International Journal of Current Advanced Research
ISSN: O: 2319-6475, ISSN: P: 2319-6505, Impact Factor: 6.614
Available Online at www.journalijcar.org
Volume 7; Issue 5(J); May 2018; Page No. 12875-12879
DOI: http://dx.doi.org/10.24327/ijcar.2018.12879.2280

Research Article

EXPLORATION OF ARTIFICIAL NEURAL NETWORK AND SUPPORT VECTOR REGRESSION FOR MALARIA INCIDENCE PREDICTION IN AMHARA REGION, ETHIOPIA

Belay Enyew*

Department of Information Technology, University of Gondar, Ethiopia

A R T I C L E I N F O

Article History:

Received 10th February, 2018
Received in revised form 6th March, 2018 Accepted 24th April, 2018
Published online 28th May, 2018

Key words:

Malaria, Prediction, Artificial Neural Network, Support Vector Regression, Incidence

A B S T R A C T

Malaria is one of the major public health problems in Ethiopia. Early prediction of a Malaria incidence is the key for control of malaria morbidity, mortality as well as reducing the risk of transmission of malaria in the community and can help policymakers, health providers, medical officers, ministry of health and other health organizations to better target medical resources to areas of greatest need. In this study, the use of Artificial Neural Networks (ANNs) and Support Vector Regression (SVR) are explored to build malaria incidence prediction models for Amahara Region using a dataset collected from 2013 to 2017. The input parameters used are elevation, monthly rainfall, monthly average temperature, monthly average humidity and number of one month lag positive malaria cases. The developed models performance evaluated and compared based on Root Mean Square Error (RMSE), mean square error (MSE), mean absolute error (MAE) and regression coefficient(R). The results indicate that the proposed SVR model provides more accurate prediction compared to the ANN model.

INTRODUCTION

Malaria is ranked as the leading communicable disease in Ethiopia, accounting for about 30% of the overall daily lost. Approximately 68% of the total populations of 78 million people live in areas at risk of malaria. According to Ethiopia's Federal Ministry of Health (FMOH), in 2008/2009 report [1]. In Ethiopia, despite the availability of interventions, malaria remains as one of the causes of maternal and child morbidity and mortality [1]. Malaria is caused by Plasmodium parasites, which are most commonly transmitted through the bite of the Anopheles mosquito. There are several factors which affect malaria severity and transmission like climate factors (temperature, rainfall, humidity, flood, drought, disasters) and non-climate factors (human migration, construction activities, and dams) [2]. There are many traditional methods used for malaria transmission prediction like "The Liverpool Malaria Model" which is a mathematical-biological model, Auto regressive (AR), Auto-Regressive Moving average (ARMA), Auto-Regressive Integrated Moving average (ARIMA) [3]. However, modeling of malaria epidemic is challenging because of disease transmission can exhibit spatial and temporal heterogeneity, spatial autocorrelation, and seasonal variation [4]. The statistical (traditional) approach uses the assumption of linearity of factors effects.

Corresponding author:* **Belay Enyew
Department of Information Technology, University of Gondar, Ethiopia

Nevertheless, computational model based systems, developed using machine learning techniques like neural network, support vector machine , random forest and others are now a days very useful to predict and diagnose many diseases [5].

The main aim of this study is to develop and compare the performance of artificial neural networks and support vector regression model for malaria incidence prediction in Amhara region, Ethiopia. The prediction models use as input of metrological factors (monthly average relative humidity, temperature and rainfall), elevation and lag confirmed malaria cases for five years (2013-2017). Environmental factors have contributed significantly to malaria prevalence and thereby affected its distribution, seasonality, and transmission intensity [6].

Overview of support vector regression

SVM can be applied not only to classification problems but also to the case of regression [7]. In a regression SVM, we estimate the functional dependence of the dependent variable Y on a set of independent variables X. It finds a functional form for f that can correctly predict new cases that the SVM has not been presented with before. This can be achieved by training the SVM model on a sample set, i.e. training set, a process that involves, like classification and the sequential optimization of an error function. All other additional information regarding error function(s) and kernels used in SVM have been described in the supplementary material [7]. This method maps data x into a high dimensional feature space using non-linear mapping and performs linear regression in this space. A set of data (x_n, y_n) is considered where x_n is the

vector of independent variables; $y_n$is the dependent variable's actual value; n=1, 2… N and N is the total number of data pairs [8].

Overview of artificial neural network

Artificial neural networks (ANN) are inspired by the architecture of the biological nervous system, which consists of a large number of relatively simple neurons that work in parallel to facilitate rapid decision-making he most important property of artificial neural networks is their ability to learn from a training set of patterns, i.e. is able to find a model that fit the data [9]. ANN consists of a large number of highly interconnected neurons. The neurons calculate a weighted sum of the input signals and this is then passed on to an activation function [10]. While the ability to capture latent intermediate factors is attractive, the power of ANNs is their flexible nonlinear modeling capability. ANNs can adapt to the features present in the data. ANN has been shown in equation (1) to be able to provide an accurate approximation to any continuous function of the inputs if there are a sufficiently large number of units in the middle (hidden) layer [11].

$$y = f\left(\sum_{i=1}^{n} w_i x_i + a\right) \tag{1}$$

Where xi, for i=1, 2, 3…n, is the input signal; wi is weight of the i^{th} input; a is a threshold; and f is the activation function. Multi-Layer Perceptron (MLP) with Back Propagation algorithm (BP) is one of the most common types of neural networks. In MLP, the neurons are situated within different layers and there are no feedback or lateral connections. In this study the Levenberg-Marquardt method is employed to implement the MLP/BP because of its confirmed performance [8]. Neural systems are naturally organized in layers as shown in Figure 1. Layers are made up of a number of artificial neurons, which hold an activation function. Patterns are offered to the network via the input layer, which transfers to one, or more hidden layers where the actual processing is done through a system of fully or partially weighted connections. The hidden layers are then connected to an output layer where the response is output. Multilayer perceptron (MLP) is the most commonly used neural network with the back-Propagation Algorithm networks. This kind of neural networks is excellent at both prediction and classification [12].

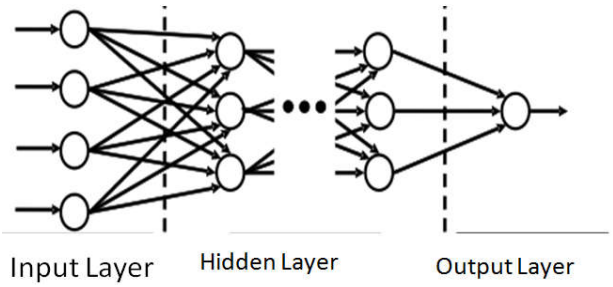

Figure 1 ANN Architecture [11]

METHODOLOGY

Methods

Study Area

The study was carried out in Amhara Region, Ethiopia. The region has a total of 17,221,976 population according to 2007 census [13].

Data

The data set used in this study includes: number of malaria cases, elevation and set of climatic factors (monthly rainfall, monthly average temperature, and monthly relative humidity, for years (2013 to 2017).

Data Preprocessing

District wise malaria data are having different population with respective number of malaria cases. To convert all those raw data into a same format, districts population data have obtain and measure all the districts on same scale.

Evaluation Criteria

In order to evaluate the performance of the developed ANN and SVR malaria prediction models quantitatively, statistical analysis involving the coefficient of determination (R), the root mean square error (RMSE), and the mean Absolut error (MAE) was conducted. RMSE provides information on the short term performance which is a measure of the variation of predicated values around themeasured data. The lower the RMSE, the more accurate is the estimation. The Mean absolute error MAE is also dependent on the scale of the dependent variable but it is less sensitive to large deviations than the usual squared loss. MAE is an indication of the average deviation of the predicted values from the corresponding measured data and can provide information on long term performance of the models; the lower MAE the better is the long term model prediction. The correlation coefficient R measure the correlation between outputs and targets value. R value of 1 and 0 means a close, random relationship respectively

The expressions for the aforementioned statistical parameters are:

The R computed as,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \tag{3}$$

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4}$$

where is the number of observation and and are the observed values, predicted values, mean value of observations and mean values of prediction of malaria cases respectively. Model adequacy was also assessed using plots of residuals (observed minus predicted) against predicted values of y to test for linear prediction bias [9].

Experimental and results

In this study, 5 climate and malaria case variables are used that have impact on malaria incidence. The main consideration for

selecting the potential variables is whether they havesignificant influence on the incidence of malaria in the next month is basedliterature reviewed and expert recommendation. The list and the description of the potential variables are given in Table 1.

Input and Output variable description

Table 1 Inputs and Output Description

Variable Name		Range	Description
Input 1	ELV	287-300 m	-The elevation here meteorology agency and Woreda is found
Input 2	RF	0-900 mm	-Monthly rainfall
Input 3	RH	20.0-96.97	-Relative Humidity
Input 4	TMP	5.4-43.6 ^{0}c	-Average monthly temperature
Input 5	LMC	0-111	-Number of positive malaria confirmed persons per month/1000 population one month lag
Output	MC	0-111	-Number of positive malaria confirmed persons per month/1000 population(current month)

ELV-Elevation, RF-Rain Fall, RH-Relative Humidity, TMP-Temperature, LMC-Lag malaria cases, MC-Malaria Case.

ANN Model Development

The proposed architecture of the multi-layer propagation (MLP) Network consists of three layers with single hidden layer as shown Figure 6. The input layer of our neural network model has 5 input nodes while the output layer consists of only one node that gives the predicted next month malaria Incidence. Empirically, 8 neurons found in the hidden layer achieved the best performance. The back propagation (BP) algorithm is used to train the MLP and update its weight. Figure 3 shows the actual and predicted malaria incidence for training cases of the developed ANN.

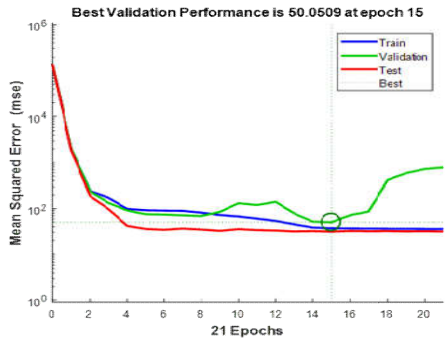

Figure 2 ANN performance Plot

The training, validation and testing error plot for the developed ANN model is shown in Figure 2 and the performance plot shows that MSE become small as number of epochs (one complete sweep of training, testing and validation) are increased.

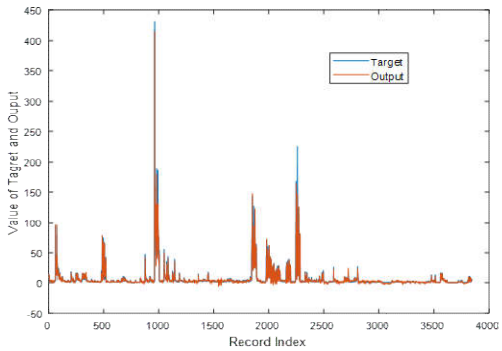

Figure 3 Targets and predicted Value of ANN model

The error histogram plot for training data is shown in Figure 4 to provide additional verification of network performance. The most data fall on zero error line which provides an idea to check the outliers to determine if the data is bad, or if those data points are different than the rest of the data set. If the outliers are valid data points, but are unlike the rest of the data, then the network is extrapolating for these points.

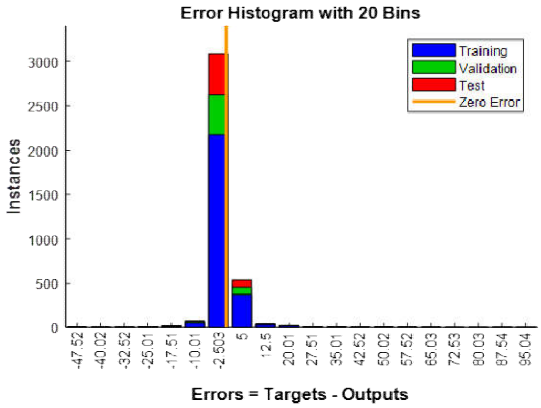

Figure 4 Error Histogram plot for Training data

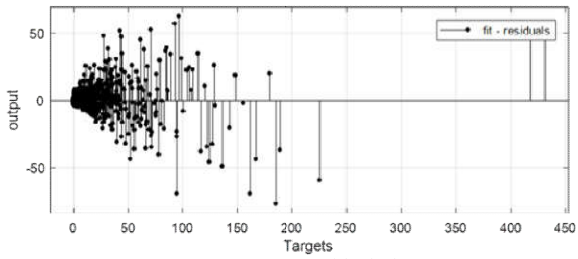

Figure 5 ANN Residual Plot.

The R-value between the predicted and the actual values of monthly malaria incidence prediction are shown in Figures 10 for training, validation, testing and the whole datasets and errors for training cities. R-values of 0.94, 0.91, 0.91 and 0.93 are obtained for the training, validation, testing and the whole dataset, respectively.

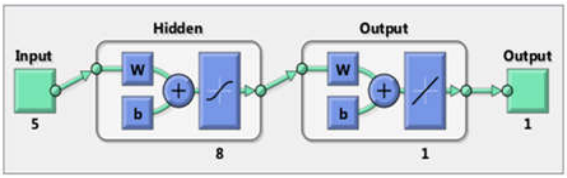

Figure 6 Trained ANN Model Layers

For Support Vector Regression Development

Figure 7 shows the actual and predicted malaria incidence of for training and testing cases of the developed SVR model with blue and yellow colors respectively. The scattered plot for the developed SVR model is shown in Figure 8.

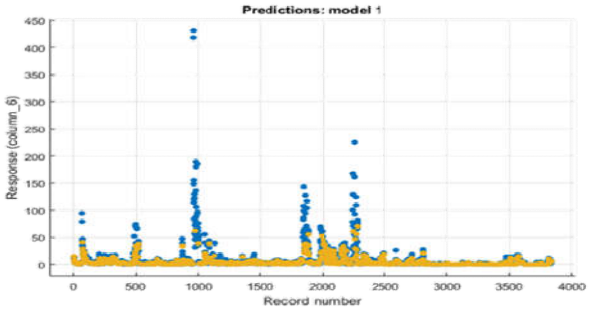

Figure 7 SVR Training data and Output plot

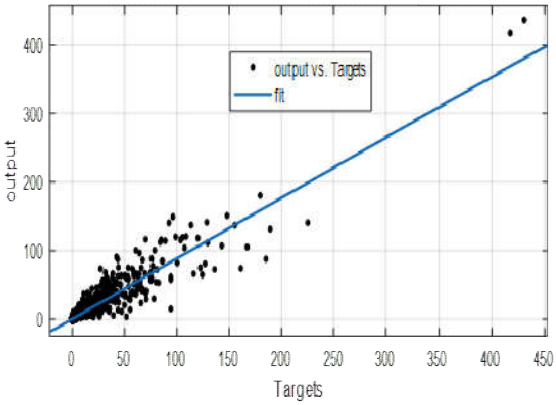

Figure 8 SVR Regression Plot

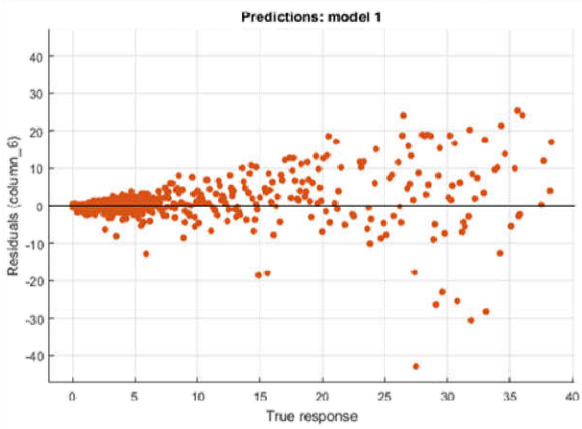

Figure 9 *SVR* Residual Plot

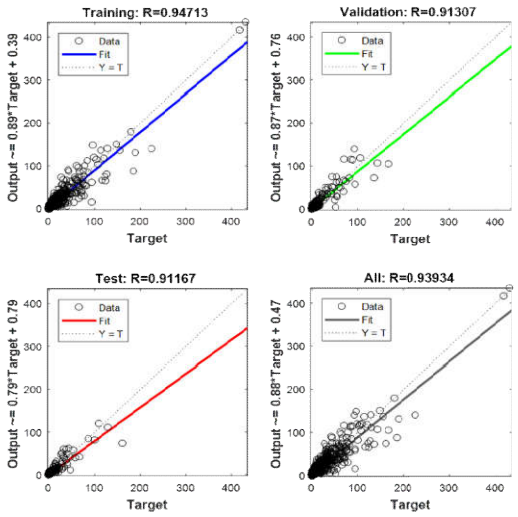

Figure 10 Regression Plot

Evaluation of the developed Models based on Testing Data

Based on table 2, the fitting of target and output correlation coefficient is 0.87 and 0.72 for ANN and SVR respectively. While RMSE, MSE and MAE are 5.57, 31.12,ans 2.51 for ANN, and 4.29, 18.17 and 1.79 for SVR respectively. Therefore, the results of this study indicate that the performance of SVR better to compare ANN model for prediction of malaria incidence using climate data (rainfall, temperature, and relative humidity), elevation and lag malaria incidence in Amhara Region, Ethiopia.

Table 2 Model performance Evaluation

Performance measure	Model	
	ANN	SVR
RMSE	5.57	4.29
MSE	31.12	18.17
MAE	2.51	1.79
R	0.87	0.72

CONCLUSION AND FUTURE WORK

After the comparative analysis, it is concluded that SVR model is more accurate and has less error rate to compare ANN model i.e. it gives us very close result. But, both methods give us a good result of prediction with one month ahead the incidence.

Malaria incidence prediction using climatic data and malaria case is complex. There will be better accuracy if hybrid machine learning algorithms employed. In addition, the accuracy of prediction may increase if other factors like chemical spray and population migration include in the prediction inputs.

Acknowledgment

My sincere gratitude goes to Amhara Regional State Public Health Institute allowed to get Malaria case data, Ethiopia Meteorology Agency cooperated to get climatic data and University of Gondar giving support in gathering data and doing the study.

Conflicts of interest

The author has no conflicts of interest to declare.

References

K. Karunamoorthi and M. Bekele, "Changes in Malaria Indices in an Ethiopian Health Centre: A Five Year," *Health Scope,* vol. 1, no. 3, pp. 118-126, 2012.

A. Alemu, G. Abebe, W. Tsegaye and L. Golassa, "Climatic variables and malaria transmission dynamics in Jimma town, South West Ethiopia," *Parasites & Vectors* , vol. 4, no. 30, 2011.

M. C. Thomson, S. J. Connor, P. J. M. Milligan and S. P. Flasse, "The ecology of malaria-as seen from Earth observation," *Annals of Tropical Medicine and Parasitology,* vol. 94, no. 3, pp. 243-2643, 1996.

R. Kasantikul, C. Rattanabumrung and P. HADAWY, "Spatiotemporal Bayesian Networks for Malaria Prediction," *EFMI,* vol. 4, 2016.

V. Sharma, A. Kumar, L. Panat and G. Karajkhede, "Malaria Outbreak Prediction Model Using Machine Learning," *IJARCET,* vol. 4, no. 12, 2015.

E. L. Darkoh, J. A. Larbi and E. A. Lawer, "A Weather-Based Prediction Model of Malaria Prevalence in Amenfi West District, Ghana," *Malaria Research and Treatment,* 2017.

R. Kaundal, A. S. Kapoor and G. P. Raghava, "Machine learning techniques in disease forecasting: a case study on rice blast prediction," *BMC Bioinformatics,* vol. 7, no. 485, 2006.

A. Shirzad, M. Tabesh and R. Farmani, " Comparison between Performance of Support Vector Regression and Artificial Neural Network in Prediction of Pipe Burst Rat in Water Distribution Networks," *KSCE Journal of Civil Engineering,* vol. 18, no. 4, pp. 941-948, 2014.

H. Ahmadi and M. Rodehutscord, "Application of Neural Netwok and Support vector Machine in predicting Metabolizable Energy in compound feed of Pigs," *Frontier Nutration,* 2017.

Mislan, Haviluddin, i. Hardwinarto and Sumaryono, "Rainfall Monthly Prediction Based on Artificial Neural Network:A Case Study in Tenggarong Station, East Kalimantan," in *ScienceDirect*, Indonesia, 2015.

G. P. Zhang, Neural Networks in business forecasting, Idea Group Publishing, 2004.

M. Zaidi, "Forecasting Stock Market Trends By Logistic Regression And Neural Networks," *IJECM,* vol. 4, no. 6, 2016.

E. S. Agencey, "Ethiopia Statistical Agencey," Addis Ababa, 2008.

How to cite this article:

Belay Enyew (2018) 'Exploration of Artificial Neural Network and Support Vector Regression For Malaria Incidence Prediction in Amhara Region, Ethiopia', *International Journal of Current Advanced Research,* 07(5), pp. 12875-12879. DOI: http://dx.doi.org/10.24327/ijcar.2018.12879.2280

****