UNIVERSITY OF NAIROBI

SCHOOL OF COMPUTING AND INFORMATICS

Machine Learning for Flood Forecasting:

Case Study: Nzoia River Basin, Western Kenya.

By

Kuria, Martin Wainaina

A research project submitted in partial fulfillment for the Requirements for the award of Degree of
Masters of Science in Computer Science at the
School of Computing and Informatics
of the University of Nairobi

August 2014

DECLARATION

This research project is my original work and has not been presented for a degree in any other University

Signed: ___________________________________ Date: _________________________

Kuria, Martin Wainaina
P58/70599/2008

APPROVAL

This research project has been submitted for examination with my approval as university Supervisor

Signed: ___________________________________ Date: _________________________

Dr. Lawrence Muchemi
Lecturer
School of Computing and Informatics
University of Nairobi

DEDICATION

This research project is dedicated first to God for the strength He gave me to carry on, my family and my good friend John Muita.

ACKNOWLEDGEMENT

I acknowledge the precious guidance throughout this research work for my supervisor, Dr. Lawrence Muchemi. Special thanks to Dr. Dan Orwa for his patience and for providing me with information I needed. I am also grateful for the guidance and useful comments from the members of the panel, Dr. Orwa, Dr. Muchemi, Prof. Waema, and Dr. Abade. I also want thank Mr. Abraham Changara, Chief Meteorologist, Kenya Meteorological Department for allowing access to the required meteorological data. Special thanks also go to NeuroDimension, Inc. for providing a neural network software package (NeuroSolutions) that was used in developing my models in this research study.

ABSTRACT

In Kenya property destruction and loss of life has occurred due to serious incidents of floods, along the Nzoia River catchment area Western Kenya. Despite having flood warning models along the Nzoia River basin; with a flood warning system at Rwambwa gauge station that sends out alerts on the river levels. These models are linear models and have overlooked the peak streamflows. A reliable intelligent nonlinear model that is capable of handling nonlinear estimation streamflow (discharge) problem is crucial in flood control operations.

This research explores applicability and performance of flood forecasting models in the Nzoia River basin, Western Kenya, using two types of artificial neural network (ANNs), namely MLP-ANN-FF a feedforward multilayer perceptron (MLP) network and GA-ANN-FF a genetic algorithm optimized multilayer perceptron feedforward neural network model. The aim of this study is to compare the performance of these two models (MLP-ANN-FF and GA-ANN-FF) and recommend the most suitable for this problem.

The historical daily rainfall, and average temperature and discharge flow, obtained from Kenya Metrological Department (KMD) were used as inputs to the two ANN models for discharge flow (streamflow) forecast for Nzoia River basin at Rwambwa river gauge. The characteristic parameters such as number of neurons within hidden layers and the selection of input variables for the MLP-ANN-FF were optimized using genetic algorithm (GA), hence yielding a GA-ANN-FF model. These two models were trained, cross verified and tested with daily rainfall, average temperature, and discharge flow.

The architectural topology that trained well on MLP-ANN-FF model was one with 9 input variables, 2 hidden layers and 1 discharge flow output; a 9:7:12:1 configuration setting. This was later optimized with the genetic algorithm (GA) to develop a GA-ANN-FF model that was able to optimize the input variables reducing them from 9 to 4 inputs, and reducing the number of neurons in the 2 hidden layers yielding a 4:6:4:1 GA-ANN-FF model.

The conventional ANN (MLP-ANN-FF) and a GA-ANN-FF model were used as the benchmark

for comparison of performances. The two models were developed using NeuroSolutions Software and were trained with 70% of the data set, 20% for cross validation and the remaining 10% was used in testing the overall performance of the models. The results revealed that the GA-ANN-FF (4:6:4:1) model was able to yield better accuracy in performance for Nzoia River basin at Rwambwa River gauge, with least input variables, and number of neurons in the hidden layers though it took longer on the computation time. With a MSE of 0.021 and an r (correlation coefficient) of the desired and estimated discharge flow of 0.887 (89%), GA-ANN-FF performed satisfactory better than MLP-ANN-FF (9:7:12:1) with 9 input variables an MSE of 0.024 and r (correlation coefficient) of 0.84 (80%).

The results showed that ANN integrated with GA has a better accuracy and therefore most suitable in developing flood forecast models with low MSE. This finding is important because it will eventually enable relevant agents in water resource planning and flood management and the public aware when a flood might occur and the areas that would be affected to avoid disaster caused by floods.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

ANN	-	Artificial Neural Networks
BP	-	BackPropagation
BPN	-	BackPropagation Network
DOC	-	Disaster Operation Center
FDFC	-	Flood Diagnostics and Forecasting Centre
FEWST	-	Flood Early Warning System Team
FF	-	Feed-Forward
GA	-	Genetic Algorithm
GA-ANN-FF	-	Genetic optimized multilayer Perceptron Feed-Forward
GDP	-	Gross Domestic Product
GIS	-	Geographic Information Systems
GRN	-	General Recurrent Neural Networks
HN	-	Hydrodynamic numerical
KMD	-	Kenya Metrological Department
LVQ	-	Learning Vector Quantization
MAE	-	Mean Absolute Error
MLFN	-	Multi layer feed forward nets
MLP	-	Multilayer Perceptron neural network
MLP-ANN-FF	-	Multilayer Perceptron Feed-Forward
MSE	-	Mean Square Error

MATLAB	-	Matrix Laboratory
NMSE	-	Normalized Mean Square Error
NN	-	Neural Networks
r	-	Correlation Coefficient
R^2	-	Coefficient of Determination
RBF	-	Radial basis function neural network
RGS	-	Rwambwa River Gauge Station
SOM	-	Self-Organizing Map
SD	-	Standard Deviation
SID	-	Station ID (Gauge station code number)

Chapter 1 - Introduction

1.1 Background

Kenya has experienced serious incidents of floods, in different parts of the country, destroying property and resulting in loss of life. Occurrence of floods is due to natural factors like flash floods, river floods and coastal floods. Torrential rainfall has been the major cause of floods in Kenya. Nzoia River experiences perennial flooding in its lower reaches affecting areas such as Budalangi and Kano flood plains. ("Flood Mitigation Strategy," 2009)

The floods affecting Kenya are becoming increasingly predictable, with major floods occurrence in year 1961, 1997, 1998, and 2003 ("Kenya water security and climate resilience project," 2013). Due to the high inter-annual and intra-annual rainfall variability that results in frequent droughts and floods, major infrastructure investments for economic growth are getting damaged with single extreme flood events ("Flood Mitigation Strategy," 2009). This has greatly impacted Kenya economy with a 2.4% of GDP cost every year, according to a 2006 World Bank report ("Kenya water security and climate resilience project," 2013).

The major rivers in Lake Victoria that experience floods have dykes installed. River Nzoia dykes measure 34.4 Km. During the long rains these dykes shield residents in Busia and Budalangi area in western Kenya from floods (Onywere et al., 2007). Although there several studies that have been done in Nzoia Basin in trying to find flood mitigation solutions effective structural and nonstructural mitigation measures have no yet being achieved (ADCL (Appropriate Development Consultants Limited), 2006).

Hydrologists are involved in the research and development of hydrological processes such as flood forecasting. The flood forecast models used in hydrology field are divided into physically based rainfall-runoff and data-driven models or a combination of both (Kia et al., 2012). Hydrodynamic models require accurate river geometric data, which might not be available in many locations. Previous studies have explored more on data-driven approach, since some of the hydrology models have been unable to understand the dynamic changes inside the river basin, and require robust optimization tools (Plate, 2009).

Flood forecasting models assist in anticipating flood occurrence thus allowing sufficient time for action. Modeling has proved to be an important item in predicting floods; from physical based models to data mining models, this has driven researchers to use different approaches in modeling flood forecasting models (Onyari and Ilunga, 2010)

In data mining lots of studies have been carried out especially in artificial neural networks (ANN) for flood forecasting modeling. According to Tom M. Mitchell Artificial neural networks (ANN) is "a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" (Mitchell, 1997). ANN is suitable for data classification, function approximation and pattern recognition, through a learning process (Mitchell, 1997). In recent years a lot of research has been done in hydrological modeling in respect to ANN to predict and forecast water resources variables. This is has been necessitated by the complexity of hydrological models unable handle excessive requirement of field data in case of physically based models rendering such models less attractive in flood forecasting. This is the reason why non-linear ANN model is being extensively used as flood forecasting to model non-linear relationships (Tingsanchali, 2000).

The variability of a good flood forecasting model depends mostly on the flood forecasting approach used. In this research the aim is to model multilayer perceptron feed-forward (MLP-ANN-FF) and a genetic algorithm (GA) optimized multilayer perceptron feed-forward (GA-ANN-FF) for flood forecasting using metrological data (rainfall, temperature, and the discharge flow) to estimate the discharge flow of Nzoia River at Rwambwa River gauge. A case study for Nzoia River Basin at water level station Rwambwa River Gauge Station (RGS) in Western Kenya will demonstrate that the two models, using historical data; daily rainfall, average temperature, and discharge flow, can estimate the streamflow. Daily rainfall, temperature and discharge data of 2000 to 2003 obtained from Kenya Metrological Department (KMD) and was used in training, testing and determining the performance of the models.

Though Kenya, has installed river-gauging stations in the rivers, for flood monitoring. There has been irregular monitoring and communication of floods situation. It's through the rainfall forecast and warning from Kenya Meteorological Department (KMD) that communicates the forecasts to the Disaster Operation Center (DOC) under the Ministry of Special Programme in the office of the President, who mobilizes various County Governments for rescue and relief operations. Due to lack of advance warning of impending floods the public is always caught unaware, leaving no time to take preventive measures. The existing flood management monitoring mechanism is in rescue and relief measures and not preventive action. Nobody is directly focused about proper and fast communication system between general public, County Governments and National Government ("Kenya water security and climate resilience project," 2013).

1.2 Problem Statement

In 2007, a Flood Early Warning System Team (FEWST) through the strategic intervention initiated by the government for Nzoia River basin was started. A number of models on early flood warning were started giving discharge forecast. Although the monitoring has improved these models are linear models and have overlooked the peak streamflows (Masibayi et al., n.d.). Since streamflows process for the daily discharge is generally recognized as nonlinear and seasonal (Guven, 2009), reliable intelligent nonlinear transfer function that is capable of handling nonlinear estimation streamflow (discharge) problem is crucial in water resource planning and flood management. These early warming models initiated by FEWST for discharge forecast lack intelligence.

In recent years a lot of research has been done in hydrological modeling in respect to ANN with BP to predict and forecast water resources variables, with a good success when used to estimate the discharge flow (Tingsanchali, 2000). The issues that limit the performance of ANN with BP are risk of network overfitting the training data, and trial and error in selecting the optimal inputs and determining the optimal number of neurons in the hidden layers.

This project is aimed at alleviating some of these issue such as removing trial and error methods of setting parameters in ANN by employing a hybrid genetic algorithm (GA) with ANN (GA-ANN-FF model) to enhance performance. The model can assist in predicting streamflows given historical data based on daily rainfall, temperature, and discharge flow. Such tools may also provide reliable intelligent estimation of Nzoia River Basin streamflow (discharge) that will enable relevant agents in water resource planning and flood management and the public aware when a flood might occur and the areas that would be affected to avoid disaster caused by floods. It's therefore important to have a model that can provide reliable discharge forecasts by applying historical data that is available from KMD. The use of machine learning can assist in developing of such models.

1.3 Objectives

The main objective of this research is to build two ANN models (MLP-ANN-FF and GA-ANN-FF) for purpose of estimating the discharge flow of Nzoia River basin at Rwambwa River gauge, by applying artificial neural network (ANN) technique and optimizing the multilayer perceptron (MLP-ANN-FF) neural network model with a genetic algorithm (GA).

The objectives are summarized below

1. Design a hybrid algorithm of ANN and GA
2. Develop a prototype based on the hybrid ANN with GA (GA-ANN-FF model) for purpose of estimating the discharge flow of Nzoia River Basin at Rwambwa River gauge.
3. To evaluate and benchmark the hybrid ANN with GA algorithm by performing test on the GA-ANN-FF prototype and compare the results from MLP-ANN-FF base prototype.
4. Recommend a suitable intelligent model based on the results of the two above objectives.

1.4 Assumption

This research has been carried out under the following assumptions. The methodology applied will predict the discharge flow of Nzoia River at Rwambwa River gauge given historical data on daily rainfall, temperature and discharge flow. The research also is based on simulations and experiments. In this research there only three input causative discharge flow factors, namely rainfall, temperature and the discharge (streamflow). The other discharge flow causative flood factors such, flood plain in the past, terrain elevation, water density, water blockage, sub basin areas, soil drainage capability, land use, were not considered. The training, verification and testing of both models is done offline. The flood forecasting models used data from Kitale metrological (SID-8834098), Kitale soil conservation service – office (SID-8834097), Leissa farm – Kitale (SID-8835039), and Rwambwa stations since they were within Nzoia River basin our catchment area of interest. As a proof of concept the research study will cover Nzoia River basin in Western Kenya, but results obtained may be used in other flood prone areas.

1.5 Limitations

ANN is data dependant, they learn well with large volumes of datasets (Babinec and Pospíchal, 2009). Although there was vast data available for period 1975 to 2012, from the three weather stations obtained from Kenya Metrological Department (KMD), and the raw discharge flow provided by Flood Diagnostics and Forecasting Centre (FDFC) at Kenya Meteorological Department for Rwambwa River Gauge Station (RGS). Only 4 year data for period 2000 to 2003 that was concurrent in all three weather stations and the one river gauge was used.

It was not be possible to integrate the developed models (MLP-ANN-FF and GA-ANN-FF) with the current early flood warning models managed by the Flood Early Warning System Team (FEWST) for Nzoia River basin.

1.6 Significance of the study

Currently, the Kenyan Government has initiated a Flood Early Warning System Team (FEWST) that has started a several models on early flood warning for the discharge forecast at Nzoia River basin. These models are linear models that issue alerts when the river levels at the river gauges reach a certain threshold of the water levels that floods might occur if flood level is between certain range (Masibayi et al., n.d.).

For lack of reliable intelligence and nonlinearity in these models, an intelligent algorithm, the GA and ANN with nonlinear transfer function, can be used in developing such models that can assist metrologists in Kenya Meteorological Department (KMD) advice the County Governments and threatened residence on when and where the next flood is going to happen and what areas are going to be inundated due to such events. The models may also benefit the Kenya RedCross who can mobilize various County Governments for rescue and relief operations. The hydrologist can easily integrate these models with their hydrological processes for flood.

Therefore there is a need to develop ANN machine learning technique trained on patterns for flood forecasting integrated with GA technique to enhance the discharge flow predictability in order to prevent loss of lives and minimize damages.

1.7 Chapter Summary

Chapter 1 presented the background of flood forecasting and discharge flow in hydrological modeling in respect to ANN to predict and forecast water resources variables. The problem statement, research objectives, assumption, limitations and significance of the study were also review and discussed.

The rest of the document is organized in four major sections. A review of current literature of study, previous work and approach of ANN and context of the GA with ANN hybrid model is discussed in Chapter 2. The methodology models architecture, data set or source, study area, flood simulation software to be used and design of the models is discussed in Chapter 3. The Evaluation of the Models, Results and discussion of the models is discussed in Chapter 4. Finally the Conclusions and recommendation is discussed in Chapter 5 followed by the reference and appendices.

Chapter 2 - Literature Review

2.1 Flood Forecasting

Over the years lots of research has been undertaken in the development of hydrological flood forecasting models. The aim into these models is to provide timely and accurate future discharge conditions at particular watershed. The most common models in flood forecasting applied by hydrologist are the physically based rainfall-runoff modeling approach, and data-driven model or a combination of both (See et al., 1997). The physically rainfall-runoff model is based on mathematical model. It uses a forecast updating intelligence; reviewing certain reference to state correction and error prediction approaches to improve its performance (Moore et al., 2001). An example of this type of physically based rainfall-runoff model is the European Hydrological System (SHE) (Maskey, 2004). In a data-driven model it tries to map data to form a pattern that best defines a certain particular data set. It has properties of linear regression model, but also boasts of complicated nonlinear models such as artificial neural network, fuzzy rule-based systems just to mention a few. (Modeling Uncertainty in Flood Forecasting Systems by Shreeda Maskey pg12)

2.2 Artificial Neural Networks

Artificial neural networks (ANN) is a mathematical model loosely designed based on the functioning of a human brain (Onyari and Ilunga, 2010). The simplest kind of neural network (NN) mainly used for illustrative purposes is known as perceptron. It's a neural network with no hidden layer, where the inputs units are directly connected to the output units; only capable to learn linearly separable functions (Mitchell, 1997). There various type of ANN architectures such as Radial basis function neural network (RBF), multilayer perceptron neural network (MLP), self-organizing map (SOM) and Learning Vector Quantization (LVQ) (Cho and Park, 2002). The most commonly used of these is the feedforward network, or multilayer feedforward network (MLFN) (LUK et al., 2001). In general it's composed of three layers; input, output and hidden layer. The input layer is made up of a set of input units, where when given some input from an example will propagate through the network producing an output. Weighted sum of the output from the input units forms the input of every hidden unit in the hidden layer. The output or target layer, which consist of a set of output units get its input from a weighted sum of the output from the hidden units. In the node of the output layer the new weighted sum from the hidden layer is computed after de-normalization of the output, the sought forecasted value might be determined. Fig 2.0 shows the topology of a feedforward network, or multilayer feedforward network (MLFN) that includes an input layer, one hidden layer and an output layer. Information is passed forward only.

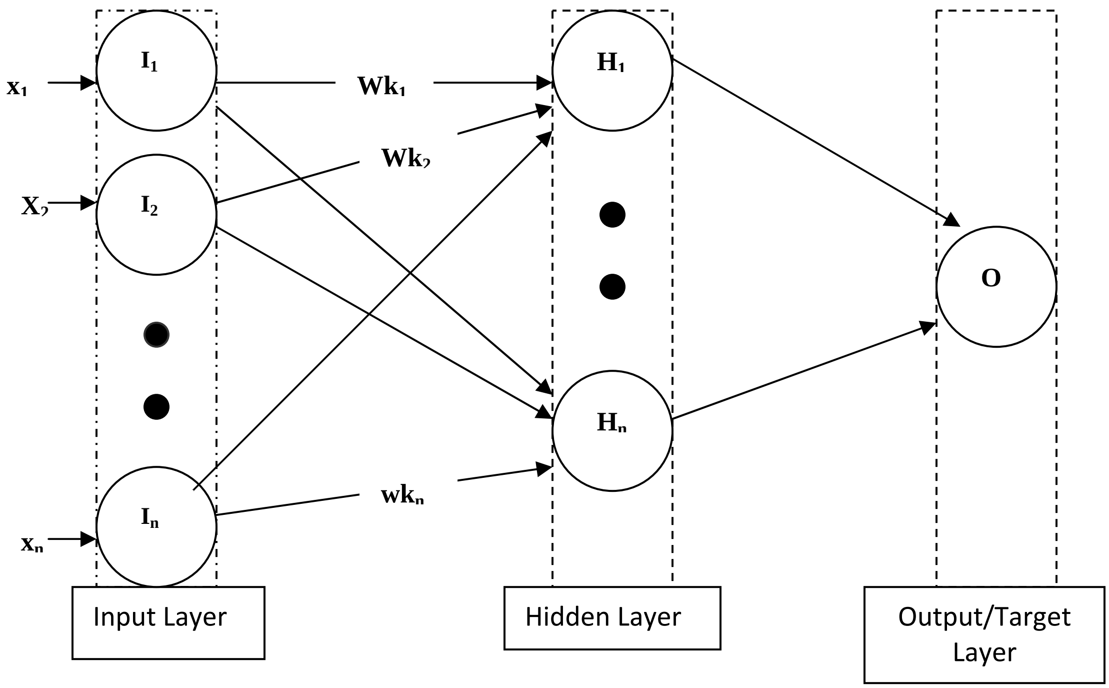


Fig 2.1 - A simple three-layer Artificial Neural Networks Topology

The main steps involved in the design and development of ANN are training and testing. ANN model should always be trained properly before it's used for testing. The training process is through adjusting weights between the nodes until the network is able to predict the target output (Heednacram, 2014). MLFN is trained in a supervised manner using a supervised backpropagation (BP) algorithm, that involves two reciprocal steps; forward pass and backward pass (Puttinaovarat et al., n.d.) MLFN uses this BP algorithm to adjusts the weights and biases of the network in order to minimize the error between its output and the target (over all output and all examples). Fig 2.0 shows the schematic diagram of the backpropagtion neural network; the hidden layer can be more than one (Onyari and Ilunga, 2010). The backpropagation algorithm details are presented in Table 2.0.

Table 2.1: Backpropagation algorithm for a Multilayer feedforward neural network (Mitchell, 1997)

Construct a Multilayer feedforward neural network with the desired number of hidden and output units
Initialize all network weights to small random values()

For each training example, present the training example to the network

Propagate the input forward through the network:

1. Input the instance x to the network and compute the output O_u of every unit u in the network. Mathematically this can be expressed with the following formula

$$O = S_1(\sum O_h.w_h+w_o),$$

Where

O is the output from the ANN,

O_h is the output value of the h_{th} hidden unit

$$O_h = S_2(\sum x_I.w_{Ih}+w_{ho}),$$

X_I are the inputs to the MLFN

w_h are the connection weights between nodes of the hidden and output layer

w_{Ih} are the connection weights between nodes of the hidden and input layer

S_1 and S_2 are activation functions. The most commonly used activation function is a logistic sigmoid function

$$S(x) = 1/1 + e$$

Propagate the error backward through the network:

2. For each network output unit k, calculate its error term δ_k

$$\delta_k \text{ -> } O_k(1-O_k)(t_k-O_k)$$

3. For each hidden unit h, calculate its error term δ_h

$$\delta_h \text{ -> } O_h(1-O_h) \sum w_{kh}\, \delta_k$$

4. Update each network weight w_{ij}

$$W_{ji} \text{ <- } w_{ji} + w_{ji}re$$

Where

$$W_{ji} = n\, \delta_j\, x_{ji}$$

Backpropagation is a gradient descent algorithm; it will always convergences towards a solution by minimizing the error of the network. The issues with BP are on the risk of the network overfitting the training data. Overfitting (overtraining) occurs when the model begins to memorize the training data rather than learning to generalize from trends (Awan et al., 2012). Some of the ways the BP networks avoids overfitting in order to improve network generalization is though use of large network enough to provide an adequate fit. Unfortunately it impossible to know ahead how network layer should be for specific application, more so with large network you might run into more complex function while building the network (Mitchell, 1997).

Some attempts that have been made to stop the BP algorithm becoming stuck in "local minima" as a learning of the network process (Che et al., 2011). Genetic Algorithm (GA) has been able to evade "local minima" by searching in several points simultaneously; determining a good set of weight through performance value with no need of gradient information (NirmalaDevi et al., 2009). GA is from the field of AI. It is a directed random search technique based on the concept of evolution. The search starts from random points and slowly converges to a solution (Yao and Liu, 1996). Incorporating of GA on ANN builds a hybrid ANN model with evolutions adaptation on the architecture, learning and connection weights of the network. Evolution of the architecture enables the ANN to adapt on topologies to different task with no external intervention. Connection weights apply adaptive and global approach to training (NirmalaDevi et al., 2009).

Table 2.2: Genetic algorithm for a Multilayer feedforward neural network (Perez, n.d.)

Genetic Algorithm pseudo-code
1). Generate initial population - Chromosome (string or individual) encoding, weights (and biases) in the NN are encoded as a list of real numbers - The GA starts by generating random generation of population (solution) of chromosomes. - We encode weights using binary weight encoding ***2). While (! solution)*** The algorithm then proceeds by performing cyclic variation and combination of initial population, searching for the best solution. a). Evaluate the fitness of all the chromosomes of the population By applying the Evaluation Function; • Assign the weights on the chromosome to the links in a network of a given architecture, run the network over the training set of examples, and return the sum of the squares of the errors. • At each evolution the output chromosomes are obtained by employing genetic operators (mutation, selection and crossover); to the input population and evaluating using fitness function the goodness of the new generated solution. The fitness function role is to give a

technique to eliminate worst chromosomes from the population (best problem solution)

b). The best chromosome will be selected to reproduce using mutation and crossover

c). Substitute the worst chromosome of the previous generation by the new produced chromosome

Finally the fittest chromosome will be selected as a solution.

The significance of training the neural networks using GA can be appreciated since GA will train the network no matter how its connected (feedforward or feedback network), unlike BP which trains certain restricted topologies and type of network (NirmalaDevi et al., 2009). When GA is incorporated into ANN the model may improve its performance by taking advantage of the characteristics of both ANN and GA.

Testing process applies an independent test data set, which has not being used in the training. The test data set is used for checking or evaluating the overall performance of the neural network (NN). Commonly used performance criterion are MSE (Mean Square Error), NMSE (Normalized Mean Square Error), r (Correlation coefficient), root mean squared error, coefficient of determination (R^2) (Deshmukh and Ghatol, 2010a).

In the water research arena, ANN has been applied to predict likelihood of impending floods and to determine water consumption. The success in these areas has lead ANN models to be extensively used in other water management areas such as river salinity, water table fluctuations, rainfall-runoff processes just to mention a few (Suliman et al., n.d.).

2.3 Review of ANN flood forecasting models

(Puttinaovarat et al., n.d.) Investigates on available techniques such as Multi Layer Perceptron (MLP) and Radial Basis Function (RBF) to improve generic ANN that only uses rainfall data for flood prediction in the Pathumani area in Thailand .RBF and MLP were used to develop two ANN flood prediction models. A comparison on both models was also done against flood data as it occurred in 2011. A Back propagation learning algorithm was used in training the models. More reliable and current GIS data derived during actual flooding in 2006, 2010, and 2011 when the city was most affected by floods. In developing the ANN model, nine flood factors (rainfall, flood plain in the past, terrain elevation, water density, water blockage, sub basin areas, soil drainage capability, land use, monthly rainfall) were selected based on public hearing from the region. During the data processing of the nine factors min-max normalization was applied, where data was rated and normalized and trained using WEKA software. The application of MLP and RBF and inclusion of nine other flood factors on top of the standard rainfall factor, improved the predictability to 70-95% accuracy. MLP did better with accuracies of 71.3, 78.1 and 80.85 percents compared to 74.45 and 81.05 of RBF. The flood forecasting ANN model in both RBF and MLP was able to improve the generalization and accuracy of the model. The improved ANN predication model has been used for flood hazard and risk assessment.

(Deshmukh and Ghatol, 2010a) developed an ANN model for short term flood forecasting for the upper area of Wardha River in India. A comparison in the performance between Jordan and general recurrent neural networks (GRN) models and their application in real time predication of short term flood was also presented. Jordan model uses past output of context unit with present inputs to create memory trace. Unlike MLP that relies on static data, GRN uses temporary data. The methodology applied in training and generalization of these two models was through three performance measures namely MSE (Mean Square Error), NMSE (Normalized Mean Square Error) and r (Correlation coefficient). In the development of the two ANN models historical data from Wardha River and real time rainfall on hourly basis from eight telematic automatic rain gauging stations was used. After splitting the data into three (training, validation and test). Jordan and GRN were trained through 5000 epochs. GRN having unlimited memory depth emerged with a better performance over Jordan since Jordan weighting over time is not flexible. This study sought GRN as a good solution in short term flood forecasting.

(Shrestha et al., 2005) investigates techniques for improving generalization of MLPN Artificial neural network by using different activation functions; sigmoidal, hyperbolic tangent, linear, and a combination of hyperbolic tangent and linear functions in the Neckar River in Germany. The data used in training the network was from historical flood data sets. This data set was divided into three; training set that consisted of flow time series from the 1998 flood events and validation and test data that consisted of flood event data from 1990 and 1993. For training the network, backpropagation algorithm with the Levenberg-Marquardt approximation was used. It also applied four different activation functions (sigmoidal, hyperbolic tangent, linear, and a combination of hyperbolic tangent and linear functions). The network was designed using MATLAB neural network toolbox, where test data set was used for evaluation of the model performance. To optimally train the ANN model the river reach was divided into three ANN blocks. ANN was found to perform better compared to hydrodynamic numerical (HN) model, providing and efficient flood flowing forecasting. In the improvement of generalization of the ANN model, a combination of a hyperbolic tangent and linear transfer functions in the hidden layer provided the best performance.

(Thirumalaiah and Deo, 1998) presents a real time neural networks flood forecasting application in Sajivali in India, to investigate flood forecast corresponding to warning times of 1, 2, and 3 hours. The ANN real-time flood forecasting was developed using hourly runoff values for 14 years individual's storms, from 1969 to 1993. The first eleven years of these storms was used in the training the rest three years were used in the testing of the ANN. Three algorithms; error backpropagation, conjugate gradient and cascade correlation were used to train the ANN model in order to reduce the global error. 560 input-output data set was used to train ANN model, and cascade algorithm was found to be more efficient compared with the other two algorithms. Although conjugate gradient algorithm involved less iteration its completion time compared to the other algorithms was higher. On performance with lead times of 1, 2, 3 hours; 162 data set was used in testing the network these yielded a satisfactory predication with a low warning time.

(Kia et al., 2012) to demonstrate ANN flood forecasting using GIS that adopts various flood causes factors to simulate flood prone areas in Johar River basin Malaysia. The final output of the study is a GIS flood map created through water levels produced by the ANN model. In the development of the ANN model, GIS, remote sensing data and field survey were used in deriving suitable thematic layers. Seven flood causes factors (topography, topographic slope, soil, land cover, lithology, and drainage) were then used as an input to the network. The data used in the network was divided into three; 60% training data, 20% validation data and 20% test data. The model architecture was on a three layer network. Input layer that consisted of seven input units each representing the seven flood causes factors two hidden layers, and an output layer that consisted of a single output unit representing river flow. Backpropagation algorithm with a 7-N-N-1 format was used to train the ANN model. N was the hidden layers 7 number of input units and 1 number of output unit. MATLAB software with ease to integrate with GIS data was adopted in training the model. First the input data was processed using Levenberg-Marquardt algorithm normalizing the data to be used in the model. A decaying trend of minimum mean square error in training and validation was used to yield an optimal learning model. The performance of the model was determined using three methods, coefficient of determination, sum squared error, mean squared error, and root mean squared error. The model yielded success with real data with coefficient of determination, but less with other methods. GIS flood maps generated with data output from ANN model were used visualization of flood coverage. In January 2007 the system was used to simulate floods that occurred in Johar River Basin.

(Masibayi et al., n.d.) Presents a Real-Time River Stage Forecasting Using Upstream Stage Approach for Flood Management, in Nzoia River Basin, Western Kenya that uses a linear regression real-time flood forecasting model to predict river stage and thus discharge flow in Nzoia River Basin, Western Kenya. The daily rainfall, river levels and rating curve data (12,357 data points) was divided into two; training (62%) and validation (38%), which was used to develop the model. These data was available on an hourly interval and not lump daily data, to avoid underestimating flood peaks, which are dependent on the lumped daily inputs. The model yielded an efficiency of (R^2) of 98% on training and 96% on validation data on the relationship between the desired and predicted river stage. This implied that linear regression with 8 hours was able to estimate stage at the station (Rwambwa) thus the discharge flow.

2.4 Context of the GA-ANN-FF hybrid model

The basic basis in the approach proposed is in building a flood forecasting model, more so in reducing the damages incurred by flood plain residents, hence enhancing their welfare. Although the main concern in this research is in creating ANN flood forecasting models, in this section a brief general description of the models components and their interrelationships will be discussed. Figure 2.2 shows a schematic framework of the ANN Flood Forecasting model based on four roles;

1. Data collection,
2. Data processing,
3. ANN model, and
4. Forecast dissemination

These will be used in simulating the two ANN Flood forecasting models; MLP-ANN-FF and GA-ANN-FF to estimate the discharge flow for Nzoia River basin at Rwambwa rive gauge.

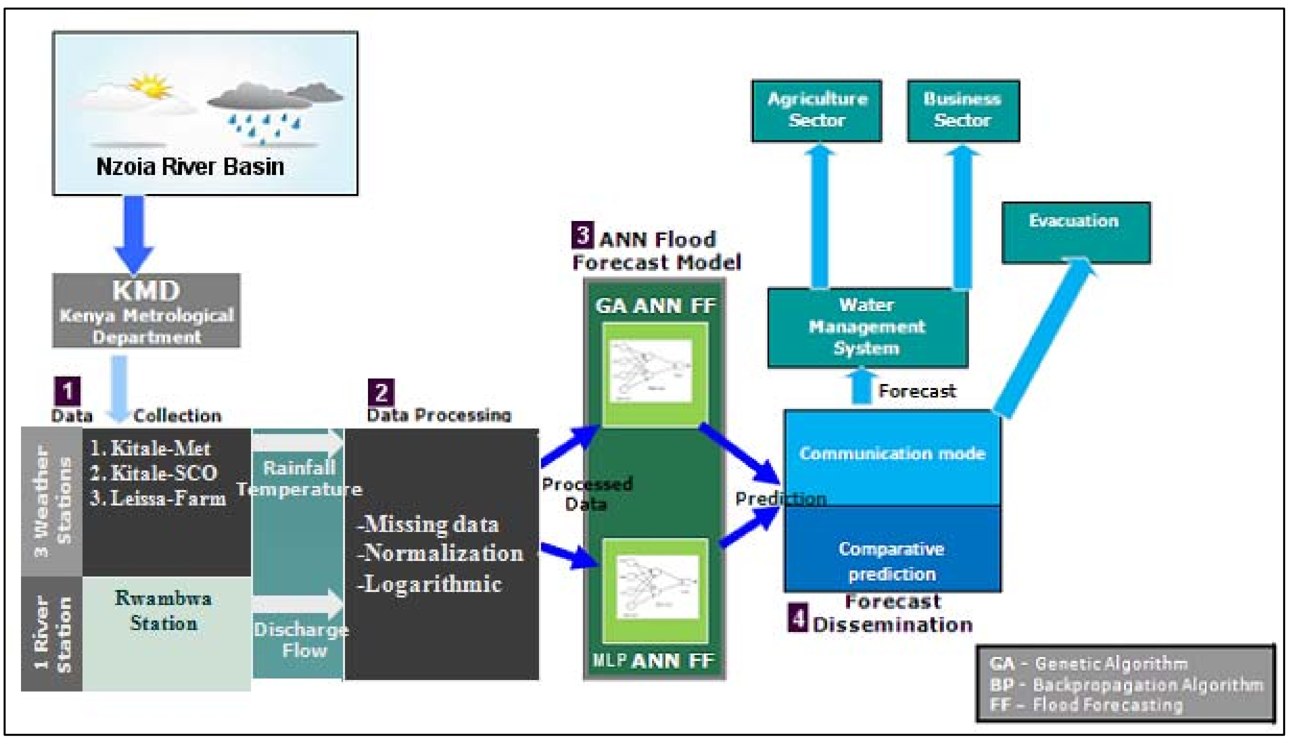


Figure 2.2 - Schematic framework for ANN Flood forecasting models (GA-ANN-FF and MLP-ANN-FF) in the Nzoia River Basin. **Based on 4 roles:** Data Collection, Data Processing, ANN Flood Model and Forecast Dissemination (Source: Author).

2.4.1 Data Collection

The catchment area of study in this research was around Nzoia River basin, Western of Kenya. Nzoia River originates from two high-ground areas of Mt. Elgon and Cherengany Hills; it gathers strength as it flows downstream to an extent of bursting as it reaches the Budalangi areas (Dulo et al., 2010).

To develop the ANN flood models historical data on daily rainfall, average temperature and the simultaneous discharge flow (streamflow) were obtained from three weather stations, and one water level station, Rwambwa River Gauge Station (RGS) within Nzoia River Basin. These data was provided by Kenya Metrological Department (KMD). Other causative flood factors that were not considered in this research study were, flood plain in the past, terrain elevation, water density, water blockage, sub basin areas, soil drainage capability and land use (Puttinaovarat et al., n.d.). The available data was for the period 1975 to 2012, from the three weather stations obtained from Kenya Metrological Department (KMD), and the raw discharge flow provided by Flood Diagnostics and Forecasting Centre (FDFC) at Kenya Meteorological Department for Rwambwa River Gauge Station (RGS). Due to the anomalies in the data available only 4 year data for period 2000 to 2003 that was concurrent in all three weather stations and the one river gauge was used.

2.4.2 Data Processing

After the data collection three data processing processes were conducted to train the Flood Forecasting models more efficiently. These methods are solving missing data values, normalizing the data and performing logarithmic transformation. The missing data are replaced by average of neighbor's values. Normalization was done to improve the performance of the models. The raw data obtained from KMD was first classified using Excel software and normalized using the Komaron formula in the excel software (Jemsi S. 2011).

$$X_{norm} = 0.5\left(\frac{X_0 - \overline{X}}{X_{\max} - X_{\min}}\right) + 0.5$$

Where, the X_{norm} is a normalized value of each conveniently measurable input (X0). The (X0) is the value of each conveniently measurable input, X is the data mean, Xmax is the maximum data and Xmin is the minimum data. The normalized properties that included daily rainfall, temperature and discharge flow data were then transferred as input in to INPUT part in the NeuroSolutions software and the normalized data transferred as network real output (discharge flow) in to OUTPUT part. Standardizing the inputs makes the training faster and reduces chances of getting stuck in local optima (Chen et al., 2013). The log transformation was applied to make the skewed discharge flow output range less skewed (Limpert et al., 2001), by harmonizing the larger values in the data set, and stretching the smaller values. Log function was the preferred option since clipping higher values could have reduced the dataset significantly. By performing a log function of base 10 on the discharge flow output data the data range gets drastically reduced.

2.4.3 ANN Flood Forecast Model

The processed data was fed into the two ANN Flood forecast models (MLP-ANN-FF and GA-ANN-FF), and were developed and simulated with NeuroSolution software separately. Backpropagation (BP) algorithm was used to train MLP-ANN-FF while Genetic Algorithm (GA) was applied on the GA-ANN-FF model. A comparative study between the two models was done to determine whether by applying GA technique the models can improve in their predication accuracy and model generalization.

MLP-ANN-FF model with backpropagation, was trained based on the daily rainfall, temperature and discharge data that was repeatedly presented to the MLP-ANN-FF. With each presentation the output of the MLP-ANN-FF was compared with the desired discharge flow output and an error was computed. This error was fed back (backpropagated) into the NN and used to adjust the weights, such that the error will decrease with each iteration and the ANN Flood forecast model got closer and closer to producing the desired flood forecast output (Priddy and Keller, 2005) Fig 2.3 shows MLP-ANN-FF flood forecast model trained with backpropagation.

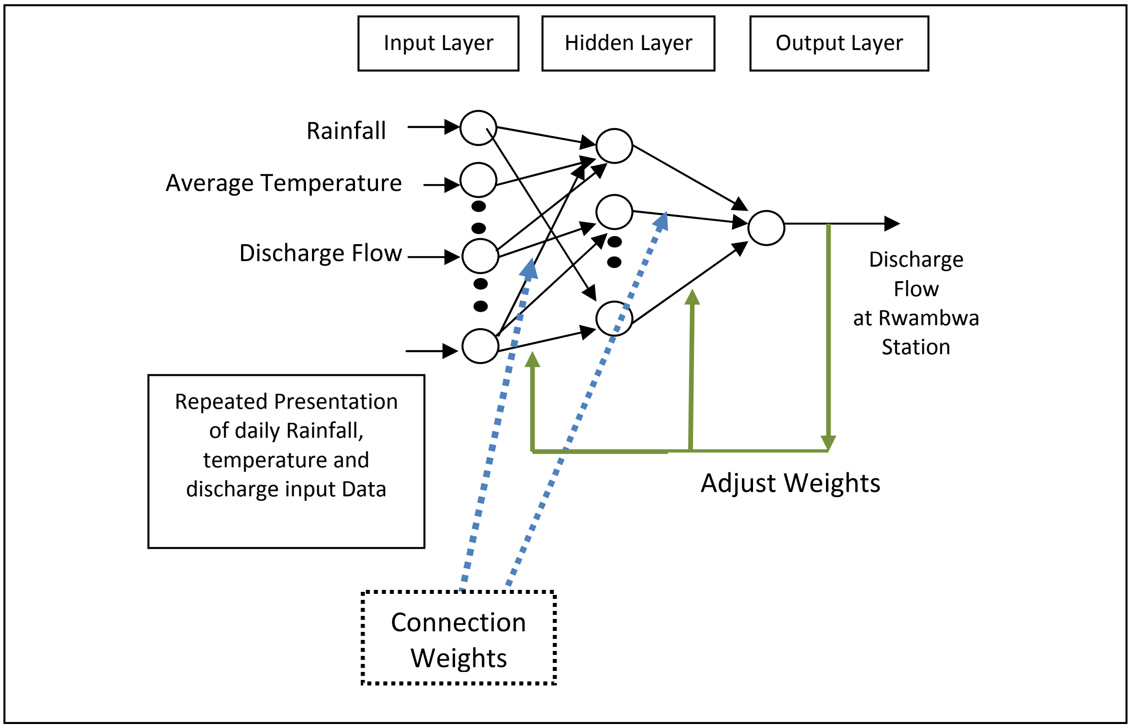


Figure 2.3 – Context model the MLP-ANN-FF Flood forecasting model in the Nzoia River (Source: Author)

Genetic algorithm is efficient in global sampling but have poor local convergence properties (NirmalaDevi et al., 2009). With genetic algorithm trained GA-ANN-FF model, the weights of the NN was joined to make on string (individual or chromosome). The string (individual) was then be used in the genetic algorithm as a member of the population. Each string represented the weights of the complete network. The weights of the initial individuals of the population was chosen at random with probability distribution (Perez, n.d.). This was different from the initial probability distribution of the weights that was given in backpropagation which were in uniform distribution between -1.0 and 1.0 (LUK et al., 2001)**.** The evaluation function returned a rating for each string, assigned weights on chromosome to the links of the GA-ANN-FF, and runs the network over training set of the daily rainfall, temperature, and the discharge data. Figure 2.4 shows GA-ANN-FF flood forecast model trained with genetic algorithm (GA).

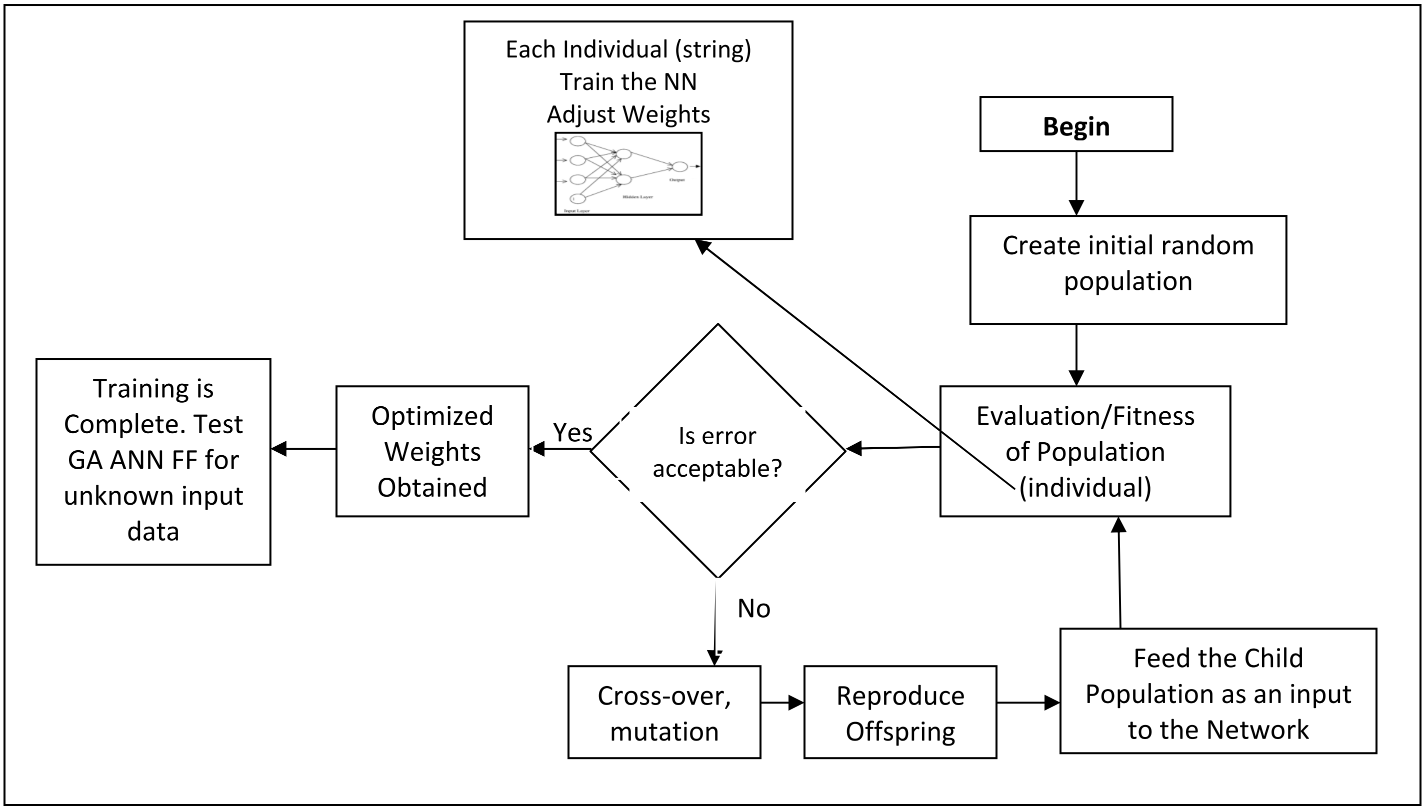


Figure 2.4 The block diagram of GA-ANN-FF procedure to predication of discharge flow in the Nzoia River Basin

2.4.4 Forecast Dissemination

In the Dissemination component, it serviced a dual purpose of generalizing the predictions into forecasts and diffusing such common information so that was beneficial for decision making in water resource management, and evacuations planning. A comparative prediction component was carried out on the basis of important performance measures such as r (correlation coefficient), and Mean-squared error (MSE) between MLP-ANN-FF and GA-ANN-FF models.

In summary Fig x illustrates the four roles (data collection, data processing, ANN model, and forecast dissemination) of the context of ANN with GA hybrid model (GA-ANN-FF) that will be used in developing the two flood forecasting models (MLP-ANN-FF and GA-ANN-FF). In this research study the main role of focus will be ANN Flood Forecast model role. This is the role that will be used in building the two models in estimating the discharge flow (streamflow) of Nzoia River Basin at Rwambwa River Gauge Station discussed in the next chapter.

Chapter 3 - Methodology

3.1 Research Design

The knowledge acquired from the context of the hybrid of ANN with GA model (GA-ANN-FF) discussed earlier in Chapter 2 section 2.4 for ANN Flood forecasting models was based on four roles (data collection, data processing, ANN model, and forecast dissemination) that were used in the model building process of estimating the discharge flow (streamflow) of Nzoia River Basin at Rwambwa River Gauge Station. The developed flood forecast ANN models can be useful to the metrologists in Kenya Meteorological Department (KMD) who can advice the County Governments and threatened residence on when and where the next flood is going to happen and what areas are going to be inundated due to such events.

To develop the two ANN flood models historical data based on the daily rainfall, average temperature and the simultaneous discharge flow (streamflow) were obtained from three weather stations, and one water level station, within Nzoia River Basin. Although the data available was from January 1975 to December 2012 from the 3 weather stations provided by Kenya Metrological Department (KMD), and the raw discharge flow from January 2000 to December 2012 obtained from Flood Diagnostics and Forecasting Centre (FDFC) at KMD. Only data from 2000 to 2003 that was concurrent in all 3 weather stations and the one river gauge was used in the developing the two models (MLP-ANN-FF and GA-ANN-FF)

The catchment area of study in this research was around Nzoia River basin, Western of Kenya. The Nzoia River originates from two high-ground areas of Mt. Elgon and Cherengany Hills; it gathers strength as it flows downstream to an extent of bursting as it reaches the Budalangi areas. (Khan et al., 2011).   The interest of Nzoia River Basin is because its where a number of models on early flood warning for discharge forecast have been initiated by the government through the Flood Early Warning System Team (FEWST)

After data collection, 3 data processing were conducted to train the two flood forecast models (MLP-ANN-FF and GA-ANN-FF) more efficiently. These methods are normalization, logarithmic transformation and use of Microsoft Excel scatter plots to identify the outliers. The use of normalization was to improve performance of the models with the normalized data. Since rainfall, temperature and the discharge flow had different units, the data was scaled between 0 and 1 before it was used as input. The normalization process was done automatically through the NeuroSolution version 6.3.1 Software. The daily discharge flow was discovered to be having a skewed data range that was not evenly distributed. The log function transformation assisted in making the skewed discharge flow range less skewed, by harmonizing the larger values in the data set, and stretching the smaller values. The Excel scatter plots were used to identify the outliers and compute the smallest minimum value in the data range. This assisted in broadening up the range for the rest of the sample, providing much better information to the networks that we were modeling, while keeping the data intact.

After the data was finally processed and reliable, the computational algorithms for both models (MLP-ANN-FF and GA-ANN-FF) was performed on commercial software; NeuroSolution, 6.3.1, presented by NeuroDimension Inc. Intelligence simulation software to develop and simulate the two models. The software with a Microsoft Excel plug-in was used to develop the two models (MLP-ANN-FF and GA-ANN-FF), train and test their performance.

In this research two learning algorithms; Genetic algorithm (GA) and Backpropagation (BP) algorithm were employed on the two models (MLP-ANN-FF and GA-ANN-FF). In order to achieve the best performance from the two models the design process was divided into two experiments. Experiment 1 was applied to model the MLP-ANN-FF with backpropagation and Multilayer perceptron feedforward network based on the daily rainfall, average temperature and the discharge flow as inputs to estimate the discharge flow. Many MLP-ANN-FF network configuration settings were examined by determining the input variables, 9 inputs were the optimal input variables opted for MLP-ANN-FF model. Determining the number of hidden layers and number of neurons in the hidden layers was also conducted and a 9:7:12:1 configuration setting was considered as optimal. Lastly checking the sensitivity of the optimal MLP-ANN-FF model (***9:7:12:1***) in regards to data splitting that included 50%, 60%, 70%, 80% and 90% MLP-ANN-FF with 70% was found to be the optimal proportion for the training data set. In the experiment 2 using the same basis defined for MLP-ANN-FF model with ***9:7:12:1*** configuration settings, genetic algorithm was applied to train and optimize the input variables and the number of neurons in the hidden layers yielding a GA-ANN-FF model with optimal configuration of 4:6:4:1.

After training the two models (GA-ANN-FF and MLP-ANN-FF), the test dataset; daily rainfall, average temperature and the discharge flow for Nzoia River Basin at Rwambwa River gauge, that was not part of training was used to determine and check the overall performance of the models. The significance of applying the test data was to determine whether the models would identify values similar as training stage .(Suliman et al., n.d.)

The accuracy of the estimation was evaluated on the basis of well known performance criteria such as r (Correlation coefficient), MSE (Mean Square Error), and Coefficient of determination (r^2) (Deshmukh and Ghatol, 2010b)

MSE (Mean Square Error):

The formula for the mean square error is given by Equation 3.0

$$MSE = \frac{\sum_{j=0}^{P}\sum_{i=0}^{N} \overline{d_{ij} - y_{ij}}^{\,2}}{NP} \quad \text{...} \; 3.1$$

Where

P = number of output Processing Elements (PEs),

N = number of exemplars in the data set,

y_{ij} *= network output for exemplar i at PE j,*

d_{iy} *= desired output for exemplar i at PE*

r (correlation coefficient):

The size of the mean square error (MSE) was used to determine how well the network output fitted the desired output, but it did not necessarily reflect whether the two sets of data moved in the same direction. For instance, by simply scaling the network output, the MSE could be changed without changing the directionality of the data. The correlation coefficient (r) solved this problem. By definition, the correlation coefficient between a network output x and a desired output d is:

$$r = \frac{\frac{\sum_i (x_i - \bar{x})(d_i - \bar{d})}{N}}{\sqrt{\frac{\sum_i (d_i - \bar{d})^2}{N}} \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{N}}} \text{ where } \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \text{ and } \bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i \qquad \text{…………………………..……………… 3.2}$$

The correlation coefficient was confined to the range [-1, 1]. When r = 1 there was a perfect positive linear correlation between x and d, that is, they co-vary, which means that they vary by the same amount.

Coefficient of determination (r^2)

The coefficient of determination based on the rainfall estimation errors will be calculated as

$$r^2 = \left(\frac{\sum_{i=1}^{N} (Q_i - \overline{Q})(\hat{Q}_i - \overline{\hat{Q}})}{\sqrt{\sum_{i=1}^{N} (Q_i - \overline{Q})^2} \sqrt{\sum_{i=1}^{N} (\hat{Q}_i - \overline{\hat{Q}})^2}} \right)^2 \qquad \text{………………………………………………3.3}$$

Where

$\hat{Q}_i$ and Q_i = *predicted and observed streamflow;*

$\overline{\hat{Q}}$ and $\overline{Q}$ = *mean predicted and observed streamflow respectively, and*

N = total number of observations.

3.2 Data Sources

The catchment around Nzoia River basin has several weather stations and water level gauging stations (Hydrometric Stations). To develop the ANN flood models historical data on daily raw rainfall, average temperature and the simultaneous discharge flow (streamflow) were obtained from three weather stations, namely Kitale metrological, Kitale soil conservation service, and Leissa farm - Kitale and one water level station, Rwambwa River Gauge Station (RGS) within Nzoia River Basin.

As shown in Table 3.1 there were 3 weather stations namely Kitale metrological (SID-8834098), Kitale soil conservation service (SID-8834097), Leissa farm - Kitale (SID-8835039) and 1 water level gauging station (Hydrometric Station) Rwambwa River Gauge Station (RGS) (SID-1EF01), from where data was collected from within the Nzoia River basin.

Although the data available was from January 1975 to December 2012 from the three weather stations provided by Kenya Metrological Department (KMD), and the raw discharge flow from January 2000 to December 2012 obtained from Flood Diagnostics and Forecasting Centre (FDFC) at Kenya Meteorological Department for Rwambwa River Gauge Station (RGS) (1EF01). Only data from 2000 to 2003 that was concurrent in all 3 weather stations and the one river gauge was used as shown in Table 3.1. The missing data was on:

- Temperature recordings from 1975 to 1999 at the Kitale metrological (SID-8834098),
- Discharge flow recordings from 1975 to 1999 at the Rwambwa River Gauge Station (RGS) (SID-1EF01)
- Rainfall recordings from 2004 to 2012 at the Kitale metrological (SID-8834098), Kitale soil conservation service (SID-8834097), and Leissa farm - Kitale (SID-8835039)

Due to these anomalies in the data only concurrent data for 4 years from 2000 to 2003 at Kitale metrological (SID-8834098), Kitale soil conservation service (SID-8834097), Leissa farm - Kitale (SID-8835039) and the one Hydrometric Station Rwambwa River Gauge Station (RGS) (SID-1EF01) was used in the development of the two ANN models (MLP-ANN-FF and GA-

ANN-FF). The sample data for the period 2000 to 2003 obtained from Kenya Metrological Department (KMD) is shown in **Appendix 1**.

Table 3.1 Summary of available data for Nzoia River catchment

No.	Station ID	Station Name	Data Available	Missing or unfixable Data part	Data used
1.	**8834098**	**KITALE MET**			
	Daily Rainfall (mm)		1975 - 2012	1975 -1978	2000 - 2003
	Daily Temperature (^{o}C)		2000 - 2012	1975 – 1999	2000 - 2003
2.	**08835020**	**KITALE SOIL CONs**			
	Daily Rainfall (mm)		1988 - 2003	1975 – 1987	2000 - 2003
3.	**8835039**	**LEISSA FARM**			
	Daily Rainfall (mm)		1975 - 2008	2009 - 2012	2000 - 2003
1.	**1EF01**	**Rwambwa RGS (River Gauge Station)**			
	Daily Discharge (m/c)		2000 - 2012	1975 - 1999	2000 - 2003

This research therefore used primary data obtained from Kenya Metrological Department (KMD) in regards to daily rainfall and average temperature. The secondary data was provided by Flood Diagnostics and Forecasting Centre (FDFC) at Kenya Meteorological Department for Rwambwa River Gauge Station (RGS).

3.2.1 Study Area

The area of interest chosen in the development of an ANN flood forecasting models was Nzoia river basin located in western of Kenya as shown in Figure 3.1. It lies between latitudes 1^{o} 30’N and 0^{o} 05’S and longitudes 34^{o} and 35^{o} 45’E, and originates from Cherangani Hills at a mean elevation of 2300 m above sea level and flows into Lave Victoria at an altitude of 1000m (Khan et al., 2011). It runs approximately South-West and measures about 334 km with a catchment area of about 12,900 km2, with a mean annual discharge of 1777 x 106 m3/year. The population within the Basin is more than 3 million (“Nzoia River Basin management initiative,” 2006).

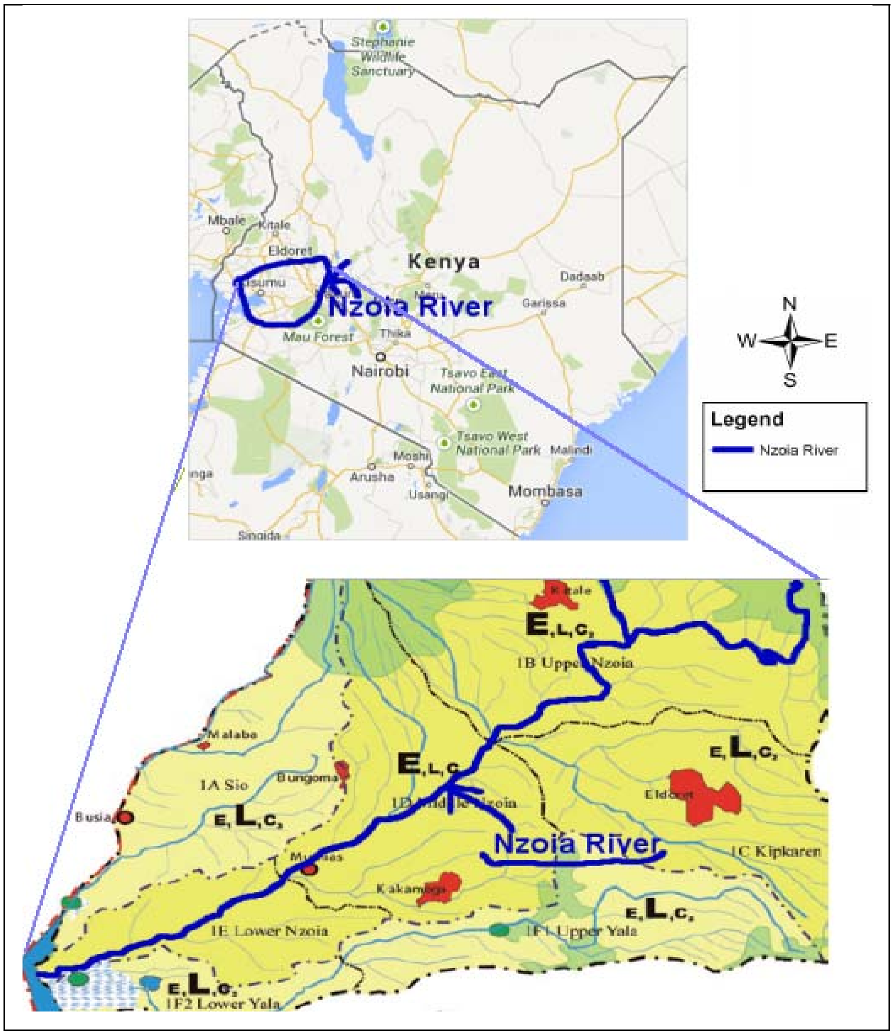


Fig 3.1 - The Nzoia River Basin, location of the study area ("Nzoia River Basin management initiative," 2006)

The interest of Nzoia River Basin is due to several damages and loss of lives from rainfall-induced floods, especially around lowland areas of Budalangi where flood deposit sediments contribute to the fertility of the soil in the area. This area is also the Kenya Government pilot basin for integrated management approach for flood management through the Western Kenya community driven development and flood mitigation project(ADCL, 2006) ("Flood Mitigation Strategy," 2009). Most sectors in commercial and agriculture within Nzoia catchment area their main source of water are from the river basin thus it's of great economic importance.

3.2.2 Data Pre-processing

In order to train the ANN models efficiently and yield valid results pre-processing of data was conducted on the daily rainfall, average temperature and the discharge flow data, and missing data was corrected. The three data processing; normalization, Logarithmic Transformation and Excel scatter plots to identify the outliers, were conducted to train the two flood forecast models (MLP-ANN-FF and GA-ANN-FF) more efficiently.

3.2.2.1 Normalization

Processing of the data was necessary before it was fed into the neural network, thus the 9 inputs and 1 output variables in regards to the daily rainfall, temperature, and discharge flow data were first normalized. Since they were of different units, the data was scaled between 0 and 1 before been used as input, otherwise there could not have been a correlation between the input and output values (Abhishek et al., 2012). Normalization of all the data was done separately on the rainfall, temperature and the discharge flow data, by taking their mean.

Mean = Sum of all values (x) /Number of Values (x)

A standard deviation (SD), for each of the input output variables were then calculated; finally normalizing each input variable.

Normalized Value(x) = (x – Mean)/SD

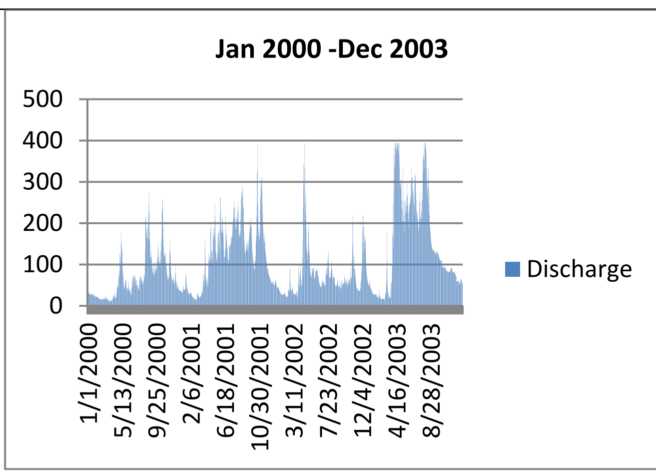
This normalization process was done automatically through the NeuroSolution version 6.3.1 Software, which was used in developing and simulating the two models. The normalized data values (x) were eventually used for training and testing the two ANN Models.

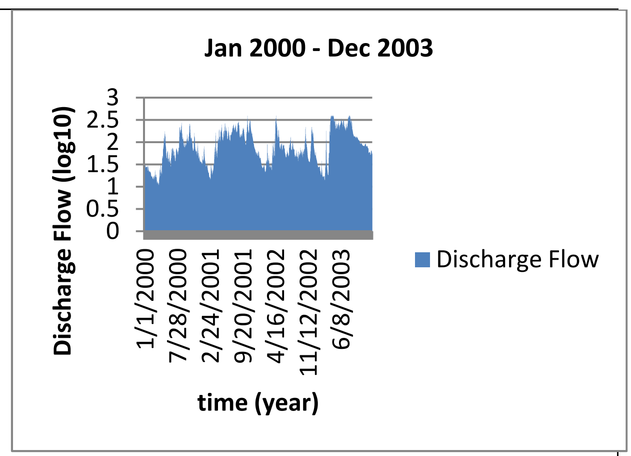
3.2.2.2 Logarithmic Transformation

The output daily discharge flow was discovered to be having a skewed data range, of between, 10.94 to 394.40; where the range was not evenly distributed, hence yielding erroneous mean square error (MSE) of over 6000 as discussed in section 3.3 during the formulation of the input variables. There were two approaches that were considered;

- Applying a log function transformation of base 10 or
- Clipping out the higher values in the discharge flow output parameter.

Log function was the preferred option since clipping higher values could have reduced our dataset significantly. By performing a log function of base 10 on the discharge flow output data the range came down to between 1.594393 to 2.595937 ranges from previous data output range of 10.94 to 394.40. The log transformation applied assisted in making the skewed discharge flow output range less skewed (Limpert et al., 2001), by harmonizing the larger values (394.40) in the data set, and stretching the smaller values (10.94). Figure 3.2 shows the log normal distribution of the discharge flow and after logarithmic with base 10 transformation. This was important in making the discharge flow more interpretable. The sample discharge flow output data transformed with Logarithmic transformation is shown in **Appendix 3.**





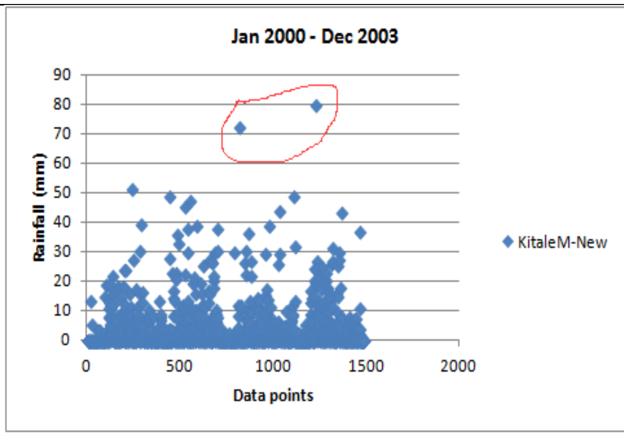
Histogram of a log-normal distribution in the Discharge Flow indices before the screening	Histogram after logarithmic transformation in the Discharge Flow indices after screening

Figure 3.2 - change in the measured discharge flow amounts from Rwambwa River Gauge Station (RGS) from January 2000 to December 2003, before and after after logarithmic transformation

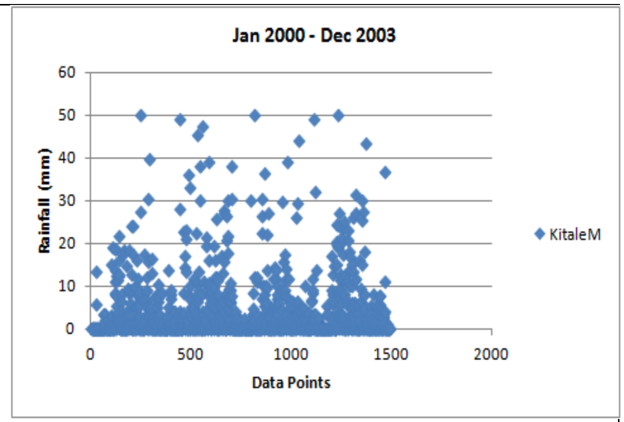
3.2.2.3 Excel scatter plots to identify the outliers

ANN is a statistical model that calculates the weights in order to estimate an outcome, thus provides useful information regarding the events patterns. The ANN results can never be better than the original data, to get good results from the ANN models the original dataset has to be reliable. Processing of the original dataset is therefore very important to treat the outliers (Steege et al., n.d.). By treating the outliers the negative on the estimation performance of the two models (MLP-ANN-FF and GA-ANN-FF) was eliminated.

The 9 input variables in regards to the rainfall, average temperature and daily discharge from 2000 to 2003 were also screened using a scatter plot from the Neurosolution Software for Excel by identifying the outliers. The inputs variables that had few outliers with an excess either in the rainfall and temperature range from the majority sample had to be screened either by locating the outliers and removing or treating the excess samples that was above the majority like they were at the range of the majority by computing the smallest minimum value in the data range. This enabled to broaden up the range for the rest of the sample, providing much better information to the networks that we were modeling, while keeping the dataset of 1488 points intact. Figure 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7 shows the daily rainfall and average temperature data used before and after scatter plot transformation, on the left of the Figures specified in the red circle are some of the outliers.

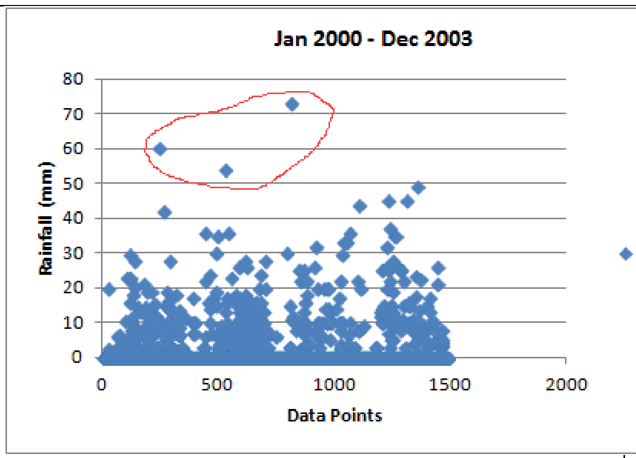


The distribution in the Rainfall indices indices before the screening. The red circle are some of the outliers

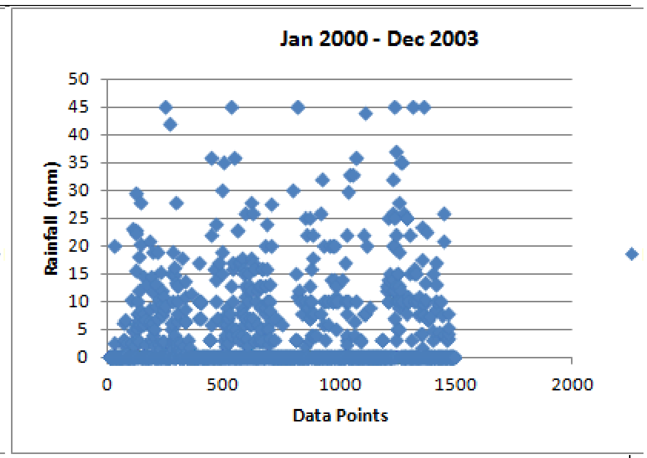


Scatter plot of distribution in the Rainfall indices after the screening

Figure 3.2 change in the measured rainfall amount from Kitale-Met Rainfall from January 2000 to December 2003, before and after screening

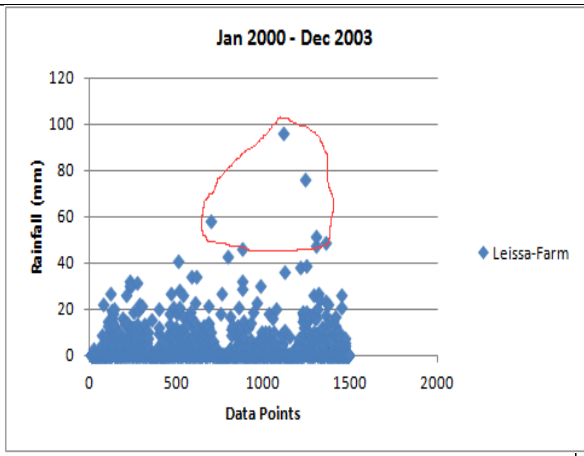


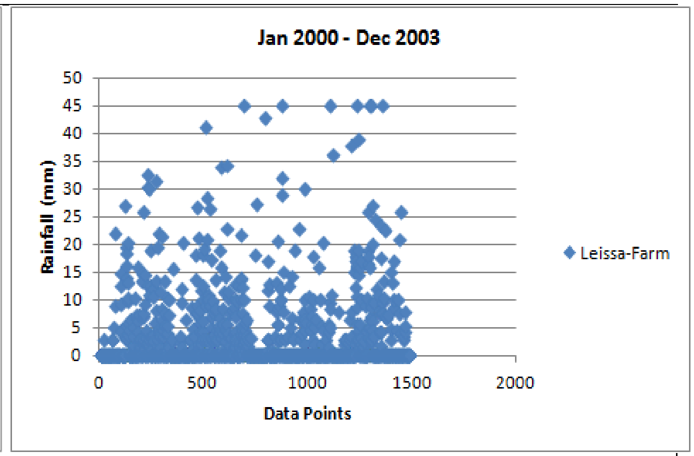
Scatter plot of distribution in the Rainfall indices before the screening. Specified in the red circle are some of the outliers



Scatter plot of distribution in the Rainfall after indices after the screening

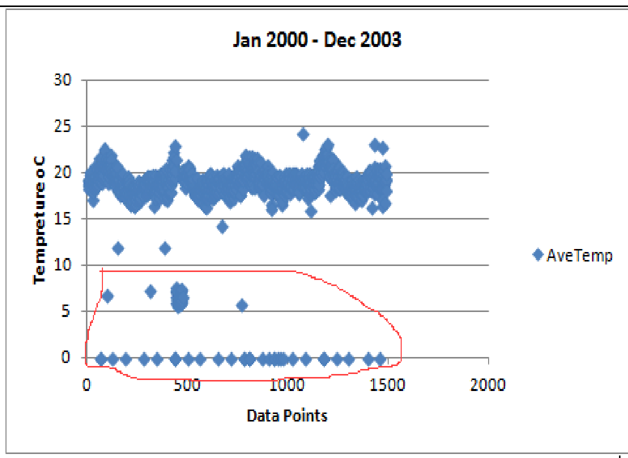
Figure 3.3 change in the measured rainfall amount from KITALE SOIL CONSERVERSION station Rainfall from January 2000 to December 2003, before and after screening

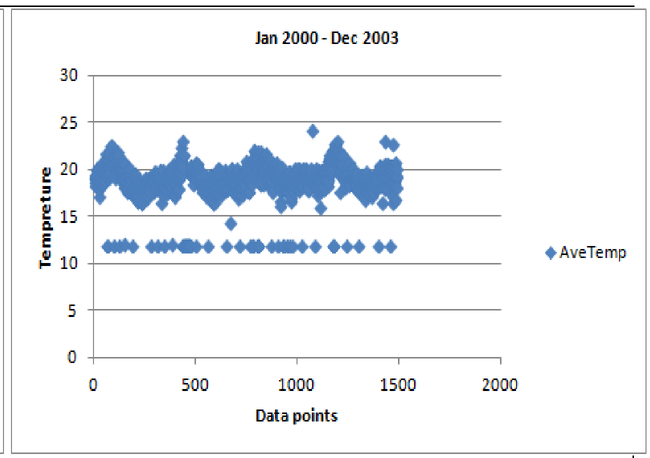




Scatter plot of distribution in the Rainfall indices before the screening. Specified in the red circle are some of the outliers	Scatter plot of distribution in the Rainfall after indices after the screening

Figure 3.4 change in the measured rainfall amount from Leissa Farm weather station Rainfall from January 2000 to December 2003, before and after screening





Scatter plot of distribution in the Temperature indices before the screening. specified in the red circle are some of the outliers	Scatter plot of distribution in the Temperature after indices after the screening

Figure 3.5 change in the temperature values from Kitale Met station from January 2000 to December 2003, before and after screening

Data pre-processing of the daily raw rainfall, and average temperature was processed, by computing the smallest minimum value in the data range, broadening up the range for the rest of the sample, thus providing much better information to the networks that we were modeling, while keeping the dataset of 1488 points intact.

It's essential to avoid numerical difficulties during the computation. In this research, TanhAxon was used as a transfer function since it's the most widely used. It was applied at the output and hidden layers in models topologies, by employing bias and hyperbolic tangent (tanh) function to each nodes between the hidden and the output layers, thus yielding values between -1 to +1 for each node in the layers. Such nonlinear units enable a network gain the ability to make soft decisions. (Chen et al., 2013).

3.2.3 Tools for modeling

In this research the main objective was to model multilayer perceptron feed-forward (MLP-ANN-FF) and a genetic algorithm (GA) optimized multilayer perceptron feed-forward (GA-ANN-FF) for flood forecasting using metrological data (rainfall, temperature, and the discharge flow) to estimate the discharge flow of Nzoia River at Rwambwa River gauge. After the data was finally processed and reliable, the data was used in NeuroSolution, 6.3.1, presented by NeuroDimension Inc. Intelligence simulation software to develop and simulate the two models (MLP-ANN-FF and GA-ANN-FF).

3.3 Designing the Proposed Models

The overall aim of this research was to develop two ANN models (MLP-ANN-FF and GA-ANN-FF) to estimate the discharge flow of Nzoia River at Rwambwa river gauge. In order to develop a better ANN model for estimating the discharge flow; this design process was divided into two groups:

- **E**xperiment **1** (MLP-ANN-FF) and
- **E**xperiment **2** (GA-ANN-FF)

Figure 3.8 shows the experiment studies of the two models, which were studied differently thus developing two different ANN models; MLP-ANN-FF and GA-ANN-FF. In the experiment **2** the model was built by optimizing the model developed from experiment **1**. The performance and accuracy were later analyzed and compared. The input output variables of each experiment were made based on the method discussed later.

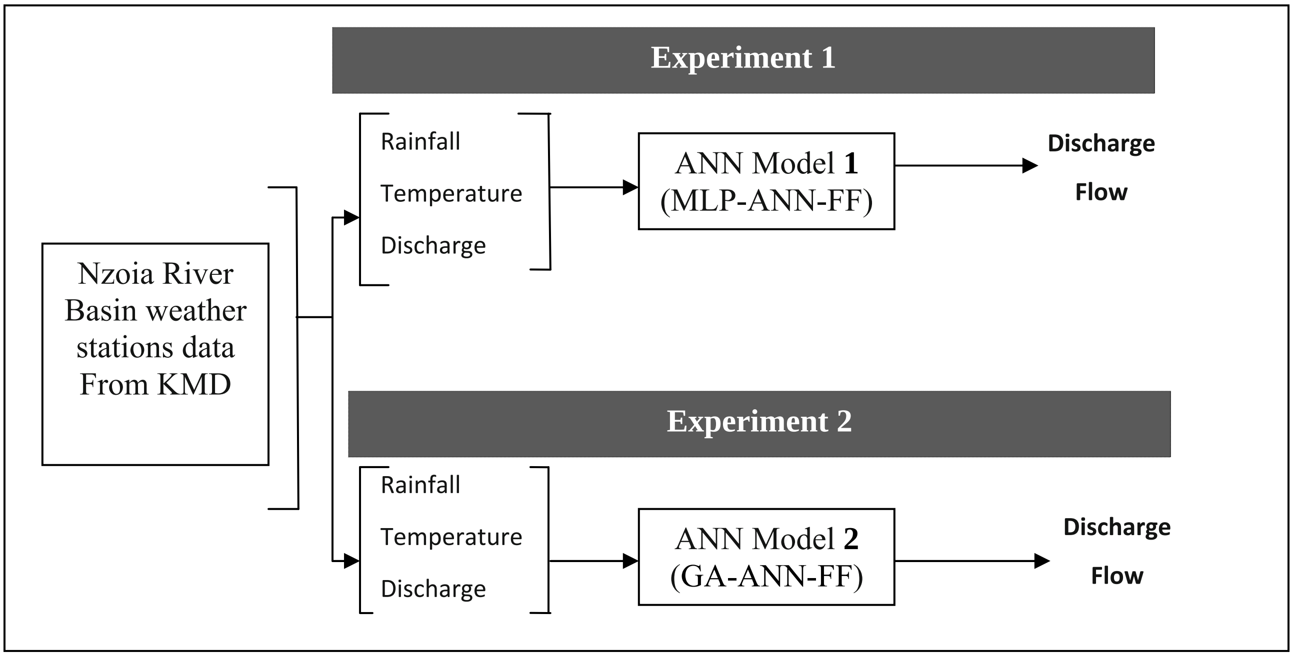


Figure 3.8 Experiments studies (Source: Author)

The constant conditions that were used in training the two models are shown in Table 3.2

Table 3.2 Constant conditions for training the two models

Training variables	Assigned value
Number of epochs	1000
Variation of hidden neurons from	2 to 30
Learning Rate	0.1
Transfer function	Tanh
Momentum factor	0.1
Network Type	Feed Forward
Learning Function	Back propagation With momentum
Weights	Randomize

3.3.1 Experiment 1 (MLP-ANN-FF Model)

An artificial neural network (ANN) has the ability to train and learn the outputs from inputs by mimicking the function of the human brain and nervous system. Mostly widely used ANN is feed-forward backpropagation network (BPN). The major drawback of the conventional BPN with the gradient descent learning is the slow convergence rate (Mitchell, 1997). In this section an MLP-ANN-FF model for predicting discharge flow of Nzoia River at Rwambwa river gauge was developed with feedfoward multilayer network. MLP-ANN-FF was trained tested without genetic algorithm (GA) optimization; instead backpropagation (BP) was used in the training of the model.

3.3.1.1 Selection of input output variables for MLP-ANN-FF model

Determination of enough model input variables is the first and important step of any modeling practice, hence the model accuracy is determined by a proper selection of input data (Abhishek et al., 2012). The main objective of this research was to model two ANN models (MLP-ANN-FF and GA-ANN-FF) for purpose predicting the discharge flow of Nzoia River Basin on Rwambwa river gauge, by applying artificial neural network (ANN) and genetic algorithm (GA) techniques.

To achieve this objective we applied used the historical data on daily raw rainfall, average temperature and discharge flow from 1975 to 2012, but only part of the data was considered useful as mentioned in section 3.2. In this research we have explored the data of 4 years from 2000 to 2003, where there was 744 entries in the input output, in this case 9 inputs and one output were considered making it a 9 * 744 matrix. The output parameter was daily discharge flow to be estimated for Nzoia River Basin at Rwambwa River gauge for period 2002 to 2003. Table 3.3 shows the 9 inputs and one input used for period 2002 to 2003 used.

Table 3.3 9 Inputs and 1 Output variables selected.

#	9 input variables	1 Output
1	Kitale-Met Rainfall 2000- 2001	**Discharge Flow 2002 - 2003**
2	Kitale-Soil Rainfall 2000- 2001	
3	Leissa-Farm Rainfall 2000- 2001	
4	Kitale-Met Temp 2000- 2001	
5	Kitale-Met Rainfall 2002- 2003	
6	Kitale-Soil Rainfall 2002- 2003	
7	Leissa-Farm Rainfall 2002- 2003	
8	Kitale-Met Temp 2002- 2003	
9	Rwambwa Discharge Flow 2000 - 2001	

The number of the 9 input variables was determined, by the experiment **1** realized after further experiments performed in section 3.3.1.2, while determining the number of variables in the MLP-ANN-FF model, where an erroneous mean square error (MSE) of over 6000 was discovered having initiated the model with 4 input parameters and 1 output as shown in the Table 3.4

Table 3.4 4 Inputs and 1 Output variables selected.

#	4 input variables	1 Output
1	Kitale-Met Rainfall 2002- 2003	**Discharge Flow 2000 - 2003**
2	Kitale-Soil Rainfall 2001- 2002	
3	Leissa-Farm Rainfall 2001- 2002	
4	Kitale-Met Temp 2002- 2003	

In order to resolve the erroneous mean square error (MSE) value to be less than zero, further experiments were done by varying the number of input parameters, and pre-processing the data values (discussed in section 3.2.2) which had outliers thus arriving at an optimal number of 9 inputs that offered the most information to our discharge flow output. The sample input output data applied with the 9 inputs and 1 output are shown in **Appendix 2**.

3.3.1.2 Determining the number of variables in the MLP-ANN-FF model

In order to determine the optimal network topology that holds a good generalization, MLP-ANN-FF the model was trained by varying the number of hidden nodes under a trial and error procedure. The proposed MLP-ANN-FF model was initiated with following baselines:

- Configuration setting of 4:20:1 (4 inputs nodes, 1 hidden layer with 20 nodes and 1 discharge flow output)

- Using 70 % training, to determine the weights
- 20 % cross-validation for validating the training stage and
- 10 % testing data to evaluate the overall performance of the model
-

The ***4:20:1*** MLP-ANN-FF topology was discarded as it exhibited worse performances yielding an mean square error (MSE) of over 6000, while a good network when trained should always have an MSE of below zero (Gonzalez et al., 2000). In fact, data collected in Table 3.5 show an r (coefficient correlation) between the predicted and the desired discharge flow was at 7.03%, this indicated the MLP-ANN-FF model could not train well, and it was likely because of the anomalies in the input output data set. A review on the input output data set was performed by reverting back to section 3.3.1.1 (Selection of input output data) and pre-processing the data further. As describe in section 3.2.2, further screening of data was required, hence leading to changing the input variables, from 9 inputs up to 4 inputs.

Table 3.5 Results from 4:20:1

Performance	Discharge
MSE	6101.59311
r	0.07037457

When the new screened data plus the 9 input variables were processed, the new proposed MLP-ANN-FF was now initiated with **two** configuration baselines:

1. **O**ne with a **9:2:1** configuration setting; 9 is the number of input neurons, 2 represent number of neurons in the **1 hidden layer**, while 1 is the node in the output layer
2. **S**econd one was setup with a **9:2:2:1** configuration setting; 9 nodes in the input layer, 2 neurons in each hidden layer and 1 node in the output layer

The two MLP-ANN-FF configuration baselines were trained separately using 70 % training data, to determine the weights 20 % cross-validation to validate the training stage and 10 % for testing, by varying the hidden parameters, to determine the optimal number of neurons per hidden layer that would yield the optimal MLP-ANN-FF architecture.

This experiment was performed with NeuroSolution software to determine the configuration setting with the lowest mean square error (MSE) and the highest r (correlation coefficient) between the training and testing data. The models were trained simultaneously by adjusting the number of neurons in the hidden layers, and testing performed to determine the optimal configuration.

The results obtained from the two models (***9:2:1*** and ***9:2:2:1***) are shown on Table 3.6 for the **1 hidden layer configuration** setting and Table 3.7 for **2 hidden layer configuration** setting, based on transformed data shown on **Appendix 2**.

Table 3.6 Results for varying the number hidden nodes model with **1 hidden layer**; optimal model ***9:20:1*** had the lowest MSE in the training and testing set data set and highest r (correlation coefficient) in the training data set.

	Training			**Cross Validation**			**Testing**		
Model Name	**MSE**	**r**	**MAE**	**MSE**	**r**	**MAE**	**MSE**	**r**	**MAE**
9:2:1	0.0800402	0.5749488	0.2264159	0.0805559	0.3914515	0.236481	0.018129	0.8573719	0.1080291
9:4:1	0.1253977	0.4463966	0.3047871	0.0521263	0.0960869	0.1856546	0.0964388	0.6448018	0.2969536
9:5:1	0.3847673	- 0.2020086	0.5513144	0.0442147	- 0.1575949	0.1811349	0.2031433	- 0.3093116	0.4373314
9:7:1	0.0835307	0.5489058	0.2321032	0.078891	0.2543722	0.2294663	0.0267411	0.8252383	0.1362502
9:10:1	0.0795199	0.5787468	0.2255599	0.0853671	0.3640532	0.2443749	0.0158739	0.8551264	0.1044595
9:14:1	0.0858865	0.5379508	0.2352832	0.0775939	0.2688148	0.2292453	0.0224148	0.8101179	0.1256643
9:17:1	0.1595401	0.377754	0.3358166	0.0387975	0.1387518	0.1652888	0.1443018	0.7239412	0.3691882
9:20:1	0.0794266	0.5791895	0.2257717	0.0885497	0.3201568	0.2531144	0.0148782	0.8314039	0.0935311
9:21:1	0.0797859	0.5767509	0.2263295	0.0784509	0.3446653	0.229554	0.0205487	0.8544861	0.1129936
9:24:1	0.2000889	0.3277104	0.3922777	0.0609927	0.0024084	0.2015982	0.1179068	0.3628258	0.332802

Table 3.7 Results for varying the number hidden nodes model with **2 hidden layers**; optimal model ***9:7:12:1*** had the lowest MSE in the training and testing data set and the highest r (correlation coefficient)

	Training			Cross Validation			Testing		
Model Name	MSE	r	MAE	MSE	r	MAE	MSE	r	MAE
9:2:7:1	0.1771167	0.4150153	0.3667779	0.0452347	0.0334711	0.1728962	0.1463524	0.6259338	0.3716803
9:5:10:1	0.0899985	0.5119762	0.2456375	0.0644381	0.253477	0.1984272	0.0388028	0.8431966	0.1661868
9:6:13:1	0.179792	0.2749852	0.3629996	0.0725573	0.0009024	0.2071176	0.1327768	0.2972192	0.344872
9:7:12:1	***0.0782435***	***0.5880668***	***0.2246031***	***0.0850055***	***0.307332***	***0.2416476***	***0.0186295***	***0.8468048***	***0.1098313***
9:9:4:1	0.4863892	0.1168536	0.6242649	0.0546319	-0.0911218	0.1801003	0.3230873	0.0593746	0.5610587
9:14:9:1	0.12457	0.4319092	0.2882586	0.0571963	0.0970688	0.1881058	0.1283893	0.7321138	0.329759
9:17:12:1	0.1895695	0.1569145	0.3777435	0.073585	0.054433	0.2093954	0.1077899	0.5224977	0.3153984
9:20:15:1	0.123923	0.4032671	0.2753434	0.0501379	0.1253049	0.1702657	0.1057649	0.764553	0.2963579
9:21:16:1	0.1307582	0.4462096	0.2950252	0.0498985	0.211071	0.1715911	0.0993123	0.7397556	0.2803737
9:24:19:1	0.1077014	0.4541041	0.2627175	0.0583737	0.2939058	0.1880984	0.0719495	0.7832782	0.1935121

Table 3.6 shows the optimal configuration setting with **1 hidden layer** was one with 9:20:1, while Table 3.7 shows that 9:7:12:1 with **2 hidden layers** was the optimal configuration setting.

The two configuration settings (***9:20:1*** and ***9:7:12:1***) were then compared and the **2 hidden layer** configuration setting of ***9:7:12:1*** emerged to be the most optimal for MLP-ANN-FF model as shown on Table 3.8

Table 3.8 Results of between for varying **1 hidden layer** (*9:20:1*) and the **2 hidden layer** (*9:7:12:1*); **optimal model** (9:7:12:1) had the highest r in training and testing data and lowest MSE in the training data set therefore the 9:7:12:1 topology was the opted architecture for the MLP-ANN-FF model.

	Training			Cross Validation			Testing		
Model Name	MSE	r	MAE	MSE	r	MAE	MSE	r	MAE
9:20:1	0.0794266	0.5791895	0.2257717	0.0885497	0.3201568	0.2531144	0.0148782	0.8314039	0.0935311
9:7:12:1	***0.0782435***	***0.5880668***	***0.2246031***	***0.0850055***	***0.307332***	***0.2416476***	***0.0186295***	***0.8468048***	***0.1098313***

After performing these experiments the MLP-ANN-FF model was finally developed using ***9:7:12:1*** configuration setting; **9** inputs variables with **2 hidden layers** one with **7** neurons and the other **12** and **1** output layer.

3.3.1.3 Determining the proportion of the training data set.

In order to check the sensitivity of the optimal MLP-ANN-FF model (***9:7:12:1***) in regards to data splitting, a percentage of the training data set was used and varied**,** to avoid the model from overfitting the training data; where the training data set closely matches the output. First, the training data set applied would generate errors prompting update on the connection weights, verification data was then applied to cross validate and supervise the training set and to evaluate the performance of the model at various stages of training the model, and finally test data set an independent data set for to evaluate accuracy of the overall of the ANN models, was used in testing the model generalization ability (Rezaeianzadeh et al., 2014)

An experiment simulation was applied by using 60%, 70%, 80% and 90 % of the total data set as training and the rest for cross validation and testing. The MLP-ANN-FF model was trained and tested. Table 3.9 shows the variations of training data set percentages, and corresponding MSE and r (correlation coefficient). The transformed data set used for this experiment is shown in **Appendix 2.**

Table 3.9 Selection of appropriate MLP-ANN-FF model in terms of data percentage

Model (% Training Data Set)	**MSE**	**Correlation Coefficient**
MLP-ANN-FF 50%	0.088918483	0.440935144
MLP-ANN-FF 60%	0.05474843	0.580985432
MLP-ANN-FF 70%	***0.023684652***	***0.835305308***
MLP-ANN-FF 80%	0.049885615	0.88409909
MLP-ANN-FF 90%	0.053452338	0.868147588

The MLP-ANN-FF (***9:7:12:1***) developed with 70% training data set emerged to be the best model that estimated the discharge flow with a r (correlation coefficient) of the desired and the estimated discharge flow of 0.835 and least MSE of 0.023, and. Therefore MLP-ANN-FF (***9:7:12:1***) trained with 70% was chosen and applied on our MLP-ANN-FF architecture modeling, 20% and the remaining 10% were used for cross validation and testing respectively.

3.3.1.4 Final MLP-ANN-FF Model

The model that was finally considered after the experiments in phase 1 was ***9:7:12:1*** MLP-ANN-FF model, using **70%** training data set, available from the daily rainfall, average and the discharge flow to estimate discharge flow for Nzoia River at Rwambwa river gauge. The model was developed trained and implemented using NeuroSolution with backpropagation algorithm. Figure 3.9 shows the basic flow that was used in developing, training and testing the MLP-ANN-FF model.

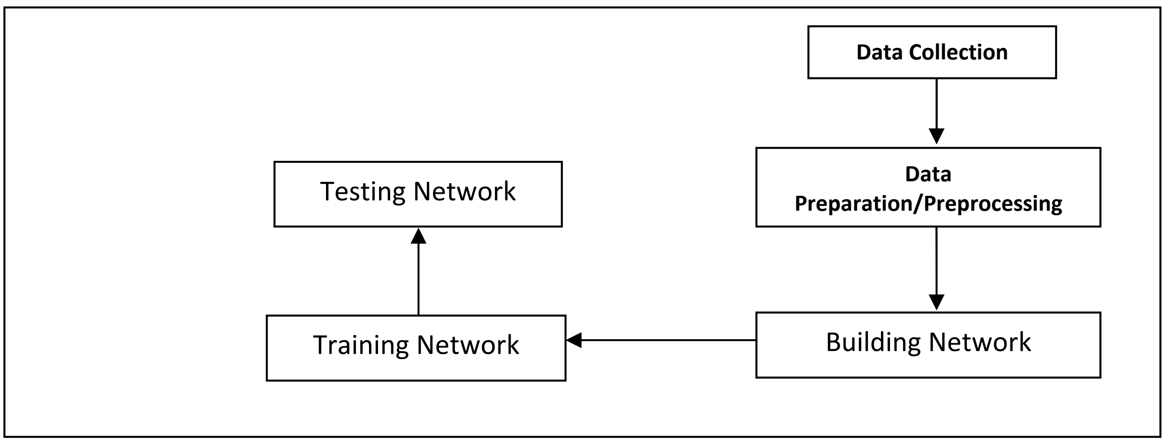


Figure 3.9 Proposed training procedure for MLP-ANN-FF model (***9:7:12:1***) flow (Source: Author)

The obtained results in this research are discussed in Chapter 4. Figure 3.10 shows a schematic of a MLP-ANN-FF (***9:7:12:1***) model. This is the model that was finally considered for training and testing the MLP-ANN-FF model. There were 9 input neurons, 2 hidden layers, with the first hidden layer made up of 7 neurons and the second layer 12 neurons, the output consisted of 1 neuron the discharge flow. The optimal MLP-ANN-FF (***9:7:12:1***) model that was finally trained and tested using the NeuroSolutions software is shown **Appendix 5; Figure 5A-1**

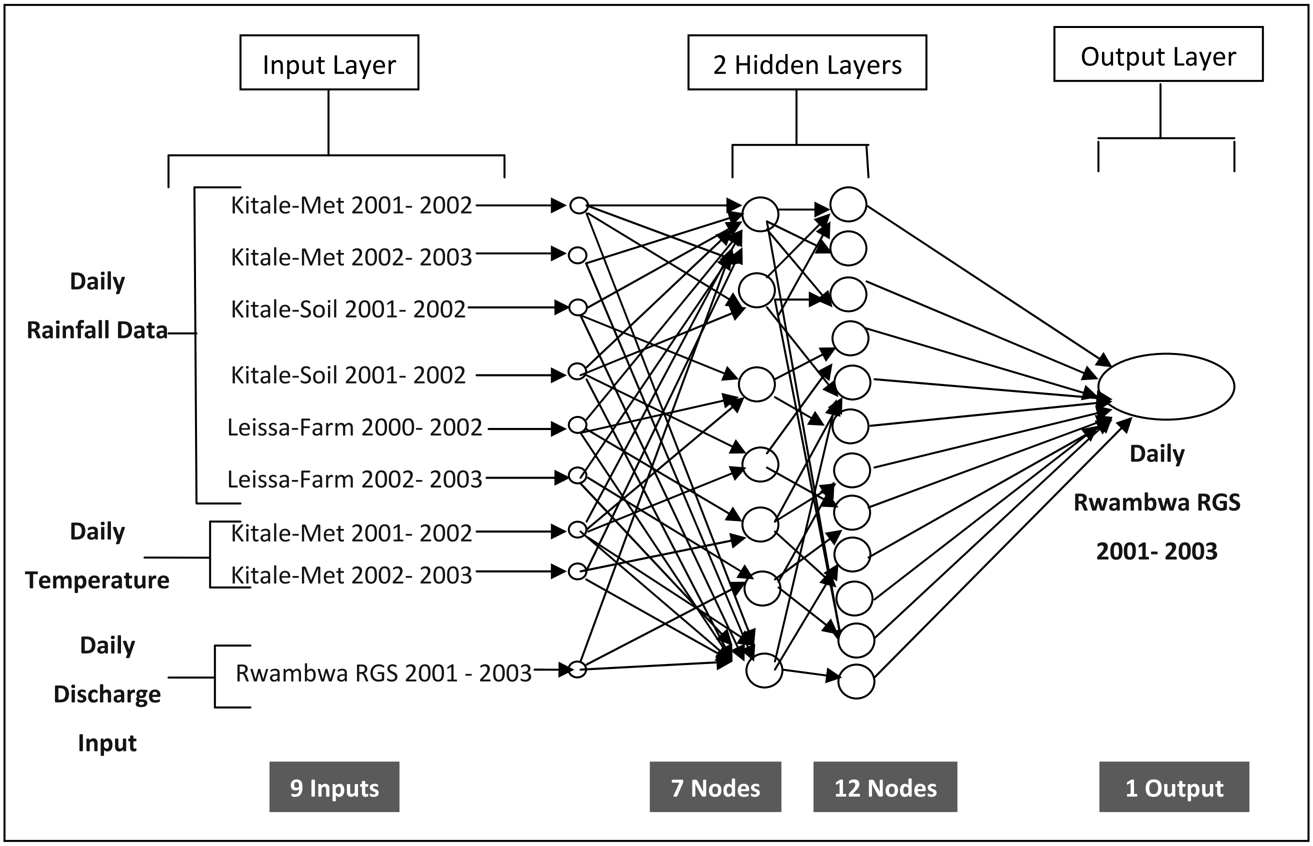


Figure 3.10 Schematic of MLP-ANN-FF (**9:7:12:1**) model (Multilayer Perceptron artificial neural network) (Source: Author)

3.3.2 Experiment 2 (GA-ANN-FF Model)

In this section GA-ANN-FF model for estimating the discharge flow of Nzoia River at Rwambwa river gauge was developed. A genetic algorithm based artificial neural network (GA-ANN-FF) model was built, by optimizing the MLP-ANN-FF model already developed from experiment **1**, to determine whether the integration of ANN and GA algorithm would yield better performance. In this research genetic algorithm was applied primarily

- To avoid the trial and error in selecting the optimal input parameters,
- To determine the number of neurons in the two hidden layers, and
- To train the network weights hence yielding an optimized GA-ANN-FF model.

Using the same basis defined for the optimal MLP-ANN-FF model with ***9:7:12:1*** configuration settings. 9 inputs and 1 output were applied on the GA-ANN-FF model, the data values were normalized to between 0 and 1, then fed into the network and trained. Genetic algorithm (GA) was used as an alternative to the backpropagation (BP) algorithm to update weight in the neural network model. Training with genetic algorithm was started by initializing the weights and neurons in the input layer, with a 70% training data set. The global error generated at the output layer of the GA-ANN-FF model was computed as the fitness to rank the potential solution. This process was iterated for 50 generation with potential solution getting to a global optimal solution after the 21 generation. The main goal in training a model is to adjust the weights between the layers (Mitchell, 1997), the weights were well updated at the 21 generation. The fitness value of the rank was then computed, with the model opting for a GA-ANN-FF model of ***4:6:4:1*** topology as shown in Fig 3.11 below. The transformed data set used for this experiment is shown in **Appendix 2.**

This new configuration setting of **4:6:4:1** was generated using NeuroSolutions for Excel Software after training GA-ANN-FF model with genetic algorithm. The model input parameters were reduced from 9 to 4 inputs, the hidden neurons in the two hidden layers were also reduced, thus reducing the complexity of the GA-ANN-FF model, providing a much leaner network, since the input variables selected by the genetic algorithm offered much information to the desired discharge flow output. Kitale-Met Temp 2001 - 2002, Leissa-Farm Rainfall 2001 - 2002, Leissa-Farm Rainfall 2002 - 2003 and Rwambwa Discharge Flow 2001 – 2002 were the selected 4 input variables by the GA offering most information to the discharge flow output.

3.3.2.1 Final GA-ANN-FF Model

The model that was finally considered after the experiment **2** was a, GA-ANN-FF model with **4:6:4:1** configuration setting using 70% training data set. Figure 3.11 shows a schematic of a GA-ANN-FF model opted for this research study. There were 4 input neurons, 2 hidden layers, with the first hidden layer made up of 6 neurons and the second hidden layer consist of 4 neurons, the output consist of 1 neuron the discharge flow. The sample data which was used for the GA-ANN-FF model with **4:6:4:1** configuration setting for this experiment is shown in **Appendix 4.** The optimized GA-ANN-FF (***4:6:4:1***) model that was finally trained and tested using the NeuroSolutions software is shown **Appendix 5; Figure A5-2**

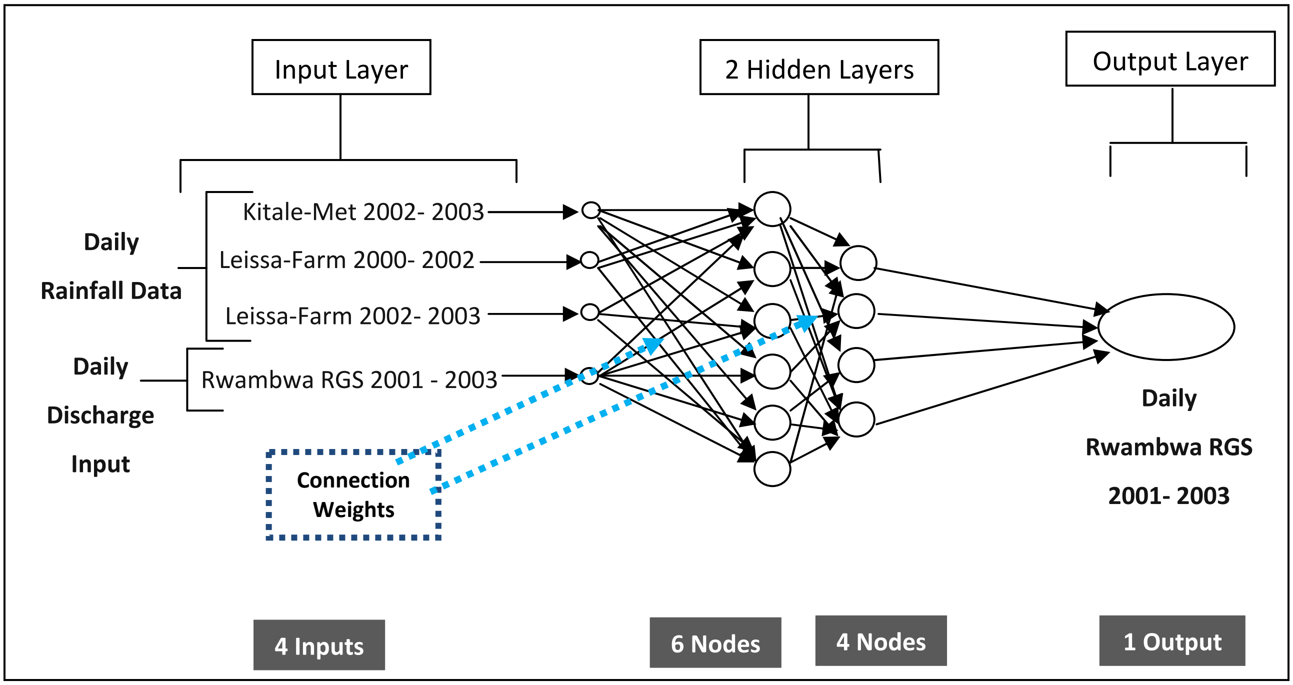


Figure 3.11 Schematic of the designed GA-ANN-FF (***4:6:4:1***) model (Source: Author)

Chapter 4 – Evaluation of the ANN Model

4.1 Evaluating the Models

Machine learning approach; whether an artificial neural network (MLP-ANN-FF), or a genetic optimized ANN (GA-ANN-FF) model, will have trained well if they had good generalization ability (Che et al., 2011). The performance of the estimating the discharge flow for Nzoia River at Rwambwa river gauge resulting from training, validation and testing the models was evaluated on the basis of the values of mean square error (MSE), r (correlation coefficient) and coefficient of determination (R^2).

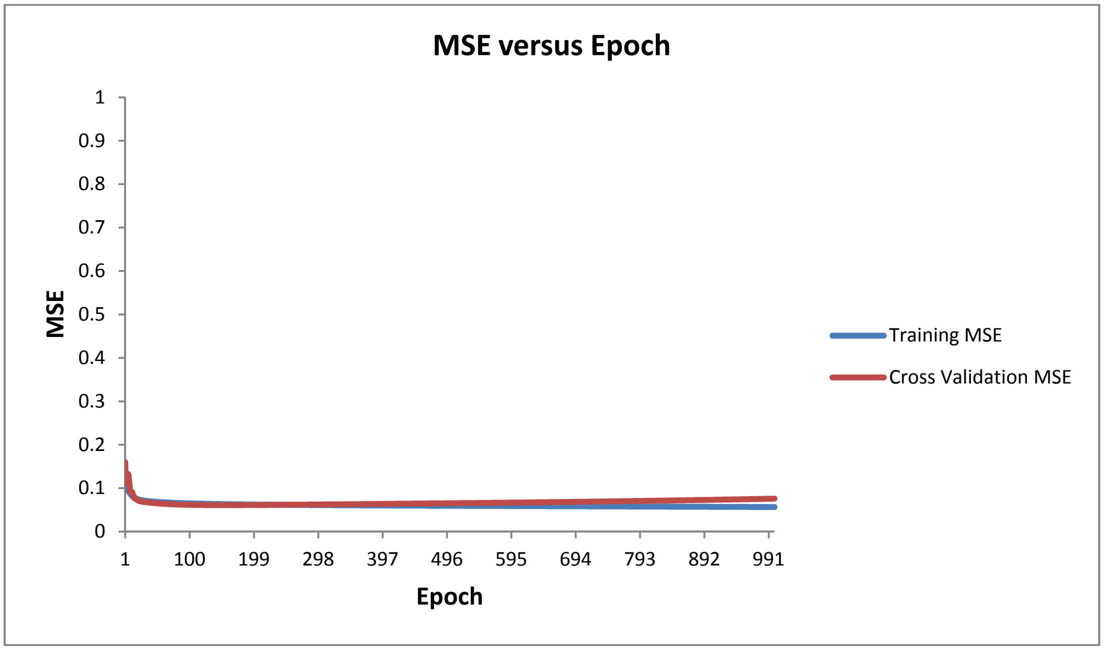
The evaluation was performed on the two models; MLP-ANN-FF and GA-ANN-FF developed with NeuroSolutions software from the two experiments in section 3.3.1 and 3.2.2. Configuration setting of **9:7*:12:1*** topology, and data percentages of 70% training, cross validation 20% and the remaining 10% testing data was applied to evaluate the overall performance of the models. GA-ANN-FF model after training it with genetic algorithm yielded a ***4:6:4:1*** configuration setting which was opted as the optimal network for the GA-ANN-FF model.

4.2 Evaluation of the MLP-ANN-FF model

The purpose of this evaluation test was examine the accuracy, and performance of the developed MLP-ANN-FF model using the 10% of the metrological test dataset (rainfall, temperature, and the discharge flow) to estimate the discharge flow of Nzoia River at Rwambwa River gauge. This was done using the developed MLP-ANN-FF model by activating the training and testing process in Neurosolution Software with backpropagation algorithm. The 10% test data used in this evaluation test is shown in **Appendix 7; Table A7-1**

4.2.1 Results

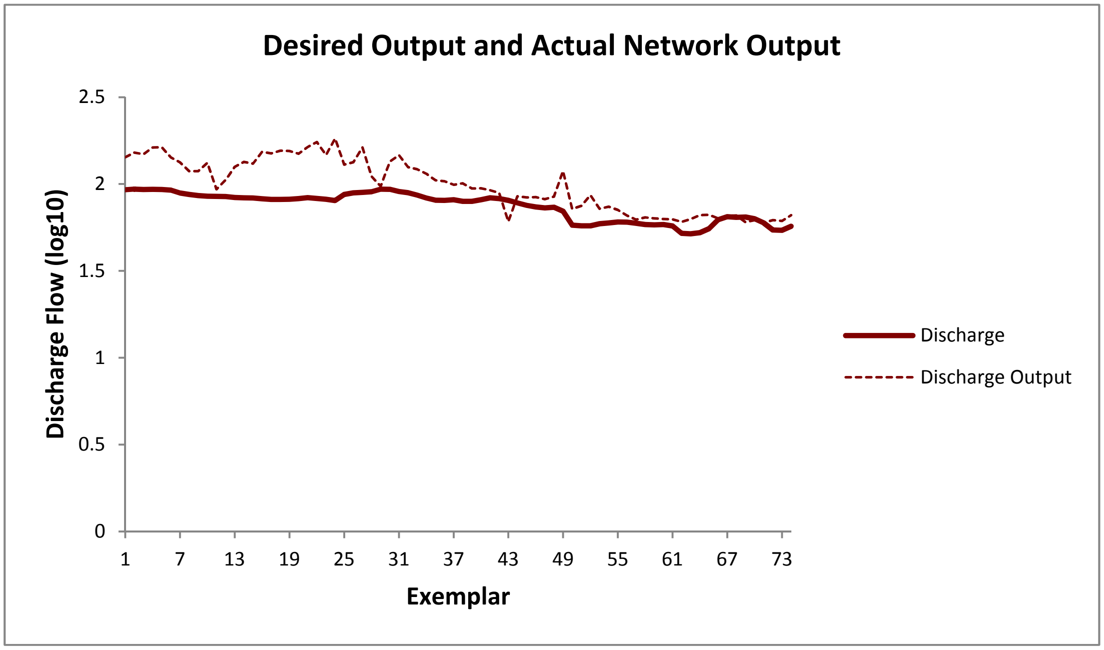
The best input parameters and the number of neurons in the hidden layers were adopted as discussed in experiment **1** (section 3.3.1) arriving at an optimal topology of **9:7*:12:1*** MLP-ANN-FF model. First the epoch of the model was fixed to 1000 for training, with a learning rate of 0.1 with a backpropagation learning function. Figure 4.1 shows the MSE trend of the **9:7:*12:1*** topology of the best MLP-ANN-FF model according to the number of 1000 epochs. The training process stopped when the MSE in the validation data set was at 0.0609 after 151 iteration. In fact increasing the number of epochs to highs of 1000 the network model started decaying in the performance, thus increasing the error to highs of 0.075 for the validation data set.



Best Networks	Training	Cross Validation
Epoch #	1000	151
Minimum MSE	0.056437	0.060977962
Final MSE	0.056437	0.075864209

Figure 4.1 MSE Analysis of 9:7:12:1 best topology for MLP-ANN-FF model over 1000 epochs for training data set.

Figure 4.2 show the comparison between the actual discharge flow and the desired output of the trained MLP-ANN-FF (9:7:12:1) model. As seen MLP-ANN-FF model estimated the discharge flow at Rwambwa river gauge for period 2002 to 2003 with an r (correlation coefficient) of 0.84 (80%) slightly lower value from 1 and a MSE of 0.024. This can be assumed as satisfactory results as compared to other researchers (Puttinaovarat et al., n.d.) Discussed in section 2.3. The test results obtained for the MLP-ANN-FF (9:7:12:1) model is shown in **Appendix 6**



Performance	Discharge
MSE	0.024031731
MAE	0.126664591
r	0.843375453

Figure 4.2 MLP-ANN-FF model; estimated data (testing stage) and error measures.

Figure 4.3 shows a scatter plot diagram showing the desired versus predicted values of the discharge flow with a good r (correlation coefficient) of 0.84 (80%). As seen, the values are evenly distributed around the regression line, indicating that there was neither overwhelming over prediction nor under prediction.

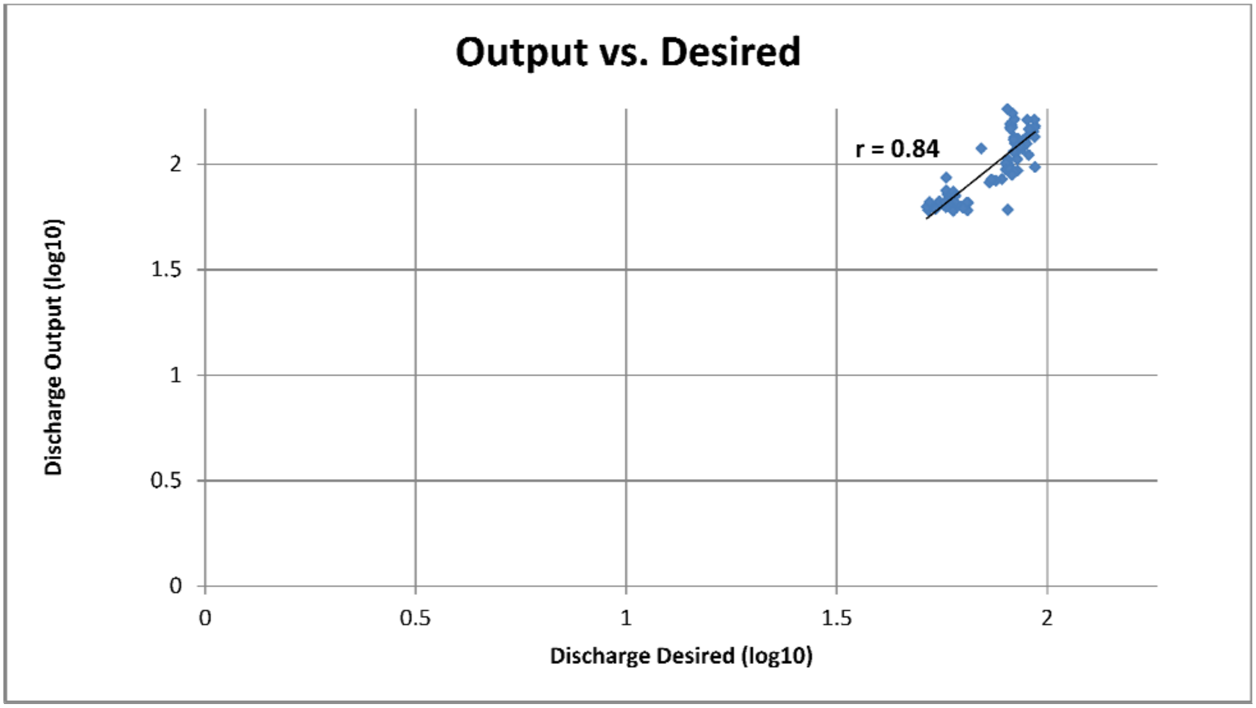


Figure 4.3 Predicted versus measured data (testing stage). The correlation between the Rwambwa River gauge daily discharge and output of MLP-ANN-FF model

4.2.2 Observations and Analysis

The MSE of the training data reduced sharply in 56 iterations and continued to decrease to levels of 0.56437 to the last 1000 epoch. This is the trend suggested in the literature review (Puttinaovarat et al., n.d.). It's clear that the learning on the training data set is best and minimum after 151 epochs for validation data set and 1000 epoch for training data with a MSE of 0.5643. On the cross validation data applied during the training the min MSE was at 0.060977 after 151 iteration, this is when the network started to learn.

As indicated in Figure 4.2 Testing data set which was not part of the training set was used to test the model predictability in order to evaluate whether the MLP-ANN-FF model after successfully training as shown in Figure 4.1 could test data well.

4.2.3 Discussions

The MLP-ANN-FF model learnt well with minimal iteration of 151 as indicated in Figure 4.1 with a minimum MSE of 0.0609 on the cross validation data. Figure 4.1 shows the trend of the error function as the number of iteration increases. As seen, the MLP-ANN-FF model estimates the discharge flow at Rwambwa river gauge for period 2002 to 2003 with MSE of 0.024 and r (correlation coefficient) of 0.84 (80%). This can be assumed as satisfactory results. With the scatter plot diagram Figure 4.3 supporting the same trend indicating an evenly distribution around the regression line, indicating that there was neither overwhelming over prediction not under prediction. The results of r of 0.84 and MSE of 0.024 obtained compare well with those obtained by other researchers (Puttinaovarat et al., n.d.) as discussed in section 2.3

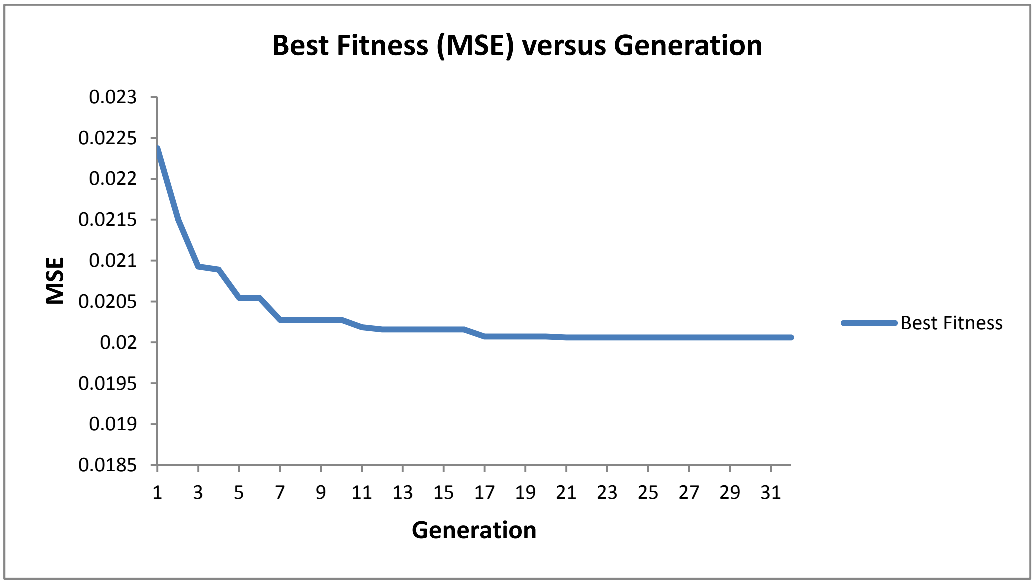
4.3 Evaluation the GA-ANN-FF model

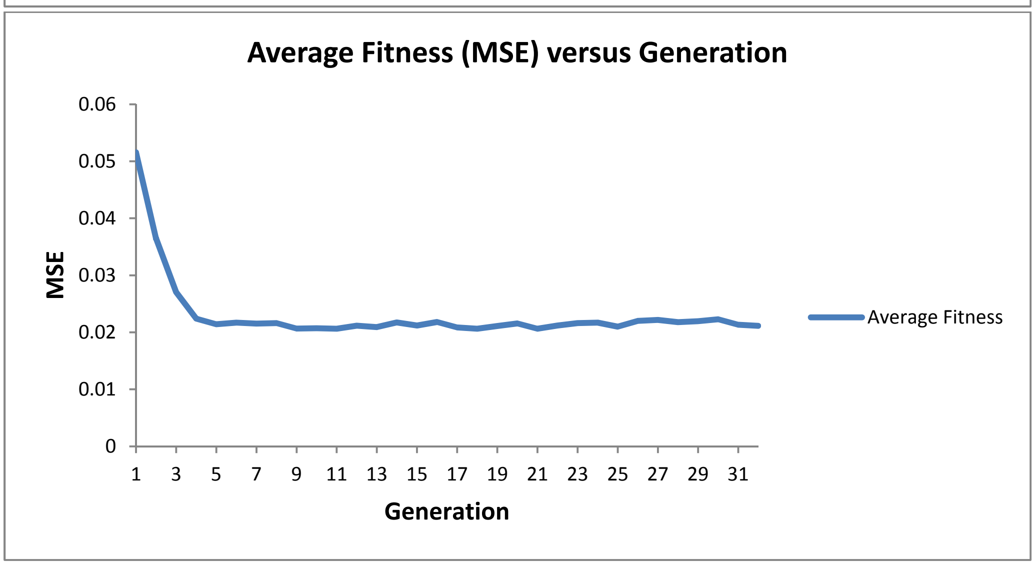
To test the accuracy, and performance of the developed GA-ANN-FF model using the 10% of the metrological test dataset (rainfall, temperature, and the discharge flow) to estimate the discharge flow of Nzoia River at Rwambwa River gauge. This was done using the developed GA-ANN-FF model trained with genetic algorithm by activating the training and testing process in Neurosolution Software. The 10% test data used in this evaluation test is shown in **Appendix 7; Table A7-2.**

4.3.1 Results

For a good comparison GA-ANN-FF model was trained and tested using the same data set applied on the MLP-ANN-FF model. The estimation of the discharge flow for Nzoia River Basin at Rwambwa River gauge based on MLP-ANN-FF was implemented with genetic algorithm technique to model GA-ANN-FF model. The model was trained using a population size of 50 and max generation of 50. The training was iterated for 50 generation with potential solution getting to a global optimal solution after the 21 generation. The GA-ANN-FF model finally yielded an optimal **4:6:4:1** configuration setting reducing the number of input variables from 9 to 4 and the number of neurons in the 2 hidden layers, from 7 to 6 in the first hidden layer and from 12 to 4 in the second hidden layer.

Figure 4.4 shows the MSE trend of the **4:6:4:1** topology of the best GA-ANN-FF model according to the number of 50 generation. The best fitness was determined when a minimum MSE of 0.02 was achieved after 21 generation.

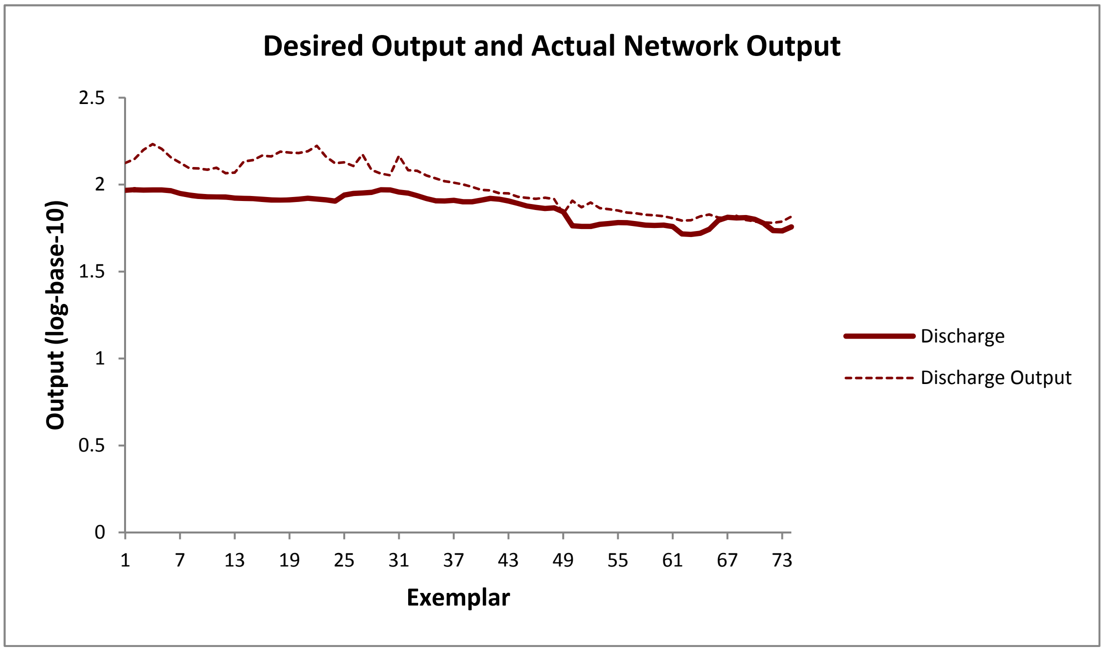




Optimization Summary	Best Fitness	Average Fitness
Generation #	21	21
Minimum MSE	0.02006049	0.020620811
Final MSE	0.02006049	0.021135689

Figure 4.4 MSE Analysis of **4:6:4:1** best topology for GA-ANN-FF model over 50 generations for training data set.

Figure 4.5 shows the predicted discharge flow of the trained GA-ANN-FF **(4:6:4:1)** model, testing data set which was not part of the training set was used in testing the model predictability in order to evaluate whether the GA-ANN-FF model after successfully training as shown in Figure 4.4 could test data well. As seen, the GA-ANN-FF model estimates the discharge flow at Rwambwa river gauge for period 2002 to 2003 with MSE of 0.021 and r (correlation coefficient) of 0.887 (89%).



Performance	Discharge
MSE	0.02176209
NMSE	3.357740688
MAE	0.123583328
r	0.887479548

Figure 4.5 GA-ANN-FF model; estimated data (testing stage) and error measures.

Figure 4.5 shows a scatter plot diagram showing the desired versus predicted values of the discharge flow with a good r (correlation coefficient) of 0.887 (89%). As seen, the values are evenly distributed around the regression line, indicating that there was neither overwhelming over prediction not under prediction. The test results obtained from GA-ANN-FF (4:6:4:1) are shown in Appendix 6; Table A6-2.

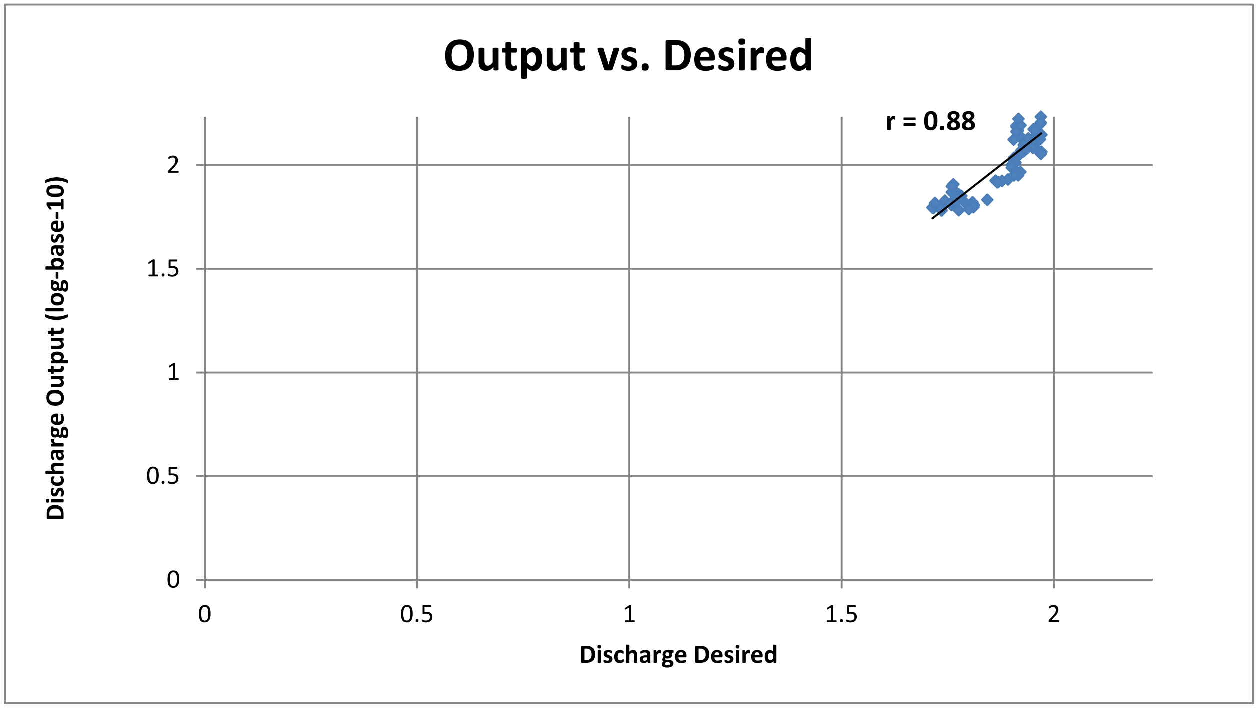


Figure 4.6 Predicted versus measured data (testing stage). The correlation between the Rwambwa River gauge daily discharge and output of GA-ANN-FF model

4.3.2 Observations and Analysis

The MSE trend of the **4:6:4:1** topology of the best GA-ANN-FF model according to the number of 50 generation arrived at the best fitness when a minimum MSE of 0.02 was achieved after 21 generation.

The discharge flow for Nzoia River Basin at Rwambwa River gauge for period 2002 to 2003 Figure 4.4 shows the trend of the error function as the number of generation increases. As seen, the error sharply decreases at 4 generations and continues to decrease and levels off around (0.0211) in 21 generations. As indicated in Figure 4.5 Training data set which was not part of the training set was used in testing the model predictability in order to evaluate whether the GA-ANN-FF model after successfully training as shown in Figure 4.4 could test data well with an r (correlation coefficient) of 0.887 (89%). As seen, the values are evenly distributed around the regression line, indicating that there was neither overwhelming over prediction not under prediction.

4.3.3 Discussions

As seen, the GA-ANN-FF model estimates the discharge flow at Rwambwa river gauge for period 2002 to 2003 with MSE of 0.021 and r (correlation coefficient) of 0.887 (89%). This can be assumed as satisfactory results. With the scatter plot diagram Figure 4.6 supporting the same trend indicating an evenly distribution around the regression line, indicating that there was neither overwhelming over prediction not under prediction.

4.4 Evaluating GA-ANN-FF performance with MLP-ANN-FF base model

The aim of this evaluation is to compare the performance of MLP-ANN-FF base model with GA-ANN-FFF model using 10% test dataset from Nzoia river basin for Rwambwa river gauge as shown in **Appendix 7**. This is done using the developed MLP-ANN-FF and GA-ANN-FF model by activating the training and testing process in Neurosolution Software.

4.4.1 Results

The results obtained by using GA-ANN-FF model are compared to those from MLP-ANN-FF base model, using the 10% test dataset of the total dataset, as shown in the Table 4.1 below. The table results shows a data column (column one) that indicates dates the actual discharge flow (column 2), the third and fourth columns shows the predicted discharge flow values of the models and their errors between the predicted and the actual discharge values. The results obtained are also illustrated using graphs as shown in figure 4.7 below.

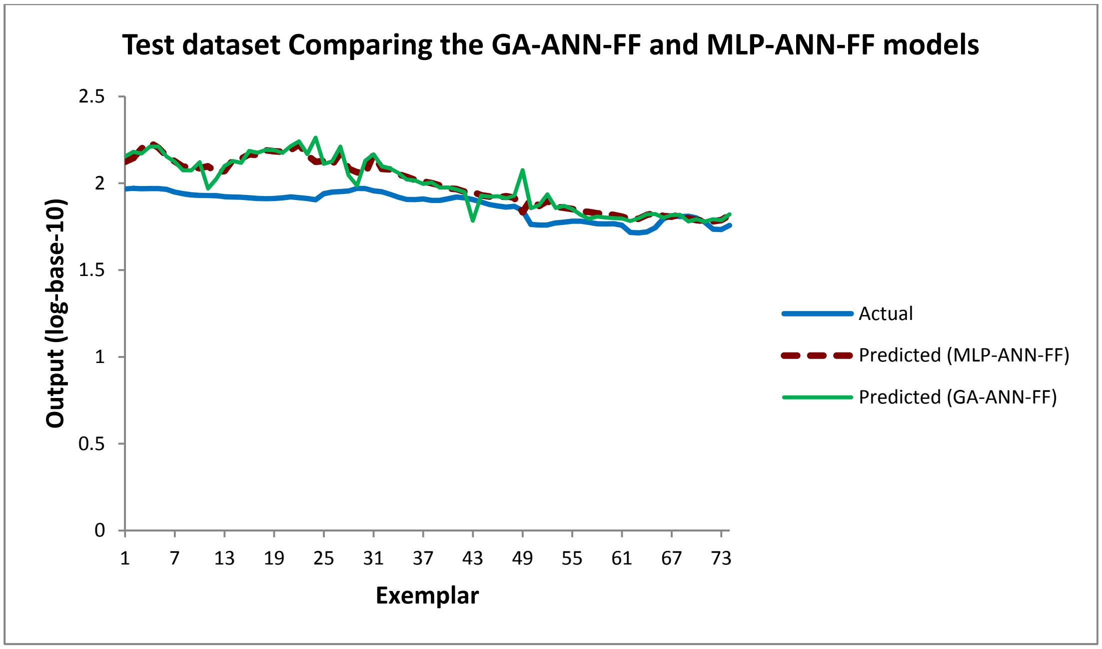


Figure 4.7 Comparison on the GA-ANN-FF and MLP-ANN-FF test results with the Actual discharge flow

Table 4.1: Test dataset (10%) of the total data Results– Comparing developed GA-ANN-FF model with MLP-ANN-FF

	Developed GA-ANN-FF			Developed MLP-ANN-FF		
Date	Actual	Predicted	Error%	Actual	Predicted	Error%
10/20/2003	1.96693916	2.1239635	7.98	1.9669392	2.154443	9.53
10/21/2003	1.97043986	2.1460355	8.91	1.9704399	2.180509	10.7
10/22/2003	1.96857634	2.199884	11.75	1.9685763	2.171373	10.3
10/23/2003	1.96922948	2.2328021	13.38	1.9692295	2.210722	12.3
10/24/2003	1.96899633	2.2048195	11.98	1.9689963	2.211793	12.3
10/25/2003	1.96501345	2.1567738	9.76	1.9650135	2.153446	9.59
10/26/2003	1.94870631	2.126117	9.10	1.9487063	2.124298	9.01
10/27/2003	1.93976878	2.094534	7.98	1.9397688	2.074249	6.93
10/28/2003	1.93318348	2.0933027	8.28	1.9331835	2.074037	7.29
10/29/2003	1.92982748	2.0859502	8.09	1.9298275	2.121682	9.94
10/30/2003	1.92890769	2.0970214	8.72	1.9289077	1.969742	2.12
10/31/2003	1.92844706	2.0654236	7.10	1.9284471	2.023615	4.93
11/1/2003	1.92251786	2.0695073	7.65	1.9225179	2.098919	9.18
11/2/2003	1.92059286	2.1329441	11.06	1.9205929	2.127686	10.8
11/3/2003	1.91970554	2.140393	11.50	1.9197055	2.117277	10.3
11/4/2003	1.91539984	2.1668135	13.13	1.9153998	2.185884	14.1
11/5/2003	1.91174338	2.1615363	13.07	1.9117434	2.175991	13.8

11/6/2003	1.91169016	2.1895741	14.54	1.9116902	2.191931	14.7
11/7/2003	1.91211573	2.1838297	14.21	1.9121157	2.190439	14.6
11/8/2003	1.91603261	2.1812808	13.84	1.9160326	2.174468	13.5
11/9/2003	1.92127019	2.191733	14.08	1.9212702	2.213124	15.2
11/10/2003	1.91661185	2.2226028	15.97	1.9166118	2.241209	16.9
11/11/2003	1.91222206	2.1600876	12.96	1.9122221	2.167013	13.3
11/12/2003	1.90509397	2.1227297	11.42	1.905094	2.261883	18.7
11/13/2003	1.93956917	2.1282523	9.73	1.9395692	2.111586	8.87
11/14/2003	1.9492924	2.1069914	8.09	1.9492924	2.125341	9.03
11/15/2003	1.95206559	2.1728318	11.31	1.9520656	2.210732	13.3
11/16/2003	1.95573584	2.0835761	6.54	1.9557358	2.043797	4.5
11/17/2003	1.97053283	2.0634057	4.71	1.9705328	1.986326	0.8
11/18/2003	1.96932271	2.0535951	4.28	1.9693227	2.129429	8.13
11/19/2003	1.95621647	2.1656787	10.71	1.9562165	2.16657	10.8
11/20/2003	1.95051089	2.0826036	6.77	1.9505109	2.096929	7.51
11/21/2003	1.93606112	2.0794301	7.41	1.9360611	2.085433	7.72
11/22/2003	1.91970554	2.0521021	6.90	1.9197055	2.059307	7.27
11/23/2003	1.90660437	2.0361006	6.79	1.9066044	2.021408	6.02
11/24/2003	1.90611946	2.0192261	5.93	1.9061195	2.016191	5.77
11/25/2003	1.909235	2.0110842	5.33	1.909235	1.995828	4.54
11/26/2003	1.90085851	2.00107	5.27	1.9008585	2.004162	5.43
11/27/2003	1.90085851	1.9874908	4.56	1.9008585	1.974819	3.89
11/28/2003	1.9099837	1.9699898	3.14	1.9099837	1.975825	3.45
11/29/2003	1.9209056	1.9669662	2.40	1.9209056	1.964299	2.26
11/30/2003	1.91576907	1.9504687	1.81	1.9157691	1.94909	1.74
31/11/2003	1.90558003	1.949718	2.32	1.90558	1.784349	-6.36
12/1/2003	1.89164894	1.9305343	2.06	1.8916489	1.928971	1.97
12/2/2003	1.87760168	1.9235019	2.44	1.8776017	1.922583	2.4
12/3/2003	1.86858567	1.9181425	2.65	1.8685857	1.925205	3.03
12/4/2003	1.86248917	1.9248801	3.35	1.8624892	1.913041	2.71
12/5/2003	1.86664172	1.9168336	2.69	1.8666417	1.928023	3.29
12/6/2003	1.84298347	1.8326756	-0.56	1.8429835	2.075329	12.6
12/7/2003	1.76297849	1.9073084	8.19	1.7629785	1.856216	5.29
12/8/2003	1.75959231	1.8691127	6.22	1.7595923	1.874467	6.53
12/9/2003	1.7594412	1.8974752	7.85	1.7594412	1.936101	10
12/10/2003	1.77144049	1.8644689	5.25	1.7714405	1.857611	4.86
12/11/2003	1.7761198	1.8585038	4.64	1.7761198	1.869813	5.28
12/12/2003	1.78161178	1.8513147	3.91	1.7816118	1.850942	3.89
12/13/2003	1.78089311	1.8388804	3.26	1.7808931	1.818504	2.11
12/14/2003	1.77458995	1.8347369	3.39	1.77459	1.795361	1.17
12/15/2003	1.76671021	1.8263404	3.38	1.7667102	1.807657	2.32

12/16/2003	1.76522137	1.8235748	3.31	1.7652214	1.80253	2.11
12/17/2003	1.76671021	1.8175714	2.88	1.7667102	1.798654	1.81
12/18/2003	1.75853342	1.8069099	2.75	1.7585334	1.796841	2.18
12/19/2003	1.71650416	1.7933446	4.48	1.7165042	1.782422	3.84
12/20/2003	1.71374248	1.7949984	4.74	1.7137425	1.798215	4.93
12/21/2003	1.71991106	1.8169762	5.64	1.7199111	1.820559	5.85
12/22/2003	1.74311763	1.8280999	4.88	1.7431176	1.822957	4.58
12/23/2003	1.7947668	1.8108608	0.90	1.7947668	1.802008	0.4
12/24/2003	1.81170903	1.8069099	-0.26	1.811709	1.816582	0.27
12/25/2003	1.80834604	1.8204014	0.67	1.808346	1.818646	0.57
12/26/2003	1.81029974	1.7969504	-0.74	1.8102997	1.780542	-1.64
12/27/2003	1.79975397	1.7877483	-0.67	1.799754	1.792119	-0.42
12/28/2003	1.77597433	1.7824216	0.36	1.7759743	1.778662	0.15
12/29/2003	1.73519955	1.7804761	2.61	1.7351995	1.791945	3.27
12/30/2003	1.73375884	1.7877008	3.11	1.7337588	1.787277	3.09
12/31/2003	1.75694024	1.8161234	3.37	1.7569402	1.82077	3.63

Performance	Discharge
MSE	0.0217621
r	0.8874795

Performance	Discharge
MSE	0.0240317
r	0.8433755

4.4.2 Observations and Analysis

The MSE and r (correlation coefficient) were used to evaluate the performance of GA-ANN-FF and MLP-ANN-FF for the test dataset. It was observed that r (correlation coefficient) for the overall performance of the models was above 80%. GA-ANN-FF performed slightly better with r (correlation coefficient) of 88% on the desired discharge flow compared to the predicted discharge output and a MSE of 0.0217 as compared to the MLP-ANN-FF model that had an 84% r (correlation coefficient) and a MSE of 0.024 on the test dataset. It should be noted that models were trained with 70% data, 20% was used for validation and the remaining 10% was used in testing the overall performance of the models. The MSE of both GA-ANN-FF and MLP-ANN-FF was close, with close prediction range where they both exhibit an 80% r (correlation coefficient).

4.4.3 Discussions

The results show the GA-ANN-FF model has a slightly superior performance in predicting the discharge flow with an r (correlation coefficient) of 0.887 (88%) (Table 4.1) as compared to MLP-ANN-FF with a r (correlation coefficient) of 0.84 (84%) (Table 4.1). We can conclude that GA-ANN-FF is superior in terms of estimating the discharge flow with minimal error and since it train well with less MSE of 0.028 it can always predict a discharge flow of Nzoia River basin of Rwambwa river gauge well with a minimal MSE.

4.5 Sensitivity of the reference discharge flow

The sensitivity testing assists in evaluating the relative importance among the input variables in the neural network model and how the output would estimate in response to the variation of an input. Results were generated giving variation of the discharge flow output with respect to the variation in each input variables. This was to determine the sensitivity of the input variables that provided much information to the desired discharge flow out. The input variables that yield low sensitivity values were disregarded or removed from the model since they are regarded to be insignificant (Gonzalez et al., 2000). This reduced the size of the network model thus reducing the complexity of the training time; more so might have improved the model performance. Therefore after training the two models (MLP-ANN-FF and GA-ANN-FF), the effect of each input variable on the discharge output was evaluated using NeuroSolution Software.

4.5.1 Results

The sensitivity of the reference discharge flow to the input variables in regards to the daily rainfall, temperature and the discharge flow was performed on both models (MLP-ANN-FF and GA-ANN-FF) differently as shown in Table 4.2, and Table 4.3.

Figure 4.8 and 4.9 indicate the plotted input sensitivity applied on the two models. This analysis assisted in explaining explains the objective in section 3.3.3.1 determining the number of input variables in the MLP-ANN-FF model, in order find out the effect of the input parameters that offered much information to the desired discharge flow output.

Table 4.2 The Sensitivity of the discharge flow output to the 9 metrological variables using the MLP-ANN-FF (9:7:12:1) model

Sensitivity	Discharge Flow
Kitale-Met Rainfall 2002- 2003	0.081845594
Kitale-Soil Rainfall 2001- 2002	0.125622744
Leissa-Farm Rainfall 2001- 2002	0.068546469
Kitale-Met Temp 2002- 2003	0.020486424
Kitale-Met Rainfall 2001- 2002	0.047385864
Kitale-Soil Rainfall 2001- 2002	0.022324005
Leissa-Farm Rainfall 2001- 2002	0.003845752
Kitale-Met Temp 2001- 2002	0.06555623
Rwambwa Discharge Flow 2001	**0.199008292**

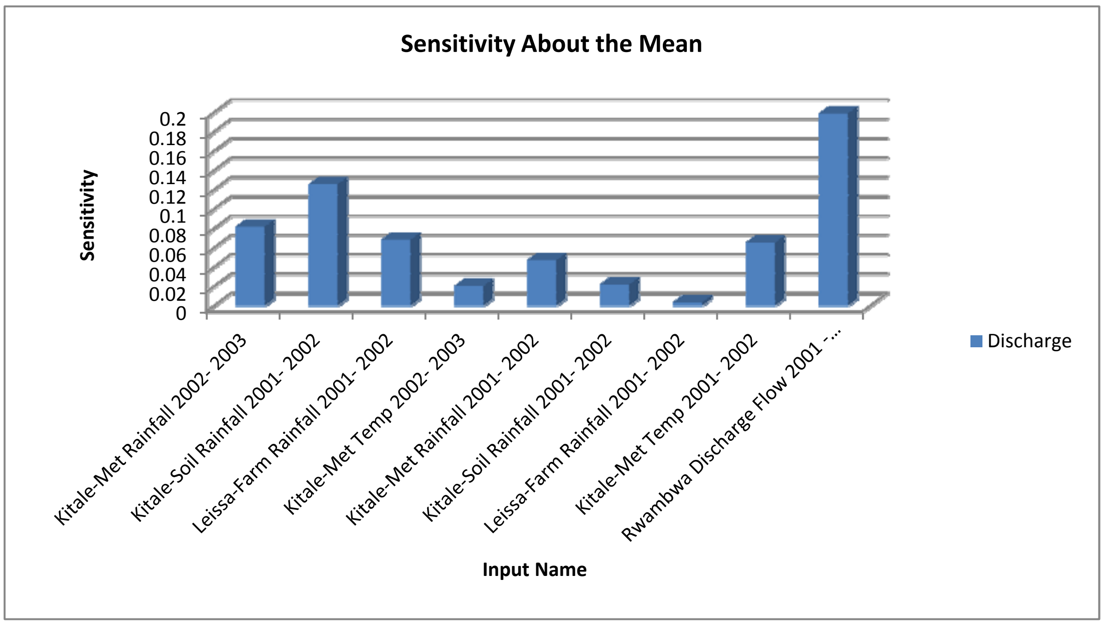


Figure 4.8 - The Sensitivity of the discharge flow output to the 9 metrological variables using the MLP-ANN-FF model

Table 4.3 The Sensitivity of the discharge flow output to the 4 optimized metrological variables using the GA-ANN-FF (**4:6:4:1)** model

Sensitivity	Discharge
Kitale-Met Rainfall 2002- 2003	0.035235
Leissa-Farm Rainfall 2001- 2002	0.012333
Leissa-Farm Rainfall 2001- 2002	0.028584
Rwambwa Discharge Flow 2001 - 2002	**0.327179**

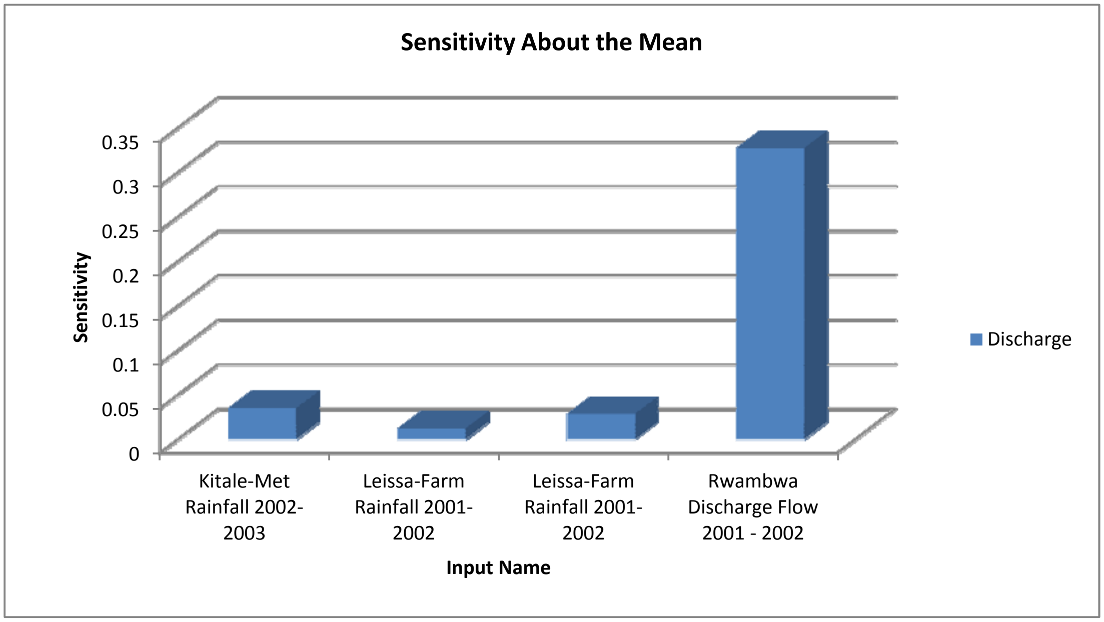


Figure 4.9 - The Sensitivity of the discharge flow output to the 4 optimized metrological variables using the GA-ANN-FF (**4:6:4:1)** model

4.5.2 Observations and Analysis

From the Figure 4.8 and 4.9 it was certain that Rwambwa Discharge Flow 2001, Kitale-Met Temp 2001- 2002, Leissa-Farm Rainfall 2001- 2002, Kitale-Met Rainfall 2002- 2003 were the most input sensitive variables. Rwambwa Discharge Flow 2001 input variable emerged the best scoring the highest in both models. The 4 optimized inputs that the GA-ANN-FF (**4:6:4:1)** model provided; Rwambwa Discharge Flow 2001 – 2002 scored the highest.

4.5.3 Discussions

This analysis was to assist explains the objective in section 3.9.1 determining the number of variables in the MLP-ANN-FF model, and find out the impact of the input parameters and remove unnecessary inputs. After performing a sensitivity analysis, it was concluded that all the selected input parameters in Figure 4.8 were necessary for modeling our MLP-ANN-FF (9:7:12:1) models, but for GA-ANN-FF (4:6:4:1) model the input variables were reduced from 9 to 4 inputs, these were the input variables that offered much information to the desired discharge flow output, reducing the complexity of the network hence better accuracy.

4.6 Summary

The results realized show it was possible to develop an ANN model optimized by genetic algorithm (GA) in the estimation of the discharge flow for Nzoia River basin at Rwambwa river gauge. There were two ANN models that were developed. MLP-ANN-FF that was developed with feedforward multilayer perceptron with BP as the training algorithm, the second model was developed using the same basis of the optimal MLP-ANN-FF model, but its inputs and number of neurons within the hidden layers were optimized with genetic algorithm (GA); GA was used as an alternative to BP in training of the model that yielded a 4:6:4:1 GA-ANN-FF model.

A comparison between the hybrid of neural network and genetic algorithm (GA-ANN-FF) and MLP-ANN-FF model and performance measures values were performed. The MSE, and r (correlation coefficient) were used to evaluate the performance of GA-ANN-FF and MLP-ANN-FF for data set. It was observed that r (correlation coefficient) for the overall performance of the models was above 70%. GA-ANN-FF performed slightly better with r (correlation coefficient) of 88% on the desired discharge flow compared to the predicted output and a MSE of 0.021 as compared to the MLP-ANN-FF model that had an 84% r (correlation coefficient) and a MSE of 0.024 on the training data. It should be noted that models were trained with 70% data, 20% was used for validation and the remaining 10% was used in testing the overall performance of the models.

GA-ANN-FF model shows a slightly superior performance in predicting the discharge flow with an r (correlation coefficient) of 0.87813 (88%) (Figure 4.5) as compared to MLP-ANN-FF with a r (correlation coefficient) of 0.84 (80%) (Figure 4.2). We can conclude that GA-ANN-FF is superior in terms of estimating the discharge flow with minimal error and since it train well with the least MSE of 0.0217 it can always predict a discharge flow well with a minimal MSE.

Chapter 5 – Conclusion and Recommendations

5.1 Conclusion

This chapter will serve to review the problem statement, objectives, and the methods applied on the objectives. The findings of the research are also discussed providing the behavior of the results obtained. Finally we conclude with the summary of the achievements and contributions.

5.2 Problem statement and objectives

There are a number of models on early flood warning initiated by the government for Nzoia River basin that give discharge forecast, they are linear models, where to some extent their performance in regards to the peak streamflows is inconsistent (Masibayi et al., n.d.). Since streamflow course for the daily discharge flow is generally recognized as nonlinear (Guven, 2009), reliable intelligent nonlinear transfer function that capable to handle nonlinearity estimation problem for streamflow (discharge) is crucial in water resource planning and flood management.

For lack of intelligence and nonlinearity in these early warming models initiated by FEWST for discharge forecast, an intelligent algorithm, the ANN with nonlinear transfer function of TanhAxon, was used in developing such models that were capable to handle nonlinearity problem. The models could eventually be used to assist in predicting streamflows given historical data based on daily rainfall, temperature, and discharge flow. Such tools can provide reliable intelligent estimation of Nzoia River Basin streamflow (discharge) that will enable relevant agents in water resource planning and flood management and the public aware when a flood might occur and the areas.

5.3 Methods used to achieve the objectives

In this research an artificial neural network (ANN) technique and optimizing the multilayer perceptron with a genetic algorithm (GA), has been studied. The objectives that were on focus to investigate whether genetic algorithm (GA) can lead to better accuracy and least errors of the ANN model, for estimating the discharge flow of Nzoia River Basin at Rwambwa River gauge. To develop two ANN models (MLP-ANN-FF and GA-ANN-FF) for purpose of estimating the discharge flow of Nzoia River Basin at Rwambwa River gauge, by applying artificial neural network (ANN) technique and optimizing the multilayer perceptron (MLP-ANN-FF) neural network model with a genetic algorithm (GA). After training the two models the research also intended test and evaluate the overall performance of the models, by comparing their performance with the Rwambwa River gauge discharge flow data for 2000 to 2003. Recommending a suitable intelligent model based on the results of the two above objectives was the final objective. The data for Nzoia River Basin at Rwambwa River gauge for period 2002 to 2003 that was concurrent in all 3 weather stations and the one river gauge was used, to achieve these objectives.

The techniques that were employed in developing the two models were ANN MLP feed forward and use of genetic algorithm (GA) to optimize the input parameters, and number of neurons in the hidden layers. This research uses two experiments to develop the two models (MLP-ANN-FF and GA-ANN-FF). In experiment 1; MLP-ANN-FF was developed with a feed forward Multilayer perceptron network with 3 layers (Input, Hidden and Output). Backpropagation with momentum was applied as the training algorithm, using nonlinear TanhAxon transfer function.

Proper selection of input data was determined through a series of experiments. High MSE were observed and these were resolved by preprocessing the data further hence arriving at an optimal 9 input variables for the MLP-ANN-FF model. Number of hidden layers and neurons in the hidden layers was also determined arriving at an optimal 9:7:12:1 configuration setting. To determine the sensitivity of the optimal MLP-ANN-FF (9:7:12:1) model in regards to data splitting that included 50%, 60%, 70%, 80% and 90% of the total used for training. MLP-ANN-FF with 70% training set performed sufficiently well in estimating the discharge flow. In experiment 2 GA-ANN-FF model was developed using the same basis for the optimal MLP-ANN-FF (9:7:12:1) model. Genetic algorithm (GA) was then applied on the GA-ANN-FF model to avoid the trial and error in selecting the optimal inputs and determining the optimal number of neurons in the hidden layers. This optimized the GA-ANN-FF model yielding a 4:6:4:1

topology. This reduced the size of the network, reducing the inputs from 9 to 4 and number of neurons within the 2 hidden layers, thus reducing the complexity of the model. It was a leaner network as compared to MLP-ANN-FF model of 9:7:12:1 topology.

5.4 Major findings of this research

After training the two models (MLP-ANN-FF and GA-ANN-FF), they were tested based on the daily rainfall, average temperature for Nzoia River Basin at Rwambwa River gauge that was not part of training to determine the overall performance of the models.

The overall performance and accuracy of the integration of ANN MLP and GA algorithm (GA-ANN-FF) model was compared with MLP-ANN-FF to find the effect of genetic algorithm on ANN MLP. The issues with ANN trained with BP algorithm not able to out of local minima (Devi et al., 2012). The main advantage of to genetic optimized ANN model (GA-ANN-FF) is it does not get stuck into the local minima and that's why it has been applied in this research study.

At the first experiment MLP-ANN-FF was developed trained and tested without any genetic optimization. The evaluation criterion that was used to compare and evaluate the results of the two models was r (correlation coefficient) and mean square error (MSE). In determining the optimal configuration setting for MLP-ANN-FF model a baseline of 9:2:1 and 9:2:2:1 was employed. The activation function of the hidden layers and output layer was set to TanhAxon. Varying the number of neurons in the hidden layers was from 2 to 30 with a learning rate of 0.1. The optimal configuration setting was realized to be 9:7:12:1. A least MSE in training of 0.078 and r (correlation coefficient) of 0.84 was realized. Also it was observed that the number of the number of neurons increased from the baseline of 9:2:2:1 to the optimal configuration of 9:7:12:1. This can be attributed by the complexity of the input data used in the training stage, starting the model with low network complexity (9:2:2:1) that yielded poor performance this necessitated an increase in of the number of neurons in the hidden layers to a 9:7:12:1 due to the complexity of the data. With data splitting that included 50%, 60%, 70%, 80%, and 90% of the total used for training. MLP-ANN-FF with 70% training set was realized to be the best model that estimated the discharge flow with a least MSE of 0.023 and an r (correlation coefficient) of desired versus predicted values of the discharge flow. This shows that MLP-ANN-FF is sensitive to data splitting

In the second experiment; using the same basis defined for MLP-ANN-FF, GA was applied to optimize the input parameters and neurons with the hidden layers for GA-ANN-FF model. Although by applying the GA there was an increased computational time there was considerable improvement in GA-ANN-FF performance. When GA optimization algorithm was applied the input variables were reduced from 9 to 4 with the hidden layers were neurons reduced to 6 and 4 yielding a GA-ANN-FF with configuration setting of 4:6:4:1. It was observed that, the number of input variables reduced to 4 this is the ideal number that offered much information to the desired discharge flow output, hence reducing the complexity of the network with lesser neurons in the hidden layers thus better accuracy. Also it can be attributed the evolution nature of the GA optimization that it's a fitness that is used to rank the potential solution.

Machine learning methods are data dependant and perform significantly well when large data set is applied (Babinec and Pospíchal, 2009). The two models (MLP-ANN-FF and GA-ANN-FF) were found to estimate the discharge flow significantly well despite inadequate training data. The predictive accuracy of GA-ANN-FF was observed to be better than that of MLP-ANN-FF developed using the same basis and the same data. GA-ANN-FF with 4:6:4:1 configuration setting estimated the discharge flow better with a MSE of 0.0217 and r (correlation coefficient) of 0.88 (90%). This can be assumed as satisfactory results as it compares well with those obtained by other researchers (Puttinaovarat et al., n.d.) As discussed in the literature review section 2.3.

5.5 Discussion

Applying the two models on Nzoia River at Rwambwa river gauge in western Kenya has demonstrated the possibility of using climatic data in a given river basin to estimate the discharge flow. GA-ANN-FF (4:6:4:1) model has shown a slightly superior performance in predicting the discharge flow for Nzoia River Basin at Rwambwa river gauge, with a r (correlation coefficient) of 0.898 (90%) (Figure 4.3.2) as compared to MLP-ANN-FF (9:7:12:1) with a r (correlation coefficient) of 0.84 (80%) (Figure 4.2.1). This can be assumed as satisfactory results as it compares well with those obtained by other researchers (Puttinaovarat et al., n.d.) As discussed in the literature review section 2.3.

Although genetic algorithm (GA) reduces the complexity of the model, it trains for long as compared to the conventional multilayer perceptron. In the literature review section 2.3 (Masibayi, and Mutua, 2010) presents a linear regression model for real-time River stage forecasting in Nzoia River Basin, Western Kenya. The linear regression presented shows a superior performance in predicting with better accuracy of coefficient of determination (R^2) of 0.987 as compared to the GA-ANN-FF model that predicts with an accuracy of 0.771 coefficient of determination (R^2). The superior performance on linear regression may attributed to the hourly input data applied with a sample size of over 12,000 entries as compared to the GA-ANN-FF model that uses inputs of lumped daily data with a sample size of 744 entries. This confirms that ANN is data dependant and performs significantly well when large data set is applied (Babinec and Pospíchal, 2009).

Due to the inadequate data collected it was not possible to make estimation based on hourly periods. Both models were found to predict and train well despite inadequate training data obtained from KMD. The GA-ANN-FF model was also able to optimize the input variables to 4 from 9 and reduced the neurons in the 2 hidden layers yielding a neural network topology with a 4:6:4:1 configuration setting, which was more optimal as compared to the MLP-ANN-FF (9:7:12:1) model though it took long in training. This confirms genetic algorithm (GA) combination with ANN can improve the model performance by optimizing the input variables and reducing the number of neurons in the hidden layers hence reducing the complexity of the model and estimating well for the desired output.

5.6 Summary of Achievements

The integration of GA with ANN MLP has shown a good effect in the discharge estimation results. It can therefore be concluded that the objectives of developing an ANN MLP that can be optimized with GA for estimating the discharge flow for Nzoia River at Rwambwa River gauge.

This confirms the integration of GA with ANN MLP can develop and reduce the complexity of the neural network architecture, where the data input variables were reduced from 9 to 4 that offered much information to the desired discharge flow output, and more so reducing the neurons in within the hidden layers. The estimation of the discharge flow of GA-ANN-FF model was found to predict well with a least MSE of 0.0217 and r (correlation coefficient) of 0.887 (88%) despite inadequate training data. This means that the proposed genetic optimized model (GA-ANN-FF) can be relied upon to yield good results despite insufficient data set.

The contribution of this research is an optimized GA-ANN-FF model with least MSE and satisfactory r (correlation coefficient) on the desired discharge flow compared to the predicted output. It's a model that can be applied on other discharge flow catchment area, with a varying configuration setting on its parameters.

5.6 Recommendations

In this research there were only three input factors, namely rainfall, temperature and discharge (streamflow). Other causative flood factors such, flood plain in the past, terrain elevation, water density, water blockage, sub basin areas, soil drainage capability, land use, are should be considered. Further training on the two models should be performed with larger data sets tom compare the performance of the models in the discharge flow. Also more training should be performed on the GA-ANN-FF model varying the population size. Developed model can be useful in decision making for metrologists and others who work with discharge flow forecast. More tests should be carried out to observe whether a multilayer neural network optimized with genetic algorithm is sensitive to data splitting.

Reference

Abhishek, K., Kumar, A., Ranjan, R., Kumar, S., 2012. A rainfall prediction model using artificial neural network, in: Control and System Graduate Research Colloquium (ICSGRC), 2012 IEEE. IEEE, pp. 82–87.

ADCL (Appropriate Development Consultants Limited), 2006. Western Kenya Community–Driven –Development and Flood Mitigation Project.

Awan, Z.K., Khan, A., Iftikhar, A., Zahid, S., Malik, A., 2012. Analysis of Hybrid Neural Networks for Improved Performance. Int. J. Comput. Appl. 50.

Babinec, Š., Pospíchal, J., 2009. Echo State and FIR Neural Networks: Comparison of Predictive Abilities, in: Proceedings of the 15th International Conference on Soft Computing, Mendel 2009. pp. 160–165.

Che, Z.-G., Chiang, T.-A., Che, Z.-H., 2011. Feed-forward neural networks training: a comparison between genetic algorithm and back-propagation learning algorithm. Int J Innov Comput Inf 7, 5839–5850.

Chen, S.M., Wang, Y.M., Tsou, I., 2013. Using artificial neural network approach for modelling rainfall–runoff due to typhoon. J. Earth Syst. Sci. 122, 399–405.

Cho, H., Park, W.S., 2002. Neural network applications in automated optical inspection: state of the arts, in: International Symposium on Optical Science and Technology. International Society for Optics and Photonics, pp. 224–236.

Deshmukh, R.P., Ghatol, A.A., 2010a. Comparative study of temporal neural networks for short term flood forecasting. Int. J. Comput. Appl. 5, 24–28.

Deshmukh, R.P., Ghatol, A.A., 2010b. Short Term Flood Forecasting using Static Neural Networks a Comparative Study. Int. J. Comput. Sci. Netw. Secur. 10, 69–74.

Devi, C.J., Reddy, B.S.P., Kumar, K.V., Reddy, B.M., Nayak, N., 2012. ANN Aproach for Weather Prediction using Back Propagation. Intenational J. Eng. Trends Technol. 3.

Dulo, S.O., Odira, P.M.A., Nyadwa, M.O., Okelloh, B.N., 2010. Integrated flood and drought management for sustainable development in the Nzoia River Basin. Nile Basin Water Sci Eng J 3.

Flood Mitigation Strategy, 2009.

Gonzalez, S., Economic, C., Branch, F.P., 2000. Neural networks for macroeconomic forecasting: a complementary approach to linear regression models. Department of Finance Canada.

Guven, A., 2009. Linear genetic programming for time-series modelling of daily flow rate. J. Earth Syst. Sci. 118, 137–146.

Heednacram, A., 2014. Suspended Sediment Forecast of Khlong Bang Yai, Phuket. Int. J. Eng. Technol. 338–345. doi:10.7763/IJET.2014.V6.723

Kenya water security and climate resilience project, 2013.

Khan, S.I., Adhikari, P., Hong, Y., Vergara, H., F Adler, R., Policelli, F., Irwin, D., Korme, T., Okello, L., 2011. Hydroclimatology of Lake Victoria region using hydrologic model and satellite remote sensing data. Hydrol. Earth Syst. Sci. 15, 107–117. doi:10.5194/hess-15-107-2011

Kia, M.B., Pirasteh, S., Pradhan, B., Mahmud, A.R., Sulaiman, W.N.A., Moradi, A., 2012. An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. Environ. Earth Sci. 67, 251–264. doi:10.1007/s12665-011-1504-z

Limpert, E., Stahel, W.A., Abbt, M., 2001. Log-normal Distributions across the Sciences: Keys and Clues On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question. BioScience 51, 341–352.

LUK, K., BALL, J.E., SHARMA, A., 2001. An Application of Artificial Neural Networks for Rainfall Forecasting. Math. Comput. Model. 33, 683–693.

Masibayi, E., Mutua, F., Otengi, S.B.B., Wakhungu, J.W., n.d. Real-Time River Stage Forecasting Using Upstream Stage Approach for Flood Management, in Nzoia River Basin, Western Kenya. Reduct. Confl. Resolut. Sustain. Dev. 422.

Maskey, S., 2004. Modelling Uncertainty in Flood Forecasting Systems. Taylor & Francis.

Mitchell, T.M., 1997. Machine Learning. McGraw-Hill, New York.

Montana, D.J., Davis, L., 1989. Training Feedforward Neural Networks Using Genetic Algorithms., in: IJCAI. pp. 762–767.

Moore, R.J., Bell, V.A., Environment Agency, 2001. Comparison of rainfall-runoff models for flood forecasting: part 1: literature review of models. Environment Agency.

NirmalaDevi, M., Mohankumar, N., Arumugam, S., 2009. Modeling and Analysis of Neuro-Genetic Hybrid System on FPGA. Electron. Electr. Eng.

Nzoia River Basin management initiative. A public private partnership programme Between water resources management authority and civil society, learning institutions and communities., 2006.

Onyari, E., Ilunga, F., 2010. Application of MLP neural network and M5P model tree in predicting streamflow: A case study of Luvuvhu catchment, South Africa, in: International Conference on Information and Multimedia Technology (ICMT 2010), Hong Kong, China. pp. V3–156.

Onywere, S.M., Getenga, Z.M., Baraza, W., Mwakalila, S.S., Twesigye, C.K., Nakiranda, J., 2007. Intensification of Agriculture as the Driving Force in the Degradation of Nzoia River Basin: the Challenges of Watershed Management, in: Publication of the Lake Abaya Research Symposium.

Perez, S., n.d. Apply genetic algorithm to the learning phase of a neural network.

Plate, E.J., 2009. HESS Opinions:“ Classification of hydrological models for flood management”. Hydrol. Earth Syst. Sci. 13.

Priddy, K.L., Keller, P.E., 2005. Artificial Neural Networks: An Introduction. Society of Photo Optical.

Puttinaovarat, S., Horkaew, P., Khaimook, K., n.d. Configuring ANN for Inundation Areas Identification based on Relevant Thematic Layers.

Rezaeianzadeh, M., Tabari, H., Arabi Yazdi, A., Isik, S., Kalin, L., 2014. Flood flow forecasting using ANN, ANFIS and regression models. Neural Comput. Appl. 25, 25–37. doi:10.1007/s00521-013-1443-6

See, L., Dougherty, M., Openshaw, S., 1997. Some initial experiments with neural network models of flood forecasting on the river ouse, in: Second Annual Conference of GeoComputation’97 & SIRC’971997.

Shrestha, R.R., Theobald, S., Nestmann, F., 2005. Simulation of flood flow in a river system using artificial neural networks. Hydrol. Earth Syst. Sci. 9.

Steege, F.-F., Stephan, V., Gro\s s, H.-M., n.d. Effects of Noise-Reduction on Neural Function Approximation.

Suliman, A., Nazri, N., Othman, M., Abdul, M., n.d. ARTIFICIAL NEURAL NETWORK AND SUPPORT VECTOR MACHINE IN FLOOD FORECASTING: A REVIEW.

Thirumalaiah, K., Deo, M.C., 1998. Real-Time Flood Forecasting Using Neural Networks. Comput.-Aided Civ. Infrastruct. Eng. 13, 101–111.

Tingsanchali, T., 2000. Forecasting model of Chao Phraya river flood levels at Bangkok, in: International Conference on Chao Phraya Delta. Bangkok. Thailand.

APPENDICES

APPENDIX 1 – Sample concurrent data from 2000 to 2003

The complete data set is made up of 1488 data points, for period 2000 to 2003.

Table A1-1: Sample data collected from 3 Weatherstations, and 1 water level station, Rwambwa Gauge Station (RGS) (SID-1EF01). Data Provided by KMD
Kitale metrological (SID-8834098), Kitale soil conservation service – office (SID-8834097), Leissa farm – Kitale (SID-8835039)

		Input							Output
		Daily Rainfall (mm) (3 Stations)				Daily Temp (Kitale-Met)		Daily Ave Temp	Daily Discharge
# of data	Date	Kitale Met	KitaleSCO Met	LeissaFarm Met	Ave Rainfall	Temp Max	Temp Min	Ave Temp	Discharge
1	1/1/2000	0	0	0	0	28.6	9.8	19.2	39.3
2	1/2/2000	0	0	0	0	28	10.9	19.45	38
3	1/3/2000	0	0	0	0	28	10.6	19.3	36.1
4	1/4/2000	0	0	0	0	28	9.8	18.9	33.9
5	1/5/2000	0	0	0	0	28.8	8.5	18.65	31.9
6	1/6/2000	0	0	0	0	28.4	8.3	18.35	30.5
7	1/7/2000	0	0	0	0	28.7	9.6	19.15	29.5
8	1/8/2000	0	0	0	0	28.5	10.2	19.35	29
9	1/9/2000	0	0	0	0	27.3	11.2	19.25	28.1
10	1/10/2000	0	0	0	0	24.9	13.2	19.05	26.8
11	1/11/2000	0	0	0	0	27.9	12.4	20.15	27.1
12	1/12/2000	0	0	0	0	29	10.9	19.95	27.3
13	1/13/2000	0	0	0	0	28.8	10.3	19.55	28
14	1/14/2000	0	0	0	0	29	9.6	19.3	27.5
15	1/15/2000	0	0	0	0	28	9.8	18.9	27.7
16	1/16/2000	0	0	0	0	29.2	8.9	19.05	27.9
17	1/17/2000	0	0	0	0	29.9	10.1	20	27.4
18	1/18/2000	0	0	3	1	28.9	11.4	20.15	28.3
19	1/19/2000	0	0	0	0	26.7	10	18.35	29.4
20	1/20/2000	0	0	0	0	27	10.7	18.85	27.9
21	1/21/2000	13.6	2.5	0	5.366666667	26.4	13.4	19.9	28
22	1/22/2000	5.7	20.1	0	8.6	27.4	11.8	19.6	25.4
23	1/23/2000	0	0	0	0	28	10.7	19.35	23.1
24	1/24/2000	0	0	0	0	28.3	10.2	19.25	23.1
25	1/25/2000	0	0	0	0	27.5	10.1	18.8	29.58
26	1/26/2000	0	0	0	0	28	6.5	17.25	28.08
27	1/27/2000	0	0	0	0	28	8.1	18.05	26.68
28	1/28/2000	0	0	0	0	29.2	7.7	18.45	24.44
29	1/29/2000	0	0	0	0	29.7	9.8	19.75	23.01
30	1/30/2000	0	0	0	0	29.8	9.4	19.6	23.01
31	1/31/2000	0	0	0	0	28.7	8	18.35	22.94

32	2/1/2000	0	0	0	0	28.5	8.8	18.65	21.97
33	2/2/2000	0	0	0	0	28.1	9.3	18.7	21.77
34	2/3/2000	0	0	0	0	28.7	13	20.85	21.24
35	2/4/2000	0	0	0	0	28.8	10.3	19.55	22.24
36	2/5/2000	0	0	0	0	28	11.5	19.75	21.98
37	2/6/2000	0	0	0	0	29.2	9.1	19.15	22.27
38	2/7/2000	0	0	0	0	29.6	8.2	18.9	21.75
39	2/8/2000	0	0	0	0	29	9.4	19.2	21.02
40	2/9/2000	0	0	0	0	28.2	10	19.1	20.88
41	2/10/2000	0	0	0	0	27.5	10.2	18.85	21.34
42	2/11/2000	0	0	0	0	29	9.5	19.25	20.9
43	2/12/2000	0	0	0	0	28.3	9	18.65	19.98
44	2/13/2000	0	0	0	0	29.9	8.8	19.35	18.68
45	2/14/2000	0	0	0	0	30	11	20.5	17.55
46	2/15/2000	0	0	0	0	29.8	11.2	20.5	17.14
47	2/16/2000	0	0	0	0	29.9	9.3	19.6	17.26
48	2/17/2000	0	0	0	0	28.8	9.7	19.25	16.7
49	2/18/2000	0	0	0	0	28.7	9.8	19.25	16.72
50	2/19/2000	0	0	0	0	28.5	9.2	18.85	16.78
51	2/20/2000	0	0	0	0	29.3	10.2	19.75	17.46
52	2/21/2000	0	0	0	0	30.2	8.6	19.4	16.52
53	2/22/2000	0	0	0	0	30	10.2	20.1	15.99
54	2/23/2000	0	0	0	0	30.6	9.6	20.1	15.31
55	2/24/2000	0	0	0	0	29.8	11	20.4	14.81

APPENDIX 2 – The sample input output data applied with the 9 inputs

The complete data set is made up of 744 data points.

Table A2-1: Input output data, 9 inputs and 1 output

9 inputs									1 output
Kitale-Met Rainfall 2002-2003	Kitale-Soil Rainfall 2002-2003	Leissa-Farm Rainfall 2002-2003	Kitale-Met Temp 2002-2003	Kitale-Met Rainfall 2000-2002	Kitale-Soil Rainfall 2000-2001	Leissa-Farm Rainfall 2000-2001	Kitale-Met Temp 2000-2001	Rwambwa Discharge Flow 2000 -2001	Rwambwa Discharge Flow 2002 - 2003
0	0	0	19.85	0	0	0	19.2	1.5943926	***1.804616***
2.1	0	18.2	20.9	0	0	0	19.45	1.5797836	***1.790144***
0	0	0	20.1	0	0	0	19.3	1.5575072	***1.775173***
0	6	0	18.95	0	0	0	18.9	1.5301997	***1.783904***
1.8	0	0	19.35	0	0	0	18.65	1.5037907	***1.739018***
0	0	0	20.4	0	0	0	18.35	1.4842998	***1.697578***
0	0	0	19.75	0	0	0	19.15	1.469822	***1.676053***
0	0	0	18.8	0	0	0	19.35	1.462398	***1.650405***
0	0	0	20.45	0	0	0	19.25	1.4487063	***1.631951***
0	0	0	20.4	0	0	0	19.05	1.4281348	***1.631241***
0.5	0	27.3	20.1	0	0	0	20.15	1.4329693	***1.66096***
1.1	0	0	19.65	0	0	0	19.95	1.4361626	***1.651181***
0	0	0	19.05	0	0	0	19.55	1.447158	***1.646502***
1.4	0	0	19.15	0	0	0	19.3	1.4393327	***1.64286***
0.6	0	0	17.7	0	0	0	18.9	1.4424798	***1.642761***
0	0	0	17.7	0	0	0	19.05	1.4456042	***1.605413***
0	0	0	19.15	0	0	0	20	1.4377506	***1.586362***
0	0	0	20.2	0	0	3	20.15	1.4517864	***1.563718***
0	0	0	20.05	0	0	0	18.35	1.4683473	***1.546296***
0	0	0	19.5	0	0	0	18.85	1.4456042	***1.530328***
2.4	0	0	20.75	13.6	2.5	0	19.9	1.447158	***1.505693***
0	0	0	5.95	5.7	20.1	0	19.6	1.4048337	***1.480869***
0	0	0	19.95	0	0	0	19.35	1.363612	***1.467312***
0	0	0	19.45	0	0	0	19.25	1.363612	***1.459392***
0	0	0	19.9	0	0	0	18.8	1.4709982	***1.453471***
0	0	0	18.45	0	0	0	17.25	1.4483971	***1.449478***
0	0	0	19.5	0	0	0	18.05	1.4261858	***1.427973***
0.5	0	0	20.85	0	0	0	18.45	1.3881012	***1.417306***
0	0	0	20.2	0	0	0	19.75	1.3619166	***1.421275***
0	0	0	19.4	0	0	0	19.6	1.3619166	***1.422426***
0	0	0	19.35	0	0	0	18.35	1.3605934	***1.416807***
0	0	0	19.7	0	0	0	18.65	1.3418301	***1.429914***
0	0	0	0	0	0	0	18.7	1.3378584	***1.466571***
0	0	0	20.9	0	0	0	20.85	1.3271545	***1.445293***
0	0	0	0	0	0	0	19.55	1.3471348	***1.451172***
0	0	0	20.6	0	0	0	19.75	1.3420277	***1.455758***
0	0	0	20.65	0	0	0	19.15	1.3477202	***1.473049***

0	0	0	21.25	0	0	0	18.9	1.3374593	***1.500785***
0	0	0	21.1	0	0	0	19.2	1.3226327	***1.494711***
0	0	0	22.05	0	0	0	19.1	1.3197305	***1.457125***
0	0	0	20.9	0	0	0	18.85	1.3291944	***1.427811***
0	0	0	21.1	0	0	0	19.25	1.3201463	***1.393926***

APPENDIX 3 – The sample discharge flow output data transformed with Logarithmic

Logarithmic transformation; column 10 was transformed with log base 10; output is shown in column 11. Complete data set is made up of 744 data points.

Table A3-1: Sample data discharge output transformed with Logarithmic transformation of log base 10

9 inputs										1 output
Kitale-Met Rainfall 2002-2003	Kitale-Soil Rainfall 2002-2003	Leissa-Farm Rainfall 2002-2003	Kitale-Met Temp 2002-2003	Kitale-Met Rainfall 2000-2002	Kitale-Soil Rainfall 2000-2001	Leissa-Farm Rainfall 2000-2001	Kitale-Met Temp 2000-2001	Rwambwa Discharge Flow 2000 -2001	Rwambwa Discharge Flow RawData 2002 - 2003	Rwambwa Discharge Flow LogBase10 Tranformed 2002 - 2003
0	0	0	19.85	0	0	0	19.2	1.594393	63.77	***1.804616417***
2.1	0	18.2	20.9	0	0	0	19.45	1.579784	61.68	***1.790144365***
0	0	0	20.1	0	0	0	19.3	1.557507	59.59	***1.775173385***
0	6	0	18.95	0	0	0	18.9	1.5302	60.8	***1.783903579***
1.8	0	0	19.35	0	0	0	18.65	1.503791	54.83	***1.739018246***
0	0	0	20.4	0	0	0	18.35	1.4843	49.84	***1.697578034***
0	0	0	19.75	0	0	0	19.15	1.469822	47.43	***1.676053125***
0	0	0	18.8	0	0	0	19.35	1.462398	44.71	***1.65040467***
0	0	0	20.45	0	0	0	19.25	1.448706	42.85	***1.631950826***
0	0	0	20.4	0	0	0	19.05	1.428135	42.78	***1.63124078***
0.5	0	27.3	20.1	0	0	0	20.15	1.432969	45.81	***1.660960292***
1.1	0	0	19.65	0	0	0	19.95	1.436163	44.79	***1.651181063***
0	0	0	19.05	0	0	0	19.55	1.447158	44.31	***1.64650175***
1.4	0	0	19.15	0	0	0	19.3	1.439333	43.94	***1.642860053***
0.6	0	0	17.7	0	0	0	18.9	1.44248	43.93	***1.642761203***
0	0	0	17.7	0	0	0	19.05	1.445604	40.31	***1.605412798***
0	0	0	19.15	0	0	0	20	1.437751	38.58	***1.586362223***
0	0	0	20.2	0	0	3	20.15	1.451786	36.62	***1.56371834***
0	0	0	20.05	0	0	0	18.35	1.468347	35.18	***1.546295835***
0	0	0	19.5	0	0	0	18.85	1.445604	33.91	***1.53032779***
2.4	0	0	20.75	13.6	2.5	0	19.9	1.447158	32.04	***1.505692508***
0	0	0	5.95	5.7	20.1	0	19.6	1.404834	30.26	***1.480868924***
0	0	0	19.95	0	0	0	19.35	1.363612	29.33	***1.467312063***
0	0	0	19.45	0	0	0	19.25	1.363612	28.8	***1.459392488***
0	0	0	19.9	0	0	0	18.8	1.470998	28.41	***1.453471234***
0	0	0	18.45	0	0	0	17.25	1.448397	28.15	***1.449478399***
0	0	0	19.5	0	0	0	18.05	1.426186	26.79	***1.427972714***
0.5	0	0	20.85	0	0	0	18.45	1.388101	26.14	***1.417305583***
0	0	0	20.2	0	0	0	19.75	1.361917	26.38	***1.421274791***
0	0	0	19.4	0	0	0	19.6	1.361917	26.45	***1.422425676***
0	0	0	19.35	0	0	0	18.35	1.360593	26.11	***1.416806872***
0	0	0	19.7	0	0	0	18.65	1.34183	26.91	***1.429913698***
0	0	0	0	0	0	0	18.7	1.337858	29.28	***1.466571072***

0	0	0	20.9	0	0	0	20.85	1.327155	27.88	***1.445292769***
0	0	0	0	0	0	0	19.55	1.347135	28.26	***1.451172158***
0	0	0	20.6	0	0	0	19.75	1.342028	28.56	***1.455758203***
0	0	0	20.65	0	0	0	19.15	1.34772	29.72	***1.473048805***
0	0	0	21.25	0	0	0	18.9	1.337459	31.68	***1.500785173***
0	0	0	21.1	0	0	0	19.2	1.322633	31.24	***1.494711025***
0	0	0	22.05	0	0	0	19.1	1.31973	28.65	***1.457124626***
0	0	0	20.9	0	0	0	18.85	1.329194	26.78	***1.427810573***
0	0	0	21.1	0	0	0	19.25	1.320146	24.77	***1.393926007***

APPENDIX 4 – The sample data which was used to develop GA-ANN-FF model

This data was applied on the ***4:6:4:1*** configuration setting; complete data set is made up of 744 data points

Table A4-1: The sample data applied on **GA-ANN-FF** model

4 Inputs				1 Output
Kitale-Met Rainfall 2002- 2003	**Leissa-Farm Rainfall 2002- 2003**	**Leissa-Farm Rainfall 2000- 2001**	**Rwambwa Discharge Flow 2000 -2001**	**Rwambwa Discharge 2002 - 2003**
0	0	0	1.59439255	***1.8046164***
2.1	18.2	0	1.579783597	***1.7901444***
0	0	0	1.557507202	***1.7751734***
0	0	0	1.530199698	***1.7839036***
1.8	0	0	1.503790683	***1.7390182***
0	0	0	1.484299839	***1.697578***
0	0	0	1.469822016	***1.6760531***
0	0	0	1.462397998	***1.6504047***
0	0	0	1.44870632	***1.6319508***
0	0	0	1.428134794	***1.6312408***
0.5	27.3	0	1.432969291	***1.6609603***
1.1	0	0	1.436162647	***1.6511811***
0	0	0	1.447158031	***1.6465018***
1.4	0	0	1.439332694	***1.6428601***
0.6	0	0	1.442479769	***1.6427612***
0	0	0	1.445604203	***1.6054128***
0	0	0	1.437750563	***1.5863622***
0	0	3	1.451786436	***1.5637183***
0	0	0	1.46834733	***1.5462958***
0	0	0	1.445604203	***1.5303278***
2.4	0	0	1.447158031	***1.5056925***
0	0	0	1.404833717	***1.4808689***
0	0	0	1.36361198	***1.4673121***
0	0	0	1.36361198	***1.4593925***
0	0	0	1.47099817	***1.4534712***
0	0	0	1.448397103	***1.4494784***
0	0	0	1.426185825	***1.4279727***
0.5	0	0	1.388101202	***1.4173056***
0	0	0	1.361916619	***1.4212748***
0	0	0	1.361916619	***1.4224257***
0	0	0	1.360593414	***1.4168069***

0	0	0	1.341830057	***1.4299137***
0	0	0	1.337858429	***1.4665711***
0	0	0	1.327154512	***1.4452928***
0	0	0	1.347134783	***1.4511722***
0	0	0	1.342027688	***1.4557582***
0	0	0	1.347720217	***1.4730488***
0	0	0	1.337459261	***1.5007852***
0	0	0	1.322632712	***1.494711***
0	0	0	1.319730494	***1.4571246***
0	0	0	1.329194415	***1.4278106***
0	0	0	1.320146286	***1.393926***
0	0	0	1.300595484	***1.3523755***
0.7	0	0	1.271376872	***1.3289909***
0	0	0	1.244277121	***1.3283796***
0	0	0	1.234010818	***1.3312248***
0	43	0	1.237040791	***1.3523755***
30.3	0	0	1.222716471	***1.3484996***
0	0	0	1.223236273	***1.3410386***
0	0	0	1.224791956	***1.3889888***
0	0	0	1.242044239	***1.4258601***
0	0	0	1.218010043	***1.5571461***
0	0	0	1.203848464	***1.5842181***
0	0	0	1.184975191	***1.5956064***

APPENDIX 5 - Optimized GA-ANN-FF (*4:6:4:1*) and MLP-ANN-FF (*9:7:12:1*) models

The figures below illustrate the two models that were finally developed using NeuroSolutions software. The pointed circles contain the neurons in the layers.

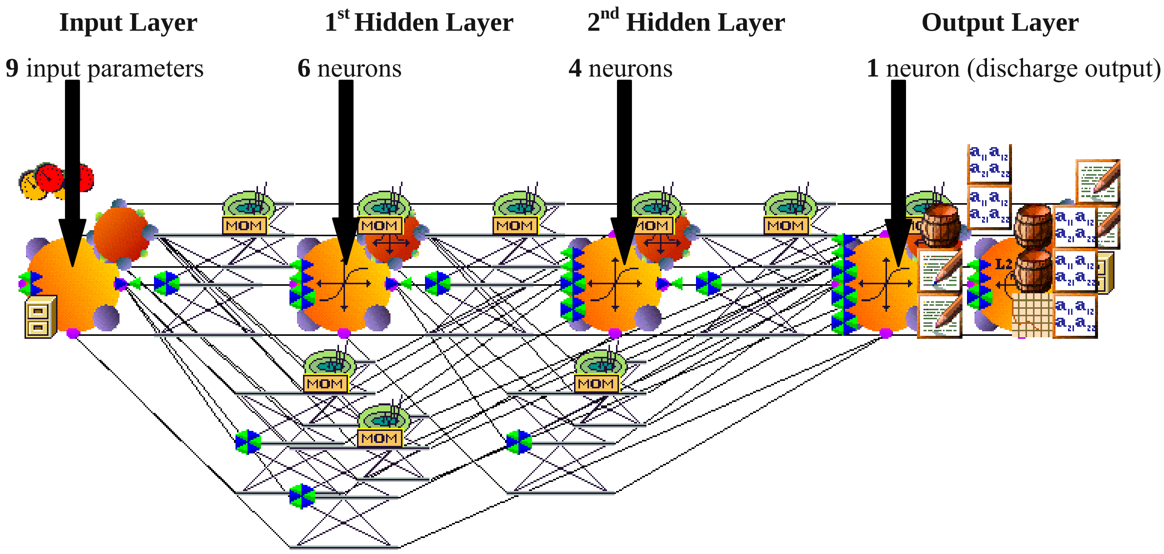


Figure A5-1: a MLP-ANN-FF (***9:7:12:1***) neural network model developed with the Excel based version of NeuroSolution Software

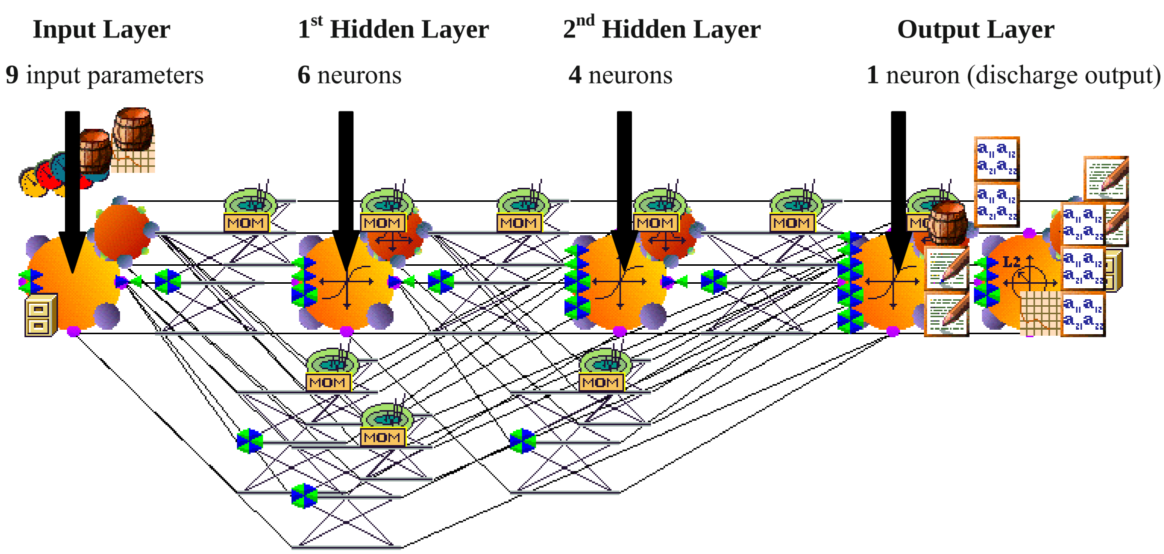


Figure A5-2: a GA-ANN-FF (***4:6:4:1***) neural network model developed with the Excel based version of NeuroSolution Software

APPENDIX 6– The sample results obtained for the MLP-ANN-FF (9:7:12:1) model

That estimated the discharge flow of Nzoia River at Rwambwa River gauge for MLP-ANN-FF (9:7:12:1) model

Table A6-1: Test results obtained for the MLP-ANN-FF (9:7:12:1) model to estimate the discharge flow of Nzoia River at Rwambwa River gauge

Date	Actual	Predicted	Error%
10/20/2003	1.9669392	2.154443	9.53
10/21/2003	1.9704399	2.180509	10.66
10/22/2003	1.9685763	2.171373	10.30
10/23/2003	1.9692295	2.210722	12.26
10/24/2003	1.9689963	2.211793	12.33
10/25/2003	1.9650135	2.153446	9.59
10/26/2003	1.9487063	2.124298	9.01
10/27/2003	1.9397688	2.074249	6.93
10/28/2003	1.9331835	2.074037	7.29
10/29/2003	1.9298275	2.121682	9.94
10/30/2003	1.9289077	1.969742	2.12
10/31/2003	1.9284471	2.023615	4.93
11/1/2003	1.9225179	2.098919	9.18
11/2/2003	1.9205929	2.127686	10.78
11/3/2003	1.9197055	2.117277	10.29
11/4/2003	1.9153998	2.185884	14.12
11/5/2003	1.9117434	2.175991	13.82
11/6/2003	1.9116902	2.191931	14.66
11/7/2003	1.9121157	2.190439	14.56
11/8/2003	1.9160326	2.174468	13.49
11/9/2003	1.9212702	2.213124	15.19
11/10/2003	1.9166118	2.241209	16.94
11/11/2003	1.9122221	2.167013	13.32
11/12/2003	1.905094	2.261883	18.73
11/13/2003	1.9395692	2.111586	8.87
11/14/2003	1.9492924	2.125341	9.03
11/15/2003	1.9520656	2.210732	13.25
11/16/2003	1.9557358	2.043797	4.50
11/17/2003	1.9705328	1.986326	0.80
11/18/2003	1.9693227	2.129429	8.13
11/19/2003	1.9562165	2.16657	10.75
11/20/2003	1.9505109	2.096929	7.51
11/21/2003	1.9360611	2.085433	7.72
11/22/2003	1.9197055	2.059307	7.27
11/23/2003	1.9066044	2.021408	6.02
11/24/2003	1.9061195	2.016191	5.77
11/25/2003	1.909235	1.995828	4.54

11/26/2003	***1.9008585***	2.004162	5.43
11/27/2003	***1.9008585***	1.974819	3.89
11/28/2003	***1.9099837***	1.975825	3.45
11/29/2003	***1.9209056***	1.964299	2.26
11/30/2003	***1.9157691***	1.94909	1.74
31/11/2003	***1.90558***	1.784349	-6.36
12/1/2003	***1.8916489***	1.928971	1.97
12/2/2003	***1.8776017***	1.922583	2.40
12/3/2003	***1.8685857***	1.925205	3.03
12/4/2003	***1.8624892***	1.913041	2.71
12/5/2003	***1.8666417***	1.928023	3.29
12/6/2003	***1.8429835***	2.075329	12.61
12/7/2003	***1.7629785***	1.856216	5.29
12/8/2003	***1.7595923***	1.874467	6.53
12/9/2003	***1.7594412***	1.936101	10.04
12/10/2003	***1.7714405***	1.857611	4.86
12/11/2003	***1.7761198***	1.869813	5.28
12/12/2003	***1.7816118***	1.850942	3.89
12/13/2003	***1.7808931***	1.818504	2.11
12/14/2003	***1.77459***	1.795361	1.17
12/15/2003	***1.7667102***	1.807657	2.32
12/16/2003	***1.7652214***	1.80253	2.11
12/17/2003	***1.7667102***	1.798654	1.81
12/18/2003	***1.7585334***	1.796841	2.18
12/19/2003	***1.7165042***	1.782422	3.84
12/20/2003	***1.7137425***	1.798215	4.93
12/21/2003	***1.7199111***	1.820559	5.85
12/22/2003	***1.7431176***	1.822957	4.58
12/23/2003	***1.7947668***	1.802008	0.40
12/24/2003	***1.811709***	1.816582	0.27
12/25/2003	***1.808346***	1.818646	0.57
12/26/2003	***1.8102997***	1.780542	-1.64
12/27/2003	***1.799754***	1.792119	-0.42
12/28/2003	***1.7759743***	1.778662	0.15
12/29/2003	***1.7351995***	1.791945	3.27
12/30/2003	***1.7337588***	1.787277	3.09
12/31/2003	***1.7569402***	1.82077	3.63

Performance	**Discharge**
Mean Square Error	0.024031731
r (correlation coefficient)	0.843375453

Table A6-2: Test results obtained for the GA-ANN-FF (4:6:4:1) model to estimate the discharge flow of Nzoia River at Rwambwa River gauge

Date	Actual	Predicted	Error%
10/20/2003	1.96693916	2.12396347	7.98318
10/21/2003	1.97043986	2.14603548	8.911493
10/22/2003	1.96857634	2.19988399	11.75
10/23/2003	1.96922948	2.23280205	13.38455
10/24/2003	1.96899633	2.20481946	11.97682
10/25/2003	1.96501345	2.15677383	9.758731
10/26/2003	1.94870631	2.12611697	9.104022
10/27/2003	1.93976878	2.09453404	7.978542
10/28/2003	1.93318348	2.09330268	8.28267
10/29/2003	1.92982748	2.08595024	8.089985
10/30/2003	1.92890769	2.09702138	8.715487
10/31/2003	1.92844706	2.06542355	7.102943
11/1/2003	1.92251786	2.06950733	7.645675
11/2/2003	1.92059286	2.13294407	11.05654
11/3/2003	1.91970554	2.14039301	11.4959
11/4/2003	1.91539984	2.1668135	13.12591
11/5/2003	1.91174338	2.1615363	13.06624
11/6/2003	1.91169016	2.18957405	14.53603
11/7/2003	1.91211573	2.18382966	14.21012
11/8/2003	1.91603261	2.18128075	13.84361
11/9/2003	1.92127019	2.19173299	14.07729
11/10/2003	1.91661185	2.22260277	15.9652
11/11/2003	1.91222206	2.16008757	12.96217
11/12/2003	1.90509397	2.12272974	11.42389
11/13/2003	1.93956917	2.12825226	9.728093
11/14/2003	1.9492924	2.10699142	8.090065
11/15/2003	1.95206559	2.17283181	11.30937
11/16/2003	1.95573584	2.08357605	6.536681
11/17/2003	1.97053283	2.06340572	4.713085
11/18/2003	1.96932271	2.05359507	4.279256
11/19/2003	1.95621647	2.16567867	10.70752
11/20/2003	1.95051089	2.08260362	6.772212
11/21/2003	1.93606112	2.07943014	7.405191
11/22/2003	1.91970554	2.05210207	6.89671
11/23/2003	1.90660437	2.03610061	6.791983
11/24/2003	1.90611946	2.01922612	5.93387
11/25/2003	1.909235	2.0110842	5.334555
11/26/2003	1.90085851	2.00106999	5.271907

11/27/2003	1.90085851	1.98749079	4.557535
11/28/2003	1.9099837	1.9699898	3.141708
11/29/2003	1.9209056	1.96696619	2.397858
11/30/2003	1.91576907	1.95046866	1.811262
31/11/2003	1.90558003	1.94971802	2.31625
12/1/2003	1.89164894	1.93053431	2.055633
12/2/2003	1.87760168	1.92350189	2.444619
12/3/2003	1.86858567	1.91814249	2.652103
12/4/2003	1.86248917	1.92488009	3.349867
12/5/2003	1.86664172	1.91683358	2.688886
12/6/2003	1.84298347	1.83267561	-0.5593
12/7/2003	1.76297849	1.9073084	8.186709
12/8/2003	1.75959231	1.86911274	6.224194
12/9/2003	1.7594412	1.8974752	7.845332
12/10/2003	1.77144049	1.86446887	5.251567
12/11/2003	1.7761198	1.85850384	4.638428
12/12/2003	1.78161178	1.85131465	3.912349
12/13/2003	1.78089311	1.83888037	3.256078
12/14/2003	1.77458995	1.83473686	3.389342
12/15/2003	1.76671021	1.82634042	3.375212
12/16/2003	1.76522137	1.82357477	3.305727
12/17/2003	1.76671021	1.81757141	2.878865
12/18/2003	1.75853342	1.80690988	2.750955
12/19/2003	1.71650416	1.79334458	4.476564
12/20/2003	1.71374248	1.7949984	4.741431
12/21/2003	1.71991106	1.81697616	5.643611
12/22/2003	1.74311763	1.82809991	4.875304
12/23/2003	1.7947668	1.81086081	0.896719
12/24/2003	1.81170903	1.80690988	-0.2649
12/25/2003	1.80834604	1.82040144	0.666654
12/26/2003	1.81029974	1.79695044	-0.73741
12/27/2003	1.79975397	1.78774832	-0.66707
12/28/2003	1.77597433	1.7824216	0.363027
12/29/2003	1.73519955	1.78047613	2.609301
12/30/2003	1.73375884	1.7877008	3.111273
12/31/2003	1.75694024	1.81612343	3.368538

Performance	**Discharge**
MSE	0.02176209
r	0.887479548

APPENDIX 7 –Test data used to evaluate MLP-ANN-FF and GA-ANN- models

This 10% test data was applied on the ***9:7:12:1*** configuration setting; the 10% test dataset is made up of 74 data points

Table A7-1: The 10% test dataset applied on **MLP-ANN-FF** model

9 Inputs									1 Output
Kitale-Met Rainfall 2002- 2003	Kitale-Soil Rainfall 2002- 2003	Leissa-Farm Rainfall 2002-2003	Kitale-Met Temp 2002-2003	Kitale-Met Rainfall 2000- 2002	Kitale-Soil Rainfall 2000- 2001	Leissa-Farm Rainfall 2000-2001	Kitale-Met Temp 2000-2001	Rwambwa Discharge Flow 2000 - 2001	Rwambwa Discharge Flow 2002 - 2003
0	0	0	19.55	0	0	0	19.85	2.332034277	***1.966939163***
0	0	0	19.3	0	0	0	20.1	2.380211242	***1.970439863***
0	0	0	19.05	0	20	0	19.05	2.50609896	***1.968576335***
0.5	0	0	19.35	26.4	13	5.5	19.25	2.592842683	***1.96922948***
0	0	0	16.45	2.5	0	3.7	19.6	2.516931809	***1.968996327***
0.7	0	0	19.25	20.8	8	14	19.8	2.401745082	***1.96501345***
0	0	0	18.65	8.5	9	0	20.1	2.336659823	***1.948706309***
5.6	0	0	19.8	7.6	10	10.2	19.45	2.28057837	***1.939768776***
0	0	0	19.95	6.7	5	21.8	18.75	2.253822439	***1.933183479***
0	0	0	18.6	21.9	0	13.7	19.55	2.239049093	***1.929827481***
1.8	4.5	4.5	19.15	4.7	24	12.3	18.9	2.226599905	***1.92890769***
0	0	0	19	17.9	16	1.7	18.75	2.208978517	***1.928447063***
0.6	0	0	20.65	30.1	0	7.5	17.7	2.211654401	***1.92251786***
3.7	10	10	19.1	0	0	4.6	17.85	2.265525335	***1.920592862***
0	0	0	20.05	2.5	10	0	18.25	2.367728546	***1.919705535***
0	0	0	20.1	9.5	0	0.9	18.3	2.426511261	***1.915399835***
0	0	0	17.95	0	0	0	18.3	2.415140352	***1.911743378***
0.5	0	0	19.25	0.7	6	0	18.2	2.48301642	***1.911690159***
0	0	0	23.15	7.4	5.1	2.1	18.4	2.465828815	***1.912115729***
0	0	0	18.85	6.2	2.3	10.1	19.3	2.457730548	***1.91603261***
1	0	0	18.95	3.7	0	3.8	19.45	2.488691698	***1.921270185***
1.1	7.5	7.5	19.6	0	0	2.7	19.15	2.494015375	***1.916611845***
7.8	0	0	19.05	2.3	10	2.2	18.55	2.443262987	***1.912222056***
0	0	0	19.45	6.4	13.2	58	18.75	2.387033701	***1.905093968***
3.4	0	0	19.35	9.2	8.6	4.5	18.8	2.349471799	***1.939569169***
5.8	0	0	19.7	10.8	3.6	6	19	2.310905629	***1.949292401***
7.7	21	21	18.55	7.9	0	1.4	18.55	2.276461804	***1.95206559***
0	0	0	20.6	30.6	20	1	18.6	2.246498581	***1.955735842***
4.8	0	0	20.1	0.6	27.7	2.9	19	2.218010043	***1.97053283***

1.6	0	0	18.8	38.2	0	2.5	19.85	2.188928484	***1.969322706***
0	26	26	20.35	0	0	1.5	18.55	2.181271772	***1.956216469***
0	4.1	4.1	17.95	0	0	0	20.05	2.206825876	***1.950510893***
0	0	0	19.9	1.1	0	0	18	2.239549721	***1.936061117***
3.6	0	0	18.6	1.2	0	3.5	18.45	2.190611798	***1.919705535***
0	0	0	19.25	0	0	6.5	19	2.144262774	***1.906604372***
0	0	0	19.1	0	0	0	18.15	2.122215878	***1.906119458***
0	0	0	20.15	0	0	0	17.05	2.106870544	***1.909235003***
0	0	0	19	0	0	0	18.8	2.088136089	***1.900858505***
0	0	0	20.5	0	0	0	17.45	2.062957834	***1.900858505***
4	0	0	19.95	0	0	0	17.5	2.040206628	***1.909983695***
0.7	0	0	18.3	0.8	0	2.9	19.25	2.021189299	***1.920905604***
0.4	0	0	20.05	0.3	0	1.8	19.05	1.992553518	***1.915769066***
3.45900068	3.21218344	2.8476386	0	3.45900068	3.21218344	2.8476386	0	1.969415912	***1.905580028***
0	0	0	20.25	0.8	0	0	18.65	1.959566047	***1.891648944***
2.8	0	0	18.6	0	0	0	18.05	1.952986065	***1.87760168***
0	0	0	18.55	0	0	0	19.6	1.937417582	***1.868585666***
0.4	3.6	3.6	19.3	0	0	0	18.4	1.916401304	***1.862489167***
1.9	4.2	4.2	19.3	0	0	0	19.55	1.899711095	***1.866641721***
36.9	0	0	18.55	0	0	0	19.25	1.884115362	***1.84298347***
0	5.4	5.4	20.45	0	7	0	18.5	1.867408557	***1.762978491***
4.1	0	0	20	4	0	0	17.65	1.858837851	***1.759592309***
11	7.9	7.9	18.75	0	0	0	17.9	1.852967691	***1.759441197***
0	0	0	18.8	0.4	0	0	18.8	1.842172229	***1.771440487***
0	0	0	22.8	0	0	0	19.55	1.831613855	***1.776119799***
0	0	0	19.4	0	0	0	19.25	1.818885415	***1.781611782***
0	0	0	18.6	0	0	0	18.05	1.796851749	***1.780893109***
0	0	0	16.55	2.4	0	3.3	17.35	1.781468143	***1.77458995***
0	0	0	17.15	0	0	0	18.35	1.77458995	***1.766710207***
0	0	0	18.65	0	0	0	18.05	1.769672664	***1.765221366***
0	0	0	19.2	0	0	0	18.15	1.758987547	***1.766710207***
0	0	0	19.8	0	0	0	18.7	1.739967697	***1.758533422***
0	0	0	19.1	0.5	0	0	18.7	1.715669142	***1.716504164***
1.3	0	0	20.45	0	0	0	19.1	1.721315881	***1.713742478***
0	0	0	20.3	0	0	0	19.5	1.757927183	***1.719911064***
0	0	0	19.5	0	0	0	19	1.777716739	***1.743117625***
0	0	0	19.55	0	0	0	18.8	1.747023177	***1.794766798***
0	0	0	20.8	0	0	0	19.85	1.739967697	***1.811709027***
0	0	0	16.9	0	0	0	19.5	1.764026608	***1.808346036***
0	0	0	19	0	0	0	18.4	1.722140125	***1.810299741***
0	0	0	19.2	0	0	0	19.75	1.705607163	***1.799753966***

0	0	0	19.15	0	0	0	19.25	1.696006715	***1.775974331***
0	0	0	19.3	0	0	0	20.2	1.692494408	***1.735199548***
0	0	0	18.2	0	0	0	19.55	1.705521614	***1.733758836***
0	0	0	20.05	0	0	0	19.6	1.756407872	***1.756940236***

This 10% test data was applied on the ***4:6:4:1*** configuration setting; the 10% test dataset is made up of 74 data points

Table A7-2: The 10% test dataset applied on **GA-ANN-FF** model

4 Inputs				**1 Output**
Kitale-Met Rainfall 2002- 2003	Leissa-Farm Rainfall 2002-2003	Leissa-Farm Rainfall 2000-2001	Rwambwa Discharge Flow 2000 - 2001	Rwambwa Discharge Flow 2002 - 2003
0	0	0	2.332034277	***1.966939163***
0	0	0	2.380211242	***1.970439863***
0	0	0	2.50609896	***1.968576335***
0.5	0	5.5	2.592842683	***1.96922948***
0	0	3.7	2.516931809	***1.968996327***
0.7	0	14	2.401745082	***1.96501345***
0	0	0	2.336659823	***1.948706309***
5.6	0	10.2	2.28057837	***1.939768776***
0	0	21.8	2.253822439	***1.933183479***
0	0	13.7	2.239049093	***1.929827481***
1.8	4.5	12.3	2.226599905	***1.92890769***
0	0	1.7	2.208978517	***1.928447063***
0.6	0	7.5	2.211654401	***1.92251786***
3.7	10	4.6	2.265525335	***1.920592862***
0	0	0	2.367728546	***1.919705535***
0	0	0.9	2.426511261	***1.915399835***
0	0	0	2.415140352	***1.911743378***
0.5	0	0	2.48301642	***1.911690159***
0	0	2.1	2.465828815	***1.912115729***
0	0	10.1	2.457730548	***1.91603261***
1	0	3.8	2.488691698	***1.921270185***
1.1	7.5	2.7	2.494015375	***1.916611845***
7.8	0	2.2	2.443262987	***1.912222056***
0	0	58	2.387033701	***1.905093968***
3.4	0	4.5	2.349471799	***1.939569169***
5.8	0	6	2.310905629	***1.949292401***

7.7	21	1.4	2.276461804	***1.95206559***
0	0	1	2.246498581	***1.955735842***
4.8	0	2.9	2.218010043	***1.97053283***
1.6	0	2.5	2.188928484	***1.969322706***
0	26	1.5	2.181271772	***1.956216469***
0	4.1	0	2.206825876	***1.950510893***
0	0	0	2.239549721	***1.936061117***
3.6	0	3.5	2.190611798	***1.919705535***
0	0	6.5	2.144262774	***1.906604372***
0	0	0	2.122215878	***1.906119458***
0	0	0	2.106870544	***1.909235003***
0	0	0	2.088136089	***1.900858505***
0	0	0	2.062957834	***1.900858505***
4	0	0	2.040206628	***1.909983695***
0.7	0	2.9	2.021189299	***1.920905604***
0.4	0	1.8	1.992553518	***1.915769066***
3.45900068	2.8476386	2.8476386	1.969415912	***1.905580028***
0	0	0	1.959566047	***1.891648944***
2.8	0	0	1.952986065	***1.87760168***
0	0	0	1.937417582	***1.868585666***
0.4	3.6	0	1.916401304	***1.862489167***
1.9	4.2	0	1.899711095	***1.866641721***
36.9	0	0	1.884115362	***1.84298347***
0	5.4	0	1.867408557	***1.762978491***
4.1	0	0	1.858837851	***1.759592309***
11	7.9	0	1.852967691	***1.759441197***
0	0	0	1.842172229	***1.771440487***
0	0	0	1.831613855	***1.776119799***
0	0	0	1.818885415	***1.781611782***
0	0	0	1.796851749	***1.780893109***
0	0	3.3	1.781468143	***1.77458995***
0	0	0	1.77458995	***1.766710207***
0	0	0	1.769672664	***1.765221366***
0	0	0	1.758987547	***1.766710207***
0	0	0	1.739967697	***1.758533422***
0	0	0	1.715669142	***1.716504164***
1.3	0	0	1.721315881	***1.713742478***
0	0	0	1.757927183	***1.719911064***
0	0	0	1.777716739	***1.743117625***
0	0	0	1.747023177	***1.794766798***
0	0	0	1.739967697	***1.811709027***

0	0	0	1.764026608	***1.808346036***
0	0	0	1.722140125	***1.810299741***
0	0	0	1.705607163	***1.799753966***
0	0	0	1.696006715	***1.775974331***
0	0	0	1.692494408	***1.735199548***
0	0	0	1.705521614	***1.733758836***
0	0	0	1.756407872	***1.756940236***