A NOVEL PREDICTION SYSTEM IN DENGUE FEVER USING NARMAX MODEL

H. Abdul Rahim[1], F. Ibrahim[2] Member IEEE, and M. N. Taib[3] Senior Member IEEE

[1]Department of Control and Instrumentation, Faculty of Electrical Engineering, Universiti Teknologi Malaysia,
81310 UTM Skudai, Johor, Malaysia.
(E-mail: herlina@fke.utm.my)
[2]Department of Biomedical Engineering, Faculty of Engineering, University of Malaya,
50603 Kuala Lumpur, Malaysia.
[3]Faculty of Electrical Engineering, Universiti Teknologi Mara,
40450 Shah Alam, Selangor, Malaysia.

Abstract - This paper describes the development of nonlinear autoregressive moving average with exogenous input (NARMAX) models in diagnosing dengue infection. The developed system bases its prediction solely on the bioelectrical impedance parameters and physiological data. Three different NARMAX model order selection criteria namely FPE, AIC and Lipschitz have been evaluated and analyzed. This model is divided two approaches which are unregularized approach and regularized approach. The results show that using Lipschitz number with regularized approach yield better accuracy by 88.40% to diagnose the dengue infections disease. Furthermore, this analysis show that the NARMAX model yield better accuracy as compared to autoregressive moving average with exogenous input (ARMAX) model in diagnosis intelligent system based on the input variables namely gender, weight, vomiting, reactance and the day of the fever as recommended by the outcomes of statistical tests with 76.70% accuracy.

Keywords: modeling, NARMAX, dengue fever.

1. INTRODUCTION

This section describes the general methodology conducted for classifying the dengue infections disease as shown in a flowchart in Figure 1 which consists of five stages.
The first stage began with the data collection on the dengue disease. At the second stage, select the model order criterion to apply the appropriate model. After that, select a model structure which is nonlinear system (NARMAX model). Once of the model structure has been chosen, next stage is to pick one particular model of this set. The model must provide the best predictions in terms of the highest AUC percentage. This process is in statistical literature known as estimation. At the fourth stage, Receiver operating characteristic (ROC) curve analysis was applied to illustrate the sensitivity, the specificity and the AUC percentage of the appropriate model. At the final stage, if the model is not good then this procedure maybe repeated from beginning or selects another structure or just looking at another model estimate.
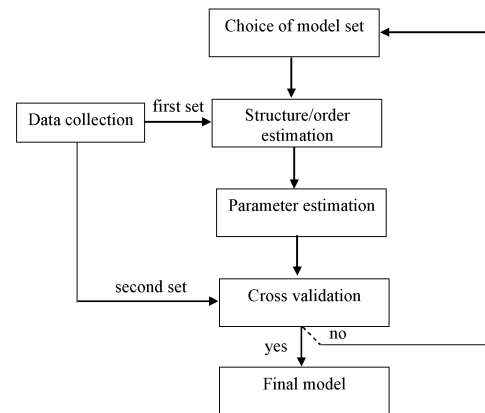

Fig. 1 Steps in identification process

It is a common practice in various scientific and engineering disciplines to represent observed discrete time random processes by nonlinear autoregressive moving average with exogenous input (NARMAX) models.
A fundamental problem in system identification is the choice of the nature of the model which should be used of the system. Some of the problems in system identification are:
i)   determining the order of the linear model
ii)   selection of a suitable criterion for   determining the accuracy of the model

iii) designing an input signal which will   maximize the
accuracy of the estimates of   the parameter of the
model.

2. RECEIVER OPERATING CHARACTERISTIC (ROC) CURVE

In 1971, Lusted [1] described how ROC curves could be
used to assess the accuracy of the test. ROC curves is a
plot of test sensitivity (plotted on the y axis) versus its 1-
specificity (plotted on the x axis). Each point on the graph
is generated by using a different threshold. The set of data
points generated from the different thresholds is the
empirical ROC curve.

The ROC plot has many advantages over single
measurements of sensitivity and specificity [2].   The
scales of the curve, that is, sensitivity and 1- specificity
are the basic measurement of accuracy and are easily read
from the plot; the values of the threshold are often labeled
on the curve as well.

One of the most popular measures of the accuracy of
diagnostic test is the area under (AUC) the ROC curve.
The ROC curve area can be chosen between the range of
0.0 to 1.0. The closer the ROC curve area is to 1.0, the
better the diagnostic test. The percentage for diagnostic
accuracy (DA) refers to the percentage of samples that
have been correctly diagnosed. In any test with a fixed
threshold, it is desirable for a decision model to produce
TPR and FPR pair nearby to this point.   Therefore,
measurement of *Euclidean Distance* (*ED*) of any
coordinate pairs in the plot to this ideal point would
distinctively differentiate performance between models
for a fixed threshold.

Figure 2 show three ROC curves representing excellent,
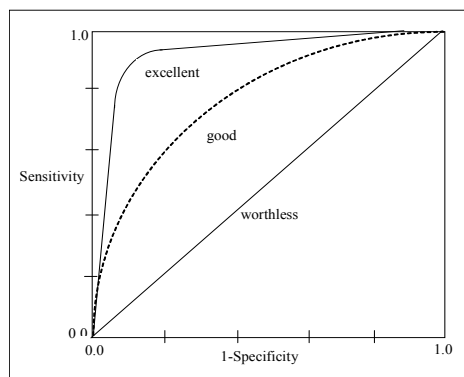good and worthless tests plotted.

Fig. 2 Graphs chows comparison of three types of ROC curves

3.   METHODS

Two hundred ten adult patients aged 12 years old and
above, suspected of DF and DHF admitted to the
Universiti Kebangsaan Malaysia Hospital (HUKM), were
studied. The five input variables used are gender, weight,
reactance (Xc), vomiting, and day of fever [3-5]. These
input variables were used to determine the order of
ARMAX model. Orders of the ARMAX model chosen in
this analysis are FPE, AIC and Lipschitz.

The accuracy value of hemoglobin based on model fitted
was observed to evaluate the ability of the 3 different
models order selection criteria chosen.

4.   RESULTS

Firstly, the best hidden layer based on the DA and the
smallest value of error was found. Figure 3 shows the
example of the Lipschitz number. This example with 2
hidden layers is selected because it has the DA of 87.91%
and a small proportion of 0.0573 for the FPE values.

Next, the best iteration also based on DA and the smallest
value of error was found. From the Figure 4, it is shown
that the best of iteration is 500 for Lipschitz number
whereas DA is 84.62% while a small value of the FPE is
0.0380.

The best regularization parameters selected are 0.0002
because the DA is 83.52% and the values of the FPE is
0.0477, hence met the maximum iteration (500) as shown
in Figure 5.

Summary of the processing steps are mentioned with 3
different types of model order criterion as shown in Table
1.

Table 1 Parameters for NARMAX Model

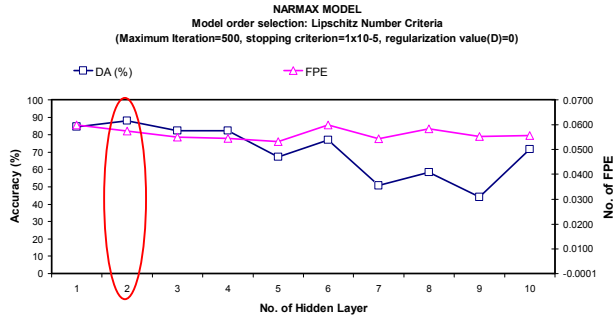| Parameter	Lipschitz	FPE	AIC
Model order	4,2,2,2	15,3,1,1	25,3,8,1
Hidden Layer	2	5	4
Maximum Iteration	500	500	500
Unregularization	0	0	0
Regularization	2×10^{-3}	4×10^{-3}	3×10^{-3}

Fig. 3 The best hidden layer of Lipschitz number for dengue patients using NARMAX model.
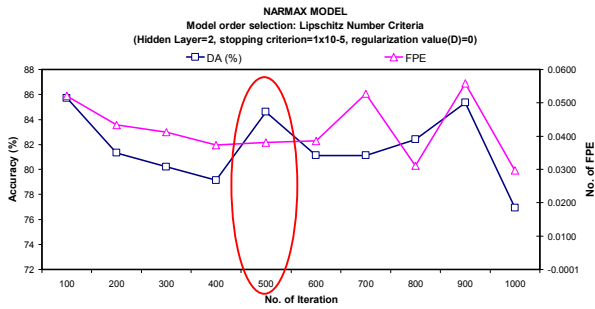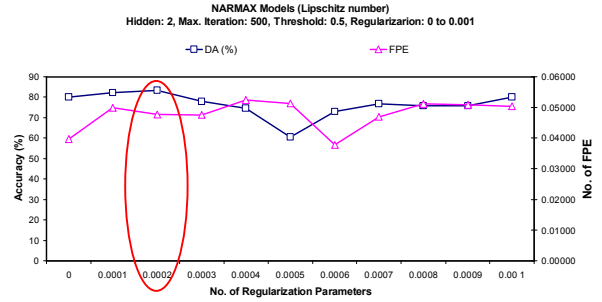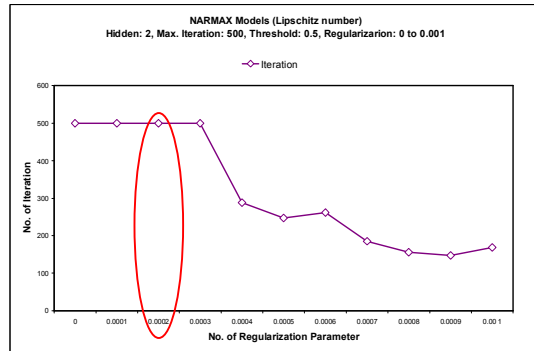

Fig. 4 The best iteration of Lipschitz number for dengue patients using NARMAX model..

(i)

(ii)

Fig. 5 (i) The best regularization parameters for regularized approach, (ii) the maximum iteration using NARMAX model.

In general, Table 2 shows the difference number of the model order using different types of criteria and the AUC performance. From this table, it was found that the Lipschitz number criterion for regularized approach produces the highest accuracy (88.40%) for the NARMAX model.

Table 2 Comparing of NARMAX models with the different number model order criteria for AUC performance.

Criteria	Model order	AUC (%) unregularized	AUC (%) regularized
Lipschitz	4,2,2,2	79.40	88.40
FPE	15,3,1,1	74.90	76.50
AIC	25,3,8,1	69.60	72.70

The model order as given by the Lipschitz number criterion was tested using Neural Network based ARMAX model. The overall performance of NARMAX model diagnosis is as shown in Table 3. The ROC curve for the respective model is shown in Figure 6. The total AUC can be derived by combining the individual area with respect to the labeled thresholds in the figure. As an example of the Lipschits number, AUC is 88.40%. The

closest ED is depicted from the ideal point (0,1) as 0.183
when the optimized model has a threshold of 0.5.

Table 3   The accuracy of the diagnostic test using
NARMAX models with different approaches.

	Lipschitz	FPE	AIC
Unregularized			
Sensitivity	87.14	83.58	83.33
Specificity	78.19	83.33	80.00
Diagnostic Accuracy	84.62	83.53	82.67
Euclidean Distance from point (0,1)	0.249	0.234	0.260
Regularized			
Sensitivity	88.57	85.07	81.67
Specificity	85.71	77.78	80.00
Diagnostic Accuracy	87.91	83.53	81.33
Euclidean Distance from point (0,1)	0.183	0.268	0.271

The diagnostic accuracy of 84.62% was achieved for the
unregularized method, whereas a small proportion of
15.38% of false classifications (diagnostic error) have
been observed for the total test group of 100 subjects. A
87.14% sensitivity, 78.19% specificity and 92.86% of
positive prediction is evaluated for the designed
classification structure. The regularized method illustrates
approximately 87.91% of accuracy in diagnosis while
12.09% is indicated diagnostic error. Overall, the designed
classification structure has about 88.57% sensitivity,
85.71% specificity and positive prediction calculation
lingers around 95.38%. The performances were measures
based on the receiver operating characteristic (ROC)
curves. The area under the ROC curve (AUC) is divided
into 79.40% of unregularized approach and 88.40% of
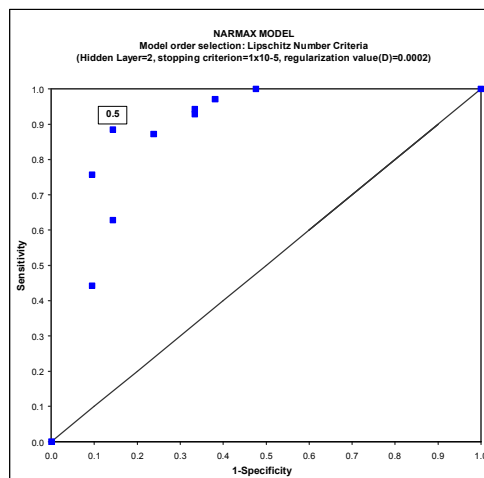regularized approach.

Fig. 6 ROC curve for NARMAX regularized model

After diagnosing the dengue patients, the following
technique is to predict the status of hemoglobin in infected
patients. This technique is able to predict Hb status with
better accuracy by using only five predictors such as
reactance, gender, weight, vomiting and the day of fever.
For example in Figure 7 the value of reactance is 52.9
ohm, gender is female, weight is 32.5 kg, and this patient
was vomit in the third day of fever. Hence, this patient is
infected and the prediction of Hb is 13.9411. The actual
value of the Hb was 13.23. On the other hand, Figure 8
illustrates a healthy patient.

Fig. 7 Infected patient

Fig. 8 Uninfected patient

5.   CONCLUSIONS

Three different NARMAX model order selection criteria namely FPE, AIC and Lipschitz have been evaluated and analyzed using FFNN. The test data set consists of 21 healthy patients and 70 dengue patients.

Lipschitz number with regularized approach for diagnosis dengue infection was chosen by analyzing the percentage successful rate of sensitivity (88.57%), specificity (85.71%), diagnostic accuracy (87.91%) and finally, the AUC (88.40%) and minimum ED value (0.18) from the respective ROC plots.   The results show that using Lipschitz number with regularized approach yield better accuracy by 88.40% to diagnose the dengue infections disease. Furthermore, this analysis show that the NARMAX model yield better accuracy as compared to autoregressive moving average with exogenous input (ARMAX) model in diagnosis intelligent system based on the input variables namely gender, weight, vomiting, reactance and the day of the fever as recommended by the outcomes of statistical tests with 76.70% accuracy [6].

After diagnosing the dengue patients, the following technique is to predict the status of hemoglobin in infected patients. This technique is able to predict Hb status with better accuracy by using only five predictors such as reactance, gender, weight, vomiting and the day of fever.

ACKNOWLEDGMENTS

REFERENCES

[1]   L. B. Lusted, "Signal detectability and medical decision-making," *Science*, vol. 171, pp. 1217-1219, 1971.

[2]   M.H. Zweigh, and G. Campbell, "Receiver-operating characteristic (ROC) plot: a fundamental evaluation tool in clinical medicine," *Clin. Chem.*, vol. 39, pp. 561-577, 1993.

[3]   F. Ibrahim, "Prognosis of dengue fever and dengue haemorrhagic fever using bioelectrical impedance," PhD Thesis, Department of Biomedical Engineering,University of Malaya, July, 2005.

[4]   F. Ibrahim, N.A. Ismail, M.N. Taib and W.A.B. Wan Abas, "Modeling of hemoglobin in dengue fever and dengue hemorrhagic fever using biolectrical impedance " *Physiol. Meas.*, vol. 25, pp. 607-615, 2004.

[5]   A.R. Herlina, I. Fatimah, and T. Mohd Nasir, "A non-invasive system for predicting hemoglobin (Hb) in dengue fever (DF) and dengue hemorrhagic fever (DHF) " in *Proc. Int. Conf. on Sensor and New Techniques in Pharmaceutical and Biomedical Research (ASIASENSE)*, Kuala Lumpur, 2005.

[6]   H. Abdul Rahim, F. Ibrahim, and M.N. Taib, "A novel dengue fever (DF) and dengue haemorrhagic fever (DHF) analysis using ARMAX model " in *Proc. Int. Conf. on Sensor and New Techniques in Pharmaceutical and Biomedical Research (ASIASENSE)*, Manila, 2007.