AN IMPROVED MULTILAYER PERCEPTRON BASED ON WAVELET APPROACH FOR PHYSICAL TIME SERIES PREDICTION

ASHIKIN BINTI ALI

A thesis submitted in partial
fulfillment of the requirement for the award of the
Degree of Master of Information Technology

Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia

FEBRUARY, 2014

ABSTRACT

The real world datasets engage many challenges such as noisy data, periodic variations on several scales and long-term trends that do not vary periodically. Meanwhile, Neural Networks (NN) has been successfully applied in many problems in the domain of time series prediction. The standard NN adopts computationally intensive training algorithms and can easily get trapped into local minima. To overcome such drawbacks in ordinary NN, this study focuses on using a wavelet technique as a filter at the pre-processing part of the ordinary NN. However, this study exposed towards an idea to develop a model called An Improved Multilayer Perceptron based on Wavelet Approach for Physical Time Series Prediction (W-MLP) to overcome such drawbacks of ordinary NN. W-MLP, a network model with a wavelet technique added in the network, is trained using the standard backpropagation gradient descent algorithm and tested with historical temperature, evaporation, humidity and wind direction data of Batu Pahat for 5-years-period (2005-2009) and earthquake data of North California for 4-years-period (1995-1998). Based on the obtained results, the proposed method W-MLP yields better performance compared to the existing filtering techniques. Therefore, it can be concluded that the proposed W-MLP can be an alternative mechanism to ordinary NN for a one-step-ahead prediction of those five events.

ABSTRAK

Set-set data pada masa kini menghadapi banyak cabaran – antaranya data hingar, variasi berkala pada skala-skala tertentu dan kecenderungan jangka panjang yang tidak pula menyela pada waktu-waktu tertentu. Sementara itu, pada masa yang sama Rangkaian Neural (*RN*) telah berjaya diaplikasikan pada kebanyakan permasalahan dalam domain jangkaan siri peramalan masa. *RN* piawai ini mengguna-pakai algoritma latihan yang dikomputasi secara intensif dan mudah pula terperangkap dalam minima tempatan. Untuk mengatasi cabaran-cabaran sebegini, maka kajian ini dijalankan bagi memfokus penggunaan teknik "*wavelet"* sebagai saringan pada peringkat pra-pemprosesan bagi *RN* piawai. Walau bagaimanapun, kajian ini juga terbuka kepada idea membangunkan sebuah model yang dipanggil *"An Improved Multilayer Perceptron based on Wavelet Approach for Physical Time Series Prediction (W-MLP)"* bagi mengatasi halangan-halangan yang dihadapi oleh *RN* piawai. *W-MLP,* sebuah model rangkaian dengan teknik *wavelet* juga telah dilatih menggunakan algoritma kecerunan menurun perambatan balik yang diuji dengan data-data historikal suhu, sejatan, kelembapan dan arah angin bagi daerah Batu Pahat bagi jangkamasa lima tahun (2005-2009) dan juga data-data gempa bumi di California Utara bagi jangkamasa empat tahun (1995-1998). Berdasarkan dapatan yang diperolehi, kaedah *W-MLP* yang dicadangkan ini menghasilkan prestasi yang lebih baik dari teknik-teknik saringan sedia ada. Oleh itu, dapat dirumuskan bahawa kaedah *W-MLP* yang dicadangkan ini boleh dijadikan mekanisme alternatif kepada *RN* piawai sebagai peramalan yang bersifat satu langkah ke hadapan bagi kelima-lima peristiwa yang disebutkan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

NN	-	Neural Network
MLP	-	Multilayer Perceptron
W-MLP	-	Wavelet Transform with Multilayer Perceptron
H-MLP	-	High Pass Filter-MLP
B-MLP	-	Band Pass Filter-MLP
L-MLP	-	Low Pass Filter-MLP
BP	-	Backpropagation
MMD	-	Malaysian Meteorological Department
NCEDC	-	Northern California Earthquake Data Centre
ANN	-	Artificial Neural Network
LPF	-	Low Pass Filter
HPF	-	High Pass Filter
BPF	-	Band Pass Filter
CWT	-	Continuous Wavelet Transform
DWT	-	Discrete Wavelet Transform
MA	-	Moving Average
ARMA	-	Autoregressive Moving Average
ARIMA	-	Autoregressive Integrated Moving Average
EEG	-	Electroencephalography
TSK	-	Takagi-Sugeno-Kang
HRR	-	High Range Resolution Radar
GRNN	-	Generalized Regression Neural Networks
IBM	-	International Business Machine
w_{ij}	-	Vector of weights
x_i	-	Vector of inputs

b-	-	Bias
φ	-	Activation function
N	-	Neurons
$x_1 \ldots x_p$	-	Input variable values
u_j	-	Weighted sum
σ	-	Transfer function
h_j	-	Output values
O_j	-	Target output
d_j	-	Desired output
η	-	Learning rate
α	-	Momentum coefficient
$\psi(.)$	-	Wave function
f	-	Frequenct
t	-	Time
$\psi(t)$	-	Mother wavelet
C	-	Normalizing factor
$y_{high}[k]$	-	Outputs of high pass g
$y_{low}[k]$	-	Outputs of low pass h
θ_j	-	Bias for the j^{th} unit

LIST OF PUBLICATIONS

Proceedings:

(i) Ashikin Ali., Rozaida Ghazali & Mustafa Mat Deris, (2011, December). The wavelet multilayer perceptron for the prediction of earthquake time series data. In *Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services* (pp. 138-143). ACM.

(ii) Ashikin Ali., Rozaida Ghazali & Lokman Hakim Ismail (2012, August). The wavelet filtering in temperature time series prediction. In *Uncertainty Reasoning and Knowledge Engineering (URKE), 2012 2nd International Conference on* (pp. 153-157). IEEE.

(iii) Ashikin Ali, Rozaida Ghazali, Yana Mazwin Mohmad Hassim (2011, November). A Review on Wavelet Pre-processing for Time Series Data. In *Proceedings of the 2nd World Conference on Information Technology (WCIT)*. (pp. 16).

CHAPTER 1

INTRODUCTION

1.1 An Overview

A time series data is a set of observations made chronologically. The study of time series data is important since data is the source of information. This delivers the validation of theories and models as well as their enhancements (Ghosh & Raychaudhuri, 2007). Data analysis sometimes can emerge of a new theory or model. Thus, a physical time series data (such as astrophysical, geophysical, meteorological and etc.) may appear as an output of an experiment or it may come out as a signal from a dynamical system or it may contain some sociological, economic or biological information. Onwards, source of a time series data always expected to embed some amount of noise in it.Revision of such data in presence of noise often misleads to a clarification of the data.Hence, the need of developing an initial platform to denoise the data is extensively requisite.However, other than these methods there are few more techniques (Bar-Joseph, 2004; Zuur, Leno & Elphick, 2010; Goldstein, 2011; Morley & Adams, 2011; Ali, Ghazali & Deris, 2011; Qu & Chen, 2012; Azam & Mohsin, 2012) that have been discovered in many studies in order to overcome problems in handling time series data.

Commonly used feedforward Neural Network (NN), namely the Multilayer Perceptron (MLP) has exposed to be a promising predicting tool (Zhang *et. al.*, 2001; Chandrasekaran *et.al.*, 2010). No hesitation that MLP provides the capability and possibilities to predict the time series events. The consumption of MLP is to overwhelm the limitation of existing prediction model as the above mentioned reason. On the other hand, MLP embraces computationally intensive training algorithm and moderately slowlearning convergence (Wilamowski, 2010).

Therefore, this study aims to predict the physical time series which furnishes motivation to develop a modified model,An Improved MLP based on Wavelet Approach for Physical Time Series Prediction (W-MLP) by combining wavelet transform as data filtering element and afterwards the filtered data is then loaded into NN which inclusive of the backpropagation algorithm. The sole purpose of this model is to overcome the hitches in MLP and time series itself. Conversely, the experimental results have shown that wavelet transform can merge well with MLP in terms of prediction. This ability has proven in providing potential applications for the study related with physical time series prediction.

1.2 Problem Statements

The standard MLP have been facing convergence and predicting problem when deals with large network architecture and huge time series datasets (Izzeldin, Asirvadam & Saad, 2010; Karlaftis & Vlahogianni, 2011; Dauphin & Bengio, 2013). There are certain challenges faced by time series and the most common are the outliers and periodicities problem (West, 1996; Mukherjee, Osuna & Girosi, 1997; Brockwell, 2005; Box, Jenkins & Reinsel, 2011; Anderson, 2011). The existing studies dealt with these challenges whereby they tend to work in particularly with single univariate datasets, for instance earthquake dataset(Deka & Prahlada, 2012), temperature dataset (Sharma & Agarwal, 2012), evaporation dataset (Abghari *et. al.*, 2012), humidity dataset (Alsadi & Khatib, 2012) and wind direction (Colak, Sagiroglu & Yesilbudak, 2012).However, this study emphasizes to focus on physical time series data which inclusive of five (5) single univariate datasets namely earthquake, temperature, humidity, evaporation and wind direction. The motivation to choose univariate is based on the problems that exist in multivariate. Multivariate has more parameters than univariate ones. It is more complex and lengthier, susceptible to errors which then affect prediction. Beside, outliers can have a more serious effect on multivariate than one univariate forecasts. Moreover, it is easier to spot and control outliers in the univariate context.

Nevertheless, filtering a time series data is always an indispensable task to deal with. There are numbers of existing methods of filtering a time series data

(Huang *et. al.*, 1998; Brockwell & Davis, 2009; Wang *et. al.*, 2013). That is, the traditional 3-point or 5-point moving average method as an initial technique to smooth the data (Stafford, 2010). Empirical Mode Decomposition inclusive of Low Pass Filtering, High Pass Filtering and Band Pass Filtering (Wu & Norden, 2009). On the other hand, wavelet analysis is a popular filtering and pre-processing technique used to overcome noise, outliers and periodicities in time series data (Cheng, 2008; Marczak &Gomez, 2012).Haar (1909) was interested in finding a basis on a functional space similar to Fourier's basis in frequency space. In physics, wavelets were used in the characterisation of Brownian motion. This work led to some of the ideas used to construct wavelet bases. Wavelets were also used for analysis of coherent states of a particular quantum system. Finally, in the signal processing field, Mallat (1989) discovered that filter banks have important connections with wavelet basis functions.

Meanwhile, wavelets have penetrated into different fields, such as image processing (Richards, 2012), signal processing (Shu & Lei, 2011; Broughton & Bryan, 2011; Nixon & Aguado, 2012), medical science (Gharabli, 2009), biotechnology (Bessero *et. al.*, 2010). The ideas behind wavelets are becoming more significant in signal processing is that it can create a suitable representation of a signal, discard the least significant pieces of that representation and thus keep the original signal largely intact. These require transformation which can separate the important parts of the signal from less important parts. Therefore, this technique compromises on fast convergence of time series data prediction (Hsu, 2010).

Looking into this adequacies, it is essential to develop a W-MLPmodel that is capable in decomposing during the pre-processing, making it possible to distinguish rapidly between source of susceptibility and sources of resistance in physical time series. In this respect, NN particularly MLP algorithm is known for their remarkable ability to derive meaning from complicated or imprecise data that are too complex to be noticed by either humans or other computer techniques. Hence, this makes the wavelet technique to be very helpful in diagnosing the physical time series data.

.

1.3 Aim of the Study

This study aims to develop a model, namely W-MLP to predict the selected physical time series data and to reduce training time of standard NN models, whilst removing the outliers from the datasets.

1.4 Objectives of the Study

This study embarks on the following objectives:

(i) To propose a Wavelet-Multilayer Perceptron (W-MLP) which can reduce the prediction error and decrease the convergence time of ordinary Multilayer Perceptron (MLP).

(ii) To develop (i) for the simulation of physical time series.

(iii) To validate out-of-sample performance of (ii) with Multilayer Perceptron (MLP), High Pass Filter-MLP, Band Pass Filter-MLP and Low Pass Filter-MLP.

1.5 Scope of the Study

This research only focuses on the use of W-MLP on the physical time series data prediction and the results are compared to the MLP. The five network models, namely W-MLP, MLP, High Pass Filter-MLP, Band Pass Filter-MLP and Low Pass Filter-MLPwere trained with standard Backpropagation (BP) algorithm. W-MLPwas tested with the 5-years daily measurement of temperature, evaporation, humidity and wind direction in Batu Pahat region, ranging from 2005 to 2009 (Malaysian Meteorological Department, 2010) and 4 years daily measurement of earthquake in North California region, ranging from 1995 to 1998, taken from the Website of Northern California Earthquake Data Center(Northern California Earthquake Data Centre, 2010).

1.6 Significance of the Study

The W-MLP model can be helpful in predicting events dealt with physical time series data. Results from the simulations can be used to design a physical time series prediction tool. In addition, this study has potential in assisting the daily prediction event for Malaysian Meteorological Department (MMD) and Northern California Earthquake Data Center (NCEDC).

1.7 Thesis Outline

The rest of the dissertation is organised as follows: Chapter 2 focuses on pertinent background of backpropagation. The discussion then endures with corresponding approaches for time series prediction. Then the discussion continues on brief explanation of wavelet transform and filtering techniques.

Chapter 3 is the illustration of research methodology which is used to present the prediction model. This chapter continues with the discussion on the architecture of W-MLP towards the proposed model.Later, explanation on the implementation of the model is briefly written.

Chapter 4 of the thesis analyses the implementation of W-MLP. Based on the acquired results, a thorough analysis related to prediction and filtering is presented in Tables and graphs. The simulation results then compared with 3 different data filtering techniques namely Low Pass Filter, Band Pass Filter and High Pass Filter and MLP itself. Obtained results were analysed based on different parameters which have been used throughout the process. Chapter 5, concludes the thesis with the work done and some fruitful recommendations are given in order to expand the proposed network model in upcoming studies.

1.8 Chapter Summary

There are varieties of applications on time series prediction that has been developed in the past. Nevertheless, the limitations are still there. Therefore, improvement on time series data prediction eventually is an upcoming research domain. Thus, this drawback has led to focus this study on physical time series and developing an alternative predicting technique. The following chapter discusses the literature on the existing approaches related to time series, the hierarchy of the feedforward NN and filtering techniques.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter explores the dominant topics of the research such as Neural Networks, Multilayer Perceptron, Backpropagation, time series, Wavelet and its applications. In these recent years, a massive amount of literature have been written on the topic of Neural Networks (Smith, 1997), which Neural Networks are applied to such a wide variety of subjects (Arbib, 1995).Brief antiquities of Neural Networks have been written to give an indulgent of where the progression of Neural Networks started. Hence, a detailed review has been written for this study. This chapter also discussed the research works on topics related to this study in order to establish the need for the proposed work in this study.

2.2 Neural Network

An Artificial Neural Network (ANN), often just called a Neural Network (NN), is a mathematical model or computational model based on biological neurons. In other words, it is an emulation of biological neural system. It consists of an interconnected group of artificial neurons and process information using a connectionist approach to computation (Yashpal, 2009). In most cases, an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase (Yashpal, 2009).

ANN can also be defined as model reasoning based on the human brain. The brain consists of a densely interconnected set of nerve cells or basic information processing units, called neurons. The human brain incorporates nearly 10 billion neurons and 60 trillion connections between them (Shepherd &Koch, 1990). Using multiple neurons simultaneously, the brain can perform its function much faster than the fastest computers in existence today.

Although each neuron has a very simple structure, such element constitutes a tremendous processing power. A neuron consists of a cell body, soma, a number of fibers called dendrites and a single long fiber called the axon. Dendrites branch into a network around the soma, the axon stretches out the dendrites and somas of other neurons. A schematic drawing of a biological neuron is shown in Figure 2.1.

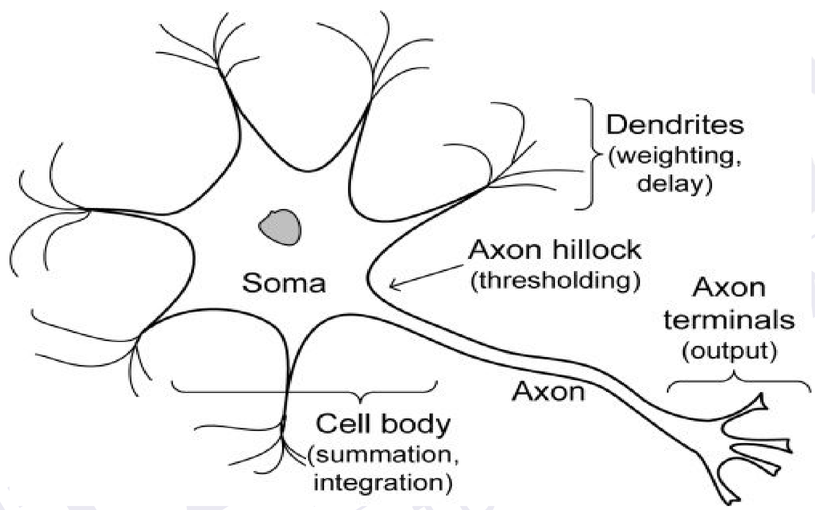

Figure 2.1: Schematic Drawing of Biological Neuron (Kravtsov*et. al.*, 2011)

Signals are propagated from one neuron to another by complex electro-chemical reactions. When the potential reaches its threshold, an electrical pulse is sent down through the axon. The pulse spreads out and eventually reaches synapses cause them to increase or decrease their potential. In response to the simulation pattern, neurons demonstrate long-term changes in the strength of their connections. Neurons also can form new connections with other neurons. Even entire collections of neurons may sometimes migrate from one place to another.

The human brain can be considered as a highly complex, nonlinear and parallel information processing. Information is stored and processed in an NN simultaneously throughout the whole network, rather than at specific locations. Connections between neurons leading to the right answer are strengthened while those leading to the wrong answer weaken. As a result, NN have the ability to learn through experience. Learning is fundamental and essential characteristic of biological neural networks. The ease and naturalness which they can learn lead to attempts to emulate a biological NNin a computer. In addition, the present ANN resembles the human brain much as a paper plane resembles a supersonic jet, it is a big step forward. Nevertheless, ANN is capable of learning, in which they use experience to improve their performance. When exposed to sufficient number of samples, ANN can generalize to others they have not yet encountered. ANN can also recognize and written characters (Perwej & Chaturvedi, 2012), identify words in human speech and detect explosives (McGarry, 1999). Moreover, ANN can observe patterns that human experts fail to recognize (Jain *et al.*, 2000).

2.3 Multilayer Perceptrons (MLP)

A single perceptron is not very useful because of its limited mapping ability. This is due to the fact that it consists of a single neuron with adjustable synaptic weights and bias and only capable to represent an oriented ridge-like function, no matter what activation function is used (Haykins, 1999). Meanwhile, a Multilayer Perceptron (MLP) consists of a set of source nodes forming the input layer, one or more hidden layers of computation nodes, and an output layer of nodes. The input signal in MLP propagates through the network layer-by-layer (Haykins, 1998). Mathematically, MLP can be written as below:

$$y = \varphi\left(\sum_{j=1}^{J} w_{jk}\varphi\left(\sum_{i=1}^{N} w_{ij} \cdot x_i + w_{oj}\right)w_{ok}\right), \qquad (2.1)$$

where w_{ij} denotes the vector of weights, x_i is the vector of inputs, w_{oj} is the bias of each hidden nodes, w_{ok} bias of output and φ is the activation function. The activation function acts as a squashing function that prevents accelerating growth throughout the network. An acceptable range of output is usually between [0, 1] or [-1, 1] (Rojas, 1996). This value is a function of the weighted inputs of the corresponding node.Figure 2.2 illustrates an MLP with three layers of neurons.

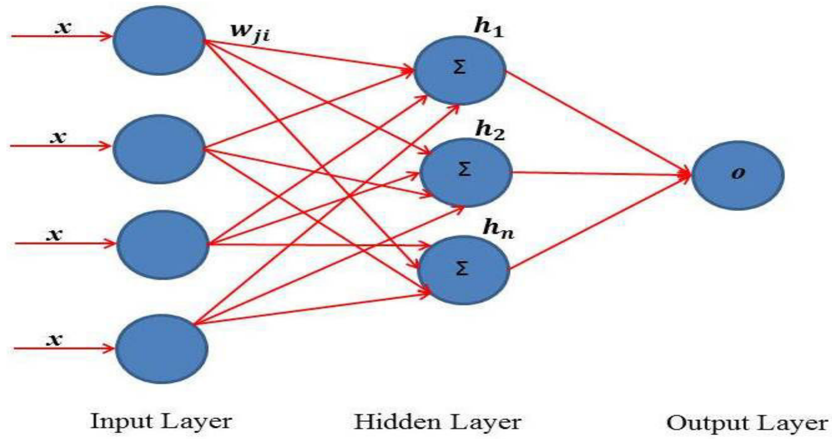

Figure 2.2: Diagram of Multilayer Perceptron

From Figure 2.2, it shows that the network has an input layer (on the left) with four neurons, one hidden layer (in the middle) with three neurons, and an output layer (on the right) with one neuron. Each neuron in the input layer represents each input variable. In the case of categorical variables, *N* neurons are used to represent the *N* categories of the variable.

Input Layer— a vector of input variable values ($x_1...x_p$) is presented to the input layer. The input layer distributes the values to each of the neurons in the hidden layer. In addition to the predictor variables, there is a constant input of 1.0, called the *bias* that is fed to each of the hidden node; the bias is multiplied by a weight and added to the sum going into the neuron.

Hidden Layer — arriving at a neuron in the hidden layer, the value from each input neuron is multiplied by a weight (w_{ji}), and the resulting weighted values are added together producing a combined value u_j. The weighted sum (u_j) is fed into a transfer function, σ, which outputs a value h_j. The outputs from the hidden layer are distributed to the output layer.

Output Layer — arriving at a neuron in the output layer, the value from each hidden layer neuron is multiplied by a weight (w_{kj}), and the resulting weighted values are added together producing a combined value v_j. The weighted sum (v_j) is fed into a transfer function, σ, which outputs a value y_k. The *y* values are the outputs of the network.

For classification problems with categorical target variables, there are *N* neurons in the output layer producing *N* values, one for each of the *N* categories of the target variable. The network is usually used in supervised learning problems, in which the training set of input-output pairs and the network must learn to model the dependency between them.

2.4 The Backpropagation Gradient Descent Algorithm

The backpropagation algorithm (Rumelhart & McClelland, 1986) is used in layered feed-forward MLP. The backpropagation algorithm uses supervised learning, where the algorithm is provided with the inputs and outputs which the network has to compute and then the error is calculated (Gershenson, 2003). The idea of the backpropagation algorithm is to reduce this error, until the MLP learns the training data. The training begins with random weights, and the goal is to adjust them so that the error will be minimal.

The weighted sum of a neuron is written as:

$$A_j(x, w) = \sum_{i-0}^{n} X_i W_{ji}, \tag{2.2}$$

where the sum of input $X_i$is multiplied by their respective weights, W$_{ji}$. The activation depends only on the inputs and the weights. If the output function would be the identity, then the neuron would be called linear. The most used output function is sigmoid function (Tommiska, 2003):

$$O_j(x, w) = \frac{1}{1 + e^{-A}(x, w)} \tag{2.3}$$

The sigmoid function is very close to one for large positive numbers and very close to zero for large negative numbers. This allows a smooth transition between the low and high output of the neuron. The output depends only in the activation, which in turn depends on the values of the inputs and their respective weights. The goal of the training process is to obtain a desired output when certain inputs are given. Since the error is the difference between the actual and desired output, the error depends on the weights and preferred to be adjusted in order to minimize the error. The error function for the output of each neuron can be defined as:

$$E_j(x, w, d) = \left(O_j(x, w) - d_j\right)^2 \tag{2.4}$$

The output will be positive and the desired target will be greater if the difference is big and lesser if the difference is small. The error of the network will simply be the sum of the errors of all the neurons in the output layer:

$$E\,(x, w, d) = \sum_j \left(O_j(x, w) - d_j\right)^2 \tag{2.5}$$

where O_j is the target output and d_j is the target or desired output. After finding this, the weights can be adjusted using the method of gradient descent:

$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} \tag{2.6}$$

This equation can be inferred in the following way: the adjustment of each weight (Δw_{ji}) will be the negative of a constant eta(η), where η is the learning rate. Multiplied by the dependence of the previous weight on the error of the network, which is derivative of E in respect to w_{ji}. The size of the adjustment will depend on η, and on the contribution of the weight to the error of the function. This is, if the weight contributes a lot to the error, the adjustment will be greater than if it contributes in a smaller amount. Equation (2.6) is used until appropriate weights with minimal error founded.

Henceforth, derivative of E in respect to w_{ji} discovered. This is the goal of the backpropagation algorithm, since the backwards need to be achieved. First, calculate the error depends on the output, which is the derivate of E in respect to O_j from Equation (2.4).

$$\frac{\partial E}{\partial O_j} = 2\left(O_j - d_j\right) \tag{2.7}$$

The reliance of the output on the activation depends on the weights from Equation (2.2) and Equation (2.3). Can be seen that from Equation (2.7) and Equation (2.8):

$$\frac{\partial O_j}{\partial w_{ji}} = \frac{\partial O_j}{\partial A_j}\frac{\partial A_j}{\partial w_{ji}} = O_j\left(1 - O_j\right)x_i \tag{2.8}$$

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial O_j}\frac{\partial O_j}{\partial w_{ji}} = 2\ \left(O_j - d_j\right)O_j\left(1 - O_j\right)x_i \tag{2.9}$$

The adjustment to each weight will begin from Equation (2.6) and Equation (2.9).

$$\Delta w_{ji} = -2\eta (O_j - d_j) O_j (1 - O_j) x_i \tag{2.10}$$

Equation (2.10) can be used as it is for training ANN with two layers. For training the network with one more layer, some considerations are needed particularly on training time which can be affected by the architecture of the network. For practical reasons, ANNs implementing the backpropagation algorithm do not have too many layers, since the time for training the networks grows exponentially (Gershenson, 2003).

2.5 Filtering Techniques

Analysis of data is a very important task since it is the source of information which will be fed into the certain techniques, namely, classification or prediction. The presence of noise often leads to a wrong interpretation of the data. Therefore, an initial platform is needed for data denoising process. Filtering can be one of denoising platform for time series data and it is an indispensable task to deal with (Ghosh & Raychaudhuri, 2007). Filtering is the process of defining, detecting and correcting errors in given data, in order to minimize the impact of errors in input data on succeeding analyses (Wedin *et al.*, 2008).There are several time series filters commonly used in research to separate the behavior of the time series. These techniques can usually be expressed using some of the commonly used filtering techniques namely, low-pass filter, high-pass filter, band-pass filter which are empirical mode decomposition, and chief among all is wavelet (Baum, 2006).

2.5.1 Low – Pass Filter (LPF)

Low – Pass Filter (LPF) is an electronic filter that passes low frequency signals but attenuates signals with frequencies higher than the cutoff frequency (Thomas *et al.*, 2000).The actual amount of attenuation for each frequency varies from filter to filter. It is sometimes called a high cut filter. A low pass filter is the opposite of a high pass filter. A band filter is a combination of a low pass and high pass.

.

2.5.2 High – Pass Filter (HPF)

High – Pass Filter (HPF) is an electric filter that passes high frequency signals but attenuates signals with frequencies lower than the cutoff frequency. A high pass filter is usually modeled as a linear time invariant system. It is sometimes called a low cut filter or bass cut filter (John, 1998). It can also be used in conjunction with a low pass filter to make a band pass filter.

2.5.3 Band – Pass Filter (BPF)

A Band – Pass Filter (BPF) is a device that passes frequencies within a certain range and rejects frequencies outside that range. Bandpass is an adjective that describes a type of filter or filtering process. An analogue electronic band pass filter is a resistor inductor capacitor circuit. These filters can also be created by combining a low pass filter with a high pass filter (Anderson *et al.*,2012).However, among all the three filtering techniques, the wavelet approach has shown some advantages over the conventional filtering techniques.

2.6 Wavelet

Wavelets are a class of functions to localize a given functions in both position and scaling (Daubechies, 2006). Wavelets are used in application such as signal processing, image processing and time series analysis (Graps, 1995; Sifuzzaman *et al.*, 2009; Starck *et al.*, 2010; Paris *et al*, 2011). Wavelets form the basis of the wavelet transforms which "cuts up data of functions or operators into different frequency components and then studies each component with a resolution matched to its scale" (Calderbank *et al.*, 1998).

A wavelet transform is a small wave function, usually denoted by ψ **(.)**. A small wave grows and decays in a finite time period, as opposed to a large wave, such as sine wave, which grows and decays repeatedly over an infinite time period. A function ψ **(.)** which is defined over the real axis **(-∞, ∞)** can be classed as a wavelet by satisfying the following three (3) properties:

(1) The integral of ψ (.) is zero:

$$\int_{-\infty}^{\infty} \psi(t)du = 0 \tag{2.11}$$

(2) The integral of the square of ψ (.) is unity:

$$\int_{-\infty}^{\infty} \psi^2(t)du = 1 \tag{2.12}$$

(3) Admissibility Condition:

$$C_\psi \equiv \int_0^\infty \frac{|\psi(a)|^2}{b} df \text{ Satisfies } 0 < C_\psi < \infty \tag{2.13}$$

where t in Equation (2.11) and Equation (2.12) denotes time, a and b in Equation (2.13) denote dilation and translation and C denotes the normalizing factor.

There are a few types of wavelet transforms. Among them are Fourier Transform, Multiresolution Discrete Wavelet Transform, Continuous Wavelet Transform, and Discrete Wavelet Transform. However, the most commonly used in time series are Continuous Wavelet Transform and Discrete Wavelet Transform (Polikar, 2001; Addison, 2010; Chaovalit *et al.*, 2011).

There are two (2) main types of wavelet transforms: Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). CWT is designed to work with functions defined over the whole real axis. Meanwhile, DWT deals with functions that are defined over a range of integers (usually $t = 1,2,\ldots,N-1$, where N denotes the number of values in the time series).

2.6.1 Continuous Wavelet Transform (CWT)

A CWT (Polikar, 2001) is designed to work with functions defined over the whole real axis. It is used to divide a continuous-time function into wavelets. Unlike Fourier Transform, the Continuous Wavelet Transform possesses the ability to construct a time frequency representation of a signal that offers very good time and frequency localisation. A mathematical representation of the Fourier Transform is as:

$$F(w) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt \tag{2.14}$$

However, the sum over all time of the signal *f(t)*, where *f* denotes frequency and *t* denotes time, multiplied by a complex exponential, and the result is the Fourier coefficients F. Meanwhile, the CWT is the sum over all time of the signal, multiplied by scaled and shifted versions of the wavelet function as given below:

$$C(a,b) = \int_{-\infty}^{+\infty} s(t)\psi_{a,b}(t)dt, \psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}) \tag{2.15}$$

where *s(t)* is the signal, *a* is the scale and *b* is the shifting. Here $\psi(t)$ is the mother wavelet, while $\psi_{a,b}(t)$ is the scaled and the shifted one. The result *C* is wavelet coefficients.

2.6.2 Discrete Wavelet Transform (DWT)

Discrete Wavelet Transform (Polikar, 2004) deals with functions that are defined over a range of integers, usually $t = 1,2,\ldots,N-1$, where *N* denotes the number of values in the time series. The wavelet series is just a sampled version of CWT and its computation may consume significant amount of time and resources, depending on the resolution required. The DWT which is based on sub-band coding is found to yield a fast computation of wavelet transform. It is easy to implement and reduces the computation time and resources required (Letelier & Weber, 2000). Similar work was done in speech signal coding which was named as sub-band coding (Vetterli &

Kovačević, 1995). In recent years, a technique similar to sub-band coding was developed which was named as pyramidal coding (Polikar, 2004). Later, many improvements were made to these coding schemes which resulted in efficient multi-resolution analysis schemes. Figure 2.3 illustrates the procedure, where $x[n]$ is the original signal to be decomposed, and $h[n]$ represents low pass, $g[n]$ represents high pass filters, respectively. The bandwidth of the signal at every level is marked on the figure below as f :

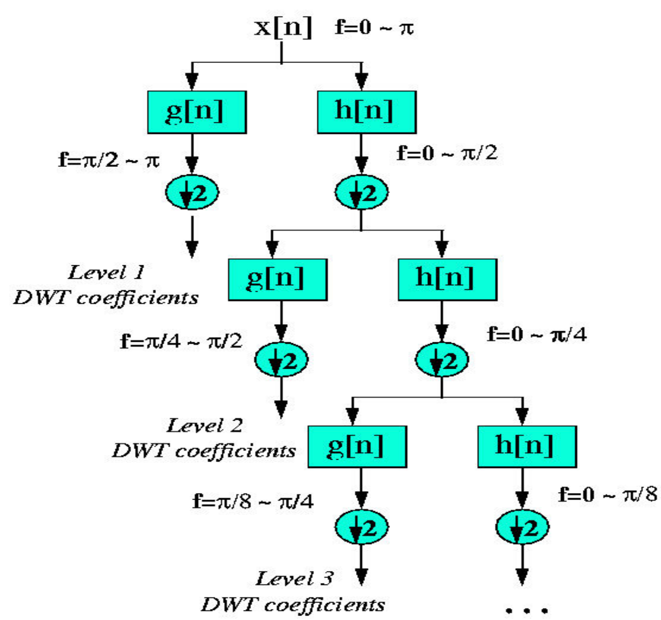

Figure 2.3: The Illustration of Sub-band Coding (Polikar, 2004)

In CWT, the signals are analyzed using a set of basic functions which relate to each other by simple scaling and translation. In the case of DWT, a time scale representation of the digital signal is obtained using digital filtering techniques. The signal to be analyzed is passed through filters with different cut off frequencies at different scales. The DWT employs two sets of functions, called scaling functions and wavelet functions, which are associated with low pass and high pass filters, respectively. They can be mathematically expressed as below:

$$y_{high}[k] = \sum_{n} x[n] \bullet g[2k-n] \tag{2.16}$$

$$y_{low}[k] = \sum_{n} x[n] \bullet h[2k-n] \tag{2.17}$$

where $y_{high}[k]$ and $y_{low}[k]$ are the outputs of the high pass, *g,* and low pass, *h,* filters after sub-sampling by 2.

2.7 Time Series

Essentially, time series can be defined as a sequence of numbers collected at regular intervals over a period of time (Ali *et al.*, 2011). There are several basic types of time series model, namely, Moving Average (MA), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), and Exponential Smoothing. ARMA models are typically applied to auto correlated time series data, while ARIMA model is a generalization of an ARMA model (Zhang, 2003). These models are fitted to time series data either to better understand the data or to predict/forecast future points in the series. Meanwhile, exponential smoothing is a technique that can be applied to time series data, either to produce smooth data for presentation, or to make forecasts. Eventually, the time series data themselves are a sequence of observations. The observed phenomenon may be an essentially random process, or it may be an orderly but noisy process. Whereas in the simple moving average the past observations are weighted equally, exponential smoothing assigns exponentially decreasing weights over time.

Time series refers to problems in which observations are collected at regular time intervals and there are correlations among successive observations. Mostly the time series applications cover virtually all areas of statistics but some of the most important include economic and financial time series, and many areas of environmental or ecological data (Chatfield, 2003; Box *et al.*, 2011; Anderson, 2011; Murphy *et al.*, 2012). Time series can be broadly categorized into three (3), namely continuous time series, interval time series and momentary time series. Continuous time series are often continuously recorded, either on the record sheet or data logger, where typically records the data either at fixed time intervals or after a certain change in the value has taken place. Meanwhile, the physical time series data comes under the umbrella of continuous time series dealt with five data types, namely hydrology, earth sciences, astronomy, oceanography and marine biology. An interval time series does not contain values for points in time but rather for particular intervals of time, these time intervals can be equidistantly or randomly dispersed in time. While, the momentary time series is the rarest form of time series, that defines a discrete set of point in time, thus it does not contain any information for the time between these points (Chatfield, 2003). The next sub-section briefly discusses the physical time series data with the respective data types.

2.7.1 Physical Time Series Data

Physical time series data consist of five data types, namely Hydrology, Earth Sciences, Astronomy, Oceanography and Marine Biology (Favali & Beranzoli, 2006; Kantardzic, 2011). Basically, any of the natural sciences that deal with nonliving materials is categorized as physical science which relates to physical time series. Hydrology is the study of the movement, distribution and quality of water on earth, including the hydrologic cycle, water resources and environmental watershed sustainability. The hydrology data consist of certain data fields namely temperature, evaporation, humidity and wind direction which is some essential elements that are needed in hydrology studies (Weber & Stewart, 2004; Karamouz *et al.*, 2012). Meanwhile, earth science which is also known as geosciences is an embracing term for the science related to the planet Earth. The formal discipline of Earth sciences may include the study of atmosphere, oceans, biosphere, as well as the solid earth. Typically earth scientists have used certain tools from varies of fields to build a quantitative understanding of how earth system works and how it evolves to its current state. The field also includes studies of earthquake effects, such as tsunamis as well as diverse seismic sources such as volcanic, tectonic, oceanic, atmospheric and artificial processes, such as explosions. Besides, astronomy is a natural sciences that pact with the study of moon, planets, stars, galaxies that originated outside the atmosphere of earth. Furthermore, oceanography is a study from a division of earth science that studies the ocean. It also covers topics including marine organisms and ecosystem dynamics. Then, marine biology is a precise lesson of organisms in the ocean or marine bodies of water.

However, this study will only be focus on 2 of physical time series which is hydrology time series data and earth sciences time series data which emphasizes on five datasets, four from the hydrology and one from the earth science. Whereas, this research will be focusing on seismology, it is the scientific study of earthquakes and the propagation of elastic waves and through earth.

2.7.2 Properties of Physical Time Series Data

Time Series occur in many different fields, economic time series, sales and marketing and physical time series. Physical time series is related to the physical science, a study which evolves nature science and its phenomena. As in most physical time series analysis, it is presumed that the data consist of random noise which usually makes the pattern difficult to identify. Physical time series analysis techniques involve some practice of filtering out noise in order to make the pattern more salient. The patterns can be described in terms of two basic classes of components: trend and seasonality (Wang & Wu, 2009). There are no proven methods to identify trend components in physical time series data, however, as long as the trend is consistently increasing or decreasing that part of data analysis is typically not difficult. If the data contain considerable error, then the first step in the process of trend identification is smoothing. Smoothing is merely used to apprehend important data while leaving out noise (Kantardzic, 2011).

The traditional techniques used for time series forecasting are Autoregressive (AR) models, Autoregressive Moving Average (ARMA) models, Autoregressive Integrated Moving Average (ARIMA) models, linear regression and exponential smoothing. None of these techniques are completely pleasing due to the nonlinear nature of most of the ordinary arising time series (De Gooijer & Hyndman, 2006; Khashei & Bijari, 2011). Other more advanced method such as neural networks has been used effectively for time series predictions (Ardalani Farsa & Zolfaghari, 2010). Literally, there are many applications and techniques has been applied which is related to time series (Kantz & Schreiber, 2003; Honaker & King, 2010;

Ratanamahatana *et al.*, 2010). Applications using wavelet preprocessing techniques and neural network are overviewed in the remainder sections.

2.8 An Overview of Wavelet Pre-processing using Time Series Data

Pre-processing is a process performed on raw data to prepare it for another processing procedure, where the data turned into easier and effective format (Cannas *et al.*, 2006). Consequently, various strategies have been used for filtering components of time series. In particular, wavelets have been applied in many fields and widely used for decomposing time series data (Ahmad, 2005). Wavelets are robust parameter free tools that cut up data to different frequency components and study each component with a resolution matched to its scale (Daubechies, 1992). Therefore, in this section, several studies that have applied the wavelet pre-processing technique on time series and outliers problems are briefly discussed.

In the study done by Mukta and Rohit (2013), comparative analysis of wavelet filters on hybrid transform domain image steganography techniques were taken into significance. Steganography has been an important area of research in recent years involving a number of applications. Image steganography is the art of hiding secret information into a cover image. In this study, Discrete Wavelet Transform (DWT) is used to transform cover image from spatial domain to frequency domain. Different wavelet filters can be used to embed secret image in these frequency components. Hybrid transform domain techniques for different wavelet filters to embed secret image into cover image were compared in this research. Peak Signal to Noise Ratio (PSNR) algorithm is compared, where it is a measure of the differences between the cover image and stego image. In future, researchers could apply the technique to different level and type of images in order to concrete their proposed method.

Zainuddin *et al.* (2012), studied on the use of wavelet neural networks (WNNs) in the task of epileptic seizure detection from electroencephalography (EEG) signals. This work investigates on the feasibility and effectiveness of WNN in the charge of epileptic seizure detection. The EEG was first pre-processed using Discrete Wavelet Transform (DWT). Followed by feature selection stage, two sets of four representative summary statistics were computed. The cross comparison shows that the classification accuracy achieved by WNNs was comparable to those of other artificial intelligence-based classifiers. Nevertheless, it is pertinent to note that experimental results from scientific and engineering applications are always subjected to outliers. On the other hand, to obtain a better accuracy more trial and errors simulations should be done in future.

Meantime, a study on time series modeling of river flow using wavelet neural network was familiarized by Krishna *et al.* (2011). A hybrid model with the combination of wavelet and artificial neural network (ANN) called wavelet neural network was proposed and applied for time series modeling of river flow. The observed time series are decomposed into sub-series using discrete wavelet transform and then appropriate sub-series is used as inputs to the neural network for forecasting hydrological variables. It is required to choose a proper resolution in order to have a worthy forecasting activity.

In the studies by Ocak (2009), automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy was introduced. It has successfully given 96% of seizure detection accuracy. However, the normal EEG without DWT as preprocessing step, where the detection rate was reduced to 73%.The new scheme was further amended by surrogate data analysis.

However, a new wavelet model called Modified Mexican Hat Wavelet was introduced by Benbrahim (2005).They essentially proposed a new algorithm based on the random projection and the principal component analysis of seismic signals. Thus, this new modified Mexican Hat Wavelet and the new proposed algorithm at certain point of architecture, it gives bad results due to the weak number of hidden nodes. Therefore, a thorough modification and suitable architecture is needed to obtain the best results.

In the work done by Zhang (2001), the combination of shift invariant wavelet transform pre-processing and neural network prediction models trained using Bayesian techniques at the different levels of wavelet scale for financial forecasting is introduced. However, additional research has to be done to overcome the outliers and improper forecasting by similar hybrid. Meanwhile, Cannas *et al.* (2006) investigates the effect of data pre-processing model performance using CWT, DWT and data partitioning. It is proven that using pre-processed data able to obtain best results. The study however still need proper division of data points in order to get best decomposing levels to get even better results with best accuracy.

Besides that, Agrawal (1995) introduced a fast similarity search in the presence of noise, scaling and translation in time series databases. They present fast search techniques to discover all similar sequences in a set of sequence. Somehow, this study should have extension on trial and error method using different data to ensure that the introduced technique is applicable with any data in order to discover the similar sequences in a time series data. In short, from all the studies shows that time series predictions can also be answered using wavelet pre-processing technique which helps a lot in term of outliers, periodicities and training time.

However, recently there has been increased interest in multiresolution decomposition techniques like the wavelet transform to deal with complex relationships in non-stationary time series (Gencay, Selcuk & Whitcher, 2002). The wavelet can produce a good local representation of a signal in both time and frequency domain and is not restrained by the assumption of stationary (Mallat, 1989). Besides, the wavelet approach has formalized old notions of decomposing a time series into trend (Ramsay, 1999). Motivated by the spatial frequency resolution property of the wavelet transform, several schemes have been developed (Aussem & Murtagh, 1997), which combines wavelet analysis machine learning approaches like neural networks for time series prediction.

Chan and Fu (1999), worked on efficient time series matching by wavelets. Haar wavelet Transform has been selected for the time series indexing. There are few contributions were mentioned, where Euclidean distance is preserved in the Haar wavelet transformed domain and no false dismissal occurs. This has proven that Haar wavelet transform can outperform discrete fourier transform through experiments, a new similarity model is suggested to accommodate vertical shift of time series. Two phase method is proposed for efficient nearest *n*-neighbor query in time series databases. However, this property has only been proven with the Haar wavelets. It would be interesting if it could be applied with different kinds of wavelets to different kinds of data series.

Meanwhile, Popoola and Khurshid (2006) have introduced the testing suitability of wavelet preprocessing for Takagi-Sugeno-Kang (TSK) fuzzy models which is an additive rule models introduced by Takagi, Sugeno and Kang in 1984. In this study, the researchers proposed a methodology that uses formal hypothesis testing to determine whether having wavelet preprocessing in prior will improve forecasting performance or not. The method evaluated on ten economic time series, and compared variance profiles of each time series with the corresponding forecast performance of fuzzy models built from raw and wavelet processed data. Somehow, for further revisions, the proposed model is recommended to be evaluated with synthetic time series with known variance characteristics and much longer real world time series data.

Huether, Gustafson & Broussard (2001) acquaint with Wavelet Preprocessing for High Range Resolution Radar (HRR) classification. In the study, a general wavelet denoising approach can overcome the HRR classifying measurements has been initiated. By choosing the best decomposition level gives the best accuracy to the results. In future, ought to do more proper degradation to adjust the denoising parameters which will be a consideration in the preprocessing part.

2.9 Application of Neural Network using Time Series Data

A neural network is a processing device, either an algorithm or actual hardware whose design was motivated by the design and functioning of human brains and components thereof. There are many types of neural networks, each of which has different strengths particular to their applications. This section attempts to compile a list of previous research on neural network, particularly applied to time series.

Gheyas & Smit (2009) proposed a neural network approach to time series forecasting. In their work they introduced new improved algorithm based on Generalized Regression Neural Networks (GRNN) which ensemble to the forecasting of time series and future volatility. This approach is proposed to overcome the lagged variables, autocorrelation and non-stationary which have been the major characteristics that distinguish time series data from spatial data. However, they face a predicament when applying the GRNN to the time series forecasting task. If provide only the most recent past value, the GRNN generated the smallest forecasting error but does not accurately forecast the correct direction of change.

Financial time series forecasting by neural network using conjugate gradient learning algorithm and multiple linear regression weight initialization successfully applied to the time series forecasting. A comparison was made between two learning algorithms and two weight initializations to find that neural network can model the time series satisfactorily, regardless which learning algorithm and weight

REFERENCES

Abghari, H., Ahmadi, H., Besharat, S., & Rezaverdinejad, V. (2012). Prediction of Daily Pan Evaporation using Wavelet Neural Networks. *Water resources management*, *26*(12), 3639-3652.

Abonyi, J., Feil, B., & Abraham, A. (2005). Computational intelligence in data mining. Informatica, 29(1), 3-12.

Addison, P. S. (2010). *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*. Taylor & Francis.

Ahmad, S., Popoola, A., & Ahmad, K. (2005). Wavelet-based multiresolution forecasting. University of Surrey, Technical Report.

Al-Gharabli, S. I. (2009). Determination of Glucose Concentration in Aqueous Solution Using ATR-WT-IR Technique. *Sensors*, *9*(8), 6254-6260.

Ali, A., Ghazali, R., & Deris, M. M. (2011, December). The wavelet multilayer perceptron for the prediction of earthquake time series data. In *Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services* (pp. 138-143). ACM.

Ali, A., Ghazali, R., & Ismail, L. H. (2012, August). The wavelet filtering in temperature time series prediction. In *Uncertainty Reasoning and Knowledge Engineering (URKE), 2012 2nd International Conference on* (pp. 153-157). IEEE.

AlSadi, S., & Khatib, T. (2012). Modeling of relative humidity using artificial neural network. *Journal of Asian Scientific Research*, *2*(2), 81-86.

Amjady, N., & Keynia, F. (2009). Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. *Energy*,*34*(1), 46-57.

Anderson, B. D., & Moore, J. B. (2012). *Optimal filtering*. Dover Publications.com

Anderson, T. W. (2011). *The statistical analysis of time series* (Vol. 19). Wiley.

Arbib, M. A. (2003). *The handbook of brain theory and neural networks*. Bradford Book.

Ardalani-Farsa, M., & Zolfaghari, S. (2010). Chaotic time series prediction with residual analysis method using hybrid Elman–NARX neural networks. *Neurocomputing*, *73*(13), 2540-2553.

Aussem, A., & Murtagh, F. (1997). Combining neural network forecasts on wavelet-transformed time series. Connection Science, 9(1), 113-122.

Azam, F., & Mohsin, S. (2012, December). Agent Based Prediction of Seismic Time Series Data. In *Frontiers of Information Technology (FIT), 2012 10th International Conference on* (pp. 269-274). IEEE.

Bar-Joseph, Z. (2004). Analyzing time series gene expression data.*Bioinformatics*, *20*(16), 2493-2503.

Benbrahim, M., Benjelloun, K., Ibenbrahim, A., Kasmi, M., & Ardil, E. (2007, January). A new approaches for seismic signals discrimination. In Proceedings of World Academy of Science, Engineering and Technology (Vol. 21).

Bowden, G. J., Maier, H. R., & Dandy, G. C. (2012). Real-time deployment of artificial neural network forecasting models: Understanding the range of applicability. Water Resources Research, 48(10), W10549

Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2013). *Time series analysis: forecasting and control*. Wiley. com.

Broughton, S. A., & Bryan, K. M. (2011). *Discrete Fourier analysis and wavelets: applications to signal and image processing*. Wiley. com.

Brockwell, P. J. (2005). *Time Series Analysis*. John Wiley & Sons, Ltd.

Brockwell, P. J., & Davis, R. A. (2009). *Time series: theory and methods*. Springer.

Broughton, S. A., & Bryan, K. M. (2011). *Discrete Fourier analysis and wavelets: applications to signal and image processing*. Wiley-Interscience.

Cannas, B., Fanni, A., See, L., & Sias, G. (2006). Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. Physics and Chemistry of the Earth, Parts A/B/C, 31(18), 1164-1171.

Calderbank, A. R., Daubechies, I., Sweldens, W., & Yeo, B. L. (1998). Wavelet transforms that map integers to integers. Applied and computational harmonic analysis, 5(3), 332-369.

Chandrasekaran, M., Muralidhar, M., Krishna, C. M., & Dixit, U. S. (2010). Application of soft computing techniques in machining performance prediction and optimization: a literature review. *The International Journal of Advanced Manufacturing Technology*, *46*(5), 445-464.

Chan, K. P., & Fu, A. W. C. (1999, March). Efficient time series matching by wavelets. In Data Engineering, 1999. Proceedings., 15th International Conference on (pp. 126-133). IEEE.

Chan, M. C., Wong, C. C., & Lam, C. C. (2000). Financial time series forecasting by neural network using conjugate gradient learning algorithm and multiple linear regression weight initialization. In Computing in Economics and Finance (Vol. 61).

Chaovalit, P., Gangopadhyay, A., Karabatis, G., & Chen, Z. (2011). Discrete wavelet transform-based time series analysis and mining. *ACM Computing Surveys (CSUR)*, *43*(2), 6.

Chatfield, C. (2003). The analysis of time series: an introduction (Vol. 59). CRC Press.

Cheng, K. O. (2008). *Pattern recognition techniques for texture retrieval and gene expression data analysis*. The Hong Kong Polytechnic University: Ph.D. Thesis.

Colak, I., Sagiroglu, S., & Yesilbudak, M. (2012). Data mining and wind power prediction: A literature review. *Renewable Energy*.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. Mathematics of control, signals and systems, 2(4), 303-314.

Daubechies, I. (1992). Ten lectures on wavelets (Vol. 61, pp. 198-202). Philadelphia: Society for industrial and applied mathematics.

Dauphin, Y. N., & Bengio, Y. (2013). *Big Neural Networks Waste Capacity*. Retrieved from: www.library.cornell.edu/

De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting.*International Journal of Forecasting*, *22*(3), 443-473.

Deka, P. C., & Prahlada, R. (2012). Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time. *Ocean Engineering*, *43*, 32-42.

Demirel, H., & Anbarjafari, G. (2010). Satellite image resolution enhancement using complex wavelet transform. *Geoscience and Remote Sensing Letters, IEEE*, *7*(1), 123-126.

Dinesh, K., Kumar, S. S., & Daniel, P. (2012). Color Image and Video Compression Based on Direction Adaptive Partitioned Discrete Wavelet Transform. *Research Journal of Applied Sciences*,4.

Favali, P., & Beranzoli, L. (2006). Seafloor observatory science: a review. Annals of Geophysics, 49(2-3).

Fidele, B., Cheeneebash, J., Gopaul, A., & Goorah, S. S. (2009). Artificial neural network as a clinical decision-supporting tool to predict cardiovascular disease. Trends in Applied Sciences Research, 4(1), 36-46.

George, T., & Thomas, T. (2010). Discrete wavelet transform de-noising in eukaryotic gene splicing. *BMC bioinformatics*, *11*(Suppl 1), S50.

Gençay, R., Selçuk, F., & Whitcher, B. An Introduction to Wavelets and Other Filtering Methods in Finance and Economics. 2002.

Gershenson, C. (2003). *Artificial neural networks for beginners*. Retrived from: arXiv.org.

Ghazali, R., Hussain, A.,El-Deredy, W., "Application of Ridge Polynomial Neural Networks to Financial Time Series Prediction," in *Proceedings of the International Joint Conference on Neural Networks, IJCNN 2006*,Vancouver, BC, 2006, pp. 913-920.

Gheyas, I. A., & Smith, L. S. (2009). A Neural Network Approach to Time Series Forecasting. In Proceedings of the World Congress on Engineering (Vol. 2, pp. 1-3).

Ghosh, K., & Raychaudhuri, P. (2007). An Adaptive Approach to Filter a Time Series Data. Retrieved from: arXi.org.

Goel, M., & Goel, R. (2013). Comparative Analysis of Wavelet Filters on Hybrid Transform Domain Image Steganography Techniques. *International Journal*,*3*(8).

Goldstein, H. (2011). *Multilevel statistical models*. Retrieved from: www.cmm.bris.ac.uk

Gottlieb, I., Miller, J. M., Arbab-Zadeh, A., Dewey, M., Clouse, M. E., Sara, L.,& Rochitte, C. E. (2010). The absence of coronary calcification does not exclude obstructive coronary artery disease or the need for revascularization in patients referred for conventional coronary angiography. *Journal of the American College of Cardiology*, 55(7), 627-634.

Granger, C. WJ, and P. Newbold. 1986. Economic Theory. In: *Forecasting economic time series*. Academic Press.

Gurley, K., & Kareem, A. (1999). Applications of wavelet transforms in earthquake, wind and ocean engineering. Engineering structures, 21(2), 149-167.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. The Journal of Machine Learning Research, 3, 1157-1182.

Haykin, S. (1999). Neural networks: A guided tour. *Soft Computing and Intelligent Systems: Theory and Applications*, 71.

Harang, R., Bonnet, G., & Petzold, L. R. (2012). WAVOS: a MATLAB toolkit for wavelet analysis and visualization of oscillatory systems. BMC research notes, 5(1), 163.

Hoffberg, S. M. (2011). U.S. Patent No. 7,974,714. Washington, DC: U.S. Patent and Trademark Office.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., & Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, *454*(1971), 903-995.

Hecht-Nielsen, R. (1989). Theory of the backpropagation neural network. In *Neural Networks, 1989. IJCNN., International Joint Conference on* (pp. 593-605). IEEE.

Honaker, J., & King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, *54*(2), 561-581.

Huether, B. M., Gustafson, S. C., & Broussard, R. P. (2001). Wavelet preprocessing for high range resolution radar classification. Aerospace and Electronic Systems, IEEE Transactions on, 37(4), 1321-1332.

Izzeldin, H., Asirvadam, V. S., & Saad, N. (2010). Enhanced conjugate gradient methods for training MLP-networks. In *Research and Development (SCOReD), 2010 IEEE Student Conference on* (pp. 139-143). IEEE.

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*,*22*(1), 4-37.

Kaastra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, *10*(3), 215-236.

Kalayci, T., & Ozdamar, O. (1995). Wavelet preprocessing for automated neural network detection of EEG spikes. Engineering in Medicine and Biology Magazine, IEEE, 14(2), 160-166.

Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*, *11*(2), 2664-2675.

Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. Wiley-IEEE Press.

Kantz, H., & Schreiber, T. (2003). *Nonlinear time series analysis* (Vol. 7). Cambridge university press.

Karamouz, M., Nazif, S., & Falahi, M. (2012). *Hydrology and Hydroclimatology: Principles and Applications*. CRC PressI Llc.

Karlaftis, M. G., & Vlahogianni, E. I. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights.*Transportation Research Part C: Emerging Technologies*, *19*(3), 387-399.

Konstantin Kravtsov, Mable P. Fok, David Rosenbluth, Paul R. Prucnal (2011). *The International Online Journal of Optics* *19 (3), 2133-2147*

Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Handling imbalanced datasets: A review. GESTS International Transactions on Computer Science and Engineering, 30(1), 25-36.

Krishna, B., Rao, Y. S., & Nayak, P. C. (2011). Time series modeling of river flow using wavelet neural networks. *Journal of Water Resource and Protection*,*3*(1).

Larochelle, H., Bengio, Y., Louradour, J., & Lamblin, P. (2009). Exploring strategies for training deep neural networks. The Journal of Machine Learning Research, 10, 1-40.

Letelier, J. C., & Weber, P. P. (2000). Spike sorting based on discrete wavelet transform coefficients. Journal of neuroscience methods, 101(2), 93-106.

Li, S. Z. (2011). Handbook of face recognition. Springerverlag London Limited.

Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T., Cohen, K. L., ... & Aguilera, A. M. (1999). Robust principal component analysis for functional data. Test, 8(1), 1-73.

Lodwich, A., Rangoni, Y., & Breuel, T. (2009). Evaluation of robustness and performance of early stopping rules with multi layer perceptrons. In Neural Networks, 2009. IJCNN 2009. International Joint Conference on (pp. 1877-1884). IEEE.

Loris, I., Simons, F. J., Daubechies, I., Nolet, G., Fornasier, M., Vetter, P., ... & Charléty, J. (2010). A new approach to global seismic tomography based on regularization by sparsity in a novel 3D spherical wavelet basis. In*EGU General Assembly Conference Abstracts* (Vol. 12, p. 6033).

Malaysian Meteorological Department (2010). *Weather forecast*. Retrieved from: http://www.met.gov.my.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 11(7), 674-693.

Marczak, M., & Gómez, V. (2012). Cyclicality of real wages in the USA and Germany:New insights from wavelet analysis. Retrieved from: http://opus.ub.uni-hohenheim.de/voltexte/2012/726/

Maxwell, T.,Giles, C. L., & Lee, Y. C. (1987, June). Generalization in neural networks: the contiguity problem. In IEEE First International Conference on Neural Networks (Vol. 2, pp. 41-46).

Melin, P., & Castillo, O. (2005). Hybrid intelligent systems for pattern recognition using soft computing: an evolutionary approach for neural networks and fuzzy systems (Vol. 172). Springer-Verlag New York Incorporated.

McClelland, J. L., Rumelhart, D. E., & PDP Research Group. (1986). Parallel distributed processing. *Explorations in the microstructure of cognition*, *2*.

McFall, K. S., & Mahan, J. R. (2009). Artificial neural network method for solution of boundary value problems with exact satisfaction of arbitrary boundary conditions. Neural Networks, IEEE Transactions on, 20(8), 1221-1233.

McGarry, K., Wermter, S., MacIntyre, J., & St Peter's Campus, S. P. S. W. (1999). Hybrid neural systems: from simple coupling to fully integrated neural networks. *Neural Computing Surveys*, *2*(1), 62-93.

Mohamad, N., Zaini, F., Johari, A., Yassin, I., & Zabidi, A. (2010). Comparison between Levenberg-Marquardt and scaled conjugate gradient training algorithms for breast cancer diagnosis using MLP. In Signal Processing and

Its Applications (CSPA), 2010 6th International Colloquium on (pp. 1-7). IEEE.

Morales, E., & Shih, F.Y. (2000).Wavelet coefficients clustering using morphological operations and pruned quadtrees. *Pattern Recognition*, *33*(10), 1611-1620.

Morley, S., & Adams, M. (2011). Graphical analysis of single-case time series data. *British Journal of Clinical Psychology*, *30*(2), 97-115.

Mehtani, P. (2011). *Pattern Classification using Artificial Neural Networks*. National Institute of Technology Rourkela: B.Tech. Thesis

Mukherjee, S., Osuna, E., & Girosi, F. (1997). Nonlinear prediction of chaotic time series using support vector machines. In *Neural Networks for Signal Processing [1997] VII. Proceedings of the 1997 IEEE Workshop* (pp. 511-520). IEEE.

Murphy, J. F., Winterbottom, J. H., Orton, S., Simpson, G. L., Shilland, E. M., & Hildrew, A. G. (2012). Evidence of recovery from acidification in the macroinvertebrate assemblages of UK fresh waters: a 20-year time series.*Ecological Indicators*.

Nixon, M., & Aguado, A. S. (2012). *Feature Extraction & Image Processing for Computer Vision*. Academic Press.

Northern California Earthquake Data Centre (2010). *Data Collections*. Retrieved on March 17, 2012, http:// http://www.ncedc.org/ncedc/

Nyquist, H. (1932). *Regeneration theory*. Bell Telephone System.

Ocak, H. (2009). Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. *Expert Systems with Applications*, *36*(2),2027-2036.

Paris, S., Hasinoff, S. W., & Kautz, J. (2011). Local Laplacian filters: edge-aware image processing with a Laplacian pyramid. *ACM Trans. Graph*, *30*(4),68

Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. Retrieved from: arXiv.org.

Perwej, Y., & Chaturvedi, A. (2012). Machine recognition of Hand written Characters using neural networks. *arXiv preprint arXiv:1205.3964*.

Polikar, R., Upda, L., Upda, S. S., & Honavar, V. (2001). Learn++: An incremental learning algorithm for supervised neural networks. Systems, Man, and

Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 31(4), 497-508.

Polikar, R. "Multiresolutionanalysis:the discrete wavelet transform," 2004.

Popoola, A., & Ahmad, K. (2006, July). Testing the suitability of wavelet preprocessing for TSK fuzzy models. In Fuzzy Systems, 2006 IEEE International Conference on (pp. 1305-1309). IEEE.

Popoola, A. O. (2007). Fuzzy-wavelet method for time series analysis (Doctoral dissertation, University of Surrey).

Qu, H., & Chen, G. (2012, July). An improved method of fuzzy time series model. In *Intelligent Control and Information Processing (ICICIP), 2012 Third International Conference on* (pp. 346-351). IEEE.

Ramsey, J. B. (1999). The contribution of wavelets to the analysis of economic and financial data. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 357(1760), 2593- 2606.

Ratanamahatana, C. A., Lin, J., Gunopulos, D., Keogh, E., Vlachos, M., & Das, G. (2010). Mining time series data. In *Data Mining and Knowledge Discovery Handbook* (pp. 1049-1077). Springer US.

Richards, J. A. (2012). *Remote sensing digital image analysis: an introduction*. Springer.

Ritter, H., Steil, J. J., Nölker, C., Röthling, F., & McGuire, P. (2003). *Neural architectures for robot intelligence*. Retrieved from: Neuroinformatics Group,Faculty of Technology, Bielefeld University.

Roh, J., & Abraham, J. A. (2004). Subband filtering for time and frequency analysis of mixed-signal circuit testing. *Instrumentation and Measurement, IEEE Transactions on*, *53*(2), 602-611.

Rojas, R. (1996). *Neural networks: a systematic introduction*. Springer.

Rumelhart, D. E. (1995). *Back Propagation: Theory, Architectures, and Applications*. Psychology Press.

Saen, R. F. (2009). The use of Artificial Neural Networks for Technology Selection in the presence of both Continuous and Categorical Data. World Applied Sciences Journal, 6(9), 1177-1189.

Sharma, A., & Agarwal, S. (2012). Temperature prediction using wavelet neural network. *Res. J. Inform. Technol*, *4*, 22-30.

Shepherd, G. M., & Koch, C. (1990). Dendritic electrotonus and synaptic integration. *The Synaptic Organization of the Brain*, 439-473.

Shu, Z., & Lei, M. (2011, March). Based on wavelet adaptive finite element analysis. In *Computer Research and Development (ICCRD), 2011 3rd International Conference on* (Vol. 4, pp. 80-82). IEEE.

Sifuzzaman, M., Islam, M. R., & Ali, M. Z. (2009). Application of wavelet transform and its advantages compared to Fourier transform. *Journal of Physical Sciences*, *13*, 121-134.

Singh, Y., & Chauhan, A. S. (2009). Neural networks in data mining. Journal of Theoretical and Applied Information Technology, 5(6), 36-42.

Stafford III, W. F. (2010). Boundary analysis in sedimentation velocity experiments. *Essential Numerical Computer Methods*, 337.

Starck, J. L., Murtagh, F., & Fadili, J. M. (2010). *Sparse image and signal processing: wavelets, curvelets, morphological diversity*. Retrieved from: Cambridge University Press.

Tan, Z., Zhang, J., Wang, J., & Xu, J. (2010). Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models.*Applied Energy*, *87*(11), 3606-3610.

Theußl, T., Hauser, H., & Gröller, E. (2000, October). Mastering windows: Improving reconstruction. In *Proceedings of the 2000 IEEE symposium on Volume visualization* (pp. 101-108). ACM.

Tommiska, M. T. (2003, November). Efficient digital implementation of the sigmoid function for reprogrammable logic. In Computers and Digital Techniques, IEE Proceedings- (Vol. 150, No. 6, pp. 403-411).

Tou, J. Y., Tay, Y. H., & Lau, P. Y. (2009). Recent trends in texture classification: a review. In Symposium on Progress in Information & Communication Technology, December (pp. 7-8).

Vetterli, M., & Kovačević, J. (1995). Wavelets and subband coding (Vol. 87). Englewood Cliffs, New Jersey: Prentice Hall PTR.

Venayagamoorthy, G. K., Moonasar, V., & Sandrasegaran, K. (September). Voice recognition using neural networks. In Communications and Signal Processing, 1998. COMSIG'98. Proceedings of the 1998 South African Symposium on (pp. 29-32). IEEE.

Wang, J., & Wu, J. (2009). Occurrence and potential risks of harmful algal blooms in the East China Sea. *Science of the Total Environment*, *407*(13), 4012-4021.

Wang, L., Wang, C., Fu, F., Yu, X., Guo, H., Xu, C., & Dong, X. (2011). Temporal lobe seizure prediction based on a complex Gaussian wavelet.*Clinical Neurophysiology*, *122*(4), 656-663.

Wang, X., Mueen, A., Ding, H., Trajcevski, G., Scheuermann, P., & Keogh, E. (2013). Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, *26*(2), 275-309.

Wang, Z., & Bovik, A. C. (2009). Mean squared error: love it or leave it? A new look at signal fidelity measures. *Signal Processing Magazine, IEEE*, *26*(1), 98-117.

Wedin, O., Bogren, J., & Grabec, I. (2008). *Data filtering methods*. Retrieved from: http://ec.europe.eu/information_society/apps/projects

Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, *78*(10), 1550-1560.

West, M. (1995). *Bayesian forecasting*. Institute of Statistics & Decision Sciences, Duke University.

Wilamowski, B. M. (2010). Human factor and computational intelligence limitations in resilient control systems. In *Resilient Control Systems (ISRCS), 2010 3rd International Symposium on* (pp. 5-11). IEEE.

Wu, Z., & Norden, E. H. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*,*1*(01), 1-41.

Xu, Q., Bai, Z., & Yang, L. (2009). An Improved Perceptron Tree Learning Model Based Intrusion Detection Approach. In Artificial Intelligence and Computational Intelligence, 2009. AICI'09. International Conference on (Vol. 4, pp. 307-311). IEEE.

Yashpal (2009) Singh, Y., & Chauhan, A. S. (2009). Neural networks in data mining. *Journal of Theoretical and Applied Information Technology*, *5*(6), 36-42

Zainuddin, Z., Huong, L. K., & Pauline, O. (2012). On the Use of Wavelet Neural Networks in the Task of Epileptic Seizure Detection from Electroencephalography Signals. *Procedia Computer Science*, *11*, 149-159.

Zhang, B. L., Coggins, R., Jabri, M. A., Dersch, D., & Flower, B. (2001). Multiresolution forecasting for futures trading using wavelet decompositions.*Neural Networks, IEEE Transactions on*, *12*(4), 765-775.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.

Zhang, H., Zhao, J., Jia, Y., Xu, X., Tang, C., & Li, Y. (2012). Exploration of artificial neural network to predict morphology of TiO nanotube. Expert Systems with Applications, 39(4), 4094-4101.

Zhang, M., Cai, W., & Shao, X. (2011). Wavelet unfolded partial least squares for near-infrared spectral quantitative analysis of blood and tobacco powder samples. Analyst, 136(20), 4217-4221.

Zhang, Y., & Wu, L. (2009). Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert systems with applications*, *36*(5), 8849-8854.

Zhan, F., Huang, Y., Colla, S., Stewart, J. P., Hanamura, I., Gupta, S., & Shaughnessy Jr, J. D. (2006). The molecular classification of multiple myeloma. Blood, 108(6), 2020-2028.

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*(1),