Environment. Technology. Resources, Rezekne, Latvia
Proceedings of the 10th International Scientific and Practical Conference. Volume III, 180-185

NDVI Short-Term Forecasting Using Recurrent Neural Networks

Arthur Stepchenko[1], Jurij Chizhov[2]
[1] *Ventspils University College*
[2] *Riga Technical University*

Abstract. **In this paper predictions of the Normalized Difference Vegetation Index (NDVI) data recorded by satellites over Ventspils Municipality in Courland, Latvia are discussed. NDVI is an important variable for vegetation forecasting and management of various problems, such as climate change monitoring, energy usage monitoring, managing the consumption of natural resources, agricultural productivity monitoring, drought monitoring and forest fire detection. Artificial Neural Networks (ANN) are computational models and universal approximators, which are widely used for nonlinear, non-stationary and dynamical process modeling and forecasting. In this paper Elman Recurrent Neural Networks (ERNN) are used to make one-step-ahead prediction of univariate NDVI time series.**

Keywords: **Artificial Neural Networks, Elman Recurrent Neural Networks, Normalized Difference Vegetation Index.**

I INTRODUCTION

Human activities affect ecosystems, including the natural vegetation cover. Vegetation cover change is important factors that affect ecosystem condition and function. A change of vegetation cover may have long term impact on sustainable food production, freshwater and forest resources, the climate and human welfare. Documenting changes occurring in vegetation cover at periodic intervals is very important to providing information about the stability of vegetation.

The use of satellite-based remote sensor data has been widely applied to provide a cost-effective means to develop land cover coverages over large geographic regions. Vegetation cover is an evident part of land cover. Change detection has become a widespread application of remotely sensed data because of repetitive wide coverage, short revisit interval*s* and good image quality. Change detection is the process of identifying differences in the state of an object or phenomenon by observing it at different times. The main prerequisite in using remote sensing data for vegetation change detection is that changes in land cover result in changes in radiance values and changes in radiance due to land cover change are large with respect to radiance change caused by others factors such as differences in atmospheric conditions, differences in soil moisture and differences in sun angles [1].

Vegetation indices calculated from satellite images can be used for monitoring temporal changes associated with vegetation. Vegetation Indices (VIs) are combinations of Digital Numbers (DNs) or surface reflectance at two or more wavelengths designed to take out a particular property of vegetation. Each of the VIs is designed to emphasize a particular vegetation property. Analyzing vegetation using remotely sensed data requires knowledge of the structure and function of vegetation and its reflectance properties. This knowledge enables the linking of vegetative structures and their condition to their reflectance behavior in an ecological system of interest [2]. The normalized difference vegetation index (NDVI) is developed for estimating vegetation cover from the reflective bands of satellite data. The NDVI is an indicator which quantifies the amount of green vegetation. Past studies have demonstrated the potential of using NDVI to study vegetation dynamics. The NDVI data layer is defined as:

$$NDVI = (NIR - R)/(NIR + R), \qquad (1)$$

where *NIR* represents the spectral reflectance in near infrared band and *R* represents red band. Greener and dense vegetation has low red light reflectance and high near infrared reflectance, and thus high NDVI values. The NDVI real values, by definition, would be between -1 and +1, where increasing positive values indicate increasing green vegetation, but low positive values and negative values indicate non-vegetated surface features such as water, barren land, rock, ice, snow, clouds or artificial materials [3]. The NDVI also has the ability to reduce external noise factors such as topographical effects and sun-angle variations.

Time series analysis of remotely sensed data, as shown earlier, has gained special attention supported

ISSN 1691-5402

DOI: http://dx.doi.org/10.17770/etr2015vol3.167

by availability of wide-coverage, high temporal satellite data. NDVI time series data has been employed to predict a NDVI variable beyond the time span. Univariate autoregressive integrated moving average (ARIMA) models are widely used for a univariate time series forecasting, also for the NDVI time series [10]. However, these models are parametric and are based on the assumption that the time series been forecasted are linear and stationary and have a limited ability to capture nonlinearity and nonstationary in time series. The difficulty of forecasting arises from the inherent non-linearity and non-stationarity in the NDVI index. Many previous studies propose that non-linear machine learning approaches such as neural network models perform better than traditional time series linear models with minimum initial assumptions and high forecasting accuracy. Therefore, neural networks are used as an alternative to traditional statistical forecasting methods.

II MATERIALS

A. Study Area

Ventspils Municipality is located in the western part of Courland, Latvia with total area of 2472 km2 (Fig. 1).

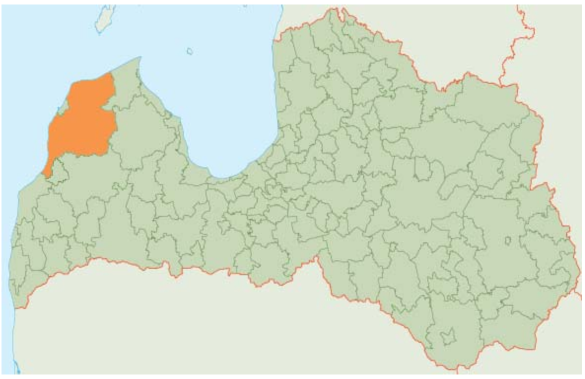

Fig. 1. Location of the Ventspils Municipality.

The climate in the study site is determined by temperate climate zone with significant maritime features. Approximately half of the area of Ventspils Municipality is covered by forests. Latvia lies on the border between two different forest types: the northern coniferous zone and the broad-leaved trees of the temperate zone, so the tree species, characteristic for the both forest types, can be found in the landscape.

B. NDVI Data Set

This study explored the use of multi-temporal, smoothed MODIS Terra and Aqua NDVI composite data with spatial resolution 250 m and produced on 16 day intervals (Fig. 2).

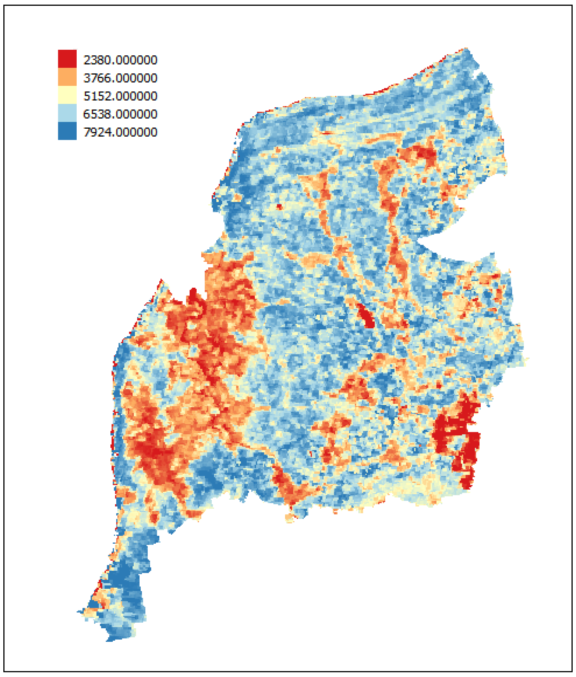

Fig. 2. MODIS Terra NDVI satellite image.

The NDVI data set consists of 624 smoothed NDVI images that obtained every 16 days over 14 years. Mean values of these images are used as NDVI time series observations (Fig. 3).

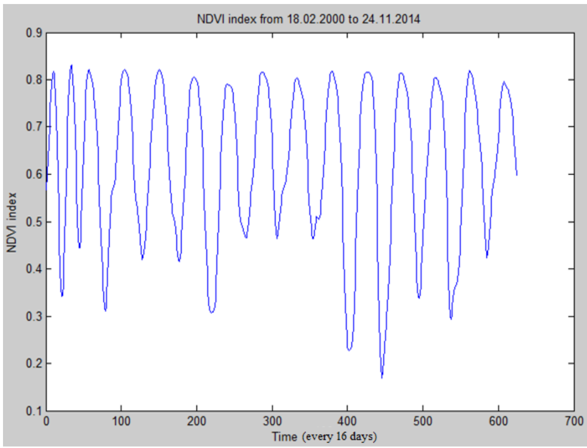

Fig. 3. Smoothed NDVI time series from 18.02.2000 till 24.11.2014.

The NDVI time series data provide a seasonal trajectory – time series show pronounced seasonal oscillations, which correspond to the vegetation phenological cycles where maximum NDVI values are observed between May and August. Variations in the NDVI values are seen to be 0.2 to 0.9 units. NDVI trends are not always monotonic but can change. A positive trend can change for example into a negative one and reversely.

III THEORETICAL BACKGROUND

A. Artificial Neural Networks

Artificial neural networks (ANNs) are a form of artificial intelligence, which attempts to mimic the function of real neurons found in the human brain [4]. ANNs are one of the most accurate and widely used forecasting models that are used in forecasting social, economic, business, engineering, foreign exchange, stock problems and others. Structure of artificial neural networks makes them valuable for a forecasting task with good accuracy.

As opposed to the traditional model-based empirical and statistical methods such as regression and Box-Jenkins approaches, which need prior knowledge about the nature of the relationships between the data, artificial neural networks are self-adaptive methods that learn from data, and only few a priori assumptions about data are needed [5].

Neural networks learn from examples and can find functional relationships among the data even if relationships are unknown or the physical meaning is the baffling [4]. Therefore, ANNs are well suited for problems, whose solutions require knowledge that is difficult to specify but for which there are enough data or observations.

Artificial neural networks can generalize. After learning from the input data (a sample or pattern), ANNs can often correctly process the early unseen sample even if the sample data are noisy. Neural networks are less sensitive to error term assumptions and they can tolerate noise, chaotic components better than most other methods. Artificial neural networks are also universal function approximators. It was proved that a neural network can approximate any continuous function with any accuracy. The ANN performs the following unknown function mapping:

$$y_{t+1} = f\left(y_t, y_{t-1}, \ldots, y_{t-p}\right), \quad (2)$$

where y_t is the observation at time t [5].

An individual neuron receives weighted inputs from previous layers, which are summed in each node using a combination function. The result of this combined summation is passed through a transfer function to produce the nodal output of the processing element. The combination function and transfer function together constitute the activation function. The most widely used activation function for the output layer is the linear function as a non-linear activation function may introduce distortion to the predicated output. The sigmoid (logistic), exponential (hyperbolic) tangent, quadratic or linear functions are often used as the hidden layer transfer function. The neural network (2) can approximate any continuous function when the number of hidden nodes q is sufficiently large [6]. In practice, if a network structure has a small number of hidden nodes, then it works well in "out-of-sample" forecasting on data that were not used in training. To improve the accuracy of the neural network, each data point in the input neurons needs to be normalized – rescaled within the range of [−1, 1] or [0, 1] and standardized to scale data and transformed to make the time series stationary. Transformation can be implemented as taking logarithmic returns of the time series, differencing the time series, etc. The multilayer perceptron is trained with error-correction learning, which means that the desired response for the system must be known. The error correction learning works in the following way: from the system response $d_i(n)$ at node i at iteration n, and the desired response $y_i(n)$ for a given input pattern, an instantaneous error $e_i(n)$ is defined by:

$$e_i(n) = d_i(n) - y_i(n). \quad (3)$$

Using the theory of gradient-descent learning and the standard Back Propagation (BP) algorithm, each weight in the network can be adapted by correcting the present value of the weight with a term that is proportional to the present input and error at the weight, i.e. the weight from node j to node i (w_{ij}) can be calculated by:

$$w_{ij}(n+1) = w_{ij}(n) + \eta \delta_i(n) x_j(n), \quad (4)$$

where, x_j is a transform function at node j, i and j indicate different layers. The local error $\delta_i(n)$ can be directly computed from $e_i(n)$ at the output node or can be computed as a weighted sum of errors at the internal nodes. The constant η is called the step size. There are some major disadvantages of this gradient descent approach, one of them is stuck into local minima, which can be mostly avoided by using a learning rate but that sometime may cause serious problem of overshooting, there also another problem of very slow convergence of the learning algorithm which severely depends upon choosing right value for learning rate [8]. Regarding to this issues, there are some more methods available to use in aid of standard back propagation learning, such as Broyden-Fletcher-Goldfarb-Shanno method or BFGS. BFGS method is a classical quasi-Newton second-derivative line search method, and also is one of the most effective and fastest algorithms of the unconstrained optimization problems at present. In quasi–Newton methods, the idea is to use matrices which approximate the Hessian matrix and/or its inverse, instead of exact computing of the Hessian matrix (as in Newton-type methods). The direction of search is based on an $n \times n$ direction matrix S which serves the same purpose as the inverse Hessian H^{-1} in the Newton method. This matrix is generated from available data and is contrived to be an approximation of H^{-1}. Furthermore, as the number of iterations is increased, S becomes progressively a more accurate representation of H^{-1}, and for convex quadratic objective functions it becomes identical to in $n + 1$ iterations. The BFGS method exposes superlinear convergence [9].

B. Elman Recurrent Neural Network

Although feed-forward neural networks are used in many forecasting applications, another type of neural networks – Elman recurrent neural network (ERNN) – is also used in forecasting applications with good accuracy. According to the general principle of the recurrent networks, there is a feedback connection from the outputs of some neurons in the hidden layer to neurons in the context layer that stores the delayed hidden layer outputs (Fig. 4).

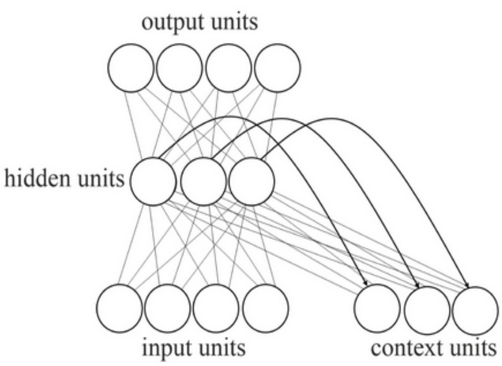

Fig. 4. Elman recurrent neural network architecture.

The most important advantage of ERNN is a robust feature extraction ability, when the context layer stores useful information about data points in past. Since ERNN contains the context layer, it is possible to improve forecasting accuracy by using ERNN instead of feedforward neural network (FNN) [7].

IV EXPERIMENTAL PROCEDURE

The aim of this experiment is to investigate the capability and accuracy of recurrent neural networks in the NDVI time series forecasting. This study implements an Elman recurrent neural network model to predict the NDVI index. In the first stage mean values of every NDVI image were calculated and the NDVI time series was created. The second stage employed ERNN for forecasting.

The data set was divided into two sets, training and testing data set by 80/20 principle, namely, 80% of the NDVI data (from February 18, 2000 to March 12, 2012, a total of 500 observations) were used as a training data set and others of the NDVI data (from March 5, 2012 to November 24, 2014, a total of 124 observations) were used as a testing data set.

There are no fixed rules for the selection of input variables for developing ANN model. ANN model used in this study is the four-layer Elman recurrent network. To achieve optimal weights of ERNN, BFGS quasi-Newton backpropogation algorithm provided by the MATLAB neural network toolbox is used to train the network. In this study, the hyperbolic tangent function and a linear function are used as activation functions for the hidden and output layers, respectively. The number of epochs that are used to train is set to 1000. The training of ERNN was stoped when the error achieves 10^{-6} or when the number of epochs reached 1000. As the number of hidden neurons is concerned, there is currently no theory to determine how many nodes in the hidden layer are optimal. In the present study, the number of hidden nodes was progressively increased from 2 to 20. A program code was written in MATLAB language for the development of the ERNN model.The optimal complexity of ERNN model, that is, the number of hidden nodes, was determined by a trial-and-error approach.

With respect to measure prediction performance of the proposed ERNN model, we introduce loss estimators, namely; RMSE, MAPE and DS. The square root of the mean of the square of all of the errors (RMSE) is given by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2}{N}}, \tag{5}$$

where, $\hat{y}_i$ – forecasted value, y_i – observed value, N – number of observations. The MAPE (Mean Absolute Percent Error) measures the size of the error in percentage terms and is given by:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|100. \tag{6}$$

Directional symmetry is a statistical measure of a model's performance in predicting the direction of change, positive or negative, of a time series from one time period to the next and is given by:

$$DS = \frac{100}{N-1}\sum_{i=2}^{N} d_i, \tag{7}$$

where,

$$d_i = \begin{cases} 1, if\,(y_i - y_{i-1})(\hat{y}_i - \hat{y}_{i-1}) \geq 0 \\ 0, else \end{cases}. \tag{8}$$

These loss estimators are used in chapter V to determine proposed ERNN model accuracy.

V RESULTS

In the several experiments were found, that optimal number of hidden nodes is six. Optimal ERNN topology is shown in Fig. 5.

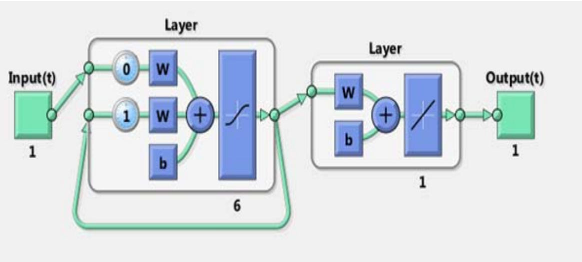

Fig. 5. Optimal ERNN topology.

ERNN network using two neurons was not converging and stopped after 1000 epochs (Fig. 6).

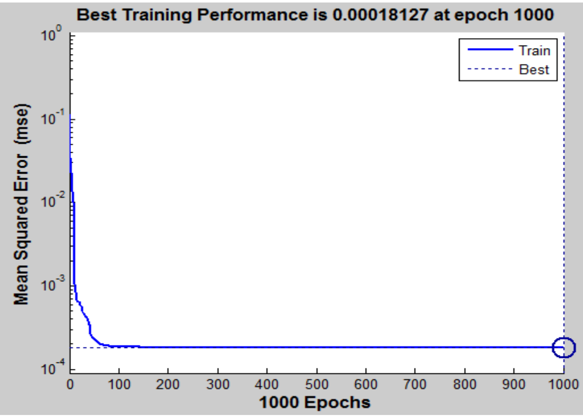

Fig. 6. ERNN convergence using 2 neurons in hidden layer.

ERNN network using three neurons was converging after 433 epochs (Fig. 7). However, this model doesn't give the best result for test data. There is an overfitting effect that can be found in the neural network modeling process. An overfitted model has a good accuracy on training data, but poor accuracy on "out of the sample" data.

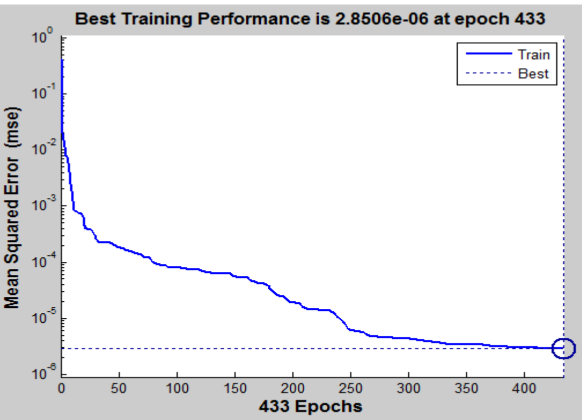

Fig. 7. ERNN convergence using three neurons in hidden layer.

ERNN convergence for best model on test data using six neurons in hidden layer was obtained after 613 epochs (Fig. 8).

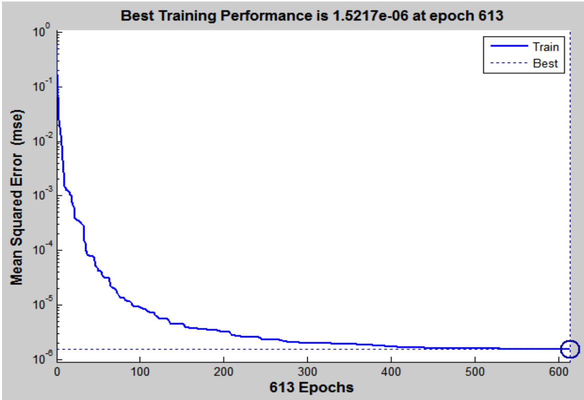

Fig. 8. ERNN convergence using six neurons in hidden layer.

For optimal ERNN architecture, the RMSE value is 0.0374, the MAPE value is 4.9959% and the DS statistic is 90.5882% for training data (table 1).

TABLE 1.

LOSS ESTIMATORS ON TRAINING DATA SET

RMSE	MAPE	DS
0.0374	4.9959%	90.5882%

Actual and predicted values of NDVI time series training data is shown in Fig. 9.

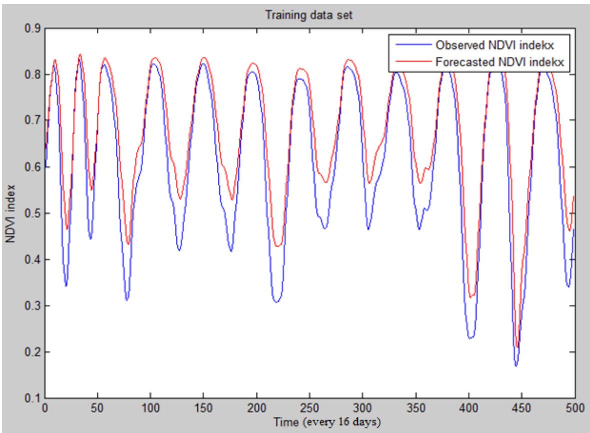

Fig. 9. Actual and predicted values of NDVI time series training data.

The RMSE value is 0.0352, the MAPE value is 1.2883% and the DS statistic is 86.2385% for testing data (table 2).

TABLE 2.

LOSS ESTIMATORS ON TESTING DATA SET.

RMSE	MAPE	DS
0.0352	1.2883%	86.2385%

Actual and predicted values of NDVI time series testing data is shown in Fig. 10.

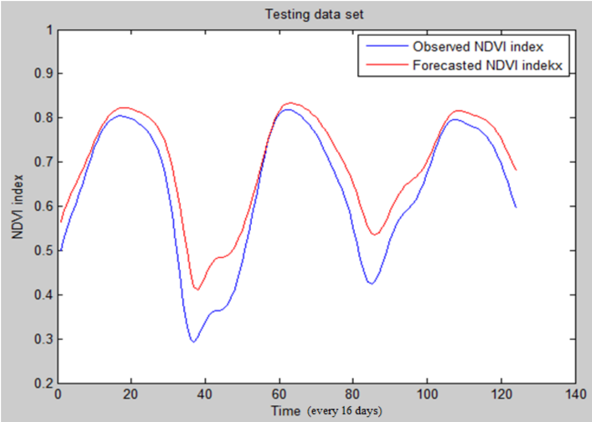

Fig. 10. Actual and predicted values of NDVI time series testing data.

VI SUMMARY AND CONCLUSIONS

In this paper one-step-ahead predictions of the Normalized Difference Vegetation Index (NDVI) data recorded by satellites over Ventspils Municipality in Courland, Latvia is obtained, using Elman recurrent neural network (ERNN). NDVI is an important variable for vegetation forecasting. Artificial Neural Networks (ANN) are computational models and universal approximators, with good generalization ability, that are widely used for nonlinear, non-stationary and dynamical process, such as NDVI time series, modeling and forecasting. Therefore ANN give better results than traditional statistical forecasting methods, such as autoregressive integrated moving average (ARIMA) parametric model, that are based on the assumption that the time series been forecasted are linear and stationary. Since ERNN contains the context layer, it is possible to improve forecasting accuracy by using ERNN instead of feed-forward neural networks. Using optimal ERNN architecture, the RMSE error is 0.0352, the MAPE error is 1.2883% and directional symmetry is 86.2385% on the test data. Therefore the study concludes that the forecasting abilities of ERNN provides a potentially very useful method for NDVI time series forecasting.

VII SUGGESTIONS FOR FUTURE RESEARCH

Some fruitful avenues for future studies are possible. First, in this study was used a univariate NDVI time series, future studies may consider multivariate NDVI time series. Second, in this study was used a simple recurrent neural network, ERNN. In contrast, future research may consider other recurrent neural network models, such as long short term memory neural network. Third, instead of one-step-ahead prediction multistep-ahead prediction can be implemented.

VIII REFERENCES

[1] M. M. Badamasi, S. A. Yelwa, M. A. AbdulRahim and S. S. Noma, "NDVI threshold classification and change detection of vegetation cover at the Falgore Game Reserve in Kano State, Nigeria," Sokoto Journal of the Social Sciences, vol. 2, no. 2, pp. 174-194.

[2] N. B. Duy and T. T. H. Giang, "Study on vegetation indices selection and changing detection thresholds selection in Land cover change detection assessment using change vector analysis," presented at International Environmental Modelling and Software Society (iEMSs), Sixth Biennial Meeting, Leipzig, Germany, 2012.

[3] E. Sahebjalal and K. Dashtekian, "Analysis of land use-land covers changes using normalized difference vegetation index (NDVI) differencing and classification methods," African Journal of Agricultural Research, vol. 8, no. 37, pp. 4614-4622, September 26, 2013.

[4] A. Shabri and R. Samsudin, "Daily crude oil price forecasting using hybridizing wavelet and artificial neural network model," Mathematical Problems in Engineering, vol. 2014, article ID 201402, July 2014.

[5] G. Zhang, B. E. Patuwo and M. Y. Hu, "Forecasting with artificial neural networks: the state of the art," International Journal of Forecasting, vol. 14, no. 1, pp. 35-62, March 1998

[6] M. Khashei and M. Bijari, "An artificial neural network (p, d,q) model for timeseries forecasting," Expert Systems with Applications, vol. 37, no. 1, pp. 479-489, January 2010.

[7] C. H. Aladag, E. Egrioglu and C. Kadilar, "Forecasting nonlinear time series with a hybrid methodology," Applied Mathematics Letters, vol. 22, no. 9, pp. 1467-1470, September 2009.

[8] S. M. Jadhav, S. L. Nalbalwar and A. A. Ghatol, "Artificial Neural Network Models based Cardiac Arrhythmia Disease Diagnosis from ECG Signal Data, "International Journal of Computer, vol. 44, no. 15, April 2012.

[9] A. Ghosh and M. Chakraborty, "Hybrid Optimized Back propagation Learning Algorithm For Multi-layer Perceptron, " International Journal of Computer Applications, vol. 60, nr. 13, December 2012.

[10] M. Manobavan, N. S. Lucas, D. S. Boyd and N. Petford, "Forecasting the interannual trends in terrestrial vegetation dynamics using time series modelling techniques, "Presented at the ForestSAT Symposium Heriot Watt University, Edinburgh, United Kingdom, 5th - 9th August 2002.