International Journal of Engineering & Technology, 8 (1.7) (2019) 253-267

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Review on Robust Artificial Neural Network Forecasting Models towards Outliers Problem

Nor Azura Md Ghani[1]*, Saadi Bin Ahmad Kamaruddin[1], Hishamuddin Hashim[2]

[1] *Center for Statistical Studies and Decision Sciences, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA,40450 Shah Alam, Selangor, Malaysia*

[2] *Pilihsepakat Sdn. Bhd., Lot 5431, Tingkat 1(B), Jalan J9, Fasa 6, Taman Melawati, Kuala Lumpur, Malaysia*

**Corresponding author E-mail: azura@tmsk.uitm.edu.my*

Abstract

Neurocomputing have been adjusted in time arrangement estimating field, however the nearness of exceptions that for the most part happen in information time arrangement might be hurtful to the information organize preparing. This is on the grounds that the capacity to consequently discover any examples without earlier suppositions and loss of all inclusive statement. In principle, the most well-known preparing calculation for backpropagation calculations inclines toward lessening ordinary least squares estimator (OLS) or all the more particularly, the mean squared error (MSE). In any case, this calculation is not completely hearty when exceptions exist in preparing information, and it will prompt false estimate future esteem. In this paper, the effect of time series outliers in backpropagation training is discussed. The comparisons and related issues of autoregressive moving average (ARIMA) to artificial neural network (ANN) are also discussed briefly in this paper. Moreover, the background of the basic backpropagation neural network (BPNN) time series models; nonlinear autoregressive (NAR) and nonlinear autoregressive moving average (NARMA), is also discussed in this paper. The critical part of the paper is the application of metaheuristics algorithms, mainly the Firefly Algorithm (FFA) to improve the backpropagation models. There are also highlights of latest research works on the robustification of backpropagation using modern optimization algorithms.

Keywords*: Outliers, Backpropagation, robust estimators, evolutionary algorithms, forecasting*

1. The Effect of Time Series Outliers in Backpropagation Training

The occurrence of outliers in time series data can range from 1% to more than 10% in usual routine data [1-2]. Outliers are data that severely deviate from the pattern set of the majority of data [3]. This refers to the sample values that are dramatically different from patterns in the remaining data, and they may be anomalous from the majority of samples, resulting from measurement error, or illustrative of significant features in the data. Previous studies [4-5] found the existence of these outliers can be a potential competitors to the standard or conventional least squares analysis.

In time series forecasting, an artificial neural network (ANN) has been applied to solve problems such as stock, electricity prices, breast cancer, rainfall-runoff [6-10]. ANN performs better than the statistical method because of its effective handling of linear and nonlinear time series data with or without noise [11]. It does not require prior information about systems of interest [12]. [13] and [14] claimed that forecasting is a major application area of ANN.

Even though ANNs have successfully captured the interest of many practitioners in many fields because of their universal ability as a function approximator, the backpropagation learning algorithm which is based on the minimisation of the mean square error (MSE) cost function is not robust in the presence of outliers which may cause error in the data training process [4]. MSE is an error measure between the actual and desired output in the popular backpropagation learning algorithm of multilayered feedforward neural networks (MFNNs). [4] agree that this popular algorithm is not completely robust when outliers are present and the neural network fit can be destroyed by a single outlier [15-16], possibly leading to wrong prediction values.

The weights of the neural network are responsible for the learning capabilities of the neural network. They need to be updated via training in order to store the experiential knowledge of the neural network. Multiple outliers will cause even greater damage to the network learning process [17]. This will soon cause false network training and inaccurate predictions of cost indices.

2. ARIMA versus ANN in Forecasting

Several studies compared between ARIMA models and ANN in forecasting using time series data. Most comparisons were data based and many used economic data. [18]discussed the possibility of using ANN in economic variables and the usability of traditional time series models and emphasised the similarities between the two methods. [19]connected the techniques on a gathering of 14 distinctive time arrangement in macroeconomics and found that the ANN strategy performed superior to different strategies in their estimating capacities. [20]looked at ANN and ARIMA in anticipating of Egypt's grain future and found that standard error of neural system estimates

are not exactly those of ARIMA models. [21]used ANN to forecast macroeconomic factors. They compared about various straight and nonlinear models by utilizing a huge example measure information and found that the execution of multivariate direct models is insignificantly superior to other univariate models. [22] found that ANN outperformed ARIMA forecasting models in six different time series originally published in [23] and more recently by [24]. The developments in the ARIMA model improved forecasting performance over standard ordinary least squares estimation by 8% to13%. In contrast, ANN achieves dramatic improvements of 10% to 40%. [25]inspected whether simulated ANN can be utilized as a part of displaying the expansion underway in view of money related and budgetary factors. The outcomes showed that ANN anticipating exhibitions were superior to those of the direct models. [26] looked at the determining capacities of both time arrangement models (ARIMA and exponential smoothing) and straight models on the one hand and ANN models on the other using the Canadian GDP data and monetary and financial variables.
[27] used the hybrid approach in a mix of ARIMA and ANN models and got empowering comes about [28] determined the GDP of the Malaysian economy in view of financial pointers. The creator thought about ANN and econometric methodologies and demonstrated that ANN would do well to brings about GDP determining. [29] applied diverse techniques for anticipating spring inflow to the Amir Kabir supply in the Karaj stream watershed. ANN, ARIMA time arrangement and relapse investigation between some hydroclimatological information and inflow were utilized to gauge the spring inflow. The models' execution were contrasted and the ANN demonstrated with be a powerful instrument for store inflow anticipating in the Amir Kabir repository utilizing snowmelt equivalent data. [30]forecasted network traffic using ARIMA Model and ANN concluding that ARIMA models are easier to use for training and forecasting, but the prediction results showed that they are not very accurate. In contrast, the ANN models are more complex in training and simulation but they give much better results compared with ARIMA models.
Writing recommends a few numerical favorable circumstances of neural systems have over conventional factual strategies. Neural systems have been appeared to be all inclusive approximators of capacities [31-32] and their subordinates [33]. They can likewise be appeared to rough customary minimum squares relapse [34], nonparametric relapse [35] and Fourier investigation [36]. Subsequently, neural systems can estimated whatever useful frame best portrays a period arrangement inferring that standard asymptotic hypothesis can be suitably appropriate to the nonlinear utilitarian structure of ANN. Nonetheless, artificial neural networks are data-driven in that there is no need to make prior assumptions about the model under study as well as the noise terms [27],[37-59], strong representation of unknowns and complex nonlinearities [60-66] simple implementation [67], and flexible black-box design methodology [63],[68-71].

3. Issues Related to ANN Forecasting

Creating ANN display for a specific anticipating application is not a trifling assignment. Albeit numerous great programming bundles exist to facilitate clients' exertion in building an ANN demonstrate, it is as yet basic for forecasters to comprehend numerous imperative issues encompassing the model building process. It is imperative to bring up that building a fruitful neural system is a blend of workmanship and science and programming alone is not adequate to tackle all issues all the while. It is a trap to indiscriminately toss information into a product bundle and after that expectation it will consequently give an attractive arrangement.

3.1. Generalizability of the ANN Prediction Model

A critical point in proficiently utilizing ANN determining is the comprehension of the issues of learning and speculation intrinsic in all ANN forecasting applications. This issues of learning and speculation can be comprehended with the ideas of model predisposition and difference [72], where inclination and change are critical measurable properties related with any observational model [27]. Display inclination measures the precise mistake of an estimating model in taking in the fundamental euphorias among factors or time arrangement perceptions, demonstrate difference identifies with the steadiness of models based on various information tests from a similar procedure and subsequently offers experiences on generalizability of the forecast show [14].

A pre-indicated or parametric model, which is less subject to the information, may distort the genuine useful relationship and, subsequently, cause an extensive inclination. Then again, an adaptable, information driven model might be excessively subject to the particular dataset and, thus, have an expansive fluctuation. Predisposition and difference are two clashing terms that effect a model's handiness. In spite of the fact that it is alluring to have both low predisposition and low difference, this examination will most likely be unable to lessen both terms in the meantime for a given dataset in light of the fact that these objectives are clashing. A model that is less reliant on the information has a tendency to have low change yet high inclination if the pre-indicated model is mistaken. Then again, a model that fits the information well has a tendency to have low inclination yet high difference when connected to various datasets. Henceforth, a great prescient model ought to have a proper harmony between model predisposition and model change.

3.2. Overfitting Problem of ANN Model

As a without model way to deal with information examination, neural system tend to fit the preparation information well and accordingly have low inclination. In any case, the cost to pay is the potential overfitting impact that causes high change. Consequently, consideration ought to be paid to address issues of overfitting and the adjust of inclination and change in neural system display building.
The real choices a neural system forecaster must take incorporate information readiness, input variable determination, decision of system sort and engineering, exchange capacity, and preparing calculation, and also display approval, assessment and choice. Some of these can be unraveled amid the model building process while others must be considered before genuine demonstrating begins.

3.3. Data Splitting in ANN Modeling

The second issue is data splitting. Commonly for neural system applications, every accessible data are partitioned into an in-test and out-of-test. The in-test information are utilized for model fitting and determination, while the out-of-test is utilized to assess the prescient capacity of the model. The in-test information once in a while are additionally part into a preparation test and an approval test. As a result of the predisposition and fluctuation issue, it is basic to test an ANN display with a free out-of-test which is not utilized as a part of the neural system preparing and show determination stage. This division of information implies that the genuine size of test utilized as a part of model building is littler than the underlying example estimate. In spite of the fact that there is no agreement on the best way to part the information, the general practice is to designate more information for model building and determination. That is, most reviews in the writing use helpful proportion of part for all through examples, for example, 70%: 30%, 80%: 20%, or 90%: 10%. It is essential to note that in information part, the issue is not about what extent of information ought to be designated in each example. In any case, rather, it is about adequate information indicates in each example guarantee sufficient learning, approval and testing. [73] proposes that for nonlinear demonstrating no less than 20% of the information ought to be kept down for an out-of-test assessment. [74] prescribes that no less than 10 information focuses ought to be in the test while [75] proposes that a considerably bigger out-of-test size is essential keeping in mind the end goal to accomplish factually noteworthy change for anticipating issues.

3.4. Data Preprocessing in ANN Modeling

Information preprocessing is another issue that is frequently prescribe highlighting imperative connections or to make more garbs information to encourage ANN learning, meeting calculation prerequisites, and stay away from calculation issues [14]. [76]compresses the strategies ordinarily utilized for info information standardization (designing term) or information institutionalization (measurable term), which are along channel standardization, crosswise over channel standardization, blended channel standardization, and outside standardization. However, the need and impact of information standardization on system learning and estimating are as yet not generally settled upon. For example, in demonstrating and determining occasional time arrangement, a few analysts [77-78] trust that information preprocessing is a bit much in light of the fact that the ANN is a widespread approximator and can catch the greater part of the fundamental examples well. Late observational reviews [22], from empirical studies, it is found that pre-depersonalization of the information is basic in enhancing anticipating execution. [79] additionally exhibit that for time arrangement containing both pattern and occasional varieties, preprocessing the information by both detrending and deseasonalisation ought to be the most suitable approach to assemble neural systems for best forecasting execution. From this exploration, [27] and [14] stress that ANN needs no earlier presumptions concerning the model and the error terms since it an all inclusive approximator which can take in any information design well, in any case the linearity and nonlinearity of the information.

3.5. Neural Network Design and Architecture

Neural system plan and design choice are critical yet troublesome errand [27]. Not exclusively are there numerous approaches to fabricate an ANN display and an expansive number of decisions to be made amid the model building and choice process, additionally various parameters and issues must be assessed and explored different avenues regarding before an acceptable model may develop. Adding to the trouble is the absence of guidelines all the while. Various dependable guidelines are accessible, however not every one of them can be connected indiscriminately to another circumstance. In building a proper model for the forecasting job that needs to be done, a few analyses are typically essential. Along these lines, a great examination configuration is required. For exchanges of numerous parts of displaying issues, perusers may counsel [80-82],[27],[14].

As expressed before, many sorts of ANN have been utilized for estimating. From this exploration, the multilayer feedforward engineering is by a long shot best created and most generally connected one for forecasting applications [27],[14]. From writing, numerous analysts utilize two-layer neural system, which comprise of info layer, shrouded layer and yield layer, primarily in light of the fact that this arrangement is equipped for evaluating any capacity [84-87]. Subsequently, the dialog will be centered around this kind of neural system, in spite of the fact that it might be connected to different sorts of ANN.

A feedforward ANN is portrayed by its design and dictated by the quantity of layers, the quantity of hubs in each layer, the exchange or enactment work utilized as a part of each layer, and also how the hubs in each layer are associated with hubs in contiguous layers. Albeit incomplete associations between hubs in neighboring layers and direct associations from info layer to yield layer are conceivable, the most ordinarily utilized ANN is the alleged 'completely associated' organize in that every hub at one layer is completely associated just to the greater part of the hubs in the adjoining layers.

3.6. Output layer of ANN

The size of the output layer is typically dictated by the way of the issue. For instance, in most anticipating issues, one yield hub is actually utilized for one-stage ahead determining, albeit one yield hub can likewise be utilized for multi-venture ahead forecasting, in which case iterative estimating mode must be utilized. That is, estimates for more than two stages ahead in the time skyline must be founded on before figures.

This may not be effective for multi-step anticipating as pointed out by [53], which is in accordance with [88] who examines the potential advantages of utilizing diverse determining models for various lead times. In this way, for multi-step forecasting, one may either utilize numerous yield hubs or build up different neural systems each for one specific stride determining.

3.7. Input Nodes of ANN

The quantity of input nodes is maybe the most vital parameter for outlining a successful neural system forecaster. For causal determining issues, it relates to the quantity of autonomous or indicator factors that forecasters accept are imperative in foreseeing the needy variable. For univariate time arrangement anticipating issues, comparable as the concentration of this exploration, it is the quantity of past slacked perceptions. Deciding a fitting arrangement of info factors is indispensable for neural systems to catch the basic fundamental relationship that can be utilized for effective determining. The number and sort of factors to be utilized as a part of the info layer will straightforwardly influence the execution of neural system in in-sample fitting and out-of-sample forecasting, resulting in the under-learning or overfitting phenomenon. Exact outcomes [89-91] likewise recommend that the information layer is more essential than the shrouded layer in time arrangement estimating issues. Subsequently, significant consideration ought to be given to decide the information factors, particularly for time arrangement forecasting.

3.8. Hidden Layers of ANN

Despite the fact that there is generous adaptability in picking the quantity of concealed layers and the quantity of shrouded hubs in each layer, most forecasting applications utilize just a single concealed layer and few shrouded hubs. By and by, the quantity of shrouded hubs is frequently controlled by exploring different avenues regarding various decisions and afterward chose by the cross-approval approach or execution on the approval set [92]. In spite of the fact that the quantity of shrouded hubs is an essential component, various reviews have found that determining execution of neural system is not exceptionally delicate to this parameter [93-94],[90].
For estimating applications, the most well known exchange work for concealed hubs is either calculated or hyperbolic and it is the direct or character work for yield hubs, albeit numerous different decisions can be utilized. On the off chance that the information, particularly the yield information, have been standardized into the scope of [0, 1], then strategic capacity can be utilized for the yield layer. By and large, extraordinary decisions of exchange capacity ought not affect much on the execution of a neural system display.

3.9. Training Algorithm of ANN

Once a specific ANN engineering is important to the forecaster, it must be prepared so that the parameters of the system can be assessed from the information. To be compelling in playing out this errand, a great preparing calculation is required. Preparing a neural system can be dealt with as a nonlinear scientific streamlining issue and distinctive arrangement methodologies or calculations can have very extraordinary consequences for the preparation comes about. Accordingly, preparing with various calculations and rehashing with numerous arbitrary introductory weights can be useful in getting a greater answer for the system preparing issue. Notwithstanding the well known and usually utilized fundamental backpropagation preparing calculation, clients ought to know about numerous different calculations, which now and then might be more viable. These incorporate supposed second-arrange methodologies, for example, conjugate inclination plunge, semi Newton, and Levenberg-Marquardt [95]. From this research, the backpropagation scheme has been extensively used as the basic supervised training algorithm for training multilayered feedforward neural networks [96-104]. [105] emphasised that BPNNs are useful tools for pattern recognition problems and found efficiently in predicting time series of financial markets. He added that neural networks have great potential to predict variations and trends in construction cost indices.
The backpropagation (BP) algorithm is the basic algorithm to train the ANN by becoming backpropagation neural networks (BPNN) [95-104],[106-115]. The BP algorithm was introduced by Arthur Earl Bryson and Yu-Chi Ho in 1969 in their paper Applied Optimal Control: Optimization, Estimation and Control and is widely considered the precursor to modern optimization algorithms for the training of neural networks [116-117].

3.10. ANN Model Selection

ANN show determination is ordinarily finished with the fundamental cross-approval handle. That is, the in-test information is part into a preparation set, an approval set and a testing set. The ANN parameter are assessed with the preparation test, while the execution of the model is assessed with the testing test, bolstered or affirmed by the outcomes from the approval set. The best model chose is the one that has the best execution on the testing test. Obviously in picking contending models, this exploration should likewise apply the rule of miserliness, where a more straightforward model that has about an indistinguishable execution from a more perplexing model ought to be favored.

3.11. ANN is a Black-box Approach

In some statistical models, the parameters estimated may be accompanied with some scientific or technological interests. This is not always the case with the ARIMA time series forecasting but it is definitely relevant in other ARIMA modelling situations such as in Intervention Analysis and broadly speaking, the regression and dynamic regression with auto-correlated errors.
The parameters contained in the neural net approach do not carry any specific scientific meaning. As a matter of fact, widely different parameter settings in this approach may lead to models which make the same virtual predictions. The lack of a probability model makes the neural net seem comparable to a purely black-box approach. Yet, by not specifying such rigorous probability and model assumptions the neural net puts forth a more flexible approach to model building. This flexibility is particularly effective in nonlinear and non-Gaussian situations.
If the goal of the modelling is to develop a rigorous understanding, explication and quantification of uncertainty, statistical models would be the best choice. Sometimes the goal is merely to predict, and this is when the neural net approach can be more helpful.

3.12 ANN versus Traditional Statistical Modeling ANN is A Black-box Approach

The ANN has stable performance progress with surprisingly low overhead [118]. Compared to traditional forecasting approaches, it is limited by its inability to forecast all the features of the history and nature of the computing system. Varying scopes of research in implementing the ANN approach conducted by [119] suggest that it is not often consistent because statistical methods are normally related to the linear data, while the ANN is associated with nonlinear data. This is why the statistical method has had a successful record in time series forecasting for a few decades. From this research, they posit that ANN performs better owing to its superior features in capturing nonlinear time series data [120]. ANN provides effective and consistent modelling of the nonlinear time series data without noise [121]. [13]agreed if a comparison is drawn between the ANN and traditional forecasting approach, the former always performs better and becomes powerful under the condition that the forecasting horizon is increased.
[122] suggest that the ARIMA models are better suited for short-term forecasts and the ANN better in long terms forecasts. According to [123], ARMA is not only better suited for short-term forecasts, but needs data of the time series in question. This leads to its ability to forecast a high frequency of time series data.
There are several advantages of the ANN compared to traditional Box-Jenkins model, in spite of the fact that the ANN model has much more complicated time series characteristics. Some also see the ANN model as having a “bad site” because it requires large number

sample data and parameters. [124] stated that this problem is similar to ARIMA's weakness of needing high frequency historical series data. This accentuates the major weakness of the ANN and ARIMA compared to hybrid and fuzzy logic models. Besides, there are non-existing sets of rule for the sample size. This is due to the fact that the training of the network depends on the structure, thresholds, and training data in hand.

3.13. ANN as A Universal Approximator

ANNs have a number of strengths as a predictive modelling tool, chief among which is their ability to act as a universal approximator. Cybenko demonstrated in 1988 that a neural system with two layers can inexact any capacity to inside any **ε**> 0 [125]. This was trailed by Cybenko's Theorem, which expresses that a feedforward coordinate with a solitary shrouded layer utilizing a sigmoid enactment capacity can surmised any continuous function on a compact subset of n to any degree of accuracy >0 [31]. This outcome has since been stretched out for self-assertive, instead of entirely the standard sigmoid initiation capacities [126-128]. From writing, with regards to neural system estimating, forecasters utilize sigmoid enactment work, eg. Tansig in the main layer, and straight actuation work eg. Purelin in the second layer [129-131].

3.14. Arguments on ANN

From this research, the ANN can also present a large variety of problems as found by [132]. Training an ANN is not always a simple task. Through the process of gradient descent during backpropagation, it is possible for the network to converge on local minima of the error function that are not the desired global minimum. The use of simulated annealing, wherein the network is slightly perturbed as convergence slows, is one suggested remedy to this problem [133]. Another serious difficulty is the possibility of overtraining on the set of training patterns [134-135]. Researched practically would like to train the network to recognise the general patterns in the set, rather than overfitting every peculiarity of the data. The problem of overtraining can be serious when the training patterns contain a large amount of noise, irrelevant factors, mislabelled data, or other facets which do not correspond with the entire population. In such cases, though training may be successful on the data given, the ANN will often fail to generalise well, causing poor performance on data outside the training set. Their finding is supported by [136], where they posit that the over fitting problem is more likely to occur in neural network models compared to other statistical models as usually other statistical models involve large parameters for estimation. To recover problems such as over fitted behaviour of the ANN, it is better to expand the number of parameters time series data for estimation. [137] found fitting behaviour by the ANN do not constitute a problem, as it is beneficial in the non-stationary complicated financial time series investigation. Indeed, it would lead the way to success for the complex financial time series analysis.

[138] proved the hybrid model is more accurate and superior in forecasting wind velocities compared to ANN and ARIMA models. As more series of training vectors of the ANN model added, accuracy and performance will improve. The ANN was found to be more precise if the data to be forecast were of a short range.

This is evident through [139] finding of forecasting hourly wind speed by using the ANN approach by employing 10 minutes data using multi-step forecasting, where the average results were used for predictions. There are two independent datasets which require testing. Every ten-minute data produced better root mean square error (RMSE), and this is four times less than other models. Furthermore, the development of the Modular Artificial Neural Network (MANN) will advance the prediction accuracy if the appropriate technique is used [140].

ANN gives better forecast with a lower RMSE in forecasting strong seasonality in time series compared to traditional forecasting approaches. From this research, a weak and changed network structure will be better. [123] added that the ANN not only has the ability to predict nonlinear time series data, but also the residual among the linear output attained by the ARMA system.

Box-Jenkins and ANN can be considered successful in making predictions in linear and nonlinear time series data. From this research, it is difficult to decide the linearity of the time series data to implement either Box-Jenkins or the ANN method. There will be opportunities to increase the probability of forecasting error [141]. From this research, in the work of [142], artificial neural network technique is suitable for both linear and nonlinear data, especially when it comes to time series forecasting. This is also proven by the work of [143]. In another resort, the hybrid model is developed due to the limitation of the Box-Jenkins and ANN method. The hybrid model offers a good strategy that it can overcome these limitations. This is because hybrid models can simultaneously model linearly or non-linearly [144-146]. [132] agree that the hybrid model produces a forecast better compared to ARIMA and ANN methods. Nonetheless, the hybrid model is hard to employ because it involves a complex model.

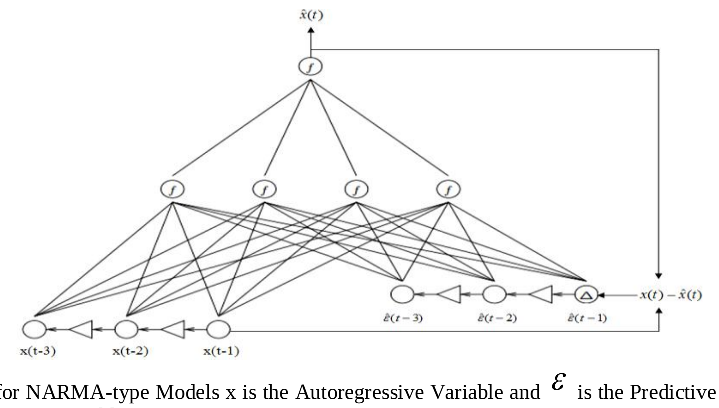

Fig. 3: Associated Predictor for NARMA-type Models x is the Autoregressive Variable and ε is the Predictive Error. $\hat{x}(t)$ Corresponds to the Prediction of the Autoregressive Variable.

In earlier studies of ANN forecasting, [147-148] directed an anticipating rivalry between ANN models and the Box-Jenkins technique utilizing 75 and 111 time arrangement information from the M-rivalry. They presumed that basic ANN models can figure and also the

Box-Jenkins strategy. Using both ANN and ARIMA models, [149] and [150] broke down a few business time arrangement and reevaluated 14 arrangement from 75 arrangement utilized as a part of [147] which are accounted for to have bigger errors. They inferred that ANNs beat the Box-Jenkins for time arrangement with short memory or with greater abnormality. [151] got comparative outcomes in a more deliberate review. [152] contrasted ANNs and ARIMA models in anticipating month to month live cows and wheat costs. Their outcomes demonstrate that ANNs conjecture are extensively and reliably more exact and can catch more defining moments than ARIMA models. [153] contrasted neural systems and six conventional factual strategies to estimate 111 M-rivalry time arrangement. Their discoveries show that the neural system models are essentially superior to conventional measurable and human judgment strategies when determining month to month and quarterly information. For the yearly information, neural systems and customary strategies are practically identical. They likewise infer that neural systems are exceptionally compelling for irregular time arrangement. [22]talk about the issue of whether ANNs can learn occasional examples in a period arrangement. [148] reason that the regularity of the time arrangement does not influence the execution of ANNs and ANNs are capable certainly to consolidate regularity. A few exact reviews observe that ANNs appear to be better in forecasting month to month and quarterly time arrangement [151],[154-155] than in estimating yearly information. This might be because of the way that month to month and quarterly information for the most part contain more inconsistencies (regularity, cyclicity, nonlinearity, clamor) than the yearly information, and ANNs are great at distinguishing the basic example covered by boisterous figures an intricate framework. [149] and [150] attempted to decide the conditions in which ANN forecasters can perform superior to anything the conventional time arrangement forecasting strategies, for example, Box-Jenkins models. The main review depends on three and the second on 16 time arrangement. Their discoveries are that (1) ANNs perform better as the gauge skyline builds, which is likewise affirmed by different reviews [151],[156],[154]; (2) ANNs perform better for short memory arrangement (see additionally [148]); and (3) ANNs give better anticipating outcomes with more information hubs. In forecasting swapping scale, [157] found that ANNs beat the direct models when week by week information are utilized and if month to month information are utilized, ANNs and straight techniques yield comparative outcomes.
There is no general guideline for the ideal specimen measure for which one can anticipate that neural nets will enhance detectably over direct models. [25] from Bank of Canada suggests that for neural network forecasting, the adequate sample size should be around 300 observations.

3.15. Time lags in ANN Forecasting

In any case, in neural system univariate time arrangement demonstrating, lags are utilized of the autonomous variable as sources of info. For quarterly and different sorts of arrangement, if the arrangement length is under 20, the satisfactory number of lags ought to associate with 1 to 4. It is proposed that the quantity of lags to be utilized as information sources ought to be founded on the time determination [158], which are 1 to 4 lags for yearly information, 1 to 6 lags for quarterly information, 1 to 15 lags for month to month information, and 1 to 6 lags for others. This govern is not tightfisted as far as the quantity of lags, however exact reviews by [158] propose than an ANN may perform preferable with more lags over would be regular for an ARIMA display. Consequently, in this examination, various time lags were executed as proposed by [159] uncommonly for ANN forecasting approaches. Conjectures are created iteratively by performing progressive one-stage ahead gauges utilizing past figures as assessments of observables.

3.16. Limitations of ANN

While the ANN is known to have many advantages, several limitations of the ANN have also been discussed. These include excessive amounts of parameters [160-162], difficulty to determine optimal network parameters (Zong-yi et al., 2010; Safak et al., 2010), computationally-expensive and time-consuming training process [65],[162],[59],[164-165],lack of model transparency [160],[162], sensitivity to initial weight values [65], over-generalisation or under-generalisation [61],[164], and its tendency to be trapped in local minima [65],[164].

4. NAR and NARMA Models

Models for time series data can have many structures and speak to various stochastic procedures. When demonstrating varieties in the level of a procedure, three wide classes of down to earth significance are the autoregressive (AR) models, the integrated (I) models, and the moving average (MA) models.
When AR and ARMA models are estimated using the ANN technique, they simply become nonlinear AR (NAR) and nonlinear ARMA (NARMA) due to nonlinear transfer function that are usually used in neurocomputing. Many endeavoured to adapt these models such as [166-174]. The NARMA show gives a capable portrayal to time arrangement examination, displaying, and expectation because of its quality to oblige the dynamic, complex and nonlinear nature of real time series applications. Using NARMA model based identification for nonlinear complex systems showed outstanding results [175]. NARMA is universal function approximators that can map any nonlinear function [176] and is a powerful method for pattern recognition, classification, and forecasting. NARMA can endure commotion, riotous segments, and substantial tails superior to most different strategies. Different points of interest incorporate more noteworthy adaptation to internal failure, power, and flexibility contrasted with master frameworks because of the extensive number of interconnected preparing components that can be prepared to learn new examples [177].
NARMA model (Fig. 3) is the extension of NAR model (Fig. 2) [178].
The NARMA model, a general representation of a nonlinear dynamical system, adopts the form of a nonlinear difference equation [179]. The NARMA methodology offers a unified solution to

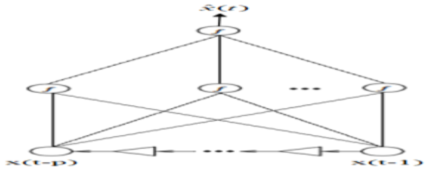

Fig. 2: Associated Predictor for NAR-type Models. $\hat{x}(t)$ Corresponds to the Prediction of the Autoregressive Variable.

the issue of finding this equation based only on experimental data recorded from one particular system. The process of identifying a NARMA model involves determining the form or structure of the unknown nonlinear equation, estimating the parameters linked with this particular structure and as the last step, checking or validating the resulting model to make sure that it details the real life system accurately.

One major purpose of the NARMA system identification is prediction. Prediction serves as a dynamic filtering, in which past values of one or more time series are used to foresee future values. Artificial neural networks include tapped delay lines work for nonlinear filtering and prediction.

From this research, in real time series data usually contain outliers. These outliers may contaminate the training data and soon cause inaccurate prediction of forecasted values. This case is worsened since the ANN learns from historical data, which might cause false model fitting. Therefore, there is a need to apply robust approach to the ANN technique to cater for outliers in the time series data automatically.

5. Robust Artificial Neural Network

Artificial neural networks is popularly applied in predictive analysis, especially the Feedforward Neural Networks (FFNs) which are frequently viewed as all inclusive devices and discover their applications in territories, for example, work estimate, design acknowledgment, or flag and picture handling. ANNs are picked in this exploration since they have demonstrated achievement in numerous zones of logical and specialized controls. One of the principle points of interest of utilizing FFNs is that they as a rule don't require, in the learning procedure, correct scientific information about information yield conditions and can learn and fabricate the information models in light of preparing illustrations. This is the reason they are mainstream and frequently thought to be anything but difficult to-utilize instruments. As it were, they might be viewed as sans model approximators [180].

They learn by limiting normal minimum square error capacity to fit preparing information as close as could reasonably be expected. Such a learning plan does not consider nature of the preparation information, so its execution depends unequivocally on the nature of preparing information. At the point when the information are tainted by the substantial commotion, or when exceptions and gross mistakes show up, the system manufactures a model that can be exceptionally mistaken. For the contaminated, noisy datasets with many outlying and erroneous patterns, the desired mapping from the input to output space cannot be properly achieved. In this case, neural networks trained on that data do not build the desired model, and instead try to fit to the noisy training examples.

In most genuine cases the suspicion that mistakes are typical and independent and identically distributed (iid), simply does not hold. The information got from nature are regularly influenced by commotion of obscure or anomalies. The amount of exceptions in routine information ranges from 1 to 10% [181]. Most of previous efforts improved only on the feedforward neuro-computing by adapting mostly the family of M-estimators.

For more details on disadvantages of M-estimators, refer to article by [182]. Below is a brief list of the disadvantages of M-estimators:

i. L_1(absolute value) estimators are not steady on the grounds that the ρ-function $|\varepsilon|$ is not entirely convex in ε . To be sure, the second derivative at $\varepsilon = 0$ is unbounded, and a vague solution may result.
ii. L_2estimators diminish the impact of large errors, however despite everything they have an impact in light of the fact that the impact work has no cut off point.
iii. $L_1 - L_2$estimators take both the upside of the L_1 estimators to lessen the impact of expansive large errors and that of L_2 estimators to be convex.
iv. The L_p (*least-powers*) function represents a group of functions. It is L_2 with $v = 2$ and $L_1$with $v = 1$. The smaller v, the smaller is the occurrence of expansive large errors in the estimate w. It appears that v must be genuinely direct to give a moderately powerful estimator or, at the end of the day, to give an estimator barely irritated by remote information. The determination of an ideal v has been explored, and for v around 1.2, a good estimate might be normal [183]. From this examination, numerous challenges are experienced in the calculation when parameter v is in the scope of intrigue $1 < v < 2$, in light of the fact that zero residuals are troublesome.
v. The function 'Fair' is among the potential outcomes offered by the Roepack package [183]. It has wherever characterized persistent subsidiaries of initial three requests, and yields a unique solution. The 95% asymptotic proficiency on the standard ordinary dissemination is acquired with the tuning constant $c = 1.3996$.
vi. Huber's function [184] is a parabola in the region of zero, and increments straightly at a given level $|x| > k$. The 95% asymptotic productivity on the standard typical dissemination is acquired with the tuning constant $k = 1.345$. This estimator is satisfactory to the point that it has been suggested for all circumstances; once in a while it has been observed to be second rate compared to some other ρ-work. From this examination, now and again, challenges are experienced, which might be because of the absence of solidness in the inclination estimations of the ρ-work due to its *discontinuous second derivative*:

$$\frac{d^2\rho(\varepsilon)}{d\varepsilon^2} = \begin{cases} 1 \; if \; |\varepsilon| \leq k, \\ 0 \; if \; |\varepsilon| \geq k. \end{cases} \tag{1}$$

vii. Cauchy's function, otherwise called the Lorentzian function, does not ensure an interesting arrangement. With a plummeting first subsidiary, such a capacity tends to yield mistaken arrangements in a way which can't be observed. The 95% asymptotic effectiveness on the standard typical dispersion is acquired with the tuning constant $c = 2.3849$.
viii. The other residual capacities have an indistinguishable issue from the Cauchy function. As can be seen from the impact work, the impact of substantial mistakes just declines directly with their size. The Geman-McClure and Welsh functions attempt to additionally lessen the impact of expansive mistakes, and the Tukey's biweight work even suppress the outliers. The 95% asymptotic effectiveness on the standard typical circulation of the Tukey's biweight capacity is acquired with the tuning constant $c = 4.6851$; that of the Welsh function, with $c = 2.9846$.

Among all M-estimators, Tukey-bisquare or biweight estimators performed most efficiently in many real life applications [185-187].

Currently the most prominent research in robustifying BPNN is by Dr. Andrzej Rusiecki from the Institute of Computer Engineering, Control and Robotic in Wroclaw University of Technology, Poland. In 2015, his colleague and him proposed trapezoid error function (TEF), Generalised Edited Nearest Neighbor (GenENN) and modified Global Anomaly Score (GAS) to enhance the performance of

backpropagation [188]. He found that in most regression datasets, the TEF methods performed especially good by obtaining RMSE on the test sets up to 40% lower than any other method.
Before that, [189] proposed the use of Last Mean Log Squares (LMLS), Median of Absolute Errors (MAE), Median Neuron Input (MIF), and Sum of Median (MedSum) performance measures. They found that hybrid MedSum performed the best. They emphasised that in most cases, modified backpropagation training methods outperformed traditional MSE approach. Rusiecki proposed the use of Least Trimmed Absolute Deviations (LTA) to enhance the performance of backpropagation training algorithm (Rusiecki, 2013). The novel LTA algorithm is compared with traditional approaches and other robust learning methods. Experimental results, presented in his work, demonstrate improved performance of the proposed training framework, especially for contaminated training datasets.
[190] introduced an alternative robust learning algorithm based on the iterated LMedS estimator. The novel approach is more powerful and fundamentally faster than the SA-LMedS strategy [4]. It likewise accomplishes better imperviousness to wrong preparing information. To make the preparation procedure more vigorous, change was made on the execution work as well as expel iteratively information suspected to be anomalies.
In addition, an inexact strategy to limit the LMedS error criterion was proposed [5]. For a notation, it is understandable that the function of LS estimator is to minimise either the mean or the sum of squared residuals, while LMedS estimator is to minimise the median of either the squared residuals or the absolute value of the residuals.
These works focused on the NAR model [191]. None considered using the robust approach to improve the NARMA model. The overall performance of the NARMA model is superior to the performance of the NAR model [192]. Here comes the novelty of the research, whereby the existing robust estimators will be implemented on BPNN of NARMA models. At the same time, none of the reviewed research examined the influence of hidden nodes increment towards the performance of the network. They did not consider the optimal lags as inputs to the network, which is crucial. Another novelty of the research is the extension of study towards the use of the firefly algorithm to minimise the LMedS error criterion, with adaptation in both NAR and NARMA models.

6. Stochastic Optimization Algorithm (Metaheuristics)

Over the last 20 years new metaheuristic algorithms have been introduced almost every year [193]. Metaheuristic algorithms frame a noteworthy block of the present day worldwide improvement calculations, computational insight and delicate processing. The nature-roused ones have turned out to be exceptionally intriguing and recognized. It is sensible that individuals started to investigate how nature takes care of the issues, since nature has developed for billions of years, and has discovered practically culminate answers for any issue it has experienced [194]. Developmental calculations are motivated by organic advancement and utilize hybrid and transformation to improve solutions or neural systems. In approximate methods such as metaheuristics [195-196], optimal solutions is guaranteed for the sake of getting good solutions in a notably reduction in the amount of time. Overlast three decades, the use of metaheuristics has received a lot of attention for its effectiveness especially in solving various noisy nonlinear mathematical optimization problems [197]. As mentioned by [197], metaheuristics are usually easier to be implemented as compared to classical gradient-based techniques, and thus do not require any gradient information [196].
Metaheuristics approaches function as global optimisers which can be defined as an undertaking of hunting down an ideal arrangement of answers for a parameter-subordinate issue, commonly by limiting a cost work identified with the current issue [198]. They are enlivened by the conduct of the nature with different connecting specialists. A subset of metaheuristics is regularly alluded to as swarm insight (SI) based calculations which have been created by mirroring the swarm knowledge attributes of natural living beings, for example, fowls, fish, people, and creepy crawlies [199]. The rise of those SI calculations are the molecule swarm enhancement which depended on the swarming conduct of winged creatures and fish [200], the firefly calculation that depended on the blazing example of tropical fireflies [195],[201], and cuckoo look calculation which was roused by the brood parasitism of some cuckoo species [202]. These calculations have demonstrated incredible execution in taking care of intense designing enhancement issues [203-206],[193-195]. From this research, among all these algorithms, the firefly algorithm is extremely effective in managing multimodal, global optimization issues [207],[197].
Optimization techniques are isolated into local and global methods [208]. Neighborhood enhancement systems look for the arrangement space in light of inclination data. This procedure profits by rapid convergence. From this exploration, it just works effectively for few parameters, and also being confined by its underlying conditions. The outcomes are particularly influenced by the nearness of discontinuities in the information [208]. GN and LM calculations are two cases of this kind of method.
Global optimization methods are better at managing multidimensional information and can take care of issues that are free of introductory qualities and under loose imperatives and presumptions [208]. They accomplish this through the stochastic optimization process, where the algorithms stress on the investigation of the arrangement space to scan for the best outcomes [209]. This kind of streamlining is prevalent in taking care of troublesome issues [210], in spite of the fact that it works slower [211]. A few calculations in this class are outstandingly the Bacteria Foraging Algorithm (BFA) [212], Artificial Bee Colony (ABC) [213-214], Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Ant Colony Optimization (ACO) [215],[209],[212],[214],[216], and latest firefly algorithm (FA or FFA) [195].
The algorithms typically perform optimization by controlling numerous fundamental self-sufficient individual units called specialists [209]. These specialists are spoken to in various structures: swarms in PSO, populaces in GA and settlements in ACO and ABC. These specialists perform fundamental undertakings in an individual way and interface with each other through some straightforward conventions [209]. Whenever joined, they exhibit a collaboration of self-coordination capable of solving optimization-type problems [209]. For global stochastic optimization, the firefly algorithm is the technique of choice because it is currently the most efficient technique for strong balance exploration and exploitation, and better convergence speed [194],[205],[217].

7. Firefly Algorithm (FFA)

[195] from Cambridge University developed the firefly algorithm. It performed better compared to PSO in handling high level of noise [197]. In this research, a new approach to make robust the backpropagation learning algorithm of nonlinear neural network time series models using a FFA-LMedS estimator is introduced. There are three idealised rules used by this algorithm [197]: (i) All the fireflies are unisex, which implies that one firefly is pulled in to different fireflies independent of their sex. (ii) Attractiveness and shine are commonly corresponding, so for any two glimmering fireflies, the less splendid one will normally move towards the brighter one. Engaging

quality and shine both diminishing as they get more distant from each other. (iii) No one brighter than different fireflies, their development will be irregular. (iv) The brightness of a firefly is affected by the perspective of the objective function.

With regards to a maximization problem, the brightness is relative to the estimation of the objective function. Different types of the splendor could be characterized likewise to the fitness function in genetic algorithms.

Average time processing or time performance of the algorithm to run the neural network system is also crucial in this research, and has been highlighted in previous works [190], [5]. Even in real world scenario, especially in business, industries and engineering fields, the time consumed for the data processing is critical and key to successful business processing [219]. That is why there exists the policy of business process reengineering among the business environment [219]. These time performance can also be considered an important performance measure [190] since it can be considered indicative, because the implementation was not optimised for speed. Since the firefly algorithm is the fastest algorithm in terms of time performance [193-195],[199],[201-202] compared to other the global optimization techniques, it is adapted in this research to minimise Least Median Squares (LMedS) for accurate and robust backpropagation algorithm computation. Many hybrid methods in which two or more methodologies are combined to enhance the final model have been introduced and attracted much attention by many researchers [220-227]. Some of these proposed hybrid approaches are very effective, robust, and show promise dealing with other hard combinatorial optimization problems.

The firefly algorithm has been successfully applied in many real life applications [228-230]. In spite of the fact that it has numerous likenesses with different calculations in light of the alleged swarm knowledge, for example, the renowned PSO, Artificial Bee Colony improvement (ABC), and Bacterial Foraging (BFA) calculations, it is considerably less difficult both in idea and execution [231],[201],[193-194]. Moreover, the algorithm is exceptionally productive and can beat other traditional calculations, for example, genetic algorithms, for taking care of numerous optimization issues [195],[231],[201],[193-194]. Its principle preference is the way that it utilizes fundamentally genuine arbitrary numbers, and depends on the worldwide correspondence among the swarming particles (the fireflies), and accordingly, is more powerful in multi objective optimizations.

[232] successfully find that hybrid artificial neural network models with firefly algorithm perform the best as compared to hybrid ANN models with particle swarm optimization algorithm when it comes to contaminated time series data with outliers problem. Some latest research works on the robustification of backpropagation using modern optimization algorithms can be seen in [233-237].

Acknowledgment

Special thanks also goes to Universiti Teknologi MARA (UiTM) for supporting this study under the Research University Grant No. 600-IRMI/MYRA 5/3/BESTARI (039/2017).

References

[1] Z. Zhang, "Parameter estimation techniques: a tutorial with application to conic fitting",Image and Vision Computing, vol. 15, no. 1, pp. 59-76, 1997.

[2] R. Pearson, "Outliers in process modeling and identification", IEEE Transactions on Control Systems Technology, vol. 10, no. 1, pp. 55-63, 2002.

[3] A.Rusiecki, “Robust learning algorithm based on LTA estimator,”Neurocomputing, vol. 120, pp. 624-632, 2013.

[4] M. El-Melegy, M. Essai, A. Ali, "Robust training of artificial feedforward neural networks" in Foundations of Computational Intelligence: Learning and Approximation (Studies in Computational Intelligence), USA, New York:Springer-Verlag, vol. 1, pp. 217-242, 2009.

[5] Rusiecki, M. Kordos, T. Kamiński, K. Greń, "Training neural networks on noisy data" in Artificial Intelligence and Soft Computing, Berlin, Germany:Springer-Verlag, vol. 8467, pp. 131-142, 2014.

[6] Abraham and B. Nath, "A neuro-fuzzy approach for modelling electricity demand in Victoria", Applied Soft Computing, vol. 1, no. 2, pp. 127-138, 2001.

[7] D. Delen, G. Walker and A. Kadam, "Predicting breast cancer survivability: a comparison of three data mining methods", Artificial Intelligence in Medicine, vol. 34, no. 2, pp. 113-127, 2005.

[8] R. Gareta, L. Romeo and A. Gil, "Forecasting of electricity prices with neural networks", Energy Conversion and Management, vol. 47, no. 13-14, pp. 1770-1778, 2006.

[9] S. Hamid and Z. Iqbal, "Using neural networks for forecasting volatility of S&P 500 Index futures prices", Journal of Business Research, vol. 57, no. 10, pp. 1116-1125, 2004.

[10] S. Srinivasulu and A. Jain, "A comparative analysis of training methods for artificial neural network rainfall–runoff models", Applied Soft Computing, vol. 6, no. 3, pp. 295-306, 2006.

[11] G. Zhang, M. Keil, A. Rai and J. Mann, “Predicting information technology project escalation: A neural network approach”,European Journal of Operational Research, vol. 146, pp. 115-129, 2003.

[12] L. Ma and K. Khorasani, "New training strategies for constructive neural networks with application to regression problems", Neural Networks, vol. 17, no. 4, pp. 589-609, 2004.

[13] H. Hippert, C. Pedreira and R. Souza, "Neural networks for short-term load forecasting: a review and evaluation",IEEE Transactions on Power Systems, vol. 16, no. 1, pp. 44-55, 2001.

[14] G. Zhang, "Business Forecasting with Artificial Neural Networks: An overview", in: G. P Zhang(Ed.), Neural Networks in Business Forecasting, Idea Group Publishing , Hershey, PA, 2004

[15] H. Allende, C. Moraga, and R. Salas, “Artificial neural networks in time Series forecasting: A comparative analysis”, Kybernetika, vol. 38, no. 6, pp. 685-707, 2002.

[16] H. Allende, C. Moraga and R. Salas, "Robust Estimator for the Learning Process in Neural Networks Applied in Time Series", Artificial Neural Networks — ICANN 2002, pp. 1080-1086, 2002.

[17] H. Jang, E. Topal and Y. Kawamura, "Unplanned dilution and ore loss prediction in longhole stoping mines via multiple regression and artificial neural network analyses", Journal of the Southern African Institute of Mining and Metallurgy, vol. 115, no. 5, pp. 449-456, 2015.

[18] A.Aliahmadi, M. Jafari-Eskandari, A. Mozafari, and H. Nozari, “Comparing Linear Regression and Artificial Neural Networks To Forecast Total Productivity Growth In Iran”, International Journal of Information, Business and Management, vol. 8, no. 1, pp. 93-113, 2016.

[19] E. Maasoumi, A. Khotanzed and A. Abaye, "Artificial neural networks for some macroeconomic series: A first report",Econometric Reviews, vol. 13, no. 1, pp. 105-122, 1994.

[20] N. Kohzadi, M. Boyd, I. Kaastra, B. Kermanshahi, andD. Scuse, “Neural networks for forecasting: an introduction”, Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie, vol. 43, no. 3, pp. 463-474, 1995.

[21] N. Swanson and H. White, "A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks", Review of Economics and Statistics, vol. 79, no. 4, pp. 540-550, 1997.

[22] J. Hansen, J. McDonald and R. Nelson, "Time Series Prediction With Genetic-Algorithm Designed Neural Networks: An Empirical Comparison With Modern Statistical Models", Computational Intelligence, vol. 15, no. 3, pp. 171-184, 1999.
[23] G. Box and D. Pierce, "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models", Journal of the American Statistical Association, vol. 65, no. 332, pp. 1509-1526, 1970.
[24] J. McDonald and Y. Xu, "Some forecasting applications of partially adaptive estimators of ARIMA models", Economics Letters, vol. 45, no. 2, pp. 155-160, 1994.
[25] G. Tkacz, S. Hu, "Forecasting GDP Growth Using Artificial Neural Networks", Working Paper 99-3 Department of Monetary and Financial Analysis Bank of Canada, 1999.
[26] G. Tkacz, "Neural network forecasting of Canadian GDP growth", International Journal of Forecasting, vol. 17, no. 1, pp. 57-69, 2001.
[27] G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", Neurocomputing, vol. 50, pp. 159-175, 2003.
[28] M. Z. Mat Junoh, "Predicting GDP Growth in Malaysia Using Knowledge-Based Economy Indicators: A Comparison Between Neural Network And Econometric Approaches", Sunway College Journal, vol. 1, pp. 39-50, 2004.
[29] K. Mohammadi, H. Eslami, and S. Dardashti, “Comparison of Regression, ARIMA & ANN Models for Reservoir Inflow Forecasting using Snowmelt Equivalent (a case study of Karaj)”,Journal of Agriculture, Science & Technology, vol. 7, no. 17, 2005.
[30] G. Rutka, "Network Traffic Prediction using ARIMA and Neural Networks Models", Electronics and Electrical Engineeging, vol. 4, pp. 47-52, 2008.
[31] G. Cybenko, "Approximation by superpositions of sigmoidal functions", Math. Control Signals Syst., vol. 2, no. 4, pp. 303-314, 1989.
[32] K. Funahashi, "On the approximate realization of continuous mappings by neural networks", Neural Networks, vol. 2, no. 3, pp. 183-192, 1989.
[33] D. White and D. Sofge, “Handbook of Intelligent Control: Neural, Fuzzy, and Adaptative Approaches”, Van Nostrand Reinhold Company, 1992.
[34] H. White and M. Stinchcombe,“Approximating and learning unknown mappings using multilayer feedward networks with bounded weights”, In Arcial neural networks: approximations and learning theory, H. White ed., Blackwell, Oxford, UK, 1992.
[35] Gallant, H. White, "On learning the derivatives of an unknown mapping with multilayer feedforward networks", Neural Networks, vol. 5, no. 1, pp. 129-138, 1992.
[36] H. White, Artificial neural networks: approximation and learning theory. Blackwell Publishers, 1992.
[37] Baccigalupi, L. Bedini, C. Burigana, G. De Zotti, A. Farusi, D. Maino, M. Maris, F. Perrotta, E. Salerno, L. Toffolatti and A. Tonazzini, "Neural networks and the separation of cosmic microwave background and astrophysical signals in sky maps", Monthly Notices of the Royal Astronomical Society, vol. 318, no. 3, pp. 769-780, 2000.
[38] R. Fierro and F. Lewis, "Control of a nonholonomic mobile robot using neural networks", IEEE Transactions on Neural Networks, vol. 9, no. 4, pp. 589-600, 1998.
[39] R. Lippmann, "An introduction to computing with neural nets", IEEE ASSP Magazine, vol. 4, no. 2, pp. 4-22, 1987.
[40] 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, 2010.
[41] M. Khashei and M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting", Applied Soft Computing, vol. 11, no. 2, pp. 2664-2675, 2011.
[42] Specht, "A general regression neural network", IEEE Transactions on Neural Networks, vol. 2, no. 6, pp. 568-576, 1991.
[43] Alon, M. Qi, R. J. Sadowsik, "Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods", J. Retailing Consumer Services, vol. 8, no. 3, pp. 147-156, 2001.
[44] M. Thompson and M. Kramer, “Modeling chemical processes using prior knowledge and neural networks”, AIChE Journal, vol. 40, no. 8, pp. 1328-1340, 1994.
[45] KTam and M. Kiang, “Managerial applications of neural networks: the case of bank failure predictions”, Management science, vol. 38, no. 7, pp. 926-947, 1992.
[46] G. Spellman, “An application of artificial neural networks to the prediction of surface ozone concentrations in the United Kingdom”, Applied Geography, vol. 19, no. 2, pp. 123-136, 1999.
[47] Chau, “Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River”, Journal of hydrology, vol. 329, no. 3, pp. 363-367, 2006.
[48] Chau, “Application of a PSO-based neural network in analysis of outcomes of construction claims”, Automation in Construction, vol. 16, no.5, pp. 642-646, 2007.
[49] Storrie-Lombardi, O. Lahav, L. Sodre and L. Storrie-Lombardi, “Morphological classification of galaxies by artificial neural networks”, Monthly Notices of the Royal Astronomical Society, vol. 259, no. 1, pp. 8P-12P, 1992.
[50] G. Foody, "Using prior knowledge in artificial neural network classification with a minimal training set", International Journal of Remote Sensing, vol. 16, no. 2, pp. 301-312, 1995.
[51] Guresen, G. Kayakutlu and T. Daim, "Using artificial neural network models in stock market index prediction", Expert Systems with Applications, vol. 38, no. 8, pp. 10389-10397, 2011.
[52] P. Atkinson and A. Tatnall, "Introduction Neural networks in remote sensing",International Journal of Remote Sensing, vol. 18, no. 4, pp. 699-709, 1997.
[53] G. Zhang, M. Y. Hu, B. Eddy Patuwo and D. C. Indro, "Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis", European Journal of Operational Research, vol. 116, no. 1, pp. 16-32, 1998.
[54] K. Fanning and K. Cogger, "A Comparative Analysis of Artificial Neural Networks Using Financial Distress Prediction", Intelligent Systems in Accounting, Finance and Management, vol. 3, no. 4, pp. 241-252, 1994.
[55] O. Nelles, “Nonlinear system identification: from classical approaches to neural networks and fuzzy models”. Springer Science & Business Media, 2001.
[56] Zhang, Artificial higher order neural networks for economics and business. Hershey, IGI Global (701 E. Chocolate Avenue, Hershey, Pennsylvania, 17033, USA), 2008.
[57] Boely and R. Botez, "New Approach for the Identification and Validation of a Nonlinear F/A-18 Model by Use of Neural Networks", IEEE Transactions on Neural Networks, vol. 21, no. 11, pp. 1759-1765, 2010.
[58] J. Michalkiewicz, "Modified Kolmogorov's Neural Network in the Identification of Hammerstein and Wiener Systems", IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no. 4, pp. 657-662, 2012.
[59] G. Wen and Y. Liu, "Adaptive fuzzy-neural tracking control for uncertain nonlinear discrete-time systems in the NARMAX form", Nonlinear Dynamics, vol. 66, no. 4, pp. 745-753, 2011.
[60] M. Shafiq and N. Butt, "Utilizing higher-order neural networks in U-model based controllers for stable nonlinear plants", International Journal of Control, Automation and Systems, vol. 9, no. 3, pp. 489-496, 2011.
[61] Z. Chen and Y. Ni, "On-board Identification and Control Performance Verification of an MR Damper Incorporated with Structure", Journal of Intelligent Material Systems and Structures, vol. 22, no. 14, pp. 1551-1565, 2011.
[62] M. Gonzalez-Olvera and Y. Tang, "Black-Box Identification of a Class of Nonlinear Systems by a Recurrent Neurofuzzy Network", IEEE Transactions on Neural Networks, vol. 21, no. 4, pp. 672-679, 2010.
[63] L. YuMin, T. DaZhen, X. Hao and T. Shu, “Productivity matching and quantitative prediction of coalbed methane wells based on BP neural network”,Science China Technological Sciences, vol. 54, no. 5, pp. 1281-1286, 2011.

[64] M. Han, J. Fan and J. Wang, "A Dynamic Feedforward Neural Network Based on Gaussian Particle Swarm Optimization and its Application for Predictive Control", IEEE Transactions on Neural Networks, vol. 22, no. 9, pp. 1457-1468, 2011.

[65] X. Zong-yi, Q. Yong, P. Xue-miao, J. Li-min and Z. Yuan, "Modelling of the Automatic Depth Control Electrohydraulic System Using RBF Neural Network and Genetic Algorithm", Mathematical Problems in Engineering, vol. 2010, pp. 1-16, 2010.

[66] Gómez-Gil, J. Ramírez-Cortes, S. Pomares Hernández and V. Alarcón-Aquino, "A Neural Network Scheme for Long-Term Forecasting of Chaotic Time Series", Neural Processing Letters, vol. 33, no. 3, pp. 215-233, 2011.

[67] X. Zhi-Bin, "Particle-swarm-optimization-based selective neural network ensemble & its application to modeling resonant frequency of microstrip antenna".inMicrostrip Antennas, pp. 69-82, 2011.

[68] Z. Chen and Y. Ni, "On-board Identification & Control Performance Verification of an MR Damper Incorporated with Structure",Journal of Intelligent Material Systems & Structures, vol. 22, pp. 1551-1565, 2011.

[69] B. Cosenza, "Development of a neural network for glucose concentration prevision in patients affected by type 1 diabetes",Bioprocess & Biosystems Engineering, pp. 1-9, 2012.

[70] S. Chatterjee, S. Nigam, J. Singh and L. Upadhyaya, "Software fault prediction using Nonlinear Autoregressive with eXogenous Inputs (NARX) network", Applied Intelligence, vol. 37, no. 1, pp. 121-129, 2012.

[71] M. Rashid, M. Frasca, A. Ali, R. Ali, L. Fortuna and M. Xibilia, "Nonlinear model identification for Artemia population motion", Nonlinear Dynamics, vol. 69, no. 4, pp. 2237-2243, 2012.

[72] S. Geman, E. Bienenstock and R. Doursat, "Neural Networks and the Bias/Variance Dilemma", Neural Computation, vol. 4, no. 1, pp. 1-58, 1992.

[73] C. Granger, "Strategies for modeling nonlinear time-series relationships",The Economic Record, vol. 69, no. 206, pp. 233-238, 1993.

[74] R. Hoptroff, "The principles and practice of time series forecasting and business modelling using neural nets", Neural Computing & Applications, vol. 1, no. 1, pp. 59-66, 1993.

[75] R. Ashley, "Statistically significant forecasting improvements: how much out-of-sample data is likely necessary?", International Journal of Forecasting, vol. 19, no. 2, pp. 229-239, 2003.

[76] Azoff,"Neural Network Time Series Forecasting of Financial Markets". Chichester, New York, Wiley, 1994.

[77] W. Gorr, D. Nagin and J. Szczypula, "Comparative study of artificial neural network and statistical models for predicting student grade point averages", International Journal of Forecasting, vol. 10, no. 1, pp. 17-34, 1994.

[78] Komarek andA. Moore, "Making logistic regression a core data mining tool with tr-irls",IEEE International Conference on Data Mining (ICDM'05), pp. 4-10, 2005.

[79] Zhang and M. Qi,"Predicting consumer retail sales using neural networks", In Smith & Gupta (Eds.), Neural Networks in Business: Techniques and Applications, Hershey, PA: Idea Group Publishing, pp. 26-40, 2002.

[80] Kaastra and M. Boyd, "Designing a neural network for forecasting financial and economic time series", Neurocomputing, vol. 10, no. 3, pp. 215-236, 1996.

[81] Zhang and M. Hu, "Neural network forecasting of the British Pound/US Dollar exchange rate", Omega, vol. 26, no. 4, pp. 495-506, 1998.

[82] Coakley and C. Brown, "Artificial neural networks in accounting and finance: modeling issues", International Journal of Intelligent Systems in Accounting, Finance & Management, vol. 9, no. 2, pp. 119-144, 1999.

[83] W. Remus and M. O'Connor, "Neural networks for time series forecasting",in Armstrong JS (ed) Principles of forecasting: a hand book for researchers & practitioners. Kluwer, Norwell, MA, pp 245–256, 2001.

[84] Y. Pan, T. Pohlen andS. Manago, "Hybrid neural network model in forecasting aggregate US retail sales", inAdvances in business & management forecasting, vol. 9, 153-170, 2013.

[85] S. Anbazhagan and N. Kumarappan, "Day-ahead deregulated electricity market price forecasting using recurrent neural network",Systems Journal, IEEE, vol. 7, no. 4, pp. 866-872, 2013.

[86] Azad, S. Mekhilef and V. Ganapathy, "Long-Term Wind Speed Forecasting and General Pattern Recognition Using Neural Networks", IEEE Transactions on Sustainable Energy, vol. 5, no. 2, pp. 546-553, 2014.

[87] Heydari and F. Keynia, "Prediction of Wind Power Generation Through Combining Particle Swarm Optimization & Elman Neural Network (El-PSO)",International Energy Journal, vol. 15, no. 2, 2015.

[88] Chatfield,"Time-Series Forecasting". Boca Raton, Chapman & Hall/ CRC, FL, 2001.

[89] Lennon, G. Montague, A. Frith, C. Gent and V. Bevan, "Industrial applications of neural networks-An investigation",Journal of Process Control, vol. 11, pp. 497-507, 2001.

[90] G. Zhang, B. Patuwo and M. Hu, "A simulation study of artificial neural networks for nonlinear time-series forecasting",Computers & Operations Research, vol. 28, no. 4, pp. 381-396, 2001.

[91] G. Zhang, "An investigation of neural networks for linear time-series forecasting",Computers & Operations Research, vol. 28, pp. 1183-1202, 2001.

[92] S. Arlot andA. Calisse, "A survey of cross-validation procedures for model selection", Statistic Surveys, vol. 4, pp. 40-79, 2010.

[93] Bakirtzis, V. Petridis, S. Kiartzis, M. Alexiadis and A. Maissis, "A neural network short term load forecasting model for the Greek power system", IEEE Transactions on Power Systems, vol. 11, no. 2, pp. 858-863, 1996.

[94] Khotanzad, R. Afkhami-Rohani, T. Lu, A. Abaye, M. Davis and D. Maratukulam, "ANNSTLF-A neural-network-based electric load forecasting system", IEEE Transactions on Neural Networks, vol. 8 no. 4, pp. 835-846, 1997.

[95] Bishop, "Neural networks for pattern recognition", Oxford university press, 1995.

[96] S. Abid, F. Fnaiech and M. Najim, "A fast feedforward training algorithm using a modified form of the standard backpropagation algorithm", IEEE Transactions on Neural Networks, vol. 12, no. 2, pp. 424-430, 2001.

[97] M. Hagan and M. Menhaj, "Training feedforward networks with the Marquardt algorithm", IEEE Transactions on Neural Networks, vol. 5, no. 6, pp. 989-993, 1994.

[98] T. Kwok and D. Yeung, "Objective functions for training new hidden units in constructive neural networks",IEEE Transactions on Neural Networks, vol. 8, no. 5, pp. 1131-1148, 1997.

[99] Y. Yamamoto andP. Nikiforuk, "A new supervised learning algorithm for multilayered and interconnected neural networks"IEEE Transactions on Neural Networks, Vol. 11, pp. 36–46, 2000.

[100] Yam and T. Chow,"Extended least squares based algorithm for training feedforward networks", IEEE Transactions on Neural Networks, Vol. 8, pp. 806–810, 1997.

[101] Yam and T. Chow, "A weight initialization method for improving training speed in feedforward neural networks", Neurocomputing, Vol. 30, pp. 219–232, 2000.

[102] Yam and T. Chow, "Feedforward networks training speed enhancement by optimal initialization of the synaptic coefficients",IEEE Transactions on Neural Networks, Vol. 12, pp. 430–434, 2001.

[103] X. Yu, M. Efe and O. Kaynak, "A general backpropagation algorithm for feedforward neural networks learning", IEEE Transactions on Neural Networks, vol. 13, no. 1, pp. 251-254, 2002.

[104] T. Kathirvalavakumar and P. Thangavel, "A Modified Backpropagation Training Algorithm for Feedforward Neural Networks*",Neural Processing Letters, vol. 23, no. 2, pp. 111-119, 2006.

[105] T. Williams, "Predicting changes in construction cost indexes using neural networks", Journal of Construction Engineering and Management, vol. 120, no. 2, pp. 306–320, 1994.

[106] V. Asari, "Training of a feedforward multiple-valued neural networks by error backpropagation with a multilevel threshold function",IEEE Transactions on Neural Networks, vol. 12, pp. 1519–1521, 2001.

[107] Chowdhury, Y. Singh and R. Chansarkar, “Dynamic tunneling technique for efficient training of multilayer perceptrons”,IEEE. Transactions on Neural Networks, vol. 10, pp. 48–55, 1999.
[108] G. Hinton, “Connectioniest Learning Procedure in Machine Learning: Paradigms & Methods”, In: J. G. Carbonell, (ed.), pp. 185–234, Cambridge MA: MIT press, 1989.
[109] T. Kathirvalavakumar andP. Thangavel, “A new learning algorithm using simultaneous perturbation with weight initialization”,Neural Processing Letters, vol. 17, pp. 55–68, 2003.
[110] Krogh and J. Hertz, “Generalization in a linear perceptron in the presence of noise”,Journal of Physics A – Mathematical & General, vol. 25, pp. 1135–1147, 1992.
[111] G. Lera andM. Pinzolas, “Neighborhood based Levenberg–Marquardt algorithm for neural network training”IEEE Transactions on Neural Networks, vol. 13, pp. 1200–1203, 2002.
[112] Parisi, E. Di Claudio, G. Orlandi and B. Rao, "A generalized learning paradigm exploiting the structure of feedforward neural networks", IEEE Transactions on Neural Networks, vol. 7, no. 6, pp. 1450-1460, 1996.
[113] Reed andR. Marks, “Neural Smithing Supervised Learning in Feedforward Artificial Neural Networks”, Cambridge: MIT, 1999.
[114] P. Thangavel and T. Kathirvalavakumar, “A new learning algorithm using simultaneous perturbation with weight initialization”, Neural processing letters, 17(1), 55-68, 2003.
[115] P. Thangavel and T. Kathirvalavakumar, “Training feedforward networks using simultaneous perturbation with dynamic tunneling”, Neurocomputing, 48(1), 691-704, 2002.
[116] Bryson, “Applied optimal control: optimization, estimation & control”,CRC Press, 1975.
[117] E. Vamsidhar, K. Varma, P. Rao and R. Satapati, “Prediction of rainfall using backpropagation neural network model”,International Journal on Computer Science & Engineering, vol. 2, no. 4, pp. 1119-1121,2010.
[118] Duy, Y. Sato and Y. Inoguchi, “Improving accuracy of host load predictions on computational grids by artificial neural networks”, International Journal of Parallel, Emergent and Distributed Systems, vol. 26, no. 4, pp. 275-290, 2011.
[119] S. Ubaidillah, R. Sallehuddin and N. Ali, “Cancer Detection Using Aritifical Neural Network & Support Vector Machine: A Comparative Study”,JurnalTeknologi, vol. 65, no. 1, 2013.
[120] Elmi, S. Ibrahim andR. Sallehuddin, “Detecting sim box fraud using neural network”IT Convergence and Security 2012, pp. 575-582, 2013.
[121] R. Alwee, S. Hj Shamsuddin and R. Sallehuddin, “Hybrid support vector regression and autoregressive integrated moving average models improved by particle swarm optimization for property crime rates forecasting with economic indicators”,The Scientific World Journal, 2013.
[122] H. Maier and G. Dandy, “Neural network models for forecasting univariate time series”, Neural Network World, vol. 5, pp. 747-772, 1996.
[123] O. Valenzuela, I. Rojas, F. Rojas, H. Pomares, L. Herrera, A. Guillén and M. Pasadas, “Hybridization of intelligent techniques and ARIMA models for time series prediction”,Fuzzy Sets and Systems, vol. 159, no. 7, pp. 821-845, 2008.
[124] Khashei, M. Bijari and G. Raissi Ardali,“Improvement of auto-regressive integrated moving average models using fuzzy logic & artificial neural networks (ANNs)”,Neurocomputing, vol. 72, no. 4, pp. 956-967, 2009.
[125] Weruaga, “Multilayer Neural Networks”, Beamer presentation, 2006.
[126] Bylander, “Universal Approximation”, 2003,http://www.cs.utsa.edu/~bylander/cs4793 /univ-approx.pdf
[127] Bonnell, “ Implementation of a New Sigmoid Function in Backpropagation Neural Networks” (Doctoral dissertation, East Tennessee State University), 2011.
[128] T. Windeatt and C. Zor, “Minimising added classification error using Walsh coefficients”,IEEE Transactions on Neural Networks, vol. 22, no. 8, pp. 1334-1339,2011.
[129] Panklib, C. Prakasvudhisarn and D. Khummongkol, “Electricity Consumption Forecasting in Thail & Using an Artificial Neural Network & Multiple Linear Regression”,Energy Sources, Part B: Economics, Planning, & Policy, vol. 10, no. 4, pp. 427-434, 2015.
[130] S. Khan, T. Ayub andS. Rafeeqi, “Prediction of compressive strength of plain concrete confined with ferrocement using artificial neural network (ANN) & comparison with existing mathematical models”, American Journal of Civil Engineering & Architecture, vol. 1, no. 1, pp. 7-14, 2013.
[131] T. Ayub, M. Nurrudin, S. Khan and F. Memon, “Prediction of Compressive Strength of Plain Concrete Confined with Ferrocement using Artificial Neural Network (ANN)”Special issues of the International Journal of Soft Computing & Software Engineering [JSCSE], vol. 3, no. 3, pp. 663-667, 2013.
[132] Maia, F. de Carvalho and T. Ludermir, “Forecasting models for interval-valued time series”,Neurocomputing, vol. 71, no. 16, pp. 3344-3352, 2008.
[133] Wang and Y. Mei, “Model for forecasting construction cost indices in Taiwan”Construction Management & Economics,vol. 16, no. 2, pp. 147-157, 1998.
[134] P. Hartono and S. Hashimoto, “Learning from imperfect data”,Applied Soft Computing, vol. 7, pp. 353-363, 2007.
[135] Pham and D. Karaboga, “Intelligent optimization techniques: genetic algorithms, tabu search, simulated annealing & neural networks”,Springer Science & Business Media, 2012.
[136] G. Zhang, M. Y. Hu, B. E.Patuwo andD. C. Indro, “Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis”, European journal of operational research, 116(1), 16-32, 1999.
[137] T. Kim, K. Oh, C. Kim and J. Do, “Artificial neural networks for non-stationary time series”,Neurocomputing, vol. 61, pp. 439-447, 2004.
[138] Cadenas and W.Rivera, “Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model”,Renewable Energy, vol. 35, no. 12, pp. 2732-2738, 2010.
[139] S. Sfetsos, “A novel approach for the forecasting of mean hourly wind speed time series”,Renewable energy, Elsevier, pp. 163-174, 2002.
[140] C. Wu, K. Chau and C. Fan, "Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques", Journal of Hydrology, vol. 389, no. 1-2, pp. 146-167, 2010.
[141] Wang, H. Zou, J. Su, L. Li and S. Chaudhry,“An ARIMA-ANN Hybrid Model for Time Series Forecasting”,Systems Research & Behavioral Science,vol. 30, no. 3, pp. 244-259, 2013.
[142] Yolcu, C. Aladag, E. Egrioglu and V. Uslu, “Time-series forecasting with a novel fuzzy time-series approach: an example for Istanbul stock market”,Journal of Statistical Computation & Simulation, vol. 83, no. 4, pp. 599-612, 2013.
[143] Claveria and S. Torra, “Forecasting tourism dem & to Catalonia: Neural networks vs. time series models”, Economic Modelling, vol. 36, pp. 220-228, 2014.
[144] K. Gairaa, A. Khellaf, Y. Messlem and F. Chellali, “Estimation of the daily global solar radiation based on Box–Jenkins & ANN models: A combined approach”,Renewable & Sustainable Energy Reviews, vol. 57, pp. 238-249, 2016.
[145] Ardakani and M. Ardehali, “Long-term electrical energy consumption forecasting for developing & developed economies based on different optimized models & historical data types”,Energy, vol. 65, pp. 452-461, 2014.
[146] J. Pati, K. K. Shukla, "A comparison of ARIMA neural network and a hybrid technique for Debian bug number prediction", Proc. Comput. Commun. Technol., pp. 47-53, 2014.
[147] R. Sharda and R. Patil, “Neural networks as forecasting experts: An empirical test” In: Proceedings of the International Joint Conference on Neural Networks. Washington, D.C., vol. 2, pp. 491-494, 1990.
[148] R. Sharda and R. Patil, “Connectionist approach to time series prediction: An empirical test”,Journal of Intelligent Manufacturing, vol. 3, pp. 317-323, 1992.

[149] Z. Tang, C. Almeida and P. Fishwick, “Time series forecasting using neural networks vs Box-Jenkins methodology”, Simulation, vol. 57, no. 5, pp. 303–310, 1991.

[150] Z. Tang and P. Fishwick, “Feedforward neural nets as models for time series forecasting”,ORSA Journal on Computing, vol.5, no. 4, pp. 374–385, 1993.

[151] S. Kang, “An Investigation of the Use of Feedforward Neural Networks for Forecasting”, Ph.D. Thesis, Kent State University, 1991.

[152] Kohzadi, M. Boyd, B. Kermanshashi and I. Kaastra, “A comparison of artificial neural network & time series models for forecasting commodity prices”,Neurocomputing, vol. 10, pp. 169-181, 1996.

[153] T. Hill, M. O’ConnorandW. Remus, “Neural Network Models for Time Series Forecasts”,Management Science, 42, 7, 1082-1092, 1996.

[154] T. Hill, L. Marquez, M. O'Connor and W. Remus, “Artificial neural network models for forecasting and decision making”,International Journal of Forecasting, vol. 10, no.1, pp. 5-15, 1994.

[155] T. Hill, M. O’Connor and W. Remus, “Neural network models for time series forecasts”Management Sciences, vol. 42, no. 7, pp. 1082-1092, 1996.

[156] Caire, G. Hatabian and C. Muller, “Progress in forecasting by neural networks”,Proceedings of the International Joint Conference on Neural Networks, vol. 2, pp. 540–545, 1992.

[157] T. Hann andE. Steurer, “Much ado about nothing? Exchange rate forecasting: Neural networks vs. linear models using monthly & weekly data”,Neurocomputing,vol. 10, pp. 323-339, 1996.

[158] S. Balkin and J. Ord, “Automatic neural network modeling for univariate time series”International Journal of Forecasting, vol. 16, no. 4, pp. 509-515, 2000.

[159] Harun, “Short term load forecasting using Artificial Neural Network (ANN) with multiple time lags, stationary time series & Principal Component Analysis (PCA)”,(Masters dissertation, Universiti Teknologi MARA), 2011.

[160] B. Barbosa, L. Aguirre, C. Martinez and Braga, “Black and Gray-Box Identification of a Hydraulic Pumping System”IEEE Trans. Control Systems Technology, vol. 19, no. 2, pp. 398-406, 2011.

[161] S. Ahmadi and M. Karrari, “Nonlinear identification of synchronous generators using a local model approach”Przeglad Elektrotechniczny (Electrical Review), vol. 8, pp. 166-170, 2011.

[162] L. Teslic, B. Hartmann, O. Nelles and I. Skrjanc, “Nonlinear system identification by gustafson–kessel fuzzy clustering and supervised local model network learning for the drug absorption spectra process” IEEE Transactions on Neural Networks, vol. 22, no. 12, pp. 1941-1951, 2011.

[163] C. Safak, V. Topuz and A. Baba, “Pneumatic motor speed control by trajectory tracking fuzzy logic controller”Sadhana, vol. 35, no. 1, pp. 75-86, 2010.

[164] K. Tanabe, B. Lučić, D. Amić, T. Kurita, M. Kaihara, N. Onodera and T. Suzuki, “Prediction of carcinogenicity for diverse chemicals based on substructure grouping & SVM modeling”,Molecular diversity, vol. 14, no. 4, pp. 789-802, 2010.

[165] K. Tanabe, B. Lučić, D. Amić, T. Kurita, M. Kaihara, N. Onodera and T. Suzuki, “Prediction of carcinogenicity for diverse chemicals based on substructure grouping & SVM modeling”, Molecular diversity, vol. 14, no. 4, pp. 789-802, 2010.

[166] T. Tsai, C. Lee and C. Wei,“Neural network based temporal feature models for short-term railway passenger demand forecasting”Expert Systems with Applications, vol. 36, no. 2, pp. 3728-3736, 2009.

[167] M. Cogollo and J. Velasquez, “Methodological advances in artificial neural networks for time series forecasting”,IEEE Latin America Transactions, vol. 12, no. 4, 764-771,2014.

[168] Yassin, A. Zabidi, M. Salleh and N. Khalid, “Malaysian tourism interest forecasting using Nonlinear Auto-Regressive (NAR) model”, 2013 IEEE 3rd International Conference on System Engineering & Technology (ICSET), pp. 32-36, 2013.

[169] Pawlus, H. Karimi and K. Robbersmyr, “Data-based modeling of vehicle collisions by nonlinear autoregressive model & feedforward neural network”,Information Sciences, vol. 235, pp. 65-79, 2013.

[170] H. Chen and X. Liu, “Melt Index Prediction Based on Two Compensation by Compound Basis Function Neural Network & Hidden Markov Model”, 2014.

[171] Saberian, H. Hizam, M. Radzi, M. Ab Kadir and M. Mirzaei, “Modelling and prediction of photovoltaic power output using artificial neural networks”International Journal of Photoenergy, pp. 1-11,2014.

[172] S. Kamaruddin, N. Ghani and N. Ramli, “Modified nonlinear neural network forecasting models on Malaysian s & costs indices”, Proceedings of 2015 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), pp. 191-196, 2015.

[173] Acuña, C. Ramirez and M. Curilem,“Comparing NARX and NARMAX models using ANN and SVM for cash demand forecasting for ATM”The IEEE 2012 International Joint Conference on Neural Networks (IJCNN), pp. 1-6, 2012.

[174] Goyal and D. Kaur, “A Novel Vehicle Classification Model for Urban Traffic Surveillance Using the Deep Neural Network Model”, I.J. Education & Management Engineering, vol.1, 18-31,2016.

[175] M. Rahiman, M. Taib and Y. Salleh, “Selection of training data for modeling essential oil extraction system using NNARX structure”,In 2007 Int. Conf. Control, Automation & Systems (COEX 2007), Seoul, Korea, pp. 1691-1695, 2007.

[176] Chiras, C. Evans and D. Rees, “Nonlinear gas turbine modeling using NARMAX structures”,IEEE Trans. Instrumentation and Measurement, vol. 50, no. 4, pp. 893-898, 2001.

[177] K. Ahn and H. P. H. Anh, “Inverse Double NARX Fuzzy Modeling for System Identification”,IEEE/ ASME Transactions on Mechatronics, 15(1), 136-148, 2010.

[178] N. Kadir, N. Md Tahir, M. assin and A. Zabidi, “Malaysian tourism interest forecasting using Nonlinear Auto-Regressive Moving Average (NARMA) model”,2014 IEEE Symposium on Wireless Technology & Applications (ISWTA), pp. 193-198, 2014.

[179] Dimogianopoulos, J. Hios and S. Fassois, “FDI for Aircraft Systems Using Stochastic Pooled-NARMAX Representations, Design and Assessment”,IEEE Trans. Control Systems Technology, vol. 17, no.6, pp. 1385-1397, 2009.

[180] Hornik, M. Stinchcombe and H. White, “Multilayer feedforward networks are universal approximators”, Neural networks, vol. 2, no.5, pp. 359-366, 1989.

[181] F. Hampel, E. Ronchetti, P. Rousseeuw and W. Stahel, “Robust Statistics, The approach based on influence functions”,Wiley, New York, 1986.

[182] C. Chen and G. Yin, “Computing the Efficiency & Tuning Constants for M-Estimation”, In Proceedings of the 2002 Joint Statistical Meetings, vol. 478, p. 482, 2002.

[183] J. William, “ Introduction to Robust & Quasi-Robust Statistical Methods”, Springer, Berlin, Heidelberg, 1983.

[184] Huber, “Robust Statistics”, John Wiley & Sons, New York, 1981.

[185] Farcomeni and L. Ventura, “An overview of robust methods in medical research”Statistical Methods in Medical Research,vol. 21, no. 2, pp. 111-133, 2012.

[186] D. Adedia, Comparison of robust regression estimators. Unpublished MPhil thesis, KNUST, Kumasi, 2014.

[187] Kantar and V. Yildirim, “Robust Estimation for Parameters of the Extended Burr Type III Distribution”Communications in Statistics-Simulation and Computation, vol. 44, no. 7, pp. 1901-1930, 2015.

[188] Kordos and A. Rusiecki, “Reducing noise impact on MLP training”, Soft Computing, pp. 1-17, 2015.

[189] Kordos and A. Rusiecki, “Improving MLP neural network performance by noise reduction. In Theory & Practice of Natural Computing”, Springer Berlin Heidelberg, pp. 133-144, 2013.

[190] Rusiecki, “Robust Learning Algorithm Based on Iterative Least Median of Squares”,Neural Process Lett, vol. 36, pp. 145-160, 2012.

[191] S. Kamaruddin, N. Ghani and N. Ramli, “Enhancing Backpropagation of ANN-NAR & ANN-NARMA using Robust Estimators with Application on Real Industrial Data”,Proceedings of The 3rd International Conference on Computer Science & Computational Mathematics (ICCSCM 2014), vol. 3, pp. 256-266, 2014.

[192] Bruna, “Short term load forecasting using non-linear models”Master Thesis Report, Measurement & Control Group, Faculty of Electrical Engineering, Eindhoven University of Technology, 1994.

[193] Yang, “Engineering Optimization: An Introduction with Metaheuristic Applications”John Wiley and Sons, USA, 2010.

[194] X. Yang, “A New Metaheuristic Bat-inspired algorithm”,Nature Inspired Cooperative Strategies for Optimization, 284, pp. 65-74, 2010.

[195] X. Yang, “Nature-Inspired Metaheuristic Algorithms” Luniver Press, UK, 2008.

[196] C. Blum, M. Aguilera, A. Roli and M.Sampels, “Hybrid Metaheuristics, An Emerging Approach to Optimization”, Springer, 2008.

[197] S. Pal, C. Rai and P. Amrit, “Comparative study of Firefly Algorithm & Particle Swarm Optimization for Noisy Nonlinear Optimization Problems”,International Journal of Intelligent Systems & Applications, vol. 10, pp. 50-57, 2012.

[198] M. Yuanbin, “Particle Swarm Optimization for Cylinder Helical GearMulti-objective Design problems”,Applied Mechanics and Materials, vol. 109, pp. 216-221, 2012.

[199] X. Yang, and X. He, “Firefly Algorithm: Recent Advances and Applications”, International Journal Swarm Intelligence, vol. 1, no.1, pp. 36-50, 2013.

[200] Kennedy and R. Eberhart, “A discrete binary version of the particle swarm algorithm”,Systems, Man, and Cybernetics, vol. 5, pp. 4104-4108, 1997.

[201] X. Yang, “Firefly algorithms for multimodal optimization”, InStochastic algorithms: foundations and applications, pp. 169-178, 2009.

[202] X. Yang and S. Deb, “Cuckoo Search via Levy Flights”,Proceedings of World Congress on Nature & Biologically Inspired Computing, pp. 210-214, 2009.

[203] Bansal and K. Deep, “Optimization of directional overcurrent relay times by particle swarm optimization”,Swarm Intelligence Symposium, pp. 1-7, 2008.

[204] C. Floudas and P. Pardalos,“State of the art in global optimization: computational methods and applications”, Springer Science & Business Media, vol. 7, 2013.

[205] Gandomi, X. Yang and A. Alavi, “Mixed variable structural optimization using firefly algorithm”Computers & Structures, vol. 89, no. 23, pp. 2325-2336,2011.

[206] Parpinelli and H. Lopes, “New inspirations in swarm intelligence: a survey”,International Journal of Bio-Inspired Computation, vol. 3, no. 1, pp. 1-16, 2011.

[207] X. Yang, “Multiobjective firefly algorithm for continuous optimization”,Engineering with Computers, vol. 29, no.2, pp. 175-184, 2013.

[208] M. Spasos, K. Tsiakmakis, N. Charalampidis and R. Nilavalan, “RF-MEMS switch actuation pulse optimization using Taguchi’s method”,Microsystem Technologies, vol. 17, pp. 1351-1359, 2011.

[209] F. Ducatelle, G. Caro and L. Gambardella, “Principles and applications of swarm intelligence for adaptive routing in telecommunications networks”,Swarm Intelligence, vol. 4, pp. 173-198, 2010.

[210] M. Arjona, M.Cisneros-González and C. Hernández, “Parameter Estimation of a Synchronous Generator Using a Sine Cardinal Perturbation and Mixed Stochastic–Deterministic Algorithms”,IEEE Trans. Industrial Electronics, vol. 58, no. 2, pp. 486-493, 2011.

[211] C. Liao, “Enhanced RBF Network for Recognizing Noise-Riding Power Quality Events”,IEEE Trans. Instrumentation & Measurement, vol. 59, no.6, pp. 1550-1561, 2010.

[212] M. Wan, L. Li, J. Xiao, C. Wang and Y. Yang, “Data clustering using bacterial foraging optimization”,Journal of Intelligent Information Systems, vol. 38, no.2, pp. 321-341, 2012.

[213] D. Karaboga and B. Basturk, “A Powerful and Efficient Algorithm for Numerical Function Optimization: Artificial Bee Colony (ABC) Algorithm”,Journal of Global Optimization, vol. 39, no. 3, pp. 459-471, 2007.

[214] Brajevic, and M. Tuba, “An upgraded artificial bee colony (ABC) algorithm for constrained optimization problems”,Journal of Intelligent Manufacturing, vol. 10 pp. 1-12, 2012.

[215] W. Yunna and S. Zhaomin, “Application of Radial Basis Function Neural Network Based on Ant Colony Algorithm in Credit Evaluation of Real Estate Enterprises”,Proc. 2008 Int. Conf. on Management Science and Engineering, pp. 1322-1329, 2008.

[216] Udhayakumar and S. Kumanan, “Sequencing & scheduling of job & tool in a flexible manufacturing system using ant colony optimization algorithm”,International Journal of Advanced Manufacturing Technology, vol. 50, pp. 1075-1084, 2010.

[217] M. Al-Meshah, “Firefly algorithms with probabilistic neural network for classification problems”, PhD Thesis. Malaysian National University, 2013.

[218] X. S. Yang and N. I. M. Algorithms, Luniver Press. Beckington, UK, 242-246, 2008.

[219] Zigiaris, “Business process reengineering BPR”,Report produced for the EC funded project, InnoRegio Project, BPR Hellas SA, January, 2000.

[220] M. Basu, “A simulated annealing-based goal-attainment method for economic emission load dispatch of fixed head hydrothermal power systems”,International Journal of Electrical Power & Energy Systems, vol. 27, no. 2, pp. 147-153, 2005.

[221] Chen and C. Huang, “Biobjective power dispatch using goal-attainment method & adaptive polynomial networks”,IEEE Transactions on Energy Conversion, vol. 19, no. 4, pp. 741–747, 2004.

[222] Bouktir, R. Labdani and L. Slimani, “Economic power dispatch of power system with pollution control using multi-objective particle swarm optimization”,Journal of Pure & Applied Sciences, vol. 4, no. 2, pp. 57–77, 2007.

[223] El-Wahed, A. Mousa, and M. Elsisy, “Solving economic emissions load dispatch problem by using hybrid ACO-MSM approach”, The Online Journal on Power and Energy Engineering, vol. 1, no. 1, pp.36-53, 2007.

[224] Xie, S. Wang and Z. Wu, “Study on economic, rapid and environmental power dispatch based on fuzzy multi-objective optimization”,Modern Applied Science, vol. 3, no. 6, pp. 38-44, 2009.

[225] Alsumait and J. Sykulski, “Solving economic dispatch problem using hybrid GA-PS-SQP method”. Proceedings of the International Conference on Computer as a Tool (EUROCON ’09), pp. 351– 356, 2009.

[226] Sudhakaran, S. Slochanal, R. Sreeram and N. Chandrasekhar, “Application of refined genetic algorithm to combined economic & emission dispatch”, Journal of the Institution of Engineers, vol. 85, no. 2, pp. 115–119, 2004.

[227] Goncalves, C. Almeida, J. Kuk, and M. Delgado, “Solving economic load dispatch problem by natural computing intelligent systems”, in Proceedings of the 15th International Conference on Intelligent System Applications to Power Systems (ISAP ’09), vol. 1, pp.257-264, 2009.

[228] E. Osaba, X. Yang, F. Diaz, E. Onieva, A. Masegosa and A. Perallos, “A discrete firefly algorithm to solve a rich vehicle routing problem modelling a newspaper distribution system with recycling policy”, Soft Computing, pp. 1-14, 2016.

[229] Alb, P. Alotto, C. Magele, W. Renhart, K. Preis and B. Trapp, “Firefly Algorithm for Finding Optimal Shapes of Electromagnetic Devices”, IEEE Transactions on Magnetics, vol. 52, no. 3, pp. 1-4, 2016.

[230] Gálvez and A. Iglesias, “New memetic self-adaptive firefly algorithm for continuous optimization”,International Journal of Bio-Inspired Computation, vol. 8, no.5, pp. 300-317, 2016.

[231] Łukasik and S. Żak, “Firefly algorithm for continuous constrained optimization tasks. In Computational Collective Intelligence”, Semantic Web, Social Networks & Multiagent Systems. Springer Berlin Heidelberg, pp. 97-106, 2009.

[232] Saadi, M. Nor Azura and M. Norazan, “The Superiority of Evolutionary Algorithms to Robustify Backpropagation Neural Network Learning Algorithm”,Journal of Applied Environmental and Biological Sciences, vol. 6, no. 25, pp. 17-31, 2016.

[233] M. M.Basri, N. M. Nawi, M. Mamat and N. A. Hamid, “A New Class of Nonlinear Conjugate Gradient for Training Back Propagation Algorithm”, In International Conference on Soft Computing and Data Mining, pp. 200-212, 2018.
[234] R. Heravi and G. A. Hodtani, “ A New Correntropy-Based Conjugate Gradient Backpropagation Algorithm for Improving Training in Neural Networks”, IEEE Transactions on Neural Networks and Learning Systems, 2018.
[235] Hubara, E. Hoffer and D. Soudry, “Quantized Back-Propagation: Training Binarized Neural Networks with Quantized Gradients”, (2018).
[236] F. Bordes, T. Berthier, L. Di Jorio, P. Vincent and Y. Bengio, “Iteratively unveiling new regions of interest in Deep Learning models”, 2018.
[237] Aljarah, H. Faris and S. Mirjalili, “ Optimizing connection weights in neural networks using the whale optimization algorithm”, Soft Computing, 22(1), 1-15, 2018.