Improved Adaptive Neuro-Fuzzy Inference System for HIV/AIDS Time Series Prediction

Purwanto[1,3,*], C. Eswaran[1], and R. Logeswaran[2]

[1] Faculty of Information Technology, Multimedia University, 63100 Cyberjaya, Malaysia
[2] Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Malaysia
[3] Faculty of Computer Science, Dian Nuswantoro University, 50131 Semarang, Indonesia
mypoenk@gmail.com, {eswaran,loges}@mmu.edu.my

Abstract. Improving accuracy in time series prediction has always been a challenging task for researchers. Prediction of time series data in healthcare such as HIV/AIDS data assumes importance in healthcare management. Statistical techniques such as moving average (MA), weighted moving average (WMA) and autoregressive integrated moving average (ARIMA) models have limitations in handling the non-linear relationships among the data. Artificial intelligence (AI) techniques such as neural networks are considered to be better for prediction of non-linear data. In general, for complex healthcare data, it may be difficult to obtain high prediction accuracy rates using the statistical or AI models individually. To solve this problem, a hybrid model such as adaptive neuro-fuzzy inference system (ANFIS) is required. In this paper, we propose an improved ANFIS model to predict HIV/AIDS data. Using two statistical indicators, namely, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), the prediction accuracy of the proposed model is compared with the accuracies obtained with MA, WMA, ARIMA and Neural Network models based on HIV/AIDS data. The results indicate that the proposed model yields improvements as high as 87.84% compared to the other models.

Keywords: Adaptive Neuro-Fuzzy Inference Systems, Neural Network, ARIMA, Moving Average.

1 Introduction

Human immunodeficiency virus (HIV) / Acquired immune deficiency syndrome (AIDS) has become a serious threat around the world due to lack of affordable effective drugs and vaccines for prevention and cure. This disease also has a long asymptomatic (without symptoms) phase. The number of cases of HIV/ AIDS has increased despite various preventive measures. No country is unaffected by this disease [1].

The spread of HIV / AIDS cases will cause an adverse effect on the development of a country. It not only affects the health sector but also the socio-economic situation. Moreover, this disease is most prevalent in the productive age group.

* Corresponding author.

A. Abd Manaf et al. (Eds.): ICIEIS 2011, Part III, CCIS 253, pp. 1–13, 2011.

Therefore, information about the development and prediction of new HIV / AIDS cases are needed to assess the magnitude of the problem for prevention and mitigation. Good and accurate prediction is very helpful in devising appropriate action plans.

Many models have been applied in time series prediction such as moving average [2] and autoregressive integrated moving average (ARIMA) [3]. Seasonal ARIMA model has been used to predict the AIDS incidence [4]. Box-Jenkins ARIMA model has been used to predict cases of incident HIV infection [5]. However, traditional statistical techniques may not produce satisfactory results for time series prediction. Recent studies have discussed the problem of time series prediction using different concepts, including artificial neural networks that have self learning capabilities, to handle non-linear data and have been used in many applications [6]-[9].

In soft computing, fuzzy logic can tolerate imprecise information, and also can make an approximate reasoning framework. Unfortunately, fuzzy logic lacks self-learning capability. The Adaptive Neuro-Fuzzy Inference System is a combination artificial neural networks and fuzzy logic that has been used to predict observed real world time series [10]-[16]. An issue that has gained much attention with regard to the ANFIS model is how to determine the appropriate input lags for univariate time series.

In this paper, we propose an improved ANFIS model to predict HIV/AIDS time series data. A new procedure is presented to determine the accurate number of input lags in ANFIS model for univariate time series prediction. The proposed model is then tested using the HIV/ AIDS data obtained from the health department of Indonesia. We also compare the proposed model with neural network and statistical models.

2 Model Used

The following is a brief description of the time series prediction models such as moving average (MA) and autoregressive integrated moving average (ARIMA), neural network and adaptive neuro-fuzzy inference systems (ANFIS) models used in this study.

2.1 Moving Average (MA) Model

A moving average model provides an efficient mechanism to obtain the value of a stationary time series prediction. The MA model is one of the most widely used models for time series prediction. In this paper, we use MA and weighted moving average (WMA) model to predict univariate time series data.

The MA model of span m at time t is calculated as [17]:

$$\hat{y}_{t+1} = \frac{1}{m} \sum_{i=t-m+1}^{t} y_i \tag{1}$$

The weighted moving average model uses different weights for the past observations as shown in Eq. (2):

$$\hat{y}_t = w_1 y_{t\text{-}1} + w_2 y_{t\text{-}2} + w_3 y_{t\text{-}3} + \ldots + w_m y_{t\text{-}m} \tag{2}$$

where $w_1, w_2, \ldots, w_m$ denote the weights associated with the past observed values.

2.2 Autoregressive Integrated Moving Average (ARIMA) Model

Box and Jenkins [3] have popularized auto regressive moving average (ARMA) and auto regressive integrated moving average (ARIMA) models for time series prediction. An ARMA model assumes that the time series data for prediction is stationary. The ARMA model is made up an AR(*p*) autoregressive part and a MA (*q*) moving average part. The ARMA (*p*,*q*) model is calculated by [18]:

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + e_t + \sum_{j=1}^{q} \theta_j e_{t-j} \tag{3}$$

where e_t is residual at time *t*, ϕ_i (i= 1, 2, …, p) are the parameters of the autoregressive part and θ_j (j= 1, 2, …, q) are the parameters of the moving average part.

The autoregressive integrated moving average (ARIMA) model is an ARMA model employed for time series data that uses ordinary differencing (*d*). General form of the ARIMA (*p, d, q*) model could be computed as [18]:

$$\phi(B)(1-B)^d x_t = \theta(B)e_t \tag{4}$$

where *p* is the number of autoregressive lags, *d* the number of differences, *q* is the number of moving average lags, B is the backward shift operator.

2.3 Neural Network Model

The neural network model used in this study is the Multilayer Perceptron (MLP). There are many fields of application for the MLP model such as classification, pattern recognition and prediction. The MLP model is the most common neural network model used in prediction [19]. It consists of an input layer, one hidden layer and an output layers. The output layer has one neuron and a variable number of neurons or nodes exist in the input and hidden layers. Each neuron has a configurable bias and the strength of each connection of a node to another is determined by a flexible weight on each connection [19]. Processing in each neuron is via the summing of the multiplication results of the connection weights by the input data. The result of the processing of the neuron is transferred to the next neuron through an activation function. There are several kinds of activation functions, such as bipolar sigmoid, sigmoid, and hyperbolic tangent

The time series prediction output, $Y(x)$, of the MLP is calculated as [19]:

$$Y(x) = \beta_0 + \sum_{j=1}^{H} \beta_j \psi(\gamma_{j0} + \sum_{i=1}^{n} \gamma_{ji} x_i) \tag{5}$$

where (β_0, β_1, …,β_H) and (γ_{10},…, γ_{Hn}) are weights of the MLP, ψ is activation function..

2.4 Adaptive Neuro-Fuzzy Inference System Model

A neuro-fuzzy system is defined as a combination of artificial neural networks and fuzzy inference system (FIS) [21]. The Adaptive Neuro-Fuzzy Inference System or

Adaptive Network-based Fuzzy Inference System (ANFIS) is a new neuro-fuzzy model reported in [22]. In neuro-fuzzy, neural network learning process with pairs of data is used to determine the parameters of fuzzy inference system. The fuzzy reasoning mechanism is shown in Fig. 1 [22].

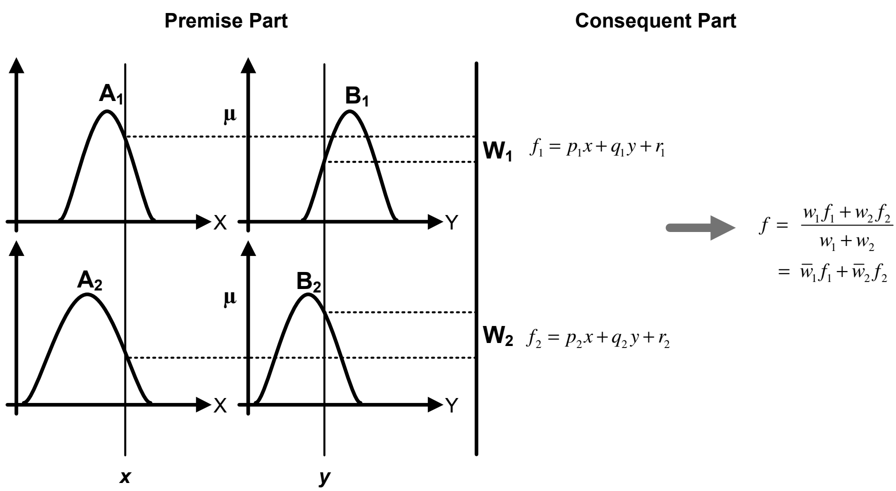

Fig. 1. Fuzzy reasoning mechanism

In Fig.1, it is assumed that the fuzzy inference system has two input variables, namely x and y, and one output f. The FIS also has two fuzzy if-then rules of Takagi and Sugeno's type which are given as [21]:

Rule 1: If x is A_1 and y is B_1 then $f_1 = p_1 x + q_1 y + r_1$

Rule 2: If x is A_2 and y is B_2 then $f_2 = p_2 x + q_2 y + r_2$.

where x and y are the input variables, A_1, A_2, B_1 and B_2 are the fuzzy sets, f_1 and f_2 are the output variables and p_1, p_2, q_1, q_2, r_1 and r_2 are the parameters.

Fig. 2 shows the structure of the ANFIS [21]. Different layers of ANFIS have different nodes. The output of each layer is used as input of the next layer.

The five layers with their associated nodes are described below:

Layer 0: The input layer. The input layer has k nodes where k is the number of inputs to the ANFIS.

Layer 1: Each node i in layer 1 is adaptive node with a function of node [21]:

$$O_{1,i} = \mu_{A_i}(x) \text{, where } i = 1, 2 \tag{6}$$

$$O_{1,i} = \mu_{B_{i-2}}(y) \text{, where } i = 3, 4 \tag{7}$$

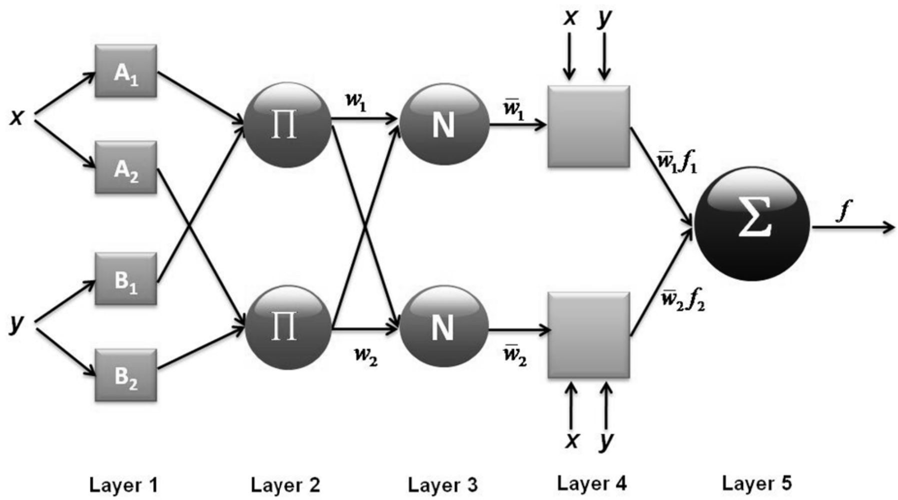

Fig. 2. Adaptive neuro-fuzzy inferences system architecture

where

x (or *y*) : input to node *I*,

A_i (or B_{i-2}) : linguistic label (low, high etc) associated with this node function,

$O_{1,i}$: membership function from fuzzy sets (A_1, A_2, B_1, B_2).

The membership function commonly used is the generalized bell [21]:

$$\mu_{A_i} = \frac{1}{1+\left[\left(\frac{x-c_i}{a_i}\right)^2\right]^{b_i}} \tag{8}$$

where $\{a_i, b_i, c_i\}$ are the set of parameters of function. The parameters in this layer are called premise parameters. Their values are adaptive by means of the back-propagation algorithm during the learning state.

Layer 2: Every node in layer 2 is a circle node, which is labeled with Π. Every node calculates the multiplication of the input values and gives the product as output, indicated by the following equation:

$$O_{2,i} = \mu_{A_i}(x)\mu_{B_i}(y)\text{, where } i=1,2 \tag{9}$$

Layer 3: Each node in layer 3 is a circle node, which is labeled with N. The *i*-th node calculates the ratio of the *i*-th rules firing strength to the sum of all rules' firing strength according to Eq. (10):

$$O_{3,i} = \overline{w}_i = \frac{w_i}{\sum_{i=1}^{2} w_i} \tag{10}$$

where w_i is firing strength of the i-th rule which is computed in layer 2. The output of this layer will be called normalized firing strength.

Layer 4: Each node i in layer 4 is a square node. The node function is given as [21]:

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i) \tag{11}$$

where $\overline{w}_i$ is the output of layer 3, $\{p_i, q_i, r_i\}$ are the set of parameters. The parameters in layer 4 are referred consequent parameters.

Layer 5: The single node in layer 5 is a circle node, which is labeled with $\sum$. This layer is as the output layer. The value of output is obtained the summation of all incoming signals. The output is calculated as:

$$O_{5,i} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{12}$$

where $\overline{w}_i f_i$ is the output of node i in layer 4.

The learning algorithm for ANFIS uses a hybrid algorithm in order to train the network, which is a combination of the least-squares estimator (LSE) and error back propagation (EBP) method [21]. Error back-propagation method is used to determine the parameters in layer 1. For training the parameters in layer 4, LSE is used.

2.5 Performance Measures

Two performance measures are used to compare the performances of obtained MA, WMA, ARIMA and ANFIS models. The following statistical indicators are used for this work: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) [23, 24]:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} \left(Y_t - \hat{Y}_t\right)^2}{n}} \tag{13}$$

$$MAE = \frac{\sum_{t=1}^{n} \left|Y_t - \hat{Y}_t\right|}{n} \tag{14}$$

where Y_t and $\hat{Y}_t$ are the observed and predicted values at time t respectively, and n is the number of data.

3 Data Used

The HIV/AIDS data used in this study for evaluating the performance were collected from the Department of Health Republic, Indonesia. The data comprises the number of HIV/AIDS for the years 1990 to 2009. The descriptive statistics of the HIV/AIDS data such as the minimum, maximum, mean and standard deviation are shown in Table. 1.

Table 1. Descriptive statistics of HIV/AIDS data

Name	Min	Max	Mean	Std. Dev.
HIV/AIDS	5.00	4969.00	9.98×10^2	1545.27

From Table 1, it is seen that the HIV/AIDS data have a high value of standard deviation. This indicates that the HIV/AIDS data are spread over a wide range of values.

4 Methodology

The input variables that are used have different patterns. The learning process will be performed based on the number of inputs. The purpose of the learning process is to study the pattern of time series data and get the value of ANFIS parameters (premise and consequent) that is used for time series prediction. In the learning process of ANFIS for univariate time series, data are divided as input and target/ output. Fig. 3 illustrates the division of univariate time series data [25].

From Fig. 3, the pattern of univariate time series data for ANFIS is presented in Table 2 as follows:

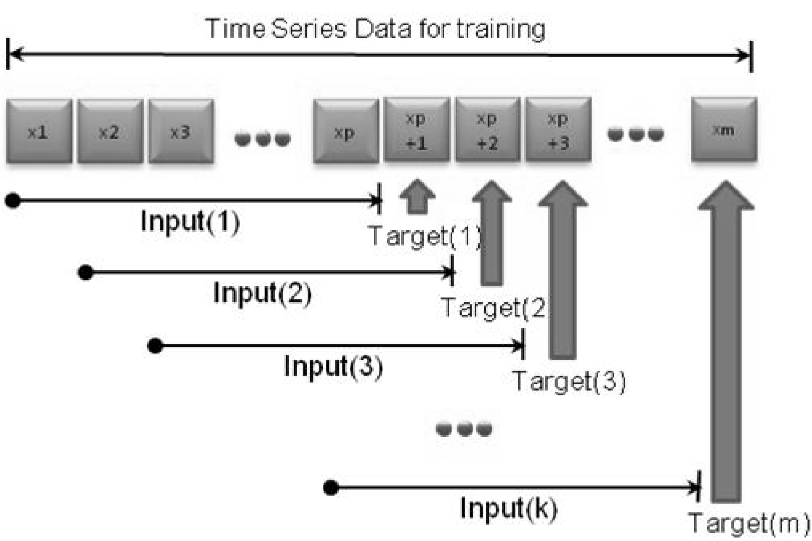

Fig. 3. The division of univariate time series data for ANFIS

Table 2. The pattern of univariate time series data for ANFIS

Pattern	Input lag	Output/ Target
1	$x_1, x_2, x_3, x_4, \ldots, x_p$	x_{p+1}
2	$x_2, x_3, x_4, x_5, \ldots, x_{p+1}$	x_{p+2}
3	$x_3, x_4, x_5, x_6, \ldots, x_{p+2}$	x_{p+3}
...	...	....
m-p	$x_{m-p}, x_{m-p+1}, x_{m-p+2}, \ldots, x_{m-1}$	x_m

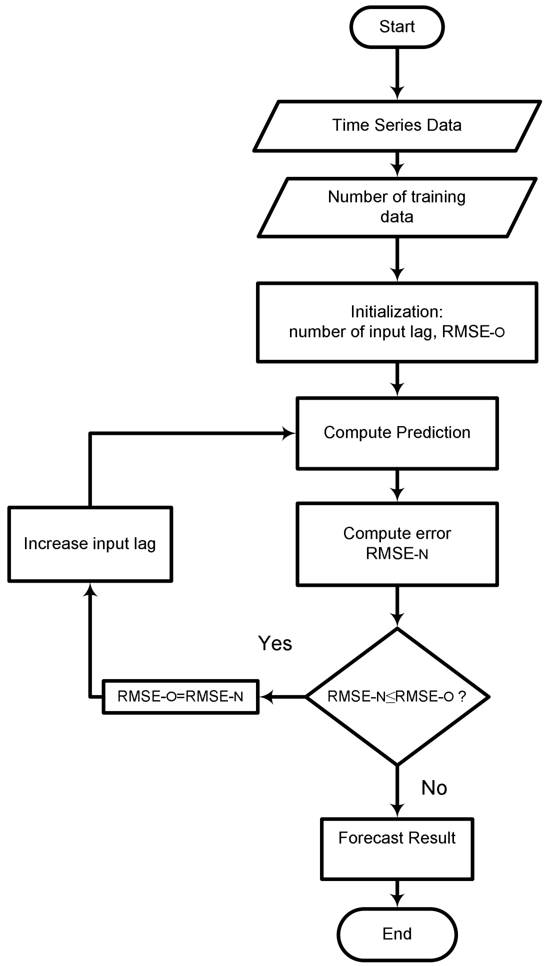

Fig. 4. The proposed procedure of ANFIS model for univariate time series prediction

ANFIS method performs learning process based on the input data that will be trained. The ANFIS algorithm that is used in this work is a hybrid learning algorithm. The algorithm uses a combination of the Least-squares estimator (LSE) method and error back propagation (EBP).

The proposed procedure of ANFIS model for univariate time series prediction in this work is shown in Fig. 4. This procedure is used to determine the optimal number of input for univariate time series prediction. In the first step, training data is applied to the ANFIS model using number of input data and target/ output, as described in Fig. 3. For Initialization, ANFIS model uses a small number of input lag. The values of premises and consequent parameters are obtained from the ANFIS model shown in Fig. 2. Furthermore, we calculate the performance of the prediction using RMSE. The next step is to test the performance of prediction ($RMSE\text{-}_{New} \leq RMSE\text{-}_{Old}$). The input lags are increased one by one to improve the performance of prediction. If value of $RMSE\text{-}_{New}$ is greater than the value of $RMSE\text{-}_{Old}$, then the iteration will be stopped. The final step, we calculate the prediction using the best configuration with optimum number of input lags.

5 Experimental Results

The performance of the proposed ANFIS model is compared with MA, WMA, ARIMA and neural network models using the HIV/AIDS data for the period 1990 to 2009 collected in Indonesia.

5.1 Moving Average Model

The moving average consists of MA and WMA models are used for time series prediction. Several numbers of inputs are made to choose the optimal moving average model. The performance measures obtained using different number of inputs (*m*) is shown in Table 3. It is found that MA (2) and WMA (2) perform better than other models with WMA (2) yielding the minimum values of RMSE and MAE.

Table 3. Performance measures using MA and WMA for HIV/ AIDS data

Model	(m)	PERFORMANCE MEASURES	
		RMSE	MAE
MA	**2**	728.028	362.222
	3	853.420	478.431
	5	1193.559	748.613
	7	1497.661	1021.978
	9	1781.642	1319.384
WMA	**2**	**696.274**	**353.741**
	3	773.803	410.745
	5	997.877	604.027
	7	1240.606	832.006
	9	1494.362	1098.236

5.2 ARIMA Model

In this section, the ARIMA models are used to predict HIV/AIDS time series. The ARIMA model with different parameter (*p, d, q*) values are computed to choose the optimal ARIMA model. It is known that an ARIMA models assume that the data are stationary. If data are not stationary, they are made stationary by performing differencing. We calculate autocorrelation of the HIV/AIDS data to check whether the data are stationary.

Table 4. Performance measures using ARIMA models for HIV/AIDS data

MODELS	PERFORMANCE MEASURES	
	RMSE	MAE
ARIMA(1,1,1)	664.510	432.039
ARIMA(2,1,1)	687.290	420.387
ARIMA(3,1,3)	**642.200**	**364.017**
ARIMA(4,2,3)	680.721	380.767
ARIMA(7,1,3)	767.718	371.315

The result of performance measures with ARIMA model is shown in Table 4. From Table 4, it is seen that the minimum values of RMSE and MAE are obtained for the ARIMA(3,1,3) model.

5.3 Neural Network Model

The MLP model is applied for HIV/AIDS time series prediction. The architecture configurations with different numbers of input and hidden layer neurons are tested to determine the optimum setup. From the experimental results, it is found that the neural network model with 7 input neurons, 12 hidden layer neurons and using hyperbolic tangent activation functions for the hidden and output layers yields the minimum values for RMSE and MAE. The result of performance measures with Neural Network model is shown in Table 5.

Table 5. Performance measures using Neural Network models for HIV/AIDS data

MODELS (input, hidden, output)	PERFORMANCE MEASURES	
	RMSE	MAE
NN(6,10,1)	216.646	125.166
NN(6,11,1)	195.520	118.957
NN(6,12,1)	150.015	88.426
NN(7,12,1)	**143.011**	**87.732**
NN(8,12,1)	181.258	110.169
NN(7,14,1)	162.745	94.508
NN(8,14,1)	145.149	91.938

5.4 ANFIS Model

The proposed ANFIS model to predict HIV/AIDS are tested using the time series of HIV/AIDS data in the period 1990 to 2009. The ANFIS model is constructed using 20 training data. The number of rules used was 2 and the number of epochs used for training was 1000. The proposed ANFIS model is tested using the procedure of Fig.4. It is found that the optimum input lag and the corresponding RMSE were obtained as 2 and 84.698 respectively. And the value of MAE using ANFIS model with optimum input lag and 1output (abbreviated as ANFIS (2, 1)) is 49.265. The values of RMSE and MAE using ANFIS(3,1) model are obtained as 90.972 and 55.456 respectively.

6 Comparison of Models

The proposed ANFIS model is compared with known statistical and artificial intelligence models such as neural network models based on the prediction performances for the HIV/AIDS data. Fig. 5 shows a comparison of RMSE and MAE values obtained using WMA, ARIMA, Neural Network and the proposed ANFIS(2,1) models.

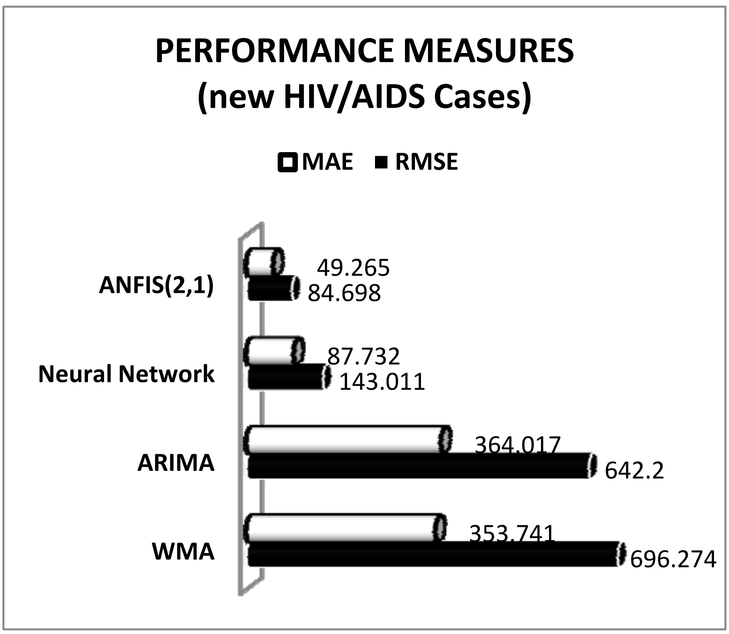

Fig. 5. Comparison of performance measures

We note that the proposed ANFIS model (ANFIS(2,1)) gives the best results compared to all other models.

The percentage improvements achieved by the proposed ANFIS with respect to RMSE and MAE values by the ANFIS model over other models are presented in

Table 6. From this table, the proposed ANFIS model is able to achieve significant performance improvements over other models for HIV/ AIDS time series prediction.

Table 6. Improvement achieved by proposed ANFIS model over the other models for HIV/AIDS data

MODELS	RMSE (%)	MAE (%)
WMA	87.84	86.07
ARIMA	86.81	86.47
Neural Network	40.78	43.85

7 Conclusion

This study has presented an improved ANFIS model for HIV/ AIDS time series prediction. The modified ANFIS model has been tested using HIV/AIDS data for a period of 20 years. The performance of the proposed model has been compared with other models using measures such as RMSE and MAE. The experimental results show that the improved ANFIS model using optimum input lag performs significantly better than MA, WMA, ARIMA and neural network models. It can be concluded that an improved ANFIS model is best suited for HIV/AIDS time series prediction.

References

1. Susilo, B., Kurniasih, N., Manullang, E., Wardah, Anam, M.S.: Istiqomah: HIV / AIDS situation in Indonesia- 1987-2006. Department of Health, Indonesia, Jakarta (2006)
2. Zhuang, Y., Chen, L., Wang, X.S., Lian, J.: A Weighted Moving Average-based Approach for Cleaning Sensor Data. In: 27th International Conference on Distributed Computing Systems, pp. 38–45 (2007)
3. Box, G., Jenkins, G.: Time Series Analysis, Forecasting and Control. Holden-Day, CA (1970)
4. Tabnak, F., Zhou, T., Sun, R., Azari, R.: Time series forecasting of AIDS incidence using mortality series. In: International Conference on AIDS (2000)
5. Aboagye-Sarfo, P., Cross, J., Mueller, U.: Trend analysis and short-term forecast of incident HIV infection in Ghana. African Journal of AIDS Research 9, 165–173 (2010)
6. Jain, B.A., Nag, B.N.: Performance Evaluation of Neural Network Decision Models. Manage Information Systems 14, 201–216 (1997)
7. Niskaa, H., Hiltunena, T., Karppinenb, A., Ruuskanena, J., Kolehmaine, M.: Evolving the Neural Network Model for Forecasting Air Pollution Time Series. Engineering Applications of Artificial Intelligence 17, 159–167 (2004)
8. Georgakarakos, S., Koutsoubas, D., Valavanis, V.: Time Series Analysis and Forecasting Techniques Applied on Loliginid and Ommastrephid Landings in Greek Waters. Fisheries Research 78, 55–71 (2006)
9. Aminian, F., Suarez, E.D., Aminian, M., Walz, D.T.: Forecasting Economic Data with Neural Networks. Computational Economics 28, 71–88 (2006)

10. Chang, F.J., Chang, Y.T.: Adaptive Neuro-Fuzzy Inference System for Prediction of Water Level in Reservoir. Advances in Water Resources 29, 1–10 (2006)
11. Tektas, M.: Weather Forecasting Using ANFIS and ARIMA Models, A Case Study for Istanbul. Environmental Research, Engineering and Management 1(51), 5–10 (2010)
12. Hernandez, S.C.A., Pedraza, M.L.F., Salcedo, P.O.J.: Comparative Analysis of Time Series Techniques ARIMA and ANFIS to Forecast Wimax Traffic. The Online Journal on Electronics and Electrical Engineering (OJEEE) 2(2), 223–228 (2010)
13. Rasit, A.T.A.: An Adaptive Neuro-Fuzzy Inference System Approach for Prediction of Power Factor in Wind Turbines. Journal of Electrical & Electronics Engineering 9(1), 905–912 (2009)
14. Caydas, U., Hascalık, A., Ekici, S.: An Adaptive Neuro-Fuzzy Inference System (ANFIS) Model for Wire-EDM. Expert Systems with Applications 36, 6135–6139 (2009)
15. Firat, M.: Artificial Intelligence Techniques for River Flow Forecasting in the Seyhan River, Catchment, Turkey. Hydrol. Earth Syst. Sci. Discuss 4, 1369–1406 (2007)
16. Atsalakis, G.S., Valavanis, K.P.: Forecasting Stock Market Short-Term Trends Using a Neuro-Fuzzy Based Methodology. Expert Systems with Applications 36, 10696–10707 (2009)
17. Makridakis, S., Wheelwright, S.C., McGee, V.E.: Metode dan aplikasi peramalan. Edisi Revisi, Jilid I, Binarupa Aksara, Jakarta (2009)
18. Brockwell, P.J., Davis, R.A.: Introduction to Time Series and Forecasting. Springer, New York (2002)
19. Suhartono: Feedforward Neural Networks Untuk Pemodelan Runtun Waktu. Gajah Mada University, Indonesia (2008)
20. Dewi, S.K.: Neuro-Fuzzy Integrasi Sistem Fuzzy Dan Jaringan Syaraf. Graha Ilmu, Indonesia, Jogjakarta (2006)
21. Jang, J.S.R., Sun, C.T., Mizutani, E.: Neuro Fuzzy and Soft Computing: A Computational Approach To Learning And Machine Intelligence. Prentice Hall International, Inc., New Jersey (1997)
22. Jang, J.: ANFIS: Adaptive Network based Fuzzy Inference System. IEEE Trans systems, Man and Cybernetics 23(3), 665–684 (1993)
23. Faruk, D.O.: A Hybrid Neural Network and ARIMA Model for Water Quality Time Series Prediction. Engineering Applications of Artificial Intelligence 23, 586–594 (2010)
24. Rojas, I., Valenzuela, O., Rojas, F., Guillen, A., Herrera, L.J., Pomares, H., Marquez, L., Pasadas, M.: Soft-Computing Techniques and ARMA Model for Time Series Prediction. Neurocomputing 71, 519–537 (2008)
25. Fariza, A., Helen, A., Rasyid, A.: Performansi Neuro Fuzzy Untuk Peramalan Data Time Series. Seminar Nasional Aplikasi Teknologi Informasi, D-77-82 (2007)