RDU1403138

REAL TIME PREDICTION FOR SUNGAI ISAP RESIDENCE USING ARTIFICIAL INTELLIGENCE TECHNIQUES

AHMAD JOHARI B MOHAMAD
MOHD RUSLLIM B MOHAMED
MAHFUZAH BT MUSTAFA
OMAR B ALIMAN
MOHD HERWAN B SULAIMAN
MOHD RAZALI B DAUD
WAN MOHD FAIZAL B WAN ISHAK

RESEARCH VOTE NO:
RDU1403138

Fakulti Kejuruteraan Elektrik & Elektronik

Universiti Malaysia Pahang

2017

REAL TIME PREDICTION FOR SUNGAI ISAP RESIDENCE USING ARTIFICAIAL
AHMAD JOHARI B MOHAMAD
FAKULTI KEJURUTERAAN ELEKTRIK & ELEKTRONIK
2017

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

____

(Student's Signature)

Full Name : Click or tap here to enter text.

ID Number :

Date : 12 October 2016

ACKNOWLEDGEMENTS

First of all, I would like to gratitude to my supervisor Dr.********** and co-supervisor, Mr. *************** for accepting and offer me a great and interesting project. Both of them guidance and support has given to explore new concepts and supported by hem is much appreciated. Them make discussion among us to exchange the opinion and obtained the better outcome is highly appreciated.

My appreciation also goes to my friends, course mate, and those who have given a helping hand to his project. Without their support, I would not able to proceed with my project..

Last, my greatest thankful would give to my parents and family for their moral support and love has brought happiness and joy in my life, thus I can totally focus on implement his project. Thanks you.

ABSTRAK

Banjir merupakan bencana yang sangat berbahaya yang dapat menghapus seluruh bandar,pantai dan kawasan luar bandar. Banjir yang disebabkan luas memusnahkan harta benda dan kehidupan yang mempunyai kuasa menghakis tertinggi dan boleh merosakkan. Penyelidikan ini meneroka penggunaan Artificial Neural Network (ANN) dan Support Vector Machine (SVM) kaedah untuk meramalkan banjir. Keseluruhan disiasat 29 data bulan meliputi dari Januari 2013 sehingga Mei 2015 di Sungai Isap Kuantan, Pahang, Malaysia. Suhu, hujan, titik embun, kelembapan, tekanan paras laut, penglihatan, angin, dan tahap sungai menganggap sebagai faktor banjir. Kajian ini mencadangkan perbandingan antara ramalan SVM dan ANN dalam banjir. Ia dijangka bahawa SVM memberikan hasil terbaik berbanding dengan ANN dalam tempoh ramalan yang tepat dengan itu hasil ramalan adalah berhampiran dengan keputusan sebenar.

ABSTRACT

A flood is an extremely dangerous disaster that able wipe away an entire city, coastline, and rural area. The flood caused wide destroy to property and life that has the supreme corrosive force and can he highly damaging. This research explores the use of Artificial Neural Network (ANN) and Support Vector Machine (SVM) method to predict on the flood. Totally investigated 29 month data covering from January 2013 until May 2015 in Sungai Isap Kuantan, Pahang, Malaysia. Temperature, precipitation, dew point, humidity, sea level pressure, visibility, wind, and river level consider as a factor of a flood. This study proposes a comparison of prediction of SVM and ANN in flood. It is expected that SVM giving best result compare to ANN in term of accurate prediction with that prediction result is near to actual result.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

σ	Sig2
γ	gamma

LIST OF ABBREVIATIONS

JPS	Jabatan Pengairan & Saliran
ANN	Artificial Neural Network
SVM	Support Vector Machine
LS-SVM	Least Square Support Vector Machine
NGO	Non Governmental Organization
MPK	Municipal Council of Kuantan
RBF	Radial Basic Function
BR	Bayesian Regularization
LM	Levenberg-Marquardt
GD	Gradient Descent
MSE	Mean Square Error
R	Regression

CHAPTER 1

INTRODUCTION

1.1 Introduction

A flood is an extremely dangerous disaster that able wipe away an entire city, coastline, and rural area. Flood causes extensive damage to life and property that has great erosive power and can be extremely destructive. It is a natural event that's happening when a piece of land or area that is usually dry land that suddenly submerges by the overflow of water. Some floods come slowly but other such as flash floods, can develop in just a few minutes and without any signs of rain. Floods are caused by rains, river overflow, temperature, strong winds in coastal areas and dam breaking. The floods hit Malaysia in 2014 that make more than 200,000 people affected while 21 killed on the floods [1]. This flood has been described as the worst floods in these decades.

The floods can have devastating consequences and can have effects on the economy, environment and people. The flood effect that can divide by two types there are primary effects and secondary long-term effect [2]. The primary effect of flooding include loss of life, damage on building and structures, roadways, bridge, canals, damage power transmission, power generation, sewerage systems, loss of clean drinking water treatment and water supply that may lack of clean water combined with human sewage in the flood water will raises the risk of waterborne diseases, which can include typhoid, giardia, cryptosporidium, cholera, e-coli, and many other diseases depending on the location of flood. Damage to roads and transport infrastructure may make it difficult to mobilize aid to those affected or to provide emergency health treatment. Flood waters typically inundate farm land, making the land unworkable and preventing crops from being planted or harvested, which can lead to shortages of food both for human and farm animals [3]. Secondary long-term effects, that affect economic hardship due to a temporary decline in tourism, rebuilding costs, or food shortages leading to price increases are a common after the effect of severe flooding. The impact affected caused psychological damage to particular where deaths, serious injuries and loss of property occur.

There are the major type of flood there are Flash floods, Rapid on-set floods, and slow on-set floods [4]. Flash floods is kind occurs within a very short time from 2 hours to 6 hours and sometimes within a few minutes that mostly caused a result of heavy rain or dam break. The rapid onset floods are taking slightly longer to develop and the flood can last for one or two days only. For rapid onset flood type, people can quickly save their valuable things and escape before flood getting very dangerous. The slow onset floods is type that usually that result of water bodies over flooding their banks. They tend to develop slowly and can last for days and week. They usually spread over many kilometers and occur more in flood plains.

This research explores the use of Artificial Neural Network (ANN) and Support Vector Machine (SVM) models to predict the floods. Temperature, precipitation, dew point, humidity, sea level pressure, visibility, wind, and river levels are considered as the flood factor with ANN and SVM were evaluated as flood prediction models. The data will be collected from January 2013 until May 2015.

1.2 Problem Statement

Flooding is the utmost common natural dangerous faced by global, causing an occurrence of death by accident or disease. Flooding occurs when the water exceeds in height over the banks of the river due to an excess rainfall that cannot be stocked in the soils or discharged fast enough by the stream network. For the last few decades, Sungai Isap is the fastest growing city in Kuantan suffer of flood during the monsoon periods (September to February) and wasted millions of Ringgits.

Due to the high rainfall volume, especially at high area (Sungai Lembing, and Kampung Panching) causes the increasing of river level at low area (Sungai Isap and Sungai Kuantan). The report from the Jabatan Pengairan Dan Saliran (JPS) Pahang, the flood happening in Sungai Isap is caused by heavy rainfalls in Sungai Kuantan valley and come with large tides. This causes flooding and the Sungai Isap area is suffering from the loss of food after the flood and the area is covered by mud and waste. Other than that, the water quality is also very extreme bad due to flooding and bacterial pollution, which is a serious health risk to the residence. Consequently, the majority of Sungai Isap residence who are settling in flood plain areas have been severely affected by flood disasters. For the long period of flood that will affect infrastructure like road communication and telecommunication networks broke down, making aid deliveries very difficult and delayed, led to dam and embankment erosion, eroded away healthy farmland with leaving it covered with sand and silt.

The most mainly issues is that currently Jabatan Pengairan & Saliran (JPS) Pahang no prediction method to predict flood at Sungai Isap. Besides that, flood warning does not avoid floods, but frequently establishes a complement of structural response as the latter cannot prevent the disaster success. It allows residence to protect themselves by building up temporary defenses or evacuating areas before the flood wave hits. Warning can prove very useful if they can be provided a long time enough ahead and if they are accurate. That is where flood prediction plays an important role. So efficient flood prediction need to apply for Sungai Isap residence can quickly save their valuable things and escape from the flood.

1.3 Objective of Research

The main objective of this research is to predict flood at Sungai Isap residence using Artificial Intelligence Technique. ANN and SVM model are studied, including their architectures and variations of associated learning rules to determine the parameters that will provide the best prediction for an impending flood. To achieve this objective, the task in the research is :-

I. To analyses historical weather data from Jabatan Pengairan & Saliran and Malaysian Meteorological Department (METMalaysia) for ANN and SVM prediction model.

II. To validate and compare the results prediction from the ANN and SVM with actual data.

III. To investigate the use of various ANN and SVM model for flood prediction along the Sungai Isap.

IV. To determine the ANN & SVM that the best prediction performance for the flood.

1.4 Scope of Research

The research is applied to one region as proof of concept, which is Sungai Isap, Kuantan area for flood prediction. The research is focused on the use of Artificial Neural Network (ANN) and Support Vector Machine (SVM) method to predict subsequent flood data.

The data were collected from Jabatan Pengairan & Saliran (JPS) and Malaysian Meteorological Department (METMalaysia). The data consists of temperature, precipitation, dew point, humidity, sea level pressure, visibility, wind, and river levels at Sungai Isap.

This research is using two methods that are ANN with Feed Forward Neural Network and SVM with Least Square Support Vector Machine (LS-SVM) at the end this two method result will be compared with actual results.

1.5 Significance of Study

This study will be a significant for Sungai Isap residence to save their valuable things quickly and escape from the flood and effectively lower the risk of harm and loss of life, if flood prediction giving the more accuracy predict. These studies will also give early preparation ahead of time to the government sector and non-governmental organization (NGO) to help Sungai Isap residence.

1.6 Study Area

Sungai Isap, Kuantan (Latitude. 3.7832°, Longitude. 103.2750°, the capital city of Pahang state is located in the east of Peninsular Malaysia, known as 'gateway to the east coast', is a fast growing commercial city in the east coast of peninsular Malaysia with the population of more than 400,000 peoples. On year 2013, Sungai Isap has faced with the worst flood caused by the destruction and loss. The report from the Jabatan Pengairan Dan Saliran (JPS) Pahang, the flood happening in Sungai Isap is caused by heavy rainfalls in Sungai Kuantan valley and come with large tides. Due to this, the flood has destroyed a big amount of properties and make the community live in discomfort. Sungai Kuantan is a river in Pahang ant it runs from Sungai Lembing through Kuantan City before flowing out to South China Sea.

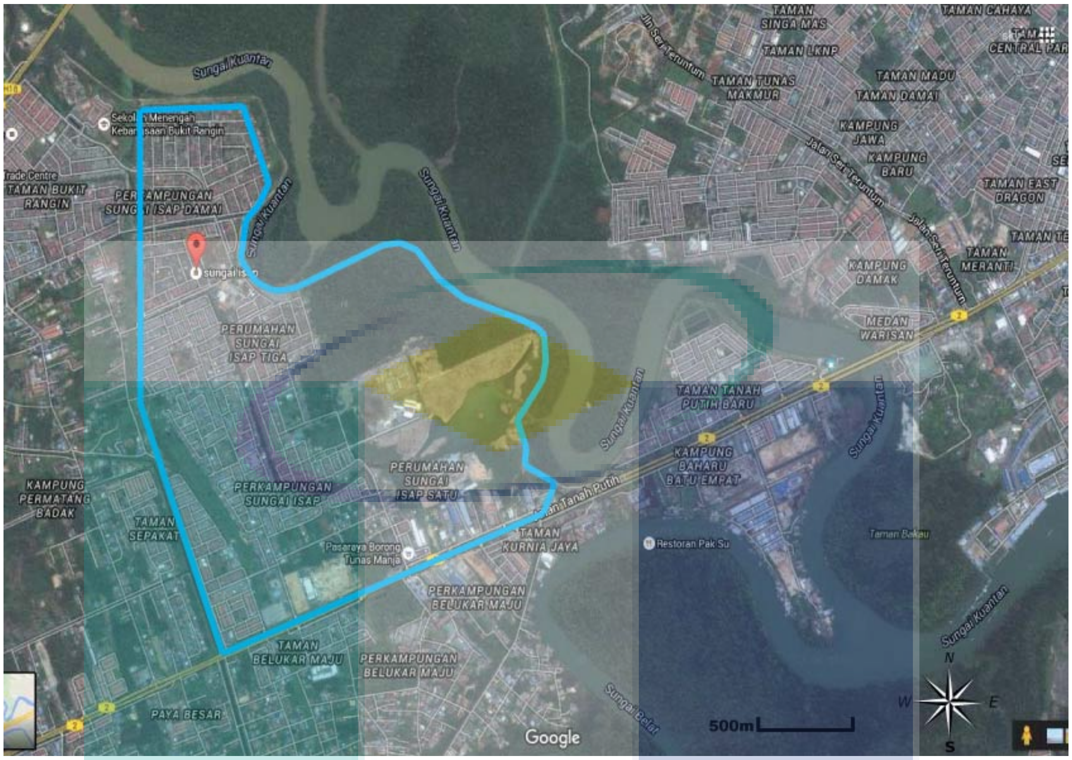

Figure 1.1 Figure satellite Sungai Isap, Kuantan

Figure 1.1 shows that the picture of the satellite from google map and Figure 1.2 shows the picture of the satellite and maps with river flow Sungai Kuantan at Sungai Isap. That clearly sees the effect of flood at Sungai Isap is from the river Sungai Kuantan. Sungai Isap is a previously swampy area and was developed into a residential area. This area has developed in more recent years, and was developed in phases which are first phase Perkampungan Sungai Isap 1, second phase Perkampungan Sungai Isap 2, third phase Perkampungan Sungai Isap Perdana, fourth phase Perkampungan Sungai Isap Uda Murni, and fifth phase Perkampungan Sungai Isap Jaya. The worst flood happened in 2013 inundated all areas of Sungai Isap as shown in Figure 1.3 The Municipal Council of Kuantan (MPK) has carried out many solutions for this problem, including the construction of a man-made lake as catchment during heavy rainfall or monsoon period bet the scenario became even worse from year to year.

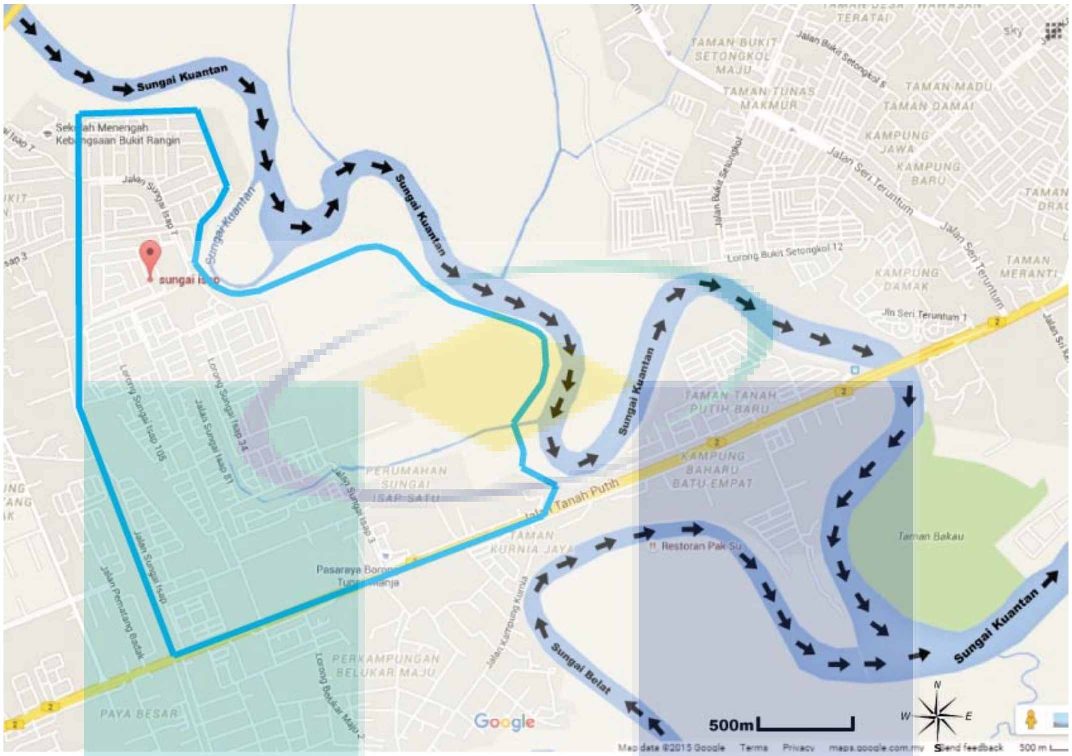

Figure 1.2 Figure Sungai Isap map with Sungai Kuantan river flow

Figure 1.3 Flood at Sungai Isap, Kuantan on 4th December 2013

CHAPTER 2

2.1 Title: Real Time Flood Prediction for Sungai Isap Residence Using Support Vector Machine

2.2 Abstrak

A flood is an extremely dangerous disaster that can wipe away an entire city, coastline, and rural area. The flood caused extensive damage to life and property that has great erosive power and can be extremely destructive. This research explores the use of Support Vector Machine (SVM) method to predict the flood in 2015. Total investigated 29 months data covering from January 2013 until May 2015 in the Sungai Isap Kuantan, Pahang, Malaysia. Rainfall, temperature and river level are considered as a factor of a flood. From the experimental, it can be concluded the SVM technique is an efficient flood prediction. The proposed of SVM prediction models is for implemented in real-time flood warning system to let the residence to take action immediately.

2.3 Pengenalan

Floods are caused by rains, river overflow, temperature, strong winds in coastal areas and dam breaking. The floods hit Malaysia in 2014 that make more than 200,000 people affected while 21 killed in the floods [1]. This flood has been described as the worst floods in these decades. The floods can have been devastating consequences and can have an effect in the economy, environment and people.

The flood effect that can divide by two types, there are primary effects and secondary long-term effect [2]. The primary effect of flooding included loss of life, damage on building and structures, damage power generation, loss of clean drinking water treatment and water supply, and raise the risk of waterborne diseases. Flood waters typically inundate farmland, making the land unworkable and preventing crops from being planted or harvested, which can lead to shortages of food both for humans and farm animals [3]. Secondary long-term effects, that affect economic hardship due to a temporary decline in tourism, rebuilding costs, and food shortages leading to a price increase common after the effect of severe flooding.

There are three major types of flood, there are flash floods, rapid on-set floods, and slow on-set floods [4]. Flash flood is kind occurs within a very short time from two hours to six hours and sometimes within a few minutes that mostly caused as a result of heavy rain or dam break. The rapid on-set floods are taking slightly longer to develop, and the flood can last for one or two days only. For rapid on-set flood's type, people can quickly save their valuable things and escape before flood getting very dangerous. The slow on-set flood is a type that usually that results of water bodies over flooding their banks. They tend to develop slowly and can last for days and week. They usually spread over many kilometers and occur more in flood plains.

This research explores the use of Support Vector Machine (SVM) models to predict the floods. Rainfall, temperature, and river levels are considered as the factor of a flood, and a number of SVM architectures were evaluated as flood prediction models. The data from the factor was noted from January 2013 until May 2015.

2.3.1 Study Area

In order to examine the performance of SVM and the impact of the prediction, Sungai Isap, Kuantan (Latitude. 3.7833°, Longitude. 103.3000°), the capital city of Pahang state is located in the east of Malaysia, known as 'gateway to the east coast', is a fast growing commercial city in the east coast of peninsular Malaysia with the population of more than 400,000 peoples. Figure 2.1 and Figure 2.2 show the picture of the satellite and maps at Sungai Isap, Kuantan that clearly see effect flood at Sungai Isap if from the river Sungai Kuantan. For the past few years, Kuantan has faced with the worst flood caused by the destruction and loss.

Figure 2.3 shows that picture flood on December 2013. The Municipal Council of Kuantan (MPK) has carried out many solutions to these problems, especially in the famous area such as Sungai Isap, including the construction of a man-made lake as catchment during heavy rainfall or monsoon period bet the scenario became even worse from year to year.

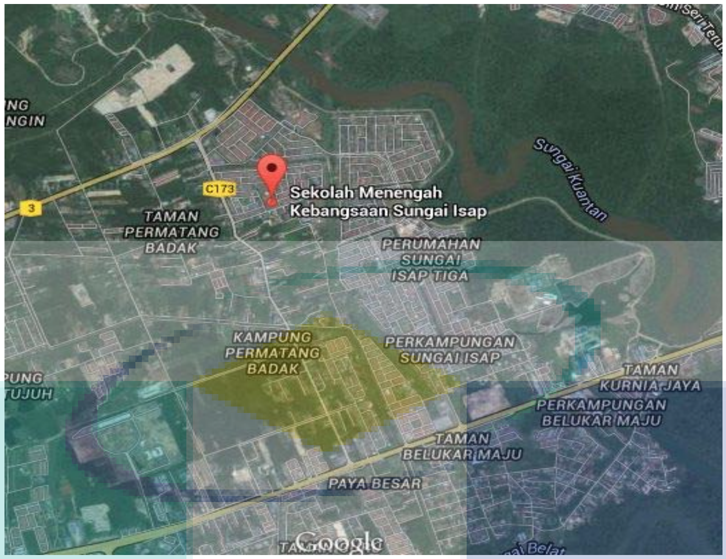

Figure 2.1 Satellite view

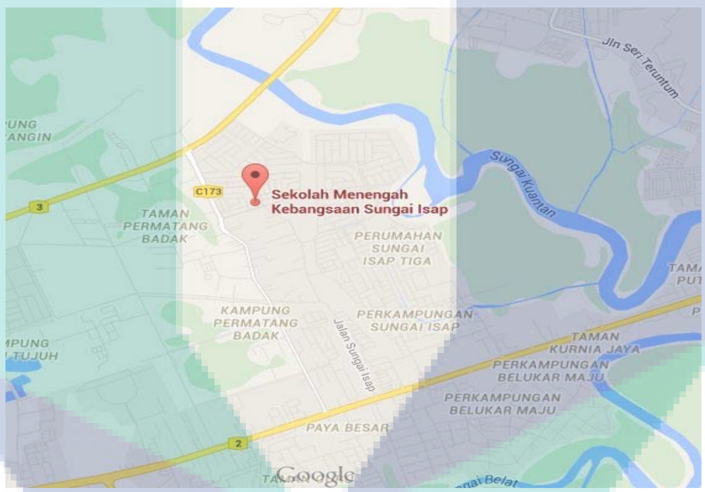

Figure 2.2 Map View

Figure 2.3 Flood 2013

2.4 Methodology

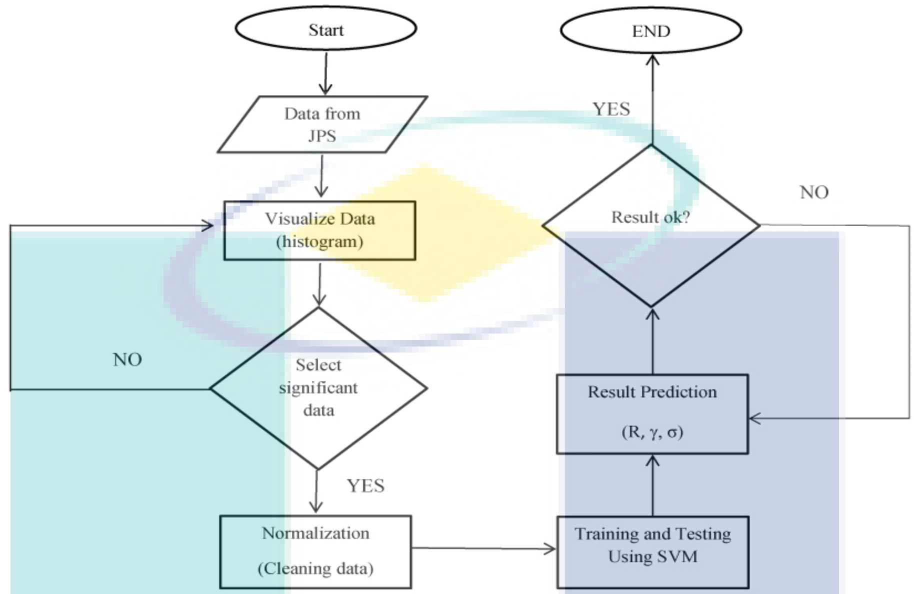

Figure 2.4 Flow chart SVM prediction

The flood prediction flow chart based on SVM is shown in Figure 2.4. Flood conditioning factor data set was constructed by factors from Jabatan Pengaliran & Saliran (JPS) is temperature, river level, rain level, dew point, humidity, sea level pressure, visibility, wind and precipitation from January 2013 until May 2015. These factors were gathered to form a visualize data histogram, then from the visualize data histogram found the factors of significant data and clearly that the data temperature, river level and rain level is most significant. Due to the large difference in the order or magnitude of the value, the available samples are scaled in 0-1 using the normalization preprocessing method. SVM is a technique [5]. The Radial Basic Function (RBF) kernel function of the SVM model is used in this paper. The training part uses 70% of the data that is from January 2013 until September 2014 (21 months) and for the testing part is use 30% of the data that start from October 2014 until May 2015 (8 months).

2.5 Result and Discussion

The performance of the SVM model is evaluated by prediction of the flood happening. All of the calculation SVM results with 100 times of calculation training. Figure 2.5, the meaning of 'R' is the percentage of SVM prediction shows, the highest percentage prediction is 94.785%, and the lowest is 85.229%. From the Figure 2.6 the graph of value γ, the max of the value is 4.55E-04 and the minimum value is 3.19E-08.

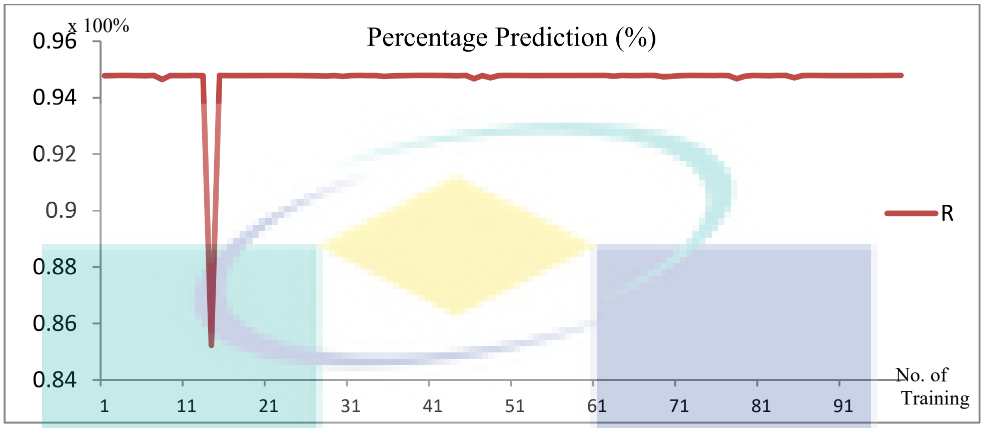

Figure 2.5 Graph percentage prediction

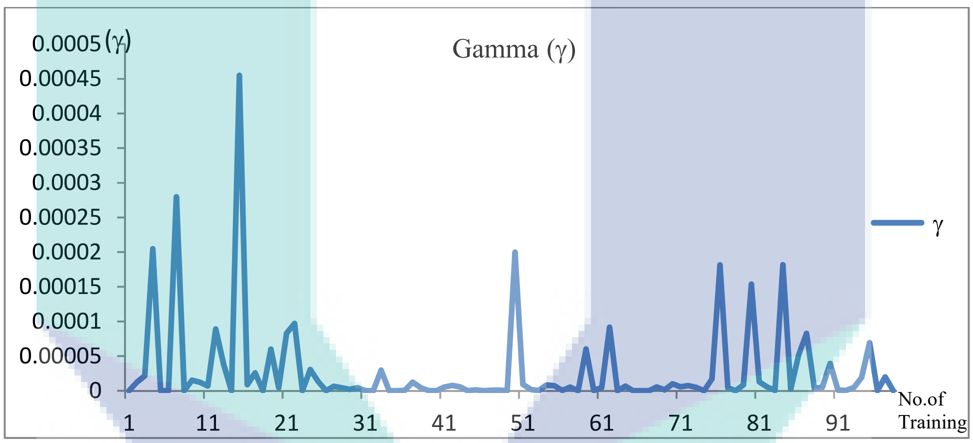

Figure 2.6 Graph gamma (γ)

From the Figure 2.7, the graph of value σ, γ the max of the value is 19458009, and the minimum value is 31.6192. The reading of highest percentage prediction from the 100 times SVM calculations is 94.785% and got 23 times same reading, so choose the best of five readings at Table 2.1 Top 5 Best ReadingsTable 2.1 with the gamma reading is not large and sig2 is not too small to prevent over-fitting problems.

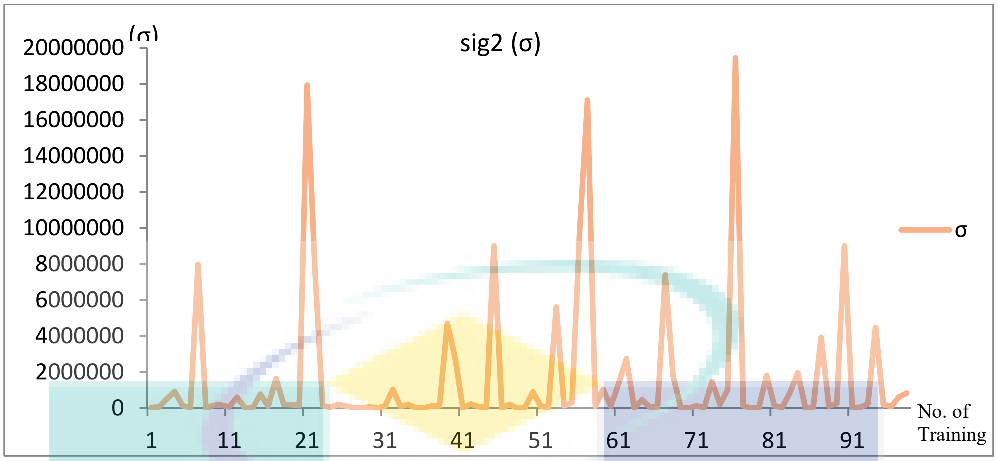

Figure 2.7 Graph sig2 (σ)

Table 2.1 Top 5 Best Readings

	R(%)	γ	σ
1	94.785	2.92E-07	4707446
2	94.785	4.28E-07	9812685
3	94.785	1.35E-06	9009867
4	94.785	5.36E-06	17095028
5	94.785	5.57E-06	1777336

2.6 Conclusion

The flood is the most damaging catastrophic phenomena in the world. Over the last decade, the flood becomes the hot topics in the world, even the literature, because this evaluation is a tough and nonlinear problem. Many methods have been tried by researchers in literature, but each method has their weakest point. Flooding on year 2013 caused serious damage in Sungai Isap, Kuantan, Pahang, Malaysia. The goal of this study was to do the accuracy flood prediction to assist having proper management over the effected area. For use SVM calculation produced 94.785% of the rate prediction. The information from current research can help JPS, Malaysian Meteorological Department, and government to perform the proper take action immediately.

CHAPTER 3

3.1 Title : Artificial Neural Network Flood Prediction for Sungai Isap Residence

3.2 Abstrak

A flood is an extremely dangerous disaster that can wipe away an entire city, coastline, and rural area. The flood can cause wide destrotion to property and life that has the supreme corrosive force and can be highly damaging. In order to decrease the damages caused by the flood, an Artificial Neural Network (ANN) model has been established to predict flood in Sungai Isap, Kuantan, Pahang, Malaysia. This model is able to initiate the same brain thinking process and avoid the influence of the predict judgment. In this paper, presentation and comparison that using Bayesian Regularization (BR) back-propagation, Levenberg-Marquardt (LM) back-propagation and Gradient Descent (GD) back-propagation algorithms will be organized and carry out the result flood prediction. The predicted result of the Bayesian Regularization indicates a satisfactory performance. The conclusions also indicate that Bayesian Regularization is more versatile than Levenberg-Marquart and Gradient Descent with that can be backup or a practical tool for flood prediction. Temperature, precipitation, dew point, humidity, sea level pressure, visibility, wind, and river level data collected from January 2013 until May 2015 in the city of Sungai Isap, Kuantan is used for training, validation, and testing of the network model. The comparison is shown on the basis of mean square error (MSE) and regression (R). The prediction by training function Bayesian Regularization back-propagation found to be more suitable to predict flood model.

3.3 Pengenalan

Floods mostly caused by heavy raining, sea level pressure, wind, and river overflow in seacoast and the lower land area. The floods hit Malaysia in 2014, that causes 21 dead and 200,000 residents is affected by this flood and has been described as the worst floods in these decades. The floods can have been devastating consequences and can have an effect in the economy, environment and people. The flood effect can be divided into two types, there are primary effects and secondary long-term effect [2]. The primary effect of flooding included loss of life, damage on building and structures, damage of power generation, loss of clean drinking water treatment and water supply, and raise the risk of waterborne diseases. Secondary long-term effect is that affect on the economic problems causes by lasting in the limit period of time in decrease in tourism sector, reconstruction costs, and insufficient food after the big impact of flood.

In the last decades, Artificial Neural Network (ANN) has been widely used for flood prediction. ANN can be used for predicting because of having the capability of examining and determining the historical data used for prediction [6]. In recent years, several types of ANN models have been developed for flood prediction in Malaysia, result from the paper [7]. exploit using Geographic Information System (GIS) and ANN method to modeling and simulate in Johor. In [8], is about compare the prediction between Non Linear Autoregressive with Exogenous Input (NARX) and Extended Kalman Filter (EKF) in flood water lever prediction. From the [9], is proposed Back-Propagation Neural Network (BPN) with to improve the prediction with extended Kalman filter is apply in output of BPN to predict the flood water level at downstream station. Result from [10], proposed an Improvement Elman Neural Network (ENN) for predict water level at Kelang River station located at Petaling Bridge, Kuala Lumpur. Previous work has been supportive of the neural networks for flood prediction at other country, India [11], show a Time Lagged Recurrent Neural Network and General Recurrent Neural Network to predict the rainfall model for upper area of Fardha River in India with the methodologies and techniques is compare of the short term runoff prediction results between both neural network. In China [12], the Optimal Subset Regression (OSR) and Back-Propagation Neural Network (BPNN) have been combined. Training and testing errors are for measure and to decide the best condition to stop training and make a prediction. In Greece [13], comparing made between the performance of Feed-Forward back-propagation ANN, Adaptive Linear Neuron Network (ADALINE), and Elman network in an attempt to assess the relative performance of existing models at the river Pinios (Greece). In Italy [14], for this model is compare the Black-Box Type Runoff Simulation Model and the Real-Time Improvement of the Discharge Forecast to rainfall-runoff model aimed at the real-time forecasting of flood events, based on integrating ANN. In Poland [15], using Radial-Basis Function Neural Networks and Multi-Layer Perceptron, extended with Linear Regression and Nearest Neighbor Approach for flood forecasting in Nysa Klodzka River.

Utilizing the advantages of ANN, in this paper aim of this work is to present and show the comparison of the three algorithms that are Bayesian Regularization (BR) back-propagation, Levenberg-Marquardt (LM) back-propagation, and Gradient Descent (GD) back-propagation algorithms of the flood prediction.

3.3.1 Study Area

Nearly every year, heavy rainfalls happened over Pahang due to the northeast monsoon period between November until February which caused flooding in a low area particularly near the flood plain area. The heavy rain at high area (Sungai Lembing, and Kampung Panching) causes the increasing of river level at low area (Sungai Isap and Sungai Kuantan). The report from the Jabatan Pengairan Dan Saliran (JPS) Pahang, the flood happening in Sungai Isap is caused by heavy rainfalls in Sungai Kuantan valley and come with large tides. Due to this, the flood has destroyed a big amount of properties and make the community live in discomfort. The worst flood happened in 2013 inundated all areas of Sungai Isap as shown in Figure 3.3.

Sungai Isap (Latitude. 3.7833°, Longitude. 103.3000°), is located approximately 10km from Kuantan town with 400,000 residents is the Capital City of Pahang and located East Malaysia. Figure 3.1 and Figure 3.2 show the picture of the satellite and maps with river flow Sungai Kuantan at Sungai Isap that clearly see effect flood at Sungai Isap if from the river Sungai Kuantan. Sungai Isap is a previously swampy area and was developed into a residential area. This area has developed in more recent years, and was developed in phases by phases.

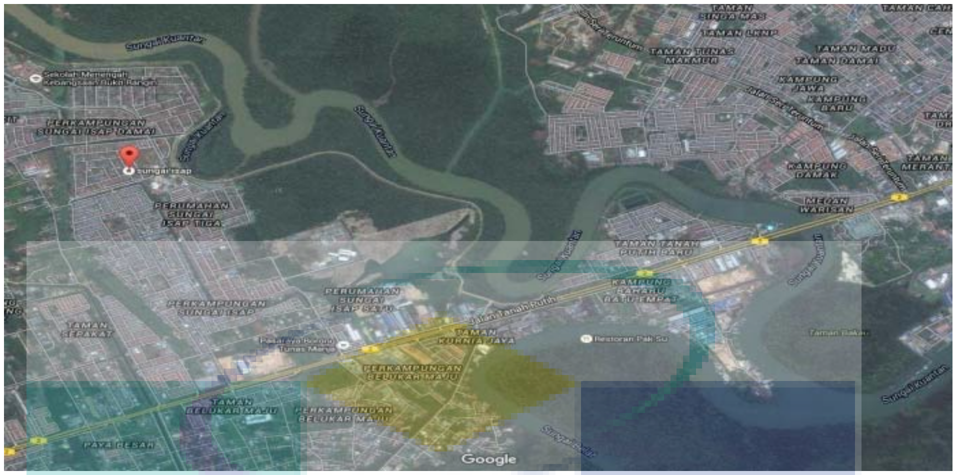

Figure 3.1 Location of Sungai Isap town (Google Maps, Satellite View)

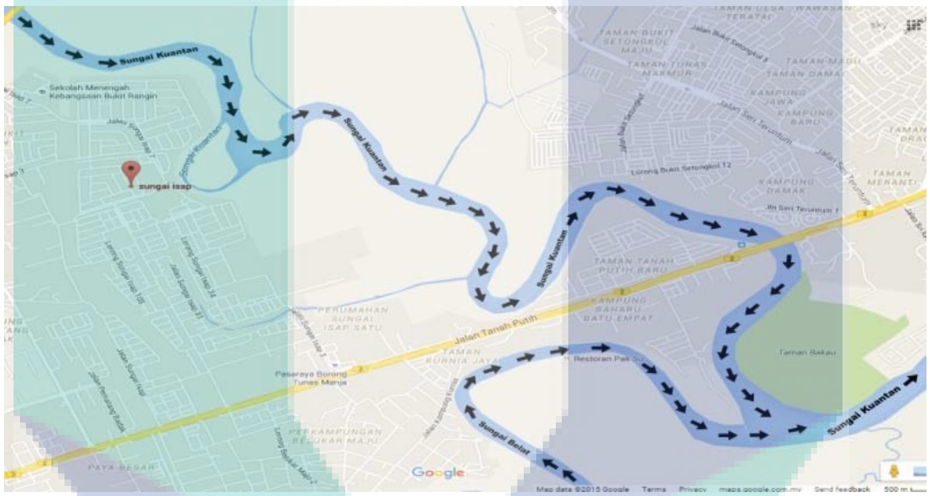

Figure 3.2 Location of Sungai Isap town with river flow Sungai Kuantan (Google Maps View)

Figure 3.3 Sungai Isap flood 2013

3.4 Methodology

ANN is is an intelligent process model that inspired biological nervous systems, for an example brain process [16]. This is caused by these network try to model the capabilities of human brain. ANN is being widely used in many field of study. World first ANN with neuron is create at 1943 with the ANN has grown faster after the first ANN training algorithm was introduce in 1958 [17]. From the past decade, ANN used a theoretically change to statistical model. A systematic methodological flow chart of the study is presented in Figure 3.4.

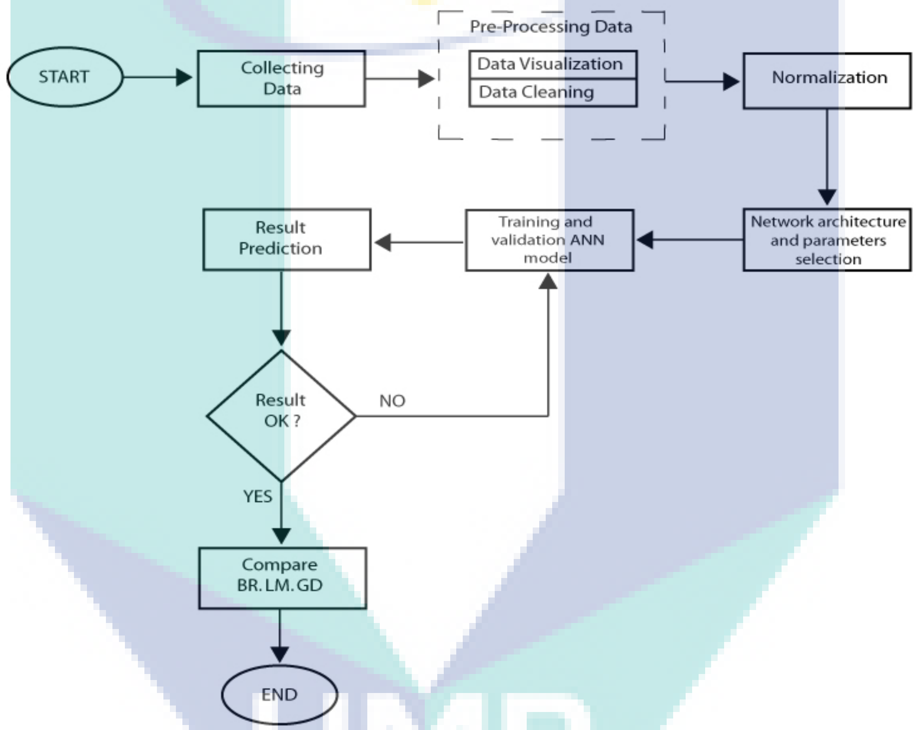

Figure 3.4 Methodological flow chart

3.4.1 Data Collevtion

Flood conditioning factor data set as constructed by factors from Jabatan Pengaliran & Saliran (JPS) Pahang is temperature, precipitation, dew point, humidity, sea level pressure, visibility, wind and river level start from January 2013 until May 2015.

3.4.2 Data Visualization

Data Visualizations method hat showing the data in pictorial or graphical format with many researcher use for represent there charts to faster and easily understanding the information from the data [18].

3.4.3 Data Cleaning

The factor data (raw data) of the 8 factors have some missing data, so the SPSS software is used to find back those missing data, this method is called as data cleaning. Data Cleaning includes the discovery and excision or correction of errors [19]. If there are incomplete data and then is either replaces, modified or deleted the data. Data cleaning process need carefully consideration because the influence the results. Cleaning data need always checks and treatment of missing responses will be done Statistical Product and Service Solution (SPSS) software will consistency check and treatment of missing responses that cleaning [20].

3.4.4 Normalization

Because of different data and different magnitude value, the samples need to be scaled in 0-1 using normalization. Normalization is method that adjusts the value measurement due to the different scale to a collaborative scale and frequently before averaging. Normalization also refers as complex adjustments where the purpose is to bring the whole probability assignation adjusted values into alignment. Equation 1 is a formula for normalization.

$$X' = \frac{X - Xmin}{Xmax - Xmin} \tag{1}$$

Where X' is scaled value, X is the sample value, X_{min} and X_{max} are the minimum value and maximum value.

3.4.5 Tranining Algorithm

ANN training quantity to iteratively adapting the weights of a neural network until the connection weights determine an input and output that function that near the relationship between input and output structure of a given training data set. In this research, three common algorithms that are Bayesian Regularization, Levenberg-

Marquardt, and Gradient Descent is used for execution the concept and the results from 3 algorithms are compared to select the lowest MSE the best results. The result in [6][21][22], that show the three common algorithms result are better than other training algorithm and that the main reason this paper using and the advantages are stated as below. Bayesian Regulation back-propagation network training function is a Levenberg-Marquardt optimization that updates according to the weight and bias values. It also can reduce the trouble in defining the optimum network parameters and also can avoid the over-fitting with the early stopping technique. The weights and squares errors combine and minimize by Bayesian Regularization, and define the proper combination to generate a network that able generalizes the proper combination. Levenberg-Marquardt back-propagation network training function is a Levenberg-Marquardt optimization that updates according to the weight and bias values and faster algorithm compares to Bayesian Regularization back-propagation. Levenberg-Marquardt also highly recommended as a first-choice manages algorithms cause not need require more memory and more time for training. Gradient Descent back-propagation network training function Gradient Descent Optimization that updates weight and bias values. From above back-propagation, the result will be compared to find the best result with the available data sets.

The program MATLAB with Neural Network Toolbox is a strong tool for prediction. The Neural Network Toolbox function to develop feed forward back-propagation model. This tool allows importing, creating, using, export neural network and data, and change parameters (number of neurons, learning rate, number of hidden layers, transfer function and performance functions). The structure of an ANN is created by weight between neuron, transfer function, and learning laws. The transfer function is used to control the generations of output in neurons. Learning laws are used to determine the relative weights for input to the neuron. In a feed-forward network, the weighted connections feed activations only in the forward direction from an input layer to the output layer.

3.4.6 Neuron

Neurons manipulate in logical parallelism and the information is sent from later to serial operations. If too few neurons will direct affect the network performance and next will outcome in under-fitting and too many neurons the network is over-fitting. For this research, the number of neurons is set by ranging from 10 neurons until 200 neurons.

3.4.7 Layer

A layer is a group that set up by neurons and a group of layer can set up a network. The ANN model usually got input layer, hidden layer and output layer. In this paper, using the single layer with prepares adequate hidden neurons in a single layer.

3.4.8 The Best Function

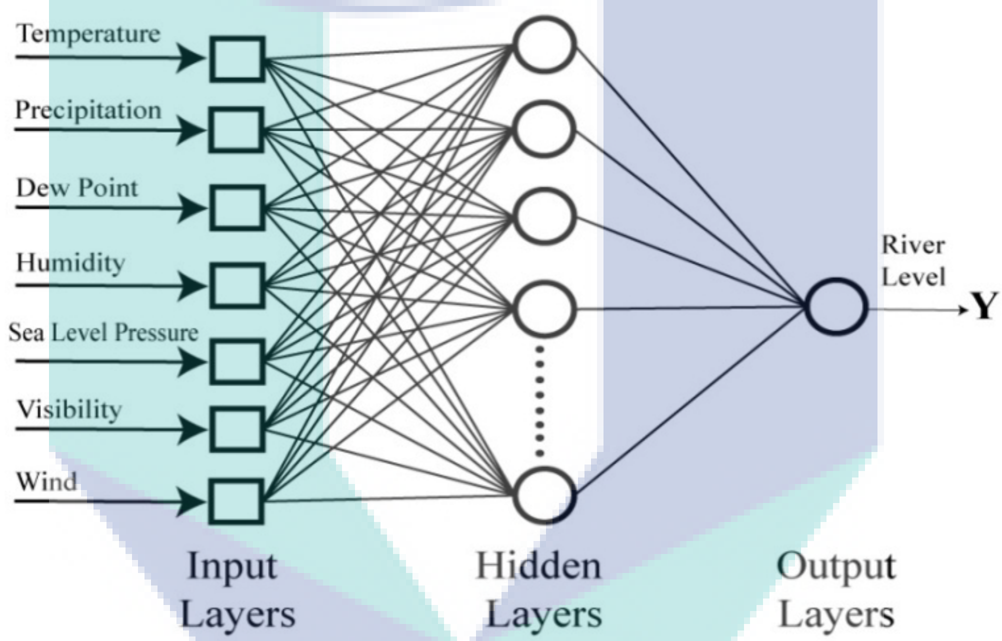

Figure 3.5 Structure of a feed-forward ANN model

To define the best structure, different number or neurons in hidden layers are testing [22]. Mean Squared Error (MSE) and Regression (R) is frequently used to excute the performance evaluation. MSE defines as the average squared difference between observed output values from ANN and targets introduced in the training samples.

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

From the Equation 2 that that n is the size of the sample set, a_i is the ANN observed output, and t_i is the corresponding target output. The (R) is defined as the correlation between the targets and output. When R=1, is meant a close relationship, vice versa R=0, there is random relationship.

The Gradient Decent with Momentum Weight and Bias Learning Function was chosen for Adaption Learning Function for neural network (7-1-1: 7 input, 1 hidden layer, and 1 output) is shown in Fig.5 structure of feed-forward ANN model and observed to be the best adaptive learning function as it shows results with minimum MSE [6]. [6]. Tan-Sigmoid Transfer Function is the best choice for hidden layers over Linear Transfer Function and Log Sigmoid Transfer Function because of its very fast learning rate and sensitivity towards change number of samples and neurons. Other than that, results bring the conclusion that Tan-Sigmoid function gives the best predictions for ANN architecture in terms of goal, epochs and comparison with the target data [23].

3.4.9 Training, Vefirying and Testing

In training part, 70% of input data from January 2013 until May 2015 were applied to train network. In verifying part, the program will stop of calculation when was 15% data are applied to determine the network structure work that was not used in training. Verifying data have checked in a different sequence of training and continued when the number of error reduced in the verifying. The last 15% data will applied for testing process after the training process and verifying process finish. Since Bayesian Regularization does not use validation samples, then the Levenberg-Marquardt and Gradient Descent is only for analysis. Different training algorithms are used to train the network and the corresponding configuration parameters are as shown in Table 3.1.

Table 3.1 Training algorithms configuration parameters

Configuration Parameters	BR	LM	GD
Maximum number of epochs to train	1000	1000	1000
Epochs between displays	25	25	25
Performance goal	0	0	0
Maximum time to train in seconds	Infiniti	Infiniti	Infiniti
Minimum performance gradient	1e-07	1e-07	1e-05
Maximum validation failures	6	6	6
Initial μ	0.005	0.001	N/A
μ decrease factor	0.1	0.1	N/A
μ increase factor	10	10	N/A
Maximum μ	1e+10	1e+10	N/A

Learning rate	N/A	N/A	0.01

3.5 Result and Discussion

ANN architecture attributes use the Bayesian Regularization, Levenberg-Marquardt and Gradient Descent for the training and testing of the proposed metrics concept. Table 3.2 captures the architecture and ANN attributes used in the experiment, Table 3.3 is the result of mean square error (MSE).

Table 3.2 Architecture and ANN attributes

Network Type	Feed-Forward Back-Propagation (FFBP)
Training Function	BR, LM, GD
Adaptive Learning Function	LEARGDM
Performance Function	MSE
Number of layers	01
Number of Neurons	10,20,30,40,50,60,70,80,90,100,110,120, 130,140,150,160,170,180,190,200
Transfer Function	Tan-Sigmoid
Epochs	1000

The lowest MSE is the better ANN attributes in maintainability prediction, from the Table 3.3 the result with the lowest MSE is 0.0007540 using Bayesian Regularization with 60 neurons is the best of ANN attributes.

Table 3.3 Results MSE with FFBP NN algorithm

Neurons	BR (MSE)	LM (MSE)	GD (MSE)
10	0.0019934	0.0047377	0.0053743
20	0.0013982	0.0080938	0.0051728
30	0.0010213	0.0027311	0.0079787
40	0.0008412	0.0032440	0.0032223
50	0.0008296	0.0075195	0.0039038
60	0.0007540	0.0036020	0.0079315
70	0.0008076	0.0029161	0.0119160
80	0.0010997	0.0020958	0.0045995
90	0.0008446	0.0089303	0.0151720
100	0.0014620	0.0035220	0.0123220
110	0.0007711	0.0020686	0.0087395
120	0.0016285	0.0040221	0.0086270
130	0.0009645	0.0040819	0.0059417
140	0.0008945	0.0029030	0.0035377
150	0.0010074	0.0022756	0.0047167
160	0.0013731	0.0043467	0.0066002
170	0.0014315	0.0055751	0.0131660
180	0.0015457	0.0039078	0.0115570
190	0.0009744	0.0043990	0.0071656
200	0.0014311	0.0027019	0.0048481

In Figure 3.6 shows that the graph of best training performance is 0.00075404 MSE at epoch 734th is considered better result that compared to [22] with 6.0579 MSE and [24] with 1.07 MSE. Figure 3.6 show the graph plot solid line represents independent testing, dash line is represents training, and the dot-dot line represents best value for this training. In Figure 3.7, shows the regression (R) graph with training, test, and all data. The dotted line in the graph shows the (perfect result – outputs = targets) and the solid line represent best fit linear regression between the target and the output for Figure 3.7(a) is training data regression, Figure 3.7(b) is testing data regression, and Figure 3.7(c) is all data regression. The regression (R) is represented the relation between the targets and the outputs that if R=1, meaning that is an accurate linear relationship between targets and outputs. Vice versa if R=0, there is no linear relationship between targets and outputs.

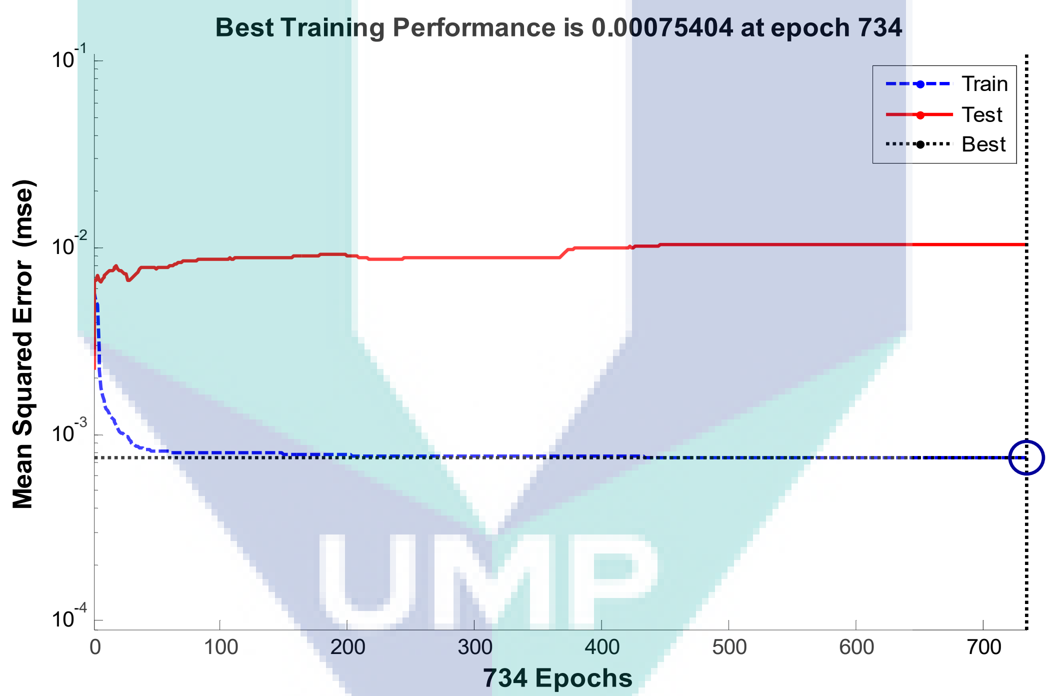

Figure 3.6 Best training performance with Bayesian Regularization using 60 neurons

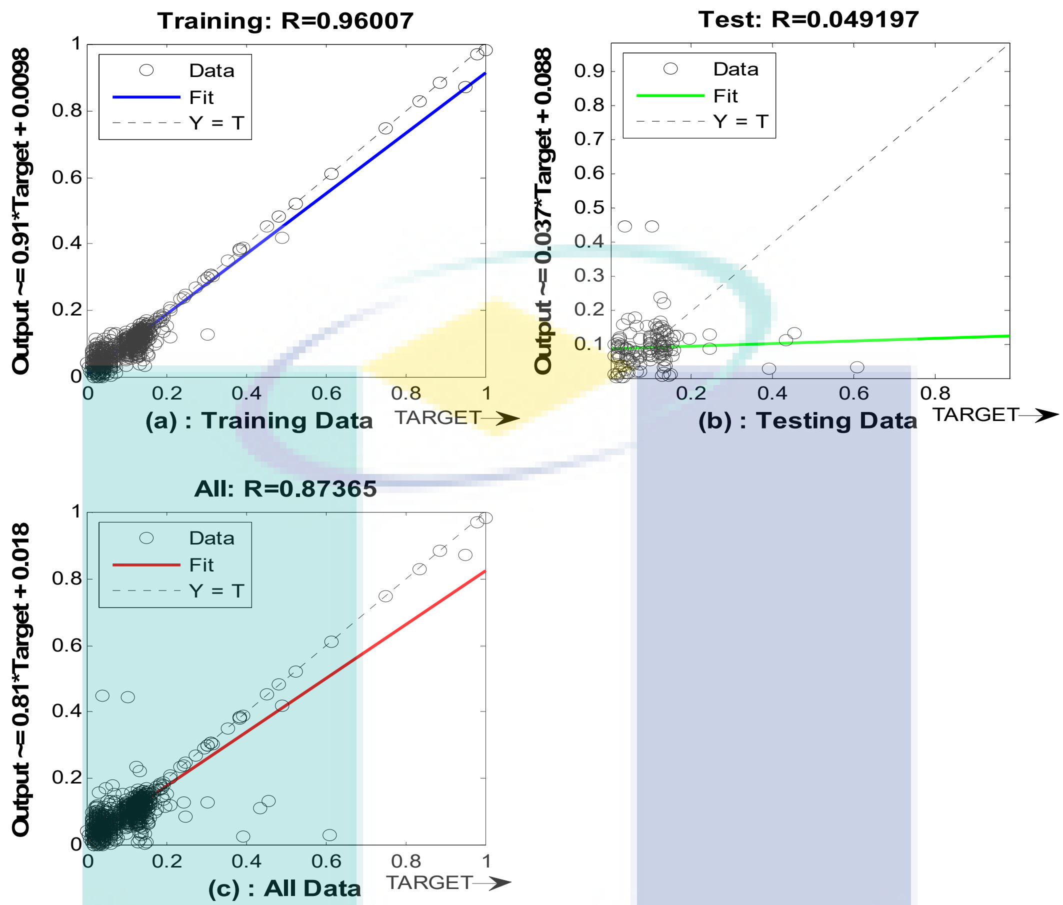

Figure 3.7 ANN regression result Bayesian Regularization (a) Training data regression (b) Testing data regression (c) All data regression

The training data, giving good results with R=0.96007 with Output~=0.91^Target+0.0098 at Figure 3.7(a) and all data R=0.87365 with Output~=0.81^Target+0.018 at Figure 3.7(c). For the Testing data part has got a poor result with R=0.0049197 with Output~=0.037^Target+0.088 Figure 3.7(b), which that mean that is near no linear relationship between targets and outputs for the Test part only. The graph is very important that can show clearly some of that data points are poor fits. After the training procedure is finished, the graph in Figure 3.8, show then the results of prediction data and the actual data are shown in the graph. The solid line represents the actual data and the dash line is showing the prediction data in Figure 3.8. The perfect way to clearly see the comparison of actual data and predict data is used graph method.

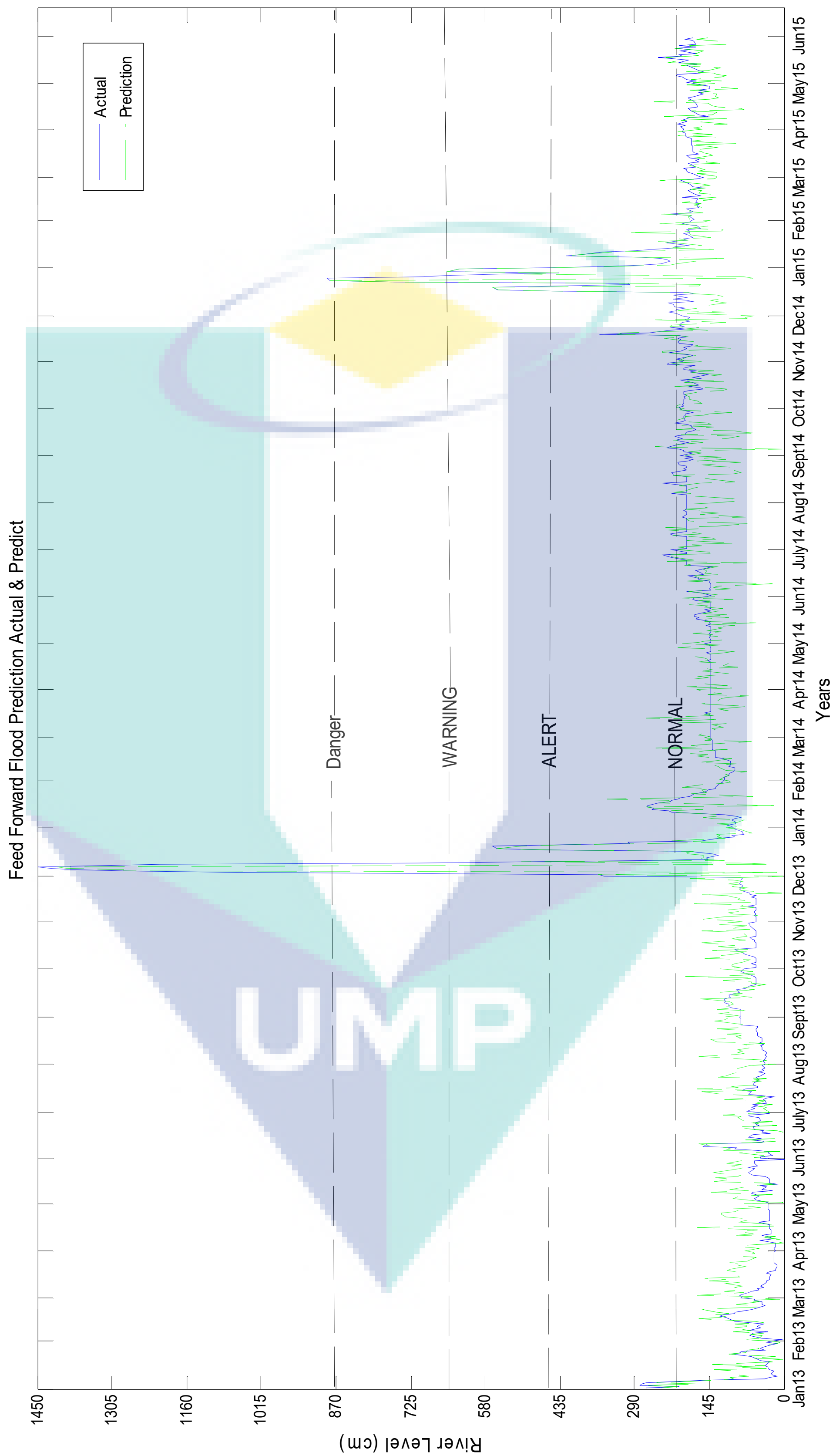

Figure 3.8 ANN Feed Forward Flood Prediction Actual & Predict

3.6 Conclusion

The flood is the most damaging catastrophic phenomena in the world. Over the last decade, the flood becomes the hot topic in the world, even the literature, because this evaluation is a tough and a nonlinear problem. Many methods have been tried by researchers in literature, but each method has their weakest point. Flooding on year 2013 caused serious damage in Sungai Isap, Kuantan, Pahang, Malaysia.

The different training algorithms and attributes of ANN architecture are used in experimentation and validation. The conclusion is based on the MSE value of the combination of neural network architecture, learning algorithm and neuron. In this case of maintainability prediction, better results are achieved with 60 neurons with the MSE is 0.0007540, using Bayesian Regularization back-propagation. The flood prediction study has been done successfully using ANN to give a model with Training R=0.96007, Test R=0.049197, and All R=0.87365. A 7-1-1 neural network with 7 inputs (temperature, precipitation, dew point, humidity, sea level pressure, visibility, and wind) and one output (river level) successfully simulated the experimental results. The information from current research can assist JPS, Malaysian Meteorological Department, and government to perform the proper take action immediately. The goal of this study was to predict flood in term of accuracy and to assist having proper management over the affected area.

CHAPTER 4

4.1 Title : Levenberg-Marquardt Flood Prediction for Sungai Isap Residence

4.2 Abstrak

The flood can cause wide destroy to property and life because of the supreme corrosive force and can be highly damaging. In order to decrease the damages cause by the flood, an Artificial Neural Network (ANN) model has been established to predict flood in Sungai Isap, Kuantan, Pahang, Malaysia. This model is able to imitate same as the brain thinking process and avoid any influence to the predict judgment. This study proposed Levenberg-Marquardt (LM) back-propagation with two different ratios that is (80%: 10%: 10%) and (70%: 15%: 15%) for training sample, testing sample, and validation sample. The data collected in terms of temperature, precipitation, dew point, humidity, sea level pressure, visibility, wind and river level data were collected from January 2013 until May 2015. The results are shown on the basic of mean square error (MSE) and regression (R). The prediction by Levenberg-Marquardt with 80% training sample was shown better result compared with 70% training sample.

4.3 Pengenalan

In 2014, it is reported as the worst flood hit Malaysia that caused 21 dead and 200,000 residents affected [1]. The factors that contributed to flood are heavy raining, high sea pressure, fast wind and river overflow in sea coast and lower land area. The floods affect the human or animal life, agriculture sector, building and its structure such as clean water supply. In the long term, the flood may affect the economic problems causing for example lasting in the limit period of time in a decrease in tourism sector, reconstruction costs, and insufficient food after the impact by flood.

The report from the Jabatan Pengairan Dan Saliran (JPS) Pahang, the flood that happened in Sungai Isap is caused by heavy rainfalls in Sungai Kuantan valley and came with large tides. Due to this, the flood has destroyed a big amount of properties and make the community live in discomfort.

Figure 4.1 Location of Sungai Isap town with river flow Sungai Isap, Kuantan (Google Maps View)

shows the maps with river flow Sungai Kuantan at Sungai Isap. That clearly sees the effect of flood at Sungai Isap is from the river, Sungai Kuantan. Sungai Isap is a previously swampy area and was developed into a residential area. This area has developed in more recent years, and was developed in phases which are first phase Perkampungan Sungai Isap 1, second phase Perkampungan Sungai Isap 2, third phase Perkampungan Sungai Isap Perdana, fourth phase Perkampungan Sungai Isap Uda Murni, and fifth phase Perkampungan Sungai Isap Jaya. The worst flood happened in 2013 inundated all areas of Sungai Isap as shown in Figure 4.2.

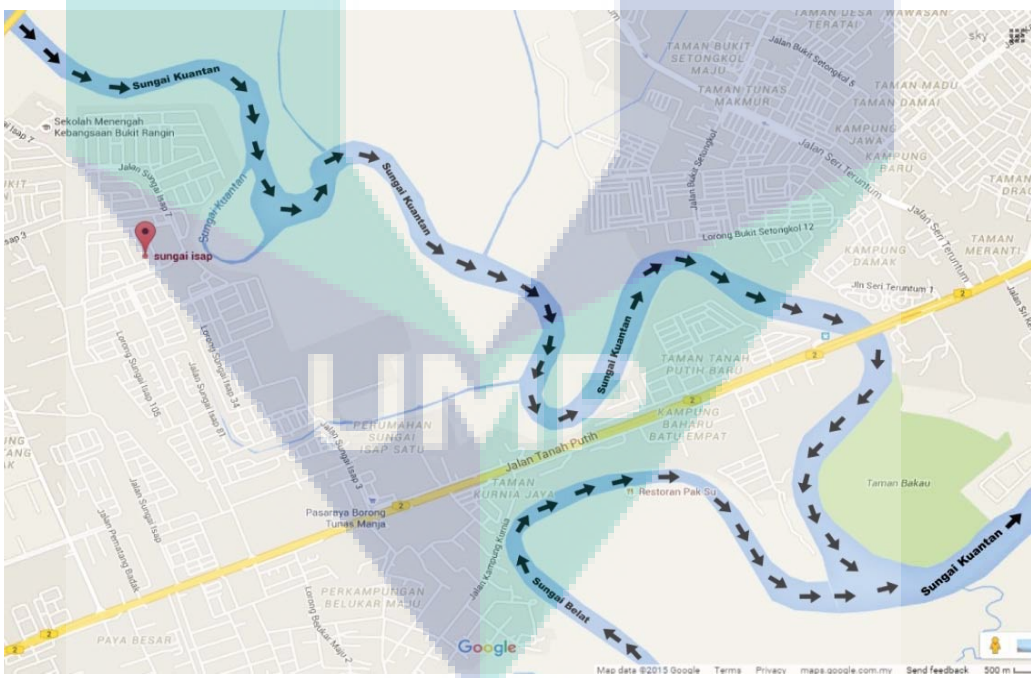

Figure 4.1 Location of Sungai Isap town with river flow Sungai Isap, Kuantan (Google Maps View)

Figure 4.2 Sungai Isap flood on December 2013

In the last decades, Artificial Neural Network (ANN) has been widely used to flood prediction. Although ANN models may be computationally intensive, the evolution in high speed computer has promoted their application. The ANN can be used for predicting because of having the capability of examining and determining the historical data used for prediction [6]. ANN flood prediction and modeling widely used in Malaysia for Johor River basin Kelang River, and Sungai Batu Pahat [7]–[10], [25]. Theresearcher employed various ANN models such as Exogenous Input (NARX), Extended Kalman Filter (EKF), Back-Propagation Neural Network (BPN) and Elman Neural Network (ENN). Based on reference [7]–[10], [25], all results showed good predicted of flood with the real records. Other than that, ANN flood prediction and modeling widely used in other country for upper Serpis Basin (Spain), Wardha River (India), Danjiangkou Reservoir (China), Pinios River (Greece), Arno River (Italy), and Nysa Klodzka River (Poland) [11]–[15], [26]. The researcher explored using PCTR-BENIARRES (***P**redicción de* ***C**recidas en* ***T**iempo* ***R**eal –* ***B**eniarrés*), Time Lagged Recurrent Neural Network, General Recurrent Neural Network, Optimal Subset Regression (OSR), Back-Propagation Neural Network (BPNN) Adaptive Linear Neuron Network (ADALINE), Elman Network, Black-Box Type Runoff Simulation Model, and Radial-Basis Function Neural Network. From the reference [11]–[15], [26], all results show good accuracy of flood predict. The researcher agrees that their result could be used to help for planning and development infrastructure to decrease flood occurred [7]. Based on the literature, it was found no specific algorithm to solve all flood predictions.

This study utilizes the advantages of ANN with Levenberg-Marquardt with two different ratios that is (80%: 10%: 10%) and (70%: 15%: 15) for training sample, validation sample, and testing sample for flood prediction at Sungai Isap.

4.4 Methodology

ANN is an intelligent process model that inspired biological nervous systems, for an example brain process [16]. This is caused by this network try to model the capabilities of the human brain. ANN is being widely used in many fields of study. World first ANN with neuron is created at 1943 with the ANN has grown faster after the first ANN training algorithm introduced in 1958 [17]. For the past decade, ANN used a theoretically change to the statistical model. A systematic methodological flow chart of the study is presented in Figure 4.3.

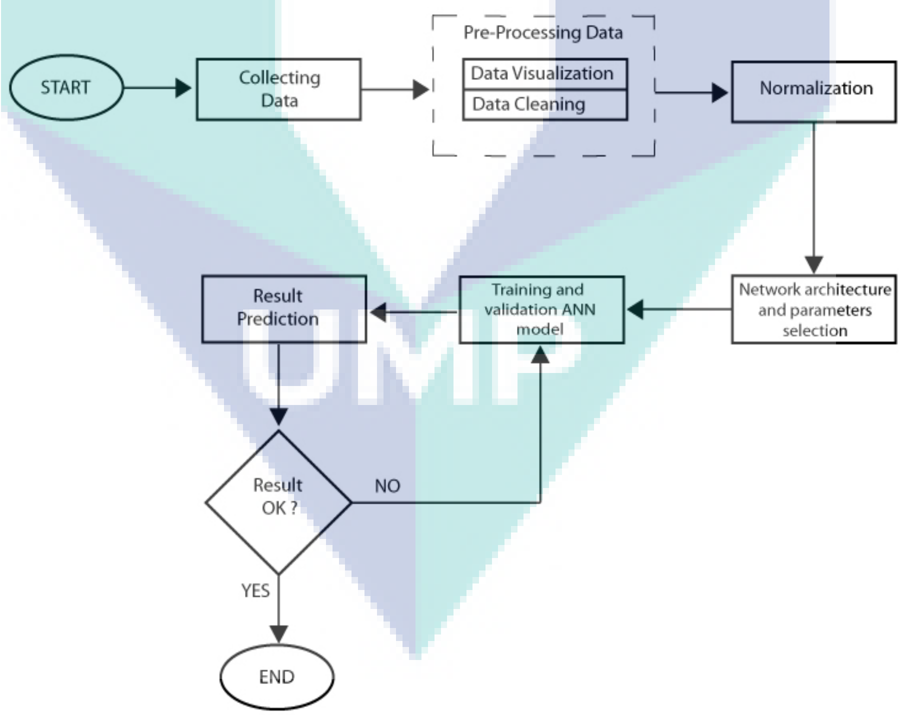

Figure 4.3 Methodological flow chart

4.4.1 Data Collection

Flood conditioning factor data set was constructed by factors is temperature, precipitation, dew point, humidity, sea level pressure, visibility, wind and river level from January 2013 until May 2015. The river level and precipitation data collect from Jabatan Pengairan & Saliran (JPS) Pahang and temperature, dew point, humidity, sea level pressure, visibility, wind is collected from the website (www.wunderground.com).

4.4.2 Data Visualization

Data Visualization is a method that shows the data in pictorial or graphical format with many researcher uses for representing their charts to faster and easily understanding the information from the data [18].

4.4.3 Data Cleaning

The factor data (raw data) of the 8 factors have some missing data, so the SPSS software is used to finds back those missing data, this method is called as data cleaning. Data Cleaning included the discovery and excision or correction of errors [19]. If there are incomplete datas and then is either replaces, modified or deleted the data. Data cleaning process need careful consideration because the influence the results. Cleaning data need always checks and treatment of missing responses will be done Statistical Product and Service Solution (SPSS) software will consistency check and treatment of missing responses that cleaning [20].

4.4.4 Normalization

Because of different data and different magnitude value, the samples need to be scaled in 0-1 using normalization. Normalization is a method that adjusts the value measurement due to the different scale to a collaborative scale and frequently before averaging. Normalization also refers as complex adjustments where the purpose is to bring the whole probability assigned adjusted values into alignment. Equation (3) is a formula for normalization.

$$X' = \frac{X - Xmin}{Xmax - Xmin}$$

(3)

Where X' is the scaled value, X is the sample value, X_{min} and X_{max} are the minimum value and maximum value.

4.4.5 Training Algorithm

ANN training quantity iteratively adapting the weights of a neural network until the connection weights determine an input and output that function that near the relationship between input and output structure of a given training data set. In this research, Levenberg-Marquardt back-propagation used for execution the concept and the lowest MSE the best results. The results in [6], [24], [27], [28], that show Levenberg-Marquardt is better that other training algorithm and the main reasons this paper. Levenberg-Marquardt back-propagation network training function is a Levenberg-Marquardt optimization that updates according to the weight and bias values and faster training algorithm for network of moderate size compare to other algorithms. Levenberg-Marquardt also highly recommended as a first-choice manages algorithms cause it does not require more memory and more time for training. Other than that, it also reduces feature for use when the sample data is large.

The program MATLAB with Neural Network Toolbox is a strong tool for prediction. The Neural Network Toolbox function to develop feed forward back-propagation model. This tool allows importing, creating, using, export neural network and data, and change parameters (number of neurons, learning rate, number of hidden layers, transfer function and performance functions). The structure of an ANN is created by weight between neuron, transfer function, and learning laws. The transfer function is used to control the generations of output in neurons. Learning laws are used to determine the relative weights for input to the neuron. In a feed-forward network, the weighted connections feed activations only in the forward direction from an input layer to the output layer.

4.4.6 Neurons

Neurons manipulate logical parallelism and the information is sent from later to serial operations. If too few neurons will direct affect the network performance and next will outcome in under-fitting and too many neurons the network is over-fitting. For this research, the number of neurons is set by ranging from 10 neurons until 200 neurons.

4.4.7 Layer

A layer is a group that set up by neurons and a group of later can set up a network. The ANN model usually got input layer, a hidden layer, and output layer. In this paper, using the single layer with prepares adequate hidden neurons in a single layer.

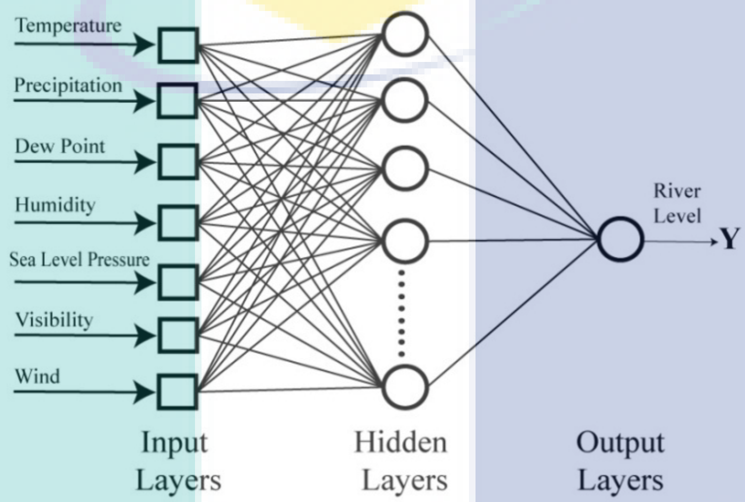

Figure 4.4 Structure of a feed-forward ANN model

4.4.8 The Best Function

To define the best structure, different number of neurons in hidden layers is testing [22]. Mean Squared Error (MSE) and Regression (R) is frequently used to execute the performance evaluation. MSE defines as the average squared difference between observed output values from ANN and targets introduced in the training samples.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(t_i - a_i)^2 \qquad (4)$$

From the Equation 4 that n is sieze of the sample set, a_i is the ANN observed output, and t_i is the corresponding target output. The (R) is defined as the correlation between the targets and output. When R=1, is meant a close relationship, vice versa R=0, there is random relationship. Smaller network weights and fewer effective parameters simplify the network size to avoid the training data overfitting caused by over-complex network structure, it also can improve the generalization of neural network in practical application

The Gradient Decent with Momentum Weight and Bias Learning Function was chosen for Adaption Learning Function for neural network (7-1-1: 7 input, 1 hidden layer, and 1 output) is shown in Fig. 4 structure of feed-forward ANN model and observed to be the best adaptive learning function as it shows results with minimum MSE [6]. Tan-Sigmoid Transfer Function is the best choice for hidden layers over Linear Transfer Function and Log Sigmoid Transfer Function because of its very fast learning rate and sensitivity towards change number of samples and neurons. Other than that, results bring the conclusion that Tan-Sigmoid function gives the best predictions for ANN architecture in terms of goals, epochs and comparison with the target data [23].

4.4.9 Training, Verifying, and Testing

In this paper, 2 different ratios are used that is for first ratio training 80% of input data from January 2013 until May 2015 were applied to training network. In verifying part, the program will stop of the calculation when 10% data are applied to determine the network structure work that was not used in training. Verifying data have checked in a different sequence of training and continued when the number of error reduced in the verifying. The last 10% data will apply to the process after the training process and verifying process finish. For second ratios are (70% Training: 15% Validation: 15% Testing). The Levenberg-Marquardt training algorithms are used to train the network and the corresponding configuration parameters are as shown in Table 4.1.

Table 4.1 The training algorithm configuration parameters

Configuration Parameters	LM
Maximum number of epochs to train	1000
Epochs between displays	25
Performance goal	0
Maximum time to train in seconds	Infiniti
Minimum performance gradient	1e-07
Maximum validation failures	6
Initial µ	0.001
µ decrease factor	0.1
µ increase factor	10
Maximum µ	1e+10
Learning rate	N/A

4.5 Results and Discussion

ANN architecture attributes use Levenberg-Marquardt for the training and testing of the proposed metrics concept. Table 4.2 captures the architecture and ANN attributes used in the experiment, Table 4.3 is the result of mean square error (MSE).

Table 4.2 Architecture and ANN attributes

Network Type	Feed-Forward Back-Propagation (FFBP)
Training Function	Levenberg-Marquardt
Adaptive Learning Function	LEARNGDM
Performance Function	MSE
Number of Layer	01
Number of Neurons	10,20,30,40,50,60,70,80,90,100,110,120, 130,140,150,160,170,180,190, 200
Transfer Function	Tan-Sigmoid
Epochs	1000

The lowest MSE is the better ANN attributes in maintainability prediction, from the Table 4.3 and Table 4.4 the result with the lowest MSE is at Table 4.3. Result MSE with ratio (80%: 10%: 10%) is 0.0019231 using Levenberg-Marquardt with 120 neurons is the best of ANN attributes.

Table 4.3 Result MSE (80:10:10)

Neurons	MSE	Neurons	MSE
10	0.0031911	**110**	0.0070726
20	0.0049203	**120**	0.0019231
30	0.0062713	**130**	0.0028143
40	0.0042255	**140**	0.0029577
50	0.0034647	**150**	0.0031299
60	0.0075899	**160**	0.0034169
70	0.0025577	**170**	0.0021680
80	0.0021512	**180**	0.0030842
90	0.0078860	**190**	0.0030574
100	0.0033275	**200**	0.0035297

Table 4.4 Result MSE (70:15:15)

Neurons	MSE	Neurons	MSE
10	0.0047377	**110**	0.0020686
20	0.0080938	**120**	0.0040221
30	0.0027311	**130**	0.0040819
40	0.0032440	**140**	0.0029030

50	0.0075195	**150**	0.0022756
60	0.0036020	**160**	0.0043467
70	0.0029161	**170**	0.0055751
80	0.0020988	**180**	0.0039078
90	0.0089303	**190**	0.0043990
100	0.0035220	**200**	0.0027019

In Figure 4.5 show that the graph best training performances is 0.0019231 MSE is considered better result that compare to [6] with 6.58 MSE, [29] with 2.75 MSE, [24] with 1.1 MSE, [30] with 0.30265, and [28] with 0.1185 MSE. Fig. 5 show the graph plot solid line represents independent train, dash line is represents test, the dash dot-dot line represent the validation, and the dot-dot line represents best value for this training. In Figure 4.6, shows the regression (R) graph with training, validation, test, and all data. The dotted line in the graph shows the $Perfect\_result - outputs = targets$ and the solid line represent best fit linear regression between the target and the output for Figure 4.6(a) is training data regression, Figure 4.6(b) is validation data regression, Figure 4.6(c) is test data regression, and Figure 4.6(d) is all data regression. The regression (R) is represented the relation between the targets and the outputs that if R=1, meaning that is an accurate linear relationship between targets and outputs. Vice versa if R=0, there is no linear relationship between targets and outputs.

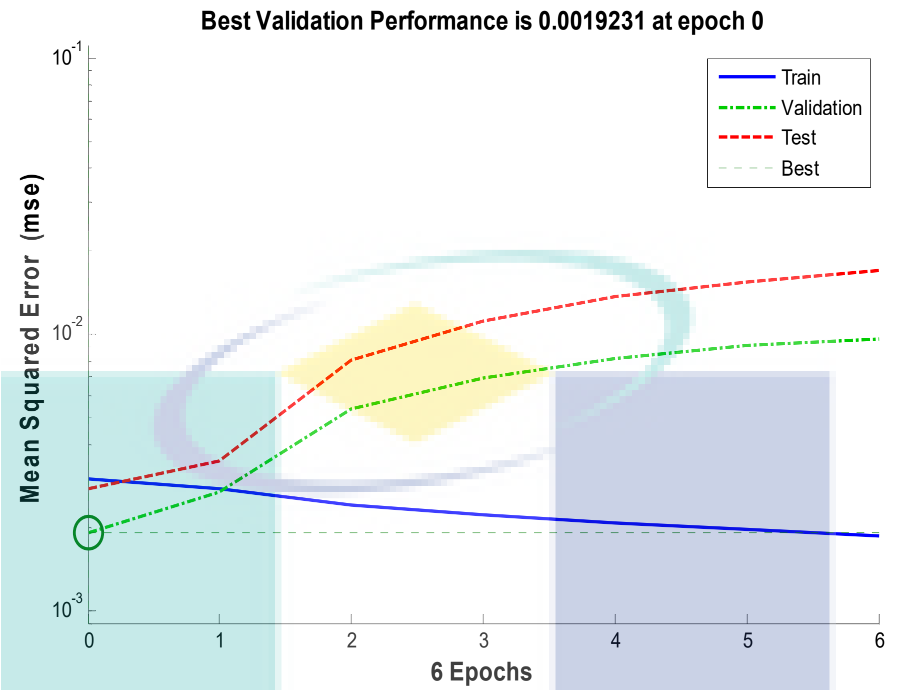

Figure 4.5 Best training performance with Levenberg-Marquardt using 120 neurons

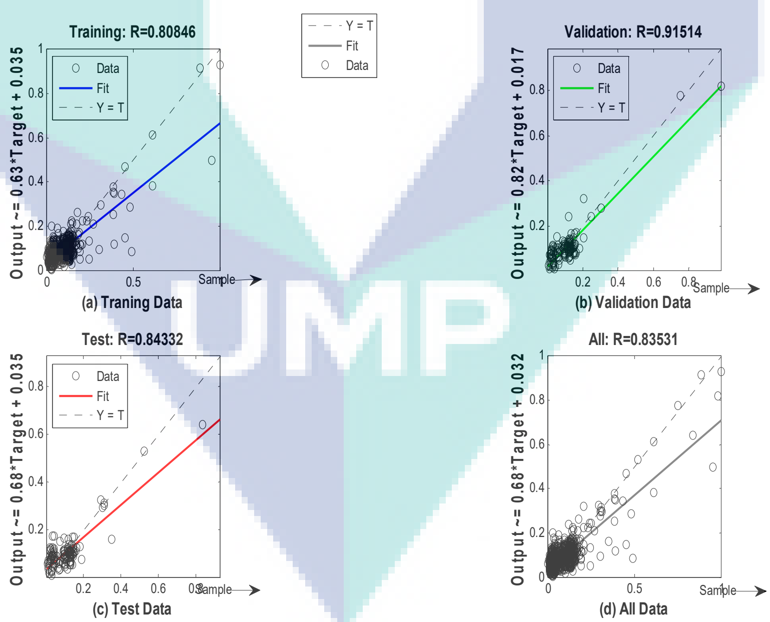

Figure 4.6 ANN regression result Levenberg-Marquardt (a) Training data regression (b) Validation data regression (c) Test data regression (d) All data regression

The training data are giving good result with R for training is 0.80846 with Ourput~=0.63^Target+0.035 at Figure 4.6(a), R for validation is 0.91514 with

Output~=0.82^Target+0.017 at Figure 4.6(b), R for test is 0.84332 with Output~=0.68^Target+0.035 at Figure 4.6(c), and R for All is 0.83531 with Output~=0.68^Target+0.032at Figure 4.6(d). The graph is very important that can to show clearly some of that data points are poor fits. After the training proses finish, the graph in Figure 4.7 shows the results of prediction data and the actual data are shown in the graph. The solid line represents the actual data and the dash line shows the prediction data in Figure 4.7. The perfect way to clearly see the comparison of actual data and predict data is used graph method.

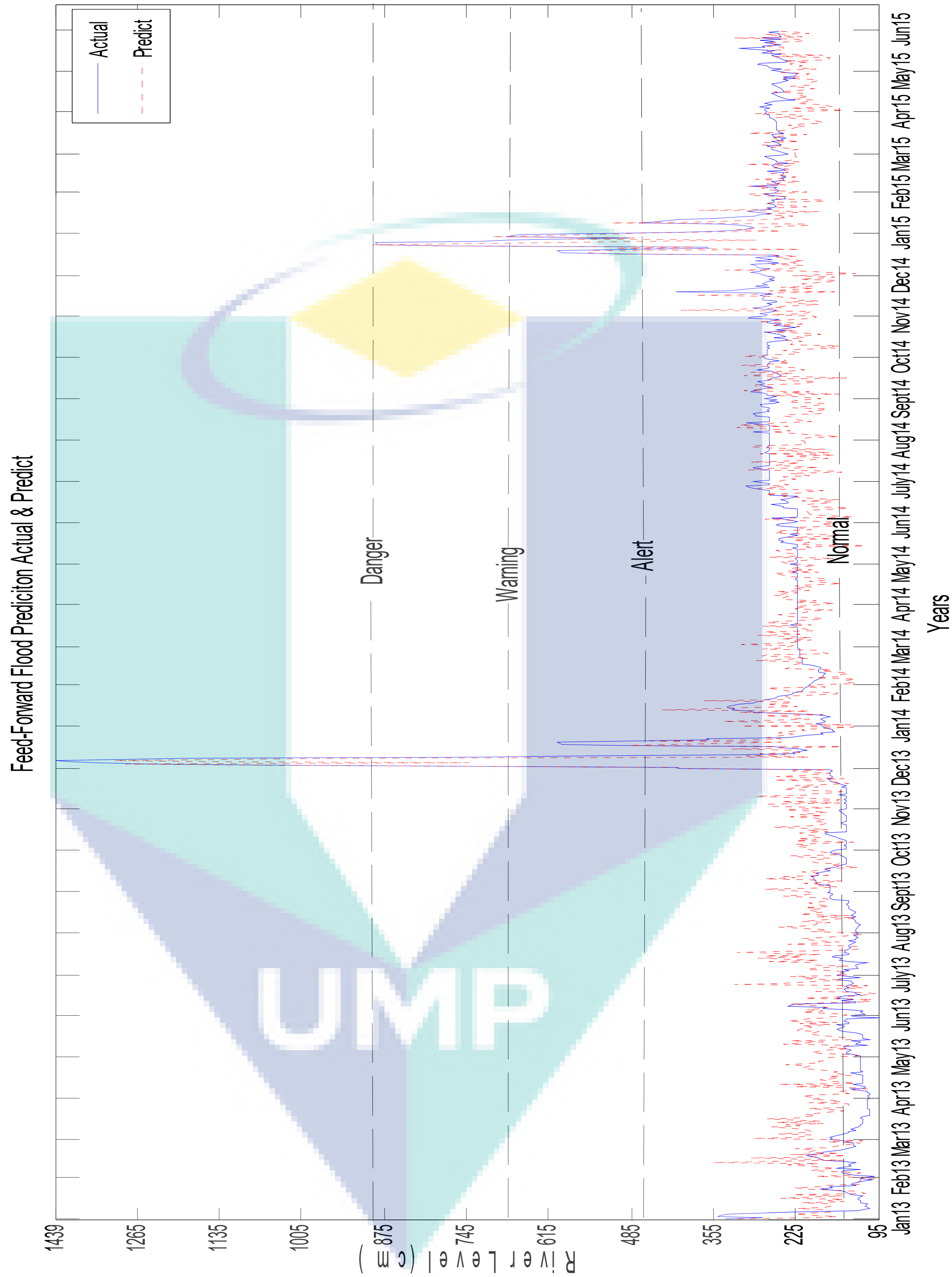

Figure 4.7 ANN Feed Forward Flood Prediction Actual & Predict

4.6 Conclusion

Artificial Neural Network ANN method is used for this project aimed to develop a good prediction result on flood prediction. In order to show the performance Levenberg-Marquardt back-propagation method, sample data from Jan 2013 until May 2015 data has been used. The best results are achieved in ratio (80%; 10%; 10%) with 120 neurons at MSE is 0.0019231, using Levenberg-Marquardt back-propagation. The prediction done successfully u with Training R=0.80846, Validation R=0.91514 and Test R= 0.84332 and All R=0.83531. On the other hand, ANN method has proven to produce satisfying results for flood prediction.

CHAPTER 5

KESIMPULAN

5.1 Kesimpulan Umum

The flood is the most damaging catastrophic phenomena in the world. Over the last decade, the flood becomes the hot topics in the world, even the literature, because this evaluation is a tough and nonlinear problem. Many methods have been tried by researchers in literature, but each method has their weakest point. Flooding on year 2013 caused serious damage in Sungai Isap, Kuantan, Pahang, Malaysia. The goal of this study was to predict flood in term of accuracy and to assist having proper management over the affected area.

For use SVM calculation produced R=0.94785 of the rate prediction. The information from current research can assist JPS, Malaysian Meteorological Department, and government to perform the proper take action immediately. For use ANN, case of maintainability prediction, results are achieved with 60 neurons with the MSE is 0.0007540, using Bayesian Regularization back-propagation. The flood prediction study has been done successfully using ANN to give a model with R=0.87365.

5.2 Cadangan

The prediction using method Support Vector Machine and Artificial Neural Network can been success predict at different state like Kelantan, Terengganu and Johor that is the famous area for flood happening.

REFERENCES

[1] A.F.Presse., “Floods kill 21 in Malaysia, water recede.,” 2014. [Online]. Available: http://news.asiaone.com/news/malaysia/floods-kill-21-malaysia-waters-recede. [Accessed: 01-Jan-2015].

[2] C.C.Attribution-ShareAlike, “Flood,” 2015. [Online]. Available: https://en.wikipedia.org/wiki/Flood#cite_note-1. [Accessed: 01-Jan-2015].

[3] L. B. S. Bratkovich., “Flooding and its effect on trees,” 1993. [Online]. Available: http://www.na.fs.fed.us/spfo/pubs/n_resource/flood/cover.htm.

[4] ESchooltoday, “Types of floods,” 2010. [Online]. Available: http://eschooltoday.com/natural-disasters/floods/types-of-floods.html. [Accessed: 01-Jan-2015].

[5] M. S. Tehrany, B. Pradhan, and M. N. Jebur, “Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS,” *J. Hydrol.*, vol. 512, pp. 332–343, 2014.

[6] A. Sharma and G. Nijhawan, “Rainfall Prediction Using Neural Network,” vol. 3, no. 3, pp. 65–69, 2015.

[7] M. B. Kia, S. Pirasteh, B. Pradhan, A. R. Mahmud, W. N. A. Sulaiman, and A. Moradi, “An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia,” *Environ. Earth Sci.*, vol. 67, no. 1, pp. 251–264, 2012.

[8] F. A. Ruslan, A. M. Samad, Z. M. Zain, and R. Adnan, “Flood prediction using NARX neural network and EKF prediction technique: A comparative study,” *2013 IEEE 3rd Int. Conf. Syst. Eng. Technol.*, pp. 203–208, 2013.

[9] R. Adnan, F. A. Ruslan, A. M. Samad, and Z. Md Zain, “Flood water level modelling and prediction using artificial neural network: Case study of Sungai Batu Pahat in Johor,” *2012 IEEE Control Syst. Grad. Res. Colloq.*, no. Icsgrc, pp. 22–25, 2012.

[10] R. A. Fazlina Ahmat Ruslan, Abd Manan Samad, Zainazlan Md Zain, “Prediction of 4 Hours Ahead Flood Water Level Using Improved ENN Structure: Case Study Kuala Lumpur,” *2014 IEEE Int. Conf. Control Syst. Comput. Eng. 28 - 30 Novemb. 2014, Penang, Malaysia*, 2014.

[11] R. P. Deshmukh, “Short Term Flood Forecasting Using General Recurrent Neural Network Modeling a Comparative Study,” vol. 8, no. 12, pp. 5–9, 2010.

[12] Y. Liu, Y. Chen, J. Hu, Q. Huang, and Y. Wang, “Long-term Prediction for Autumn

Flood Season in Danjiangkou Reservoir Basin Based on OSR-BP Neural Network," no. 2006, pp. 2006–2009, 2010.

[13] D. F. Lekkas, C. Onof, M. J. Lee, and E. a Baltas, "Application of Artificial Neural Networks for Flood Forecasting," *Glob. Nest Int. J. 6*, vol. 6, no. 3, pp. 205–211, 2004.

[14] E. Toth and A. Brath, "Flood Forecasting Using Artificial Neural Networks in Black-Box and Conceptual Rainfall-Runoff Modelling," pp. 166–171.

[15] a. Piotrowski, J. J. Napiórkowski, and P. M. Rowiński, "Flash-flood forecasting by means of neural networks and nearest neighbour approach – a comparative study," *Nonlinear Process. Geophys.*, vol. 13, no. 4, pp. 443–448, 2006.

[16] M. G. Naik and V. S. B. Radhika, "Time and Cost Analysis for Highway Road Construction Project Using Artificial Neural Networks," pp. 26–31, 2015.

[17] R. Adnan, F. A. Ruslan, A. M. Samad, and Z. M. Zain, "New Artificial Neural Network and Extended Kalman Filter hybrid model of flood prediction system," *Signal Process. its Appl. (CSPA), 2013 IEEE 9th Int. Colloq.*, pp. 252–257, 2013.

[18] S. Institutte Inc., "Data Visualization.What it is and why it is important." [Online]. Available: http://www.sas.com/en_us/insights/big-data/data-visualization.html.

[19] BetterEval:World, "Data Cleaning," 2015. [Online]. Available: http://betterevaluation.org/evaluation-options/data_cleaning. [Accessed: 01-Jan-2015].

[20] U. Spss, "Cleaning your data Using SPSS to clean your data," pp. 22–25.

[21] L. V Kamble, D. R. Pangavhane, and T. P. Singh, "Neural network optimization by comparing the performances of the training functions -Prediction of heat transfer from horizontal tube immersed in gas–solid fluidized bed," *Int. J. Heat Mass Transf.*, vol. 83, pp. 337–344, 2015.

[22] S. kumari Sharma, "Anticipating milk yield using artificial neural network," vol. 3, no. 3, pp. 690–695, 2014.

[23] M. Barello, D. Manca, R. Patel, and I. M. Mujtaba, "Neural network based correlation for estimating water permeability constant in RO desalination process under fouling," *Desalination*, vol. 345, pp. 101–111, 2014.

[24] S. Pandey, D. A. Hindoliya, and R. Mod, "Artificial neural networks for predicting indoor temperature using roof passive cooling techniques in buildings in different climatic conditions," *Appl. Soft Comput. J.*, vol. 12, no. 3, pp. 1214–1226, 2012.

[25] F. A. Ruslan and R. Adnan, “Flood modelling using Artificial Neural Network,” *2013 IEEE 4th Control Syst. Grad. Res. Colloq.*, pp. 116–120, 2013.

[26] R. García-Bartual, “Short term river flood forecasting with neural networks,” *iEMSs 2002 Int. Congr. Integr. Assesment Decis. Support*, pp. 160–165, 2002.

[27] V. Kumar, R. Kumar, and A. Sharma, “Maintainability Prediction from Project Metrics Data Analysis Using Artificial Neural Network: An Interdisciplinary Study,” *Middle-East J. Sci. Res.*, vol. 19, no. 10, pp. 1412–1420, 2014.

[28] F. Dalipi and S. Y. Yayilgan, “The Impact of Environmental Factors to Skiing Injuries : Bayesian Regularization Neural Network Model for Predicting Skiing Injuries,” 2015.

[29] K. Abhishek, M. P. Singh, S. Ghosh, and A. Anand, “Weather Forecasting Model using Artificial Neural Network,” *Procedia Technol.*, vol. 4, pp. 311–318, 2012.

[30] M. Nasr, M. Ateia, and K. Hassan, “Artificial intelligence for greywater treatment using electrocoagulation process,” *Sep. Sci. Technol.*, vol. 6395, no. July 2015, p. 150707112432002, 2015.

APPENDIX A
SAMPLE APPENDIX 1

For Appendices Heading, use TITLE AT ROMAN PAGES style.