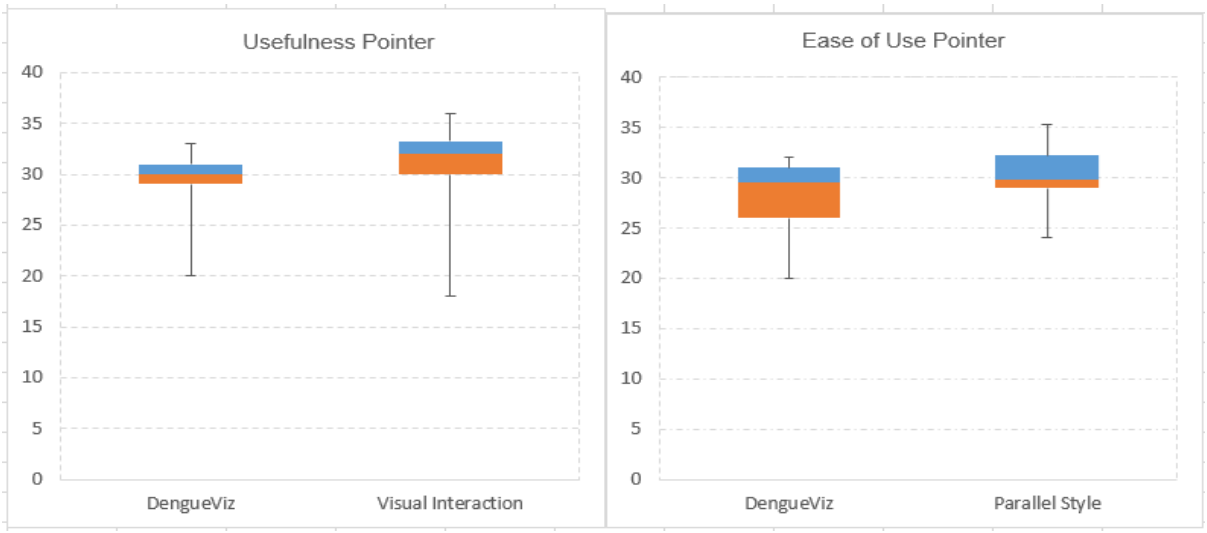

Fig. 9: TAM result distribution box-plot

Overall, the results are aligned with the responses of the participants, whom had returned mostly favorable feedback on the interactions between the two aspects of DengueViz, which the visual interactions made it easier to explore the characteristics of the dataset, and that they become more interactive with the system for better data insights, even for new users. However, participants had commented on the usability glitches such as cluttered data and slowed interaction responses when a larger dataset is imported into DengueViz [31].

5.0 CONCLUSION

In this research, the knowledge-based expert system from DengueViz has been extended to include machine learning classifiers such as the MLP, SVM, and TSE to aid in dengue diagnosis and assessing the severity of dengue. The results of the diagnosis and assessment are visually shown in parallel coordinates along with the visualization interactions for information visualization purposes. The robustness of the system is tested through the machine learning classifiers with training and validation with MLP and SVM, with outliers verified through TSE. In this case, the term probability of diagnosis is used as the system is still in the prototyping stage; this system is still only able to assist medical care experts in diagnosis and assessing the severity of dengue. Additional tests within the real-world environments are still required to verify the competence and performance of DengueViz.

ACKNOWLEDGMENT

KDU College (PG) Sdn Bhd has funded the work as an internal research grant scheme to the Department of Computing to conduct the computational medical research. We thank KDU Penang University College, Intelligent Processing Applications (IPA) research cluster under the Department of Computing has provided the venue to conduct and complete the research work. This analysis work has registered underneath National Medical analysis Register (NMRR) ID: NMRR-17-94-34388, with the analysis titled Computerized Expert System Tool to diagnose dengue fever. The article has been presented and received best paper award in IVIC2017. The author's wish to thank the IVIC2017 committee members for their LNCS publication.

REFERENCES

[1] World Health Organization (WHO), "Dengue and severe dengue," *World Health Organization (WHO)*, 15-Apr-2019. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue. [Accessed: 07-Sep-2019].

[2] F. F. Loh, "Dengue cases at an all-time high," *The Star Online*, 2019. [Online]. Available: https://www.thestar.com.my/news/nation/2019/08/10/dengue-cases-at-an-all-time-high. [Accessed: 07-Sep-2019].

[3] World Health Organization (WHO) Regional Office for South-East Asia, "Dengue: guidelines for diagnosis, treatment, prevention, and control," 2009.

[4] J. Y. L. Ooi and J. J. Thomas, "DengueViz: A knowledge-based expert system integrated with parallel coordinates visualization in the dengue diagnosis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10645 LNCS, pp. 50–61.

[5] World Health Organization (WHO), "Dengue Control," *World Health Organization (WHO)*. [Online]. Available: https://www.who.int/denguecontrol/human/en/. [Accessed: 07-Sep-2019].

[6] S. R. S. Hadinegoro, "The revised WHO dengue case classification: Does the system need to be modified?," *Paediatr. Int. Child Health*, vol. 32, no. SUPP1, pp. 33–38, May 2012.

[7] G. Watt, K. Jongsakul, C. Chouriyagune, and R. Paris, "DIFFERENTIATING DENGUE VIRUS INFECTION FROM SCRUB TYPHUS IN THAI ADULTS WITH FEVER," *Am. J. Trop. Med. Hyg.*, vol. 68, no. 5, pp. 536–538, 2003.

[8] K. Chang *et al.*, "Dengue Fever Scoring System: New Strategy for the Early Detection of Acute Dengue Virus Infection in Taiwan," *J. Formos. Med. Assoc.*, vol. 108, no. 11, pp. 879–885, 2009.

[9] S. Mitra, I. Gautam, M. Jambugulam, K. P. Abhilash, and V. Jayaseeelan, "Clinical score to differentiate scrub typhus and dengue: A tool to differentiate scrub typhus and dengue," *J. Glob. Infect. Dis.*, vol. 9, no. 1, p. 12, 2017.

[10] C. Sa-ngamuang, P. Haddawy, V. Luvira, W. Piyaphanee, S. Iamsirithaworn, and S. Lawpoolsri, "Accuracy of dengue clinical diagnosis with and without NS1 antigen rapid test: Comparison between human and Bayesian network model decision," *PLoS Negl. Trop. Dis.*, vol. 12, no. 6, pp. 1–14, Jun. 2018.

[11] C. Xuan, T. Phuong, N. T. Nhan, and R. Kneen, "Clinical Diagnosis And Assessment Of Severity Of Confirmed Dengue Infections In Vietnamese Children: Is The World Health Organization Classification System Helpful?," *Am. J. Trop. Med. Hyg.*, vol. 70, no. 2, pp. 172–179, 2004.

[12] S. Pongpan *et al.*, "Clinical Study Development of Dengue Infection Severity Score," *ISRN Pediatr.*, 2013.

[13] S. M. Pone *et al.*, "Clinical and laboratory signs associated to serious dengue disease in hospitalized children," *J Pediatr (Rio J)*, vol. 92, no. 5, pp. 464–471, 2016.

[14] S. Mishra, R. Ramanathan, and S. K. Agarwalla, "Clinical Profile of Dengue Fever in Children: A Study from Southern Odisha, India," *Scientifica (Cairo).*, vol. 2016, pp. 1–6, 2016.

[15] K. Phakhounthong *et al.*, "Predicting the severity of dengue fever in children on admission based on clinical features and laboratory indicators: Application of classification tree analysis," *BMC Pediatr.*, vol. 18, no. 1, p. 109, Mar. 2018.

[16] R. Gour, "What is Expert System in Artificial Intelligence — How it Solve Problems," *Medium*, 2018. [Online]. Available: https://medium.com/@rinu.gour123/what-is-expert-system-in-artificial-intelligence-how-it-solve-problems-83bbaf3f93c3. [Accessed: 07-Sep-2019].

[17] T. Sharma, N. Tiwari, and D. Kelkar, "STUDY OF DIFFERENCE BETWEEN FORWARD AND BACKWARD REASONING," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 10, pp. 1–3, 2012.

[18] E. H. Shortliffe, "Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases," in *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 1977, pp. 66–69.

[19] J. Brownlee, "Crash Course On Multi-Layer Perceptron Neural Networks," *Machine Learning Mastery*, 2019. [Online]. Available: https://machinelearningmastery.com/neural-networks-crash-course/. [Accessed: 08-Sep-2019].

[20] C. Nicholson, "A Beginner's Guide to Multilayer Perceptrons (MLP)," *Skymind*, 2018. [Online]. Available: https://skymind.ai/wiki/multilayer-perceptron. [Accessed: 08-Sep-2019].

[21] J. D. Mello-Román, J. C. Mello-Román, S. Gómez-Guerrero, and M. García-Torres, "Predictive Models for the Medical Diagnosis of Dengue: A Case Study in Paraguay," *Comput. Math. Methods Med.*, vol. 2019, pp. 1–7, Jul. 2019.

[22] T. Faisal, M. N. Taib, and F. Ibrahim, "Neural network diagnostic system for dengue patients risk classification," *J. Med. Syst.*, vol. 36, no. 2, pp. 661–676, Apr. 2012.

[23] A. Laureano-Rosario *et al.*, "Application of Artificial Neural Networks for Dengue Fever Outbreak Predictions in the Northwest Coast of Yucatan, Mexico and San Juan, Puerto Rico," *Trop. Med. Infect. Dis.*, vol. 3, no. 1, p. 5, Jan. 2018.

[24] R. Gandhi, "Support Vector Machine — Introduction to Machine Learning Algorithms," *Medium*, 2018. [Online]. Available: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47. [Accessed: 08-Sep-2019].

[25] X. Dang, H. Peng, X. Wang, and H. Zhang, "The Theil-Sen Estimators in a Multiple Linear Regression Model," *Manuscript, Available at URL: ...*, no. 1993. pp. 1–30, 2009.

[26] L. Chittaro, "Information visualization and its application to medicine," *Artif. Intell. Med.*, vol. 22, no. 2, pp. 81–88, 2001.

[27] T. Munzner, *Visualization analysis & design*. Boca Raton: CRC Press, 2014.

[28] S. Steinparz, R. Abmair, A. Bauer, and J. Feiner, "InfoVis – parallel coordinates," *Technology*. pp. 1–21, 2010.

[29] G. Riley, "A Tool for Building Expert Systems," *CLIPS*. 2013.

[30] F. D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Q.*, vol. 13, no. 3, p. 319, Sep. 1989.

[31] Ooi, J. Y. L., & Thomas, J. J. (2017, November). DengueViz: A Knowledge-Based Expert System Integrated with Parallel Coordinates Visualization in the Dengue Diagnosis. In International Visual Informatics Conference (pp. 50-61). Springer, Cham