# Modelling and Forecasting Foreign Direct Investment: A Comparative Application of Machine Learning Based Evolutionary Algorithms Hybrid Models

**by**

**Mr. Mogari Ishmael Rapoo**

**orcid.org/0000-0003-0771-7461**

Thesis submitted for the degree *Doctor* in Business Statistics in Economic and Management Sciences on Mafikeng Campus at the North-West University

Promoter: Prof Elias Munapo

Co-promoter: Prof Martin M. Chanza

Examination: November 2022

Student number: 23809213

# DECLARATION

I declare that this thesis is the result of my own investigation unless stated otherwise. I also declare that it has never been submitted previously as a whole or in part for any other degree whatsoever to the North-West University or any institution.

| | | |
|---|---|---|
| Mogari Ishmael Rapoo | ….Rapoo MI… | ……06/12/2022… |
| (23809213) | Signature | Date |

# ACKNOWLEDGEMENTS

I would want to thank the Almighty God who enabled me and taken me under his mighty wings to understand, undertake and complete this work through everything else as part of the academic requirements for my PhD programme. His love and mercy have showered me throughout this difficult yet fulfilling journey. My sincere gratitude goes to both my promoters; Prof E Munapo and Dr MM Chanza, for providing such mind stimulating guidance and for supporting me throughout this thesis. Without their commitment in making sure that this thesis is of good quality, I would never have been able to finish this thesis. Moreover, to Mr Monchwe, thank you for dedicating your time with your coding skills. Secondly, I would like to extend my gratitude to my family, especially my father, my mother, my girlfriend and son, brothers and uncles, for their support throughout this thesis.

# DEDICATION

To my entire family, those who departed this world and those who are still alive, the love and support you show me is, indeed, second to none.

# PREFACE

The purpose of this thesis is to determine whether a genetic algorithm (GA) can improve the forecasting performance of nonlinear supervised machine learning models in modelling and forecasting foreign direct investment. Hybrid models of genetic algorithm-based artificial neural network and genetic algorithm-based support vector regression employed to improve the forecasting performance and accuracy of the models.

The past three years have been a challenging time with both ups and downs, a time of great benefit for me even through the hardest moments. Fortunately, I was never alone in this whole time, and I was accompanied by expertise in my promoters who were always coaching me and helping me and for that, I am truly thankful. To my entire family, the support and love you all showed me were deeply noticed.

To my promoters, your knowledge, guidance and persistent pushing have enabled me to finish this thesis and I am truly thankful. Many thanks to everyone, Colette my girlfriend, Tshiamo my son, Mr Maleke my father, Ms Rapoo my mother, Mr Rapoo my brother and Mr Legae my uncle for all the encouragement and support.

This thesis will contribute to the academic discipline and to those in the field of research in machine learning, evolutionary algorithms and for time series forecasting. It should also be of interest to scholars of machine learning and to those who want to contribute to the field of machine learning.

ABSTRACT

The study aims to determine whether genetic algorithms can improve the forecasting accuracy of machine learning models in modelling and forecasting foreign direct investment (FDI). The use of evolutionary algorithms for FDI modelling and forecasting is relatively unexplored. The research therefore employs benchmark models (ANN, SVR, and LSSVR) to assess the forecasting performance of hybrid models (ANN-GA and SVR-GA) in modelling and forecasting FDI time series data. Monthly time series data of FDI (dependent variable) and GDP, inflation rate, and exchange rate (explanatory variables) were collected from World Bank and South African Reserve Bank spanning January 1970 to June 2019, with 594 observations.

The nonlinearity tests conducted in the study confirmed that the variables under consideration exhibit nonlinear characteristics. The nonlinearity tests considered were Brock, Dechert and Scheinkman (BDS) test, Cumulative Sum (CUSUM) test, Keenan's test, White and Terasvirta neural network test, Tsay test and McLeod-Li test. In traditional machine learning models, artificial neural networks (ANN) outperformed support vector regression (SVR) and least squares support vector regression (LSSVR), showing lower error measures. It was further found that the support vector regression (SVR) was the second-best performing model and the worst being the least squares support vector regression (LSSVR). Among hybrid models, support vector regression based genetic algorithm (SVR-GA) performed better than artificial neural network based genetic algorithm (ANN-GA). However, overall, ANN proved to be the best performing model in terms of error measurements (MSE, RMSE, MAE, MAPE, and MASE) with the lowest error values reported. Genetic algorithm (GA) did not significantly improve the forecasting performance of the hybrid models, which still fell short compared to traditional models.

The decision of the error measurements was supported by Diebold Mariano test, concluding that artificial neural network (ANN) performed best in forecasting accuracy, followed by support vector regression (SVR) and least squares support vector regression (LSSVR). Both ANN and SVR exhibited similar forecasting accuracy for testing predictive power. Among hybrid models, ANN-GA showed higher predictive accuracy than SVR-GA. Overall, machine learning models proved effective in modelling and forecasting foreign direct investment. The study recommends exploring other evolutionary algorithm(s) methods to optimise hyperparameters and suggests employing a wider range of machine learning models for FDI forecasting.

**Key Words: Error measurements, evolutionary algorithm(s), foreign direct investment, hybrid models, modelling and forecasting, machine learning.**

# LIST OF ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| ADE-BPNN | Adaptive Differential Evolution based Backpropagation Neural Network |
| AIC | Akaike Information Criterion |
| ANFIS | Adaptive Neuro-Fuzzy Inference Systems |
| ANFIS-GA | Adaptive Neuro-Fuzzy Inference Systems based Genetic Algorithm |
| ANN | Artificial Neural Network |
| ANN-GA | Artificial Neural Network based Genetic Algorithm |
| AR | Autoregressive |
| ARIMA | Autoregressive Integrated Moving Average |
| BIC | Bayesian Information Criterion |
| BDS test | Brock- Dechert-Scheinkman test |
| BPNN | Backpropagation Neural Network |
| CFE | Comision Federal de Electricidad |
| CGA | Chaotic Genetic Algorithm |
| CNY | Chinese Yuan |
| CUSUM | Cumulative Sum of Squared |
| DLSSVR | Dynamic Least Squares Support Vector Regression |
| DBNN | Difference Boosting Neural Network |
| DS | Directional Symmetry |
| DM | Diebold-Mariano |
| EA | Evolutionary Algorithms |
| EMD | Empirical Mode Decomposition |
| EEMD | Ensemble Empirical Mode Decomposition |

| | |
|---|---|
| EUR | Euro |
| EEMD-RBFN | Ensemble Empirical Mode Decomposition based Radial Basis Function Network |
| EMD-SVR | Empirical Mode Decomposition based Support Vector Regression |
| FA | Firefly Algorithm |
| FDI | Foreign Direct Investment |
| FIS | Fuzzy Inference Systems |
| FTSE | Financial Time Stock Exchange |
| GA | Genetic Algorithm |
| GARCH | Generalised Autoregressive Conditional Heteroskedasticity |
| GDP | Gross Domestic Product |
| GP | Genetic Programming |
| GRNN | Generalised Regression Neural Network |
| IMF | International Monetary Fund |
| JPY | Japanese Yen |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| MAE | Mean Absolute Error |
| MAD | Mean Absolute Deviation |
| MAPE | Mean Absolute Percentage Error |
| MASE | Mean Absolute Scaled Error |
| MNC | Multinational Corporations |
| MLP | Multilayer Perceptron |
| NNARX | Neural Network Autoregressive with Exogenous Inputs |

| | |
|---|---|
| NEAT | Neuro-Evolution of Augmenting Topologies |
| NIFTY | National Stock Exchange Fifty |
| PSO | Particle Swarm Optimisation |
| PSO-ANN | Particle Swarm Optimisation based Artificial Neural Network |
| PSO-SVM | Particle Swarm Optimisation based Support Vector Regression |
| SA | Simulated Annealing |
| SVR | Support Vector Regression |
| S&P | Standard and Poor's |
| TAIEX | Taiwan Stock Exchange Capitalisation Weighted Stock Index |
| TCS | Tata Consultancy Services |
| TLNN | Time Lag Neural Network |
| USD | United States Dollar |
| WD | Wavelet Decomposition |
| ZAR | South African Rand |

# TABLE OF CONTENTS

**CHAPTER 3**

CHAPTER 5

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

## 1.1 Introduction and background

According to Khashei and Bijari (2011), considerable amount of attention has been given to time series forecasting in application to the variety of areas. Chujai et al. (2013) define time series as a group of data collected over time according to a discrete period of time or an order of historical data similar to an observation. Various models have been employed in the past for modelling and forecasting time series data. These conventional models include regression analysis, classical decomposition methods, Box and Jenkins methods and smoothing techniques. However, Wheelwright et al. (1998) state that several factors are questioned in consideration of an appropriate model to be used.

The most extensively applied model for time series modelling and forecasting has been the Box-Jenkins ARIMA model. This model has been applied successfully in various disciplines and particularly economic time series forecasting, such as for exchange rate forecasting (Nyoni, 2018, Nwankwo, 2014, Newaz, 2008), and for dependence between successive times and failure (Walls and Bendell, 1987). A linear structure is usually assumed among the time series values in the existing studies. However, Wang et al., (2013) states that in the real-world phenomenon, linear relationship is not factual as there exists a nonlinear relationship. Furthermore, Zhang et al. (1998) also states that real world systems are nonlinear in nature. Therefore, this presents a limitation to the ARIMA model since it is a linear model.

According to Chaudhry et al. (2000), knowledge-based decision support is provided through the development of applied artificial intelligence system amongst others artificial neural networks (ANNs). These machine learning models are data driven and can approximate any function because of their exceptional properties, e.g., data drive, and can learn more information on the data. The other advantage of these models is that they are nonparametric, meaning no or fewer prior assumptions are needed in the model building process. These models are also nonlinear in nature. The models include, but not limited to, artificial neural networks (ANNs), support vector regression (SVR), adaptive neuro-fuzzy inference systems (ANFIS) and neural networks autoregressive with exogenous inputs (NNARX).

In recent years, these models have been efficiently carried out in various research areas, such as for rainfall run-off simulation (Talei et al., 2010, Nourani et al., 2009); water quality modelling

(Yan et al., 2010, Singh et al., 2009); electricity forecasting (Kaytez et al., 2015) and financial time series forecasting (Tay and Cao, 2001). These models have obtained exceptional results in many research areas. However, a particular discovery was made that even though that might be the case, perhaps time series data is made up of both the linear and nonlinear characteristics.

Therefore, hybrid models were introduced, in which case, these models can be either homogenous or heterogeneous in their nature. Various studies amongst others have applied hybrid models for time series forecasting such as Sarıca et al., (2018), Shi et al., (2012), Zeng et al., (2008), and Zhang, (2003) and it was found that these models improve forecasting accuracy.

Artificial intelligence has a branch which has been developed based on the theory of evolution. This branch is computational methods called evolutionary algorithms. The Darwin's evolution and genetics was used in the development of computational methods called evolutionary algorithms, amongst others, there is genetic algorithm. These methods are used to search for optimum solutions and are applied in a hybrid setting together with the various machine learning models.

The prime objective of this approach is that the algorithm can help in the optimal selection of the hyper-parameters of the machine learning models. Furthermore, it has been found that machine learning models depend heavily on the right choices of the hyper-parameters to improve forecasting accuracy. Various scholars have employed these hybrid models for time series forecasting, for example, in water quality prediction (Liu et al., 2013); system reliability (Chen, 2007); tourism demand forecasting (Chen and Wang, 2007); long span suspension bridges (Cheng, 2010) and electric load forecasting (Hong, 2009). Thereby, it was found that these algorithms can improve the forecasting accuracy of machine learning models.

Foreign direct investment (FDI) is a type of investment in which foreign firms/companies make an investment into a host country. This type of investment is important as it plays a role in creating employment, developing managerial skills, and it promotes improvement in productivity and transferring technology to local firms, in which case, this will boost the local economy (Majavu and Kapingura, 2016).

Furthermore, in developing economies, domestic savings are supplemented, and employment is generated. Thus, there will be economic growth because of receiving FDI. In the recent past, researchers have studied FDI particularly in terms of determinants that drive foreign direct investment to a receiving economy such as the studies of (Majavu and Kapingura, 2016b, Wafure and Nurudeen, 2010) amongst others. The results showed that amongst other determinants, FDI is driven by the following factors; namely, GDP, economic openness, inflation rate, exchange rate, corporate tax and a measure of financial crises.

This research seeks to investigate the application of machine learning based evolutionary algorithms hybrid models in terms of their forecasting efficiency and performance of modelling and forecasting foreign direct investment. Furthermore, to determine if the evolutionary algorithms can improve the forecasting accuracy of the included machine learning models in modelling and forecasting foreign direct investment, the study uses traditional machine learning models as benchmark models.

## 1.2 The statement of the problem

The existing literature has successfully documented the application of both the traditional statistical models and machine learning models in terms of their applications in forecasting time series data. Various models, especially conventional statistical models have been used from the past to the present for modelling foreign direct investment (FDI) such as the model of Vector Error Correction model employed by (Masipa, 2018, Majavu and Kapingura, 2016), Cointegration and Error Correction Model (Mabule, 2012) for identifying determinants of FDI. Moreover, machine learning models have also been efficiently used to forecast time series data. However, there is a problem that accompanies these models.

There should be a proper setting of hyper-parameters for these models because their forecasting performance is dependent on the proper settings or selections of the hyper-parameters. Therefore, there is a need to employ computational methods known as evolutionary algorithms to select the optimal parameters of the models for improving forecasting performance and accuracy of the models.

Thus, this study seeks to assess whether genetic algorithm (GA) will improve the forecasting accuracy and performance of machine learning models in a hybrid system in modelling and forecasting foreign direct investment. Furthermore, traditional machine learning models, for example, Artificial Neural Networks and Support Vector Regression, will be employed as benchmark models to assess forecasting performance of these proposed hybrid models.

## 1.3 Purpose of the study

The study seeks to develop an accurate model(s)/model(s) that can be employed in the analysis of foreign direct investment. This will help in various ways, not only in the academia but also to those who are investors of foreign direct investment. The model / models can also be used to extrapolate foreign direct investment and based on the extrapolated values from the accurate model, right decisions can and will be made which will be used to the benefit of the masses.

## 1.4 Significance of the study

Machine learning models are used in different disciplines and this study will also seek to recommend the model(s) to the academia discipline, shareholders, regulators and financial institutions which can also be in a position to use the model(s) recommended by the current study. This study also uses an evolutionary algorithm to select optimal values of machine learning models.

## 1.5 Research aim(s) and objectives

Time series forecasting needs the application of an accurate model / models. These accurate models will ensure that the results are not biased and can be trusted. Therefore, the study seeks to assess whether genetic algorithm (GA) can improve forecasting efficiency and performance of the included machine learning models in a hybrid system in modelling and forecasting foreign direct investment. The structured study objectives are as follows:

- To statistically test the linearity/nonlinearity nature of foreign direct investment time series.
- To statistically test the forecasting efficiency and performance of traditional machine learning models on foreign direct investment time series.
- To statistically test whether genetic algorithm improves forecasting accuracy of machine learning models on foreign direct investment.
- To compare the efficiency and performance of traditional machine learning models and hybrid machine learning models on foreign direct investment.
- To compare the predictive power of both the traditional machine learning models and the hybrid machine learning models on foreign direct investment.

## 1.6 Literature review

In the existing published research on time series modelling and forecasting, a lot of research has been done. Over the years, the research has been undertaken and over emphasised on using both the conventional linear and nonlinear models which have obtained a great deal of results.

The incredible work of the scholars enabled researchers to employ machine learning (ML) techniques in time series modelling and forecasting which has led to the improvement of forecasting accuracy. Moreover, there is a good literature on the application of evolutionary algorithms applied in hybrid settings with machine learning models. Therefore, the study will also highlight the success that is documented in the literature of these hybrid models. These methods are used in the selection of hyper-parameters for these machine learning models.

There are several evolutionary algorithms such as genetic algorithms, firefly algorithm, particle swarm optimization etc. Therefore, this section focuses on presenting the literature of these ML models with evolutionary algorithms and empirical decomposition method on their application to time series data. The models are good approximates and can handle complex and nonlinear time series. It is somewhat good to use hybrid models as they can perform better and be more effective. This is the standout point that the current study would like to follow.

### 1.6.1 Artificial neural networks based evolutionary algorithm

McCulloch and Pitts (1943) which developed the artificial neural network model and other variants of the ANNs model was developed by other scholars thereafter, Vapnik (1995) proposed support vector regression and other scholars came with other ML models. Therefore, Cadenas and Rivera (2009) applied artificial neural networks technique for forecasting short term wind speed in the region of La Venta, Qaxaca, Mexico. Comision Federal de Electricidad (CFE) collected the data through a network of measurement stations located in the place of interest for 7 years. A number of models were estimated and a simple model of two layers with two inputs neurons and one output neuron was selected as the best model for the short-term wind speed forecasting. Mean squared error and mean absolute percentage error values of 0.0016 and 0.0399 were found for the model respectively. The study recommended the model to be used by the Electric Utility Control Center in Qaxaca for the energy supply.

The study used a computing technique known as ANN with backpropagation for short-term forecasting implementation of system marginal price. The data used was that of SMP real data from the deregulated Victorian power system for training and testing purposes. The results show that ANNs can be a good model for forecasting SMP (Szkuta et al., 1999).

In the city of Yasouj in Iran, a hybrid model was developed to predict natural gas consumption. The model developed is an artificial neural network-based genetic algorithm, where GA is used to optimise the hyper-parameters and architecture of the artificial neural network. The results demonstrate that the developed model has a good accuracy with real data based on absolute deviation and correlation coefficient. The study recommends that the model can be used in predicting natural gas consumption and that it can even assist authorities in making policies in line with natural gas consumption distribution (Karimi and Dastranj, 2014).

Panigrahi et al. (2013) assessed the effectiveness in terms of forecasting ability of time series of the three models, the models are ANN-GD, ANN-GA and ANN-DE. Furthermore, the study benchmarks these models using two-time series datasets of Wisconsin employment and monthly milk production respectively. According to the results, both ANN-GA and ANN-DE outperformed ANN-GD. Furthermore, ANN-DE outperformed ANN-GA in terms of forecasting both time series data.

In the study, researchers developed a new type of neural network model called the single neuron model. To train this model, they used a technique called particle swarm optimization (PSO). Additionally, they improved the standard PSO method by introducing a more advanced version called cooperative random learning particle swarm optimization (CRPSO). Three well known time series were used to train the model of CRPSO, PSO, back-propagation and genetic algorithm. It was shown that CRPSO-based neuron model is superior to the other algorithms used (Zhao and Yang, 2009).

In the study, a hybrid model of EMD-SVR was proposed for stock indices forecasting. The dataset used is of US and UK stock indices. The series are decomposed into IMFs by the EMD method, and after that, BPNN is used to forecast each IMF. The resulting forecasts are then aggregated to form the final forecasting results. The result is that the proposed model has a superior forecasting ability (Zhang and Pan, 2015).

In the study, Yang and Lin (2011), forecasted exchange rate by employing a hybrid model of EMD-SVR. The study sought to explore the feasibility of applying the hybrid model for exchange rate forecasting. EMD is employed to decompose the series into small intrinsic mode functions (IMFs) and then ANN model is employed to forecast the IMF and aggregate the resulting forecast. Based on the error measurement of MAPE, the study found that the hybrid model performs better than the random walk model.

### 1.6.2 Support vector regression based evolutionary algorithm

Nguyen et al. (2015) employed machine learning models of LASSO, Random Forest and Support Vector Regression to assess the application to forecasting daily water levels of Thakhek station on the Mekong River. Results demonstrate that support vector regression is the accurate model to be applied, with the reported mean absolute error of 0.486 (m) and an acceptance error between 0.5 (m) and 0.75 (m).

The study examined weather support vector regression can be applied to system reliability and compared the SVR model with neural networks approaches and autoregressive integrated moving average model. The genetic algorithm technique was used to search for optimal parameters of the SVR model and those optimal parameters are adopted to construct SVR models. The data set employed was that of 40 suits of turbocharges. The results demonstrate that GA-SVR hybrid model outperforms both neural network approaches and ARIMA model based on normalised root mean squared error and mean absolute percentage error (Chen, 2007).

Chen and Wang (2007) examined the feasibility of applying SVR model to tourism demand forecasting and compared the model with back-propagation neural networks and autoregressive integrated moving average model. The study adopted genetic algorithm to search for optimal parameters of SVR model. China's tourist data set covering the period of 1985 – 2001 was analysed. The results show that (GA)-SVR model outperforms both back-propagation neural networks and autoregressive integrated moving average model based on normalised root mean squared error and mean absolute percentage error.

Hong (2009) forecasted the electric load employing a hybrid model of support vector regression with chaotic genetic algorithm (CGA). CGA is used to optimise the hyper-parameters of support vector regression model. Based on the results, it is evident that SVR-CGA has a superior forecasting accuracy as compared to SVR-GA, regression model and ANNs model. Liu et al. (2013) proposed a support vector regression model to solve aquaculture water prediction problem. The study adopted real value genetic algorithm to search for optimal parameters of the SVR model. The data used was collected from YiXing, China and it was for aquaculture water quality. Based on the results, random genetic algorithm-based support vector regression (RGA-SVR) outperforms traditional SVR, and back-propagation neural network based on root mean squared error and mean absolute percentage error. The study recommends that the model may be used to predict aquaculture water quality.

In the study, a model of support vector machine with adaptive particle swarm optimisation algorithm was presented for short term load forecasting. Furthermore, in terms of increasing the accuracy of SVM, the study used APSO to search for optimal parameters of SVM model. Dataset used was that of a certain electricity load forecasting problem. The results demonstrate that APSO-SVM has the highest forecasting accuracy than SVM, BPNN and regression model. APSO can also be used to search for optimal values of parameters with the smallest prediction error (Yue et al., 2009).

Fei et al. (2009) forecasted dissolved gases content in power to transform oil by applying a hybrid model of PSO-SVM. To determine SVM parameters, the study employed PSO algorithm. The time series data used was from China and it was that of several electric power companies. Based on the results, it was found that, PSO-SVM has a higher forecasting accuracy than grey model and ANN model respectively in the circumstances of smaller sample.

The study employed a hybrid technique of genetic algorithm with support vector regression to predict CNY exchange rates using various kernel function. The study employed BP neural networks as a benchmark model. The kernel functions used are radial basis, polynomial, sigmoid and linear respectively. The data set used is exchange rate data of United States dollar against Chinese Yuan (USD/CNY), Euro against Chinese Yuan (EUR/CNY) and Chinese Yuan against Japanese Yen (CNY/JPY). The results demonstrate that the hybrid model is effective in predicting Chinese Yuan (CNY) exchange rate (Jiang and Wu, 2016).

Al-hnaity and Abbod (2016) introduced a hybrid model of support vector regression, support vector machine and back-propagation neural networks for predicting financial time series data. For improving SVR and SVM prediction output, the study adopted quantisation factor. Genetic algorithm is also employed to determine the weights of the proposed model. Data set used is that of daily index closing prices of FTSE 100, S&P 500 and Nikkei 225. The experimental results show that the hybrid model is superior to single model and traditional AR model.

Fu (2010) proposed EMD-SVR model for EURO/RMB exchange rate series. EMD method is employed to decompose the series which is non-stationary and irregular into finite and small number of sub-signals. Furthermore, each sub-signal was modelled by SVR and the prediction results were aggregated. The results show that the proposed model has a superior forecasting capability than a normal support vector regression (SVR) model.

The study employed a hybrid model of EMD-SVR for predicting electricity load demand. EMD method is used as a denoising method on the training data. Three various electricity load time series data from various countries were used as dataset. The results demonstrate that EMD-SVR model outperformed SVR model and non-feature denoised SVR model for predicting electricity load (Yaslan and Bican, 2017).

The study employs a hybrid model of empirical mode decomposition (EMD) and support vector regression (SVR) for accurately forecasting electric load. EMD was used for decomposing the training set into finite intrinsic mode functions (IMFs) and used SVR to forecast each IMF. After the forecast of IMF, the forecasting results are aggregated to get the final forecasting results. Based on the results, it is found that the proposed hybrid model has a greater generalisation capability and high forecasting precision (Zhu et al., 2007).

### 1.6.3 Neural network autoregressive with exogenous inputs based evolutionary algorithm

Chatterjee et al., (2012) in their study investigated NARX neural network for predicting software faults and the investigation of applicability of NARX network in software reliability was carried out. The data used was that of the two real software failure data set. The model was compared to some parametric software reliability model and Elman network and TDNN based SRGM. The results shows that NARX model has a good predictive accuracy as compared to other models.

The study used an advanced neural network technique (NNARX) to come up with a model that can forecast level of flood water 5 hours ahead of time. The data used was obtained from the Department of Irrigation and Drainage Malaysia. The obtained results demonstrate that NNARX model can forecast level of flood water with the highest accuracy (Adnan et al., 2014).

Nor and Gharleghi (2011) compared NNARX model, ANN model, GARCH model and ARIMA model in predicting the Singapore dollar against US dollar exchange rate in three-time horizons. Results obtained reveal that NNARX model had a superior forecasting accuracy than the other models and among the nonlinear models, NNARX model outperformed ANN model, and both the models had a superior forecasting accuracy than the GARCH model.

Ruslan et al. (2016) proposed a nonlinear dynamic neural network model (NNARX) for predicting flood with 7 hours prediction time. The study used levels of river water at upstream stations as inputs and level of river water at downstream stations as output. The data used were collected from 15 December 2011 until 19 December 2011, with a total of 970 observations, of which 542 were used for validating the model, and 428 were used for testing the model. The results show that the model (NNARX) successfully forecast level of flood water 7 hours ahead of time.

Zhang et al. (2016), in the study, proposed EMD-NNARX hybrid model for exchange rate forecasting. Firstly, the series will be decomposed into small number of intrinsic mode functions (IMFs) and one residue then NNARX model will be used to model and forecast the series thereafter. The resulting forecasts will be aggregated to form the final forecasts of the series. The results show a very high stability of the proposed model for exchange rate forecasting.

### 1.6.4 Adaptive neuro-fuzzy inference system based evolutionary algorithm

The study forecasted regional electricity load in Taiwan using adaptive network based fuzzy inference system (ANFIS) model and shows the performance in terms of forecasting of this model. The results show that ANFIS model has the best forecasting accuracy of regional electricity load of Taiwan based on mean absolute percentage error than regression model, ANN, GA-SVR, RSVMGA, and HEFST. Furthermore, the study suggests that ANFIS is better for forecasting electricity loads and is a promising alternative. (Ying and Pan, 2008a).

Rajalakshmi et al. (2012) compared Neural Network Auto-Regressive model with exogenous inputs and Adaptive Neuro Fuzzy Inference Systems model in identifying nonlinear dynamic process. To identify model structure and suitable model of dynamic systems which are nonlinear is the main contribution of the study. The results demonstrate that ANFIS model is accurate in identifying dynamic systems which are nonlinear.

Tehran Stock Exchange is forecasted by comparing both the static neural networks models (ANFIS and MFNN) and dynamic neural networks (NNARX) model in determining the model which is efficient and accurate in forecasting the series. The data used covered the period from 25 March 2009 till 22 October 2011. According to the results based on MSE and RMSE, ANFIS model has a more accurate forecast than the other models included (Abounoori et al., 2012).

In improving the performance of ANFIS model, the application of evolutionary algorithms (GA, ACOR and DE) was investigated in the study in terms of evaluating the quality parameters of Gorganroud River water. Sensitivity analysis was firstly employed to estimate suitable parameters. ANFIS-DE was the best model for predicting EC and TH in the test stage. Moreover, ANFIS-DE and ANFIS-GA were the best in predicting SAR in the test stage according to the results. Thereby, the ANFIS model is the best model for predicting the parameters (water quality). The included algorithms are also improving the accuracy performance of ANFIS model (Azad et al., 2018).

The study used a hybrid model of ANFIS-GA to estimate the amount of regional rainfall. The model is coded in the MATLAB environment. The data used was that of meteorological data with 3650 observations covering the period 2008-2018. The data was analysed using ANN model and ANFIS model respectively. The obtained results show that ANFIS model in both training and testing sets obtained R-values of 0.9920 and 0.9840 respectively with the error of 0.0011. Therefore, it shows that hybrid models of ANFIS-GA can be employed to analyse the meteorological events (CALP, 2019).

Wei (2016) forecasted stock prices in the Taiwan Stock Exchange Capitalisation Weighted Stock Index (TAIEX) and Hang Seng Stock Index (HSI) using a hybrid model of EMD based adaptive neuro-fuzzy inference system (EMD-ANFIS). The proposed hybrid model was compared with Chen's model, Yu's model, the autoregressive (AR) model, the ANFIS model, and the support vector regression (SVR) model to measure its performance. Based on the error measurement of root mean squared error, it was found that the proposed model outperformed the other models in the study.

### 1.6.5 Least square support vector regression based evolutionary algorithm

Kaytez et al. (2015) forecasted the electricity consumption of Turkey using LSSVM model. The model was compared with regression analysis and artificial neural network. The predictor variables used are gross electricity generation, installed capacity, total subscription and population covering the period of 1970 to 2009. The results demonstrate that LSSVM is superior in terms of accuracy and prediction than regression analysis and artificial neural network.

Zou (2009) applied least squares support vector regression model to forecast the condition of hydro turbine generating units. The study used a real-value genetic algorithm for the optimal search of hyper-parameters of the LSSVR model, and the parameters are then adopted to construct a new optimal LSSVR model. The dataset employed is the peak-peak value time series data of the stator vibration signals in HGUs. It is demonstrated in the results that GA-LSSVR model has outperformed BP neural networks and conventional LSSVR model using mean absolute percentage error.

In the study, genetic algorithm/least squares support vector regression was used to model pKa values of the 107 indicators. It is shown that GA-LSSVR had the most accuracy with squared correlation coefficient of 0.90 and 0.89 for both the training and testing set respectively. The model outperformed the traditional LSSVR which then shows the importance of proper selection of the hyper-parameters of the model (Goodarzi et al., 2010).

Xu (2013) proposed least square support vector regression trained with genetic algorithm as a forecasting method for forecasting tourism population. Genetic algorithm is used to optimise the parameters of LSSVR model. The data used is that of China’s tourist population from 1990 to 2005 and from 1991 to 2009 for Jiangsu tourist population. The results demonstrate that GA-LSSVR has a higher forecasting accuracy than SVR.

The study forecasted crude oil prices using a hybrid algorithm of grid-GA-LSSVR. The grid method was used to determine the boundaries of parameters for LSSVR model and GA is used to select the best parameters for the model. The results show that grid-GA-LSSVR model is outperforming LSSVR and other forecasting techniques and it can be employed for crude oil prices forecasting (Yu et al., 2016).

In the study, water demand was forecasted using the proposed hybrid model of EMD-LSSVM. The study used a monthly water demand record as a dataset from Batu Pahat city in Johor of

Peninsular Malaysia. The data series was decomposed to a small intrinsic mode function (IMFs) and one residue by the EMD method and the LSSVM was used to model the IMFs and one residue. The resulting forecasts were aggregated to form the final forecasts. Based on the results, it is found that the proposed model has a better forecasting performance than the traditional LSSVM and ANN without EMD. The study further stipulates that EMD-LSSVM is an effective model for water demand forecasting (Shabri et al., 2015).

Lin et al. (2012) forecasted the exchange rate by employing a proposed hybrid model of EMD-LSSVR. LSSVR will be employed to model and forecast the intrinsic mode functions (IMFs) and one residue of the series after the EMD has decomposed the series and the resulting forecast form LSSVR will be aggregated to produce a final forecasted value. The results demonstrated that the proposed model outperforms EMD-ARIMA, LSSVR and ARIMA without EMD.

### 1.7 Proposed methodology

The study which is quantitative in nature uses monthly time series data of foreign direct investment as a dependent variable and exchange rate, GDP and inflation rate are independent variables from the period of January 1970 to June 2019, sourced from World Bank (https://data.worldbank.org/indicator/BX.KLT.DINV.CD.WD?locations=ZA) and South African Reserve Bank (https://www.resbank.co.za/Research/Statistics/Pages/Statistics) respectively. The reason for selection of the variable is that, as outlined in the introduction section, FDI is good for the host country as it can create jobs and there can be economic growth amongst others. The independent variables (exchange rate, GDP, and inflation rate) amongst others, are chosen because they are the determinants of FDI and data for them is available from 1970 to 2019. The data is in South African context. Furthermore, the series is tested for nonlinearity. The tests which are employed to assess nonlinearity are BDS test, CUSUM test, Keenan's test, White and Terasvirta neural network test, Tsay test and McLeod-Li test.

Three supervised nonlinear machine learning models of artificial neural network (ANN), support vector regression (SVR), and least square support vector regression (LSSVR) and one evolutionary algorithm (genetic algorithm) are employed to model and forecast foreign direct investment. The evolutionary algorithm technique (genetic algorithm) is used in order to optimise the hyper-parameters of the included machine learning models and to improve forecasting accuracy of the models. The mathematical presentation of the models and algorithm are briefly shown in this section.

**Artificial neural network (ANN):** ANNs are high performance, non-linear analytical tools that can establish a relationship between the dependent and independent variables without prior knowledge of the system. They are also nonparametric in nature. Multilayer perceptron (MLP) is a commonly used ANN approach for solving problems. The following structure shows the architecture of the MLP model.

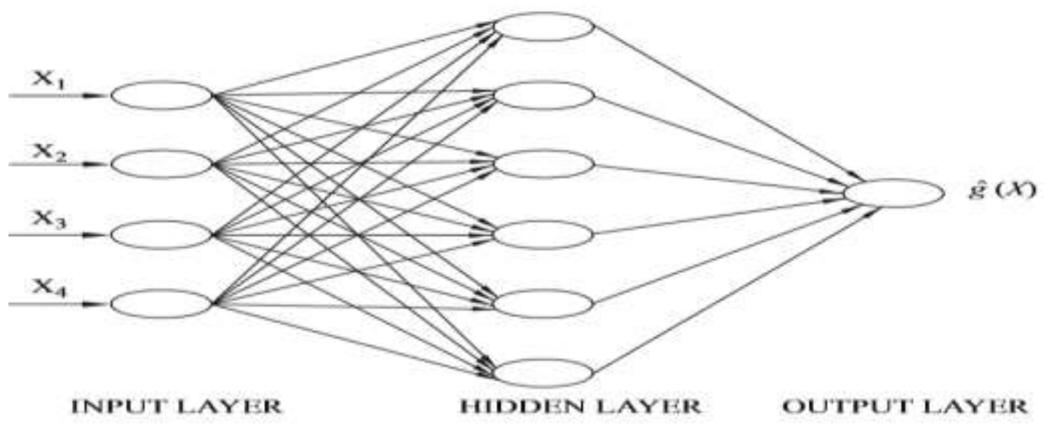


**Figure 1. 1: Artificial neural network structure**
**Source: Cheng (2010)**

There are three layers in the structure of the neural network: input layer (independent variables), output layer (dependent variables) and one or more hidden layers. These layers have transfer functions. In general, one hidden layer with sufficient neurons is good for solving problems, however, this will depend on the complexity of the problem. Input and output layers respectively will adopt purline, which is a linear transfer function, and in the hidden layer, a nonlinear transfer function of sigmoid transfer function is employed (Karimi and Dastranj, 2014). Hyper-parameters that need to be optimised so that the models forecasting performance may improve are the number of neurons in the hidden layer, weights and bias.

**Support vector regression (SVR):** the model support vector machine (SVM) was initially intended to address classification problems (Vapnik, 2013). Drucker et al. (1997) proposed the SVR model which is the version of SVM. Support vector regression follows a principle of structural risk minimisation which minimises an upper bound of the generalisation error than the prediction error on the training set (the principle of empirical risk). The linear regression function (in the feature space) using mathematical notation, is approximated using the following expression:

$$y = f\left(x\right) = \omega^T \varphi\left(x_i\right) + b, \varphi : R^n, \omega \in F, \qquad 1.1$$

given that  and  are coefficients,  denotes the high dimensional feature space which is nonlinearly mapped from the input space . The study will adopt the Gaussian function as the kernel function. Smola and Schölkopf (1998) stated that Gaussian function yields better prediction accuracy. Hyper-parameters that must be optimised are regularisation parameter C which determine the trade-off cost between minimising the training error and minimising the model's complexity, bandwidth of the kernel function which represents the variance of the Gaussian kernel function and the tube size of the $(\varepsilon)$ -insensitive loss function which is equivalent to approximation accuracy placed on the training data points, $(\varepsilon)$ -insensitive loss function is a key component of SVR, a type of machine learning model used for regression tasks. Furthermore, is designed to measure the deviation or error between the predicted output of the SVR model and the true target value.

**Least square support vector regression (LSSVR):** least square support vector machine is an alternative of the standard support vector machine and it was improved by (Suykens et al., 2002). Least squares loss function is employed by LSSVM to construct the optimisation problem based on the equality constraints. LSSVM is used commonly for optimal control, classification and regression problem (Suykens et al., 1998). The LSSVR technique consists of fitting an ambiguous function using a specific sample of the training data set. The regression function can be formulated as feature space representation. Given by the following mathematical notation:

$$y = f(x) = \omega^T \varphi(x_i) + b, \varphi : R^n, \omega \in F, \quad 1.2$$

given that and are coefficients, denotes the high dimensional feature space which is nonlinearly mapped from the input space . The current study will adopt radial basis function as a kernel function. LSSVR has equal constraints unlike the standard SVR. Hyper-parameters to be optimised are.

**Genetic algorithm (GA):** A genetic algorithm is a probabilistic search technique that can be used to search large and complex spaces. The ideas of natural genetics and evolutionary principles are employed by GA. GA has two types: real and binary coded GA respectively. The study will adopt real value GA to optimise the hyper-parameters of the machine learning models because it is straight forward, fast, and efficient.

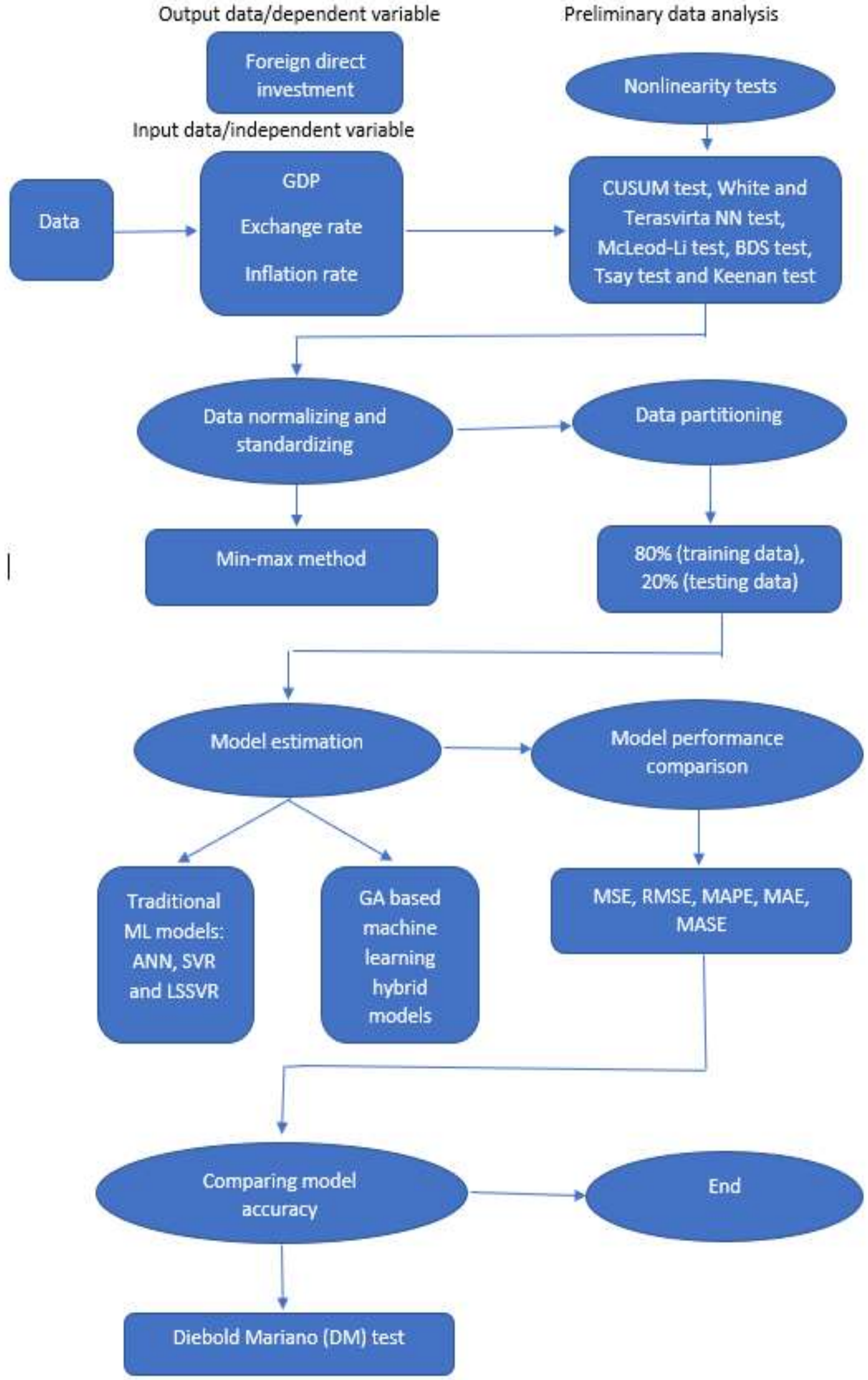


**Figure 1. 2: Methodology diagram**

## 1.8 Expected contribution

On the ongoing research, machine learning models have been applied to model time series data. Various machine learning models have been compared with traditional time series models whereby machine learning models have had superior forecasting performance. Studies by (Domingos et al., 2019, Liu et al., 2018, Khosravi et al., 2018a, Phaladisailoed and Numnonda, 2018, Deb et al., 2017, Chatterjee and Bandopadhyay, 2012, Krollner et al., 2010, Ahmed et al., 2010, Fei et al., 2009, Hong, 2008), have applied machine learning models in a comparative analysis, however, the models included in the current study have not been compared in the existing literature to the knowledge of the researcher.

This being the case, the study seeks to contribute by comparing the machine learning models in terms of modelling and forecasting foreign direct investment into South Africa by assessing the forecasting performance of the models. However, setting of appropriate parameters of the models is vital as it influences the performance of the model. Parameter setting is usually done using the trial-and-error method which can be time consuming.

Thus, the study will employ hybrid setting evolutionary algorithms for the purpose of selecting optimum hyper-parameters and improving forecasting performance of the models employed to model and forecast foreign direct investment. Recommendations will also be made in terms of models which have the accurate forecasting performance that can be used for selecting optimum hyper-parameters and improving forecasting or prediction performance.

## 1.9 Ethical consideration

This study will not include human and animal participation, it tends to only focus on the analysis of the secondary data sourced from World Bank website and South African Reserve Bank respectively, using statistical techniques. A postgraduate manual will be referred to when conducting this study to ensure that rules and regulations of the university are adhered to. Any information that is obtained in connection with this study will remain confidential and will not be disclosed. Ethical clearance will be sought from the Faculty of Economics and Management Science Ethical clearance committee.

## 1.10 Layout of the thesis

- Chapter one is an introduction and background of the study
- Chapter two is literature review which will review the empirical literature of the machine learning models and machine learning hybrid models.

- Chapter three is the methodology of the models together with the tests that will be employed to address the aim(s) and objectives of the study.
- Chapter four comprises of data analysis which will be about the preliminary analysis, estimation of traditional machine learning models and estimation of machine learning hybrid models together with the interpretations and discussions of the results.
- Chapter five comprises of the conclusion of the study, recommendations, and suggestions for further research.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1 Introduction

This chapter reviews empirical literature on both machine learning models and evolutionary algorithm in modelling and forecasting time series data. Advantages and disadvantages of machine learning models will also be reviewed; and empirical literature on the hybrid models of machine learning based evolutionary algorithms will also be presented. This chapter is organised as follows: Section 2.2 gives overview of foreign direct investments and its determinants, while section 2.3 presents empirical literature on machine learning models and evolutionary algorithms. Lastly, section 2.4 offers the concluding remarks of the chapter.

## 2.2 Foreign direct investment and its determinants

Mathur and Singh (2013) state that “foreign investors care about economic freedoms, rather than political freedoms, in making decisions about where to locate capital.”. Therefore, democratic countries may receive less foreign direct investment flows if economic freedoms are not guaranteed.

Foreign direct investment (FDI) is defined as a type of investment in which foreign firms/companies make an investment into a host country. This type of investment is important as it plays a role in creating employment, developing managerial skills, and promoting improvements in productivity while transferring technology to local firms, which will boost the local economy (Majavu and Kapingura, 2016a). Moreover, Borensztein et al. (1998) postulate that foreign direct investments (FDIs) are considered an important channel for multinational companies to access advanced technologies in developing countries. Furthermore, in transferring technology, FDI serves as an important vehicle and contributes in a larger measure to growth than domestic investment. Borensztein et al. (1998), in line with Levine and Renelt (1992), show that there is a robust relationship between economic growth, FDI, and human capital.

Furthermore, in developing economies, domestic savings are supplemented, employment is generated thus, there will be economic growth because of receiving FDI In the recent past, researchers have studied determinants that drive FDI to a receiving economy; such as the studies of (Majavu and Kapingura, 2016c, Wafure and Nurudeen, 2010) amongst others. The results show that FDI is driven by the following factors: GDP, openness, inflation rate, exchange rate, corporate tax and a measure of financial crises.

Moreover, foreign direct investment (FDI) has been intensely researched over the past and some important findings have been made in terms of its important determinants, from Latin America, Asia, Europe and even Africa, several important determinants of FDI have been found and some of these determinants are similar. For instance, these studies were made in the recent past; Resnick (2001) and Li and Resnick (2003) found out that foreign capital flows are impacted negatively by the level of democracy. Nonnenberg and Mendonca (2004) used 38 developing countries panel data covering the period from 1975 to 2000. The results of the panel model revealed that FDI is positively and significantly affected by gross national product, the average rate of growth in previous years, the level of schooling, and the degree of openness.

In the cross-country analysis conducted by Alfaro (2003), it was revealed that FDI exerted an ambiguous effect on the host country's economic growth, however in the primary sector, FDI tended to have negative effect on growth. Blonigen (2005) stated that "FDI can therefore enhance growth by allowing host countries access to advanced technologies not available domestically". The study conducted by Johnson (2006) showed that in the developing countries, FDI inflow is boosting economic growth, but this is not the case in the developed nations.

Onyeiwu and Shrestha (2004) found that economic growth, inflation rate, openness of the economy, international reserves, and natural resource availability plays a major role in the allocation of foreign direct investment (FDI) into Africa. Oke et al. (2012) employed generalised method of moments (GMM) with autoregressive error technique to investigate locational determinants of FDI in Nigeria. The time span on data employed covered the period from 1975 to 2008. The study revealed that index government expenditure, index of energy consumption, and political stability indicator are positive and significant determinants of FDI, while inflation rate, exchange rate, market size, infrastructure and human capital, are not significant determinants of FDI determinants in Nigeria.

Erdogan and Goksu (2014) investigated the determinants of FDI in the period from 1985 to 2011 for 88 countries. The study found that social security spending, urbanisation role, the ratio of population over 65 years, and health spending have a negative relationship and statistically significant impact to FDI, whereas GDP per capita, GDP growth, market size, inflation rate, unemployment rate, labour force growth, credit to private sector, market capitalisation, and corruption control have a statistically significant and positive impact to FDI.

Akinsete (2016) found that South African consumer buying power and high GDP growth rate are determinants of market seeking FDI. Furthermore, the study found that FDI is inhibited by uncertainty in government policy, unstable regulatory framework, an unfavourable labour regime and high labour costs relative to productivity. Oro and Alagidede (2021) investigated the claim that market size and petroleum resources are regarded widely as drivers of FDI using panel and time series data respectively for forty-nine African countries. The results of the study revealed the mining of mineral ore in support of cheap labour as a key driver of FDI. Zhang (2021) found out that in the study, “C-FDI are much of greater benefits to Africa than the FDI from the North, and further found that uniquely the large Chinese investment in Africa's infrastructure not only contributes to overall economic growth, but also enhance host-country's absorptive capacity in attracting and utilising foreign capital”.

Asafo-Agyei and Kodongo (2022) state that, there must be minimum capacity to absorb the growth-enhancing benefits of FDI, for FDI to have an appreciable impact on economic growth. Sylvaire et al. (2022) in their study, aimed at determining the impact of CFDI on inclusive development in Africa. The panel data used was of 48 African countries from 2003 to 2020. The empirical results demonstrated that African economy is promoted through CFDI by direct investment.

The study employed Bayesian Model Averaging (BMA) to re-investigate the determinants of FDI for which data is available for 53 African countries. The study found out that out of the 23 explanatory variables, the following variables are substantive predictors of FDI in the African continent, these variables are; government consumption expenditure, inflation, GDP per capita, capital openness and credit to the private sectors (Ajide and Ibrahim, 2022). Furthermore, it was found that there are differences in the determinants of FDI across the African regions therefore assuming general policy-framework for attracting FDI may not be an efficient way of trying to attract FDI to other regions.

## 2.3 Literature on machine learning models and evolutionary algorithms

Shalev-Shwartz and Ben-David (2014) states that “machine learning is one of the fastest growing areas of computer science, with far-reaching applications”. Machine learning (ML) refers to the automated detection of meaningful patterns in the data. In the recent past, machine learning has been extensively employed in almost any task that needs information extraction from large data sets. Machine learning is also employed in various fields including in scientific applications, for example, in bioinformatics, medicine, and astronomy. Machine learning tools or algorithms are concerned with providing programs with the ability to learn and adapt.

In time series prediction, various traditional time series models have been employed in the recent past such as for financial time series prediction, electric utility load forecasting, weather and environmental state prediction. However, these models are generated from the assumption of linearity, and this is a shortcoming for predicting nonlinear time series data. Furthermore, time series data generating processes are generally complex for these traditional time series models. As Sapankevych and Sankar (2009) state, “accurate and unbiased estimation of the time series data produced by these systems (e.g. ARIMA) cannot always be achieved using well known linear techniques, thus the estimation process requires more advanced time series prediction algorithms”. Therefore, this leads to various machine learning algorithms to be used and employed for time series forecasting. These machine learning models include, but not limited to, multilayer perceptron (MLP), recurrent neural network, adaptive neuro-fuzzy inference systems (ANFIS), neural network autoregressive with exogenous inputs (NNARX), generalised regression neural network (GRNN), support vector regression (SVR), least square support vector regression (LS-SVR), dynamic least squares support vector regression (DLS-SVR), fuzzy inference systems (FIS), time lag neural network (TLNN) and Bayesian networks (BN).

These various machine learning models have been used for time series forecasting in the recent past. In the study, support vector regression (SVR) model was compared with traditional neural network architectures including feed-forward multilayer perceptron employing back-propagation and radial basis functions for stock price indices prediction. Based on the results, it was found that SVR model demonstrates superior forecasting accuracy than the other model for a very small time window of three samples and a very small prediction horizon of one sample (Trafalis and Ince, 2000).

In a comparative study, Australian national electricity market prices were predicted by comparative forecasting performance of support vector regression (SVR) and multilayer perceptron (MLP). The results demonstrated that MLP was outperformed by SVR in training time, however, both models obtained similar results in forecasting accuracy, which the study believes might be due to the way training data was selected and used (Sansom et al., 2002)

Ongsritrakul and Soonthornphisaj (2003) predicted gold price by combining various models including MLP, decision trees, and support vector regression. In the study, decision tree was used to feed factors into SVR model which were then employed as input to linear regression model, MLP, and SVR to predict gold price. According to the results, which were based on MSE, MAD, and MAPE, SVR model outperformed all the models in predicting gold price. Abraham et al. (2004), in a one-step ahead, predicted Nasdaq 100 index and the NIFTY index using the comparative study ANN employing Levenberg-Marquardt algorithm, SVM, Takagi-Sugeno neuro-fuzzy model and a Difference Boosting Neural Network (DBNN). Results showed that SVM has marginally outperformed all other included models.

Remesan et al. (2008) proposed a dynamic neural network model of Neural Network Auto-regressive with exogenous input (NNARX) and a static neural network model of adaptive neuro-fuzzy inference system (ANFIS) model for rainfall runoff phenomenon effectively from antecedent rainfall and runoff information. The study employed Gamma test for investigating the best input structure from the various structures that were constructed, for the selection of the length of the training data, and for carrying out the best input combination. It was found that both the models provide a better forecasting accuracy of rainfall runoff and that they work efficiently in rainfall runoff predictions.

For annual rainfall in Zhengzhou, the study proposed a model of Generalized Regression neural network (GRNN) and compared it with back-propagation neural network (BPNN) model and stepwise regression model. The results demonstrate that GRNN has higher forecasting accuracy than both BPNN and stepwise regression model respectively. On the overall, it was also found that both GRNN and BPNN have higher prediction accuracy than stepwise regression (Wang and Sheng, 2010).

Kajornrit et al. (2012), for monthly rainfall prediction in the northeast region of Thailand, proposed fuzzy inference system (FIS) model. For measuring the forecasting performance of the proposed model, the study compared the model with conventional Box-Jenkins model and artificial neural networks model. Results illustrate that the modular FIS model gives better results than Box-Jenkins model and artificial neural networks models respectively.

Faulina et al. (2012) contributed to the literature by proposing ensemble hybrid model of adaptive neuro-fuzzy inference system (ANFIS) and auto-regressive integrated moving average for monthly rainfall forecasting at certain areas of Indonesia, Pujon and Wagis area. The model was compared with singles models of ARIMA and ANFIS. The study used two sets of rainfall data from January 1975 to December 2010. Results showed that single models are more superior in forecasting accuracy than hybrid model. These results were also obtained in the M3 competition that showed that more complicated models do not always yield better results.

The study predicted for a one month ahead monsoon rainfall in the country based upon the Indian Ocean dipole parameter using a proposed model of Time lag neural network with gamma memory processing neuron. A FTLNN three layer with the following architecture (4-6-1) with gamma memory and conjugate gradient back-propagation learning algorithm were designed for forecasting of rainfall series with delay taps equal to 3 and gamma memory coefficient equal to 0.5. In the testing period, the forecasting accuracy was found to be 93% (Charaniya and Dudul, 2012).

Although machine learning (ML) models or techniques have shown a great deal of success in modelling and forecasting time series data, there are some shortcomings inherent in their application. The shortcomings make it difficult for these models and researchers to realise the true potential of forecasting accuracy that these models can achieve. For instance, in neural network models, the following weaknesses are present; large number of free parameters to be estimated, selection of free parameters usually are calculated empirically, prone to get trapped in a local minima, and can be computationally expensive (Sapankevych and Sankar, 2009). Moreover, disadvantages of ANFIS model, although it is a robust modelling technique in different applications, are artificiality, robustness and irregularity and needs to be trained well to work successfully (Sarkheyli et al., 2015) cited from (Zhang et al., 2012).

Machine learning models, including support vector regression (SVR), least square support vector regression (LSSVR) and neural network autoregressive with exogenous inputs (NNARX) need for their parameters to be successfully optimised for the models and other machine learning models to have higher forecasting accuracy and improved generalisation. As Sarkheyli et al. (2015) reiterated, the adaptive modelling parameters of these machine learning models must be carefully set or selected.

Over the recent past, there has been quite a lot of effort in finding optimal values for machine learning modelling parameters to decrease training error and increase modelling accuracy, and this lead led to the development of various computational methods called evolutionary algorithms. “Evolutionary algorithms are considered as random search methods” (Mühlenbein, 1993).

Evolutionary algorithms are defined as algorithms that performs optimisation or learning tasks with ability to evolve (Yu and Gen, 2010). Evolutionary algorithms have three main characteristics which are population based where a group of solutions is maintained by evolutionary algorithms called a population to optimize or learn the problem in a parallel way. The basic principle of the evolutionary process is the population. There is fitness orientated in which every solution in the population is called an individual. Every gene represents every individual called its code and performance evaluation called its fitness function. Lastly, variation driven by individuals mimic a genetic gene change by undergoing several variations, which is important in searching the solution space. Many algorithms based on population, fitness orientated, and variation driven have been developed since the 1960s. The algorithms include, but not limited to, genetic algorithms (GAs), particle swarm optimisation (PSO), chaotic genetic algorithms (CGAs), differential evolution (DE), firefly algorithm (FA), adaptive differential evolution (ADE), simulated annealing (SA), and others. Figure 2.1 illustrates some of these algorithms with their timelines.

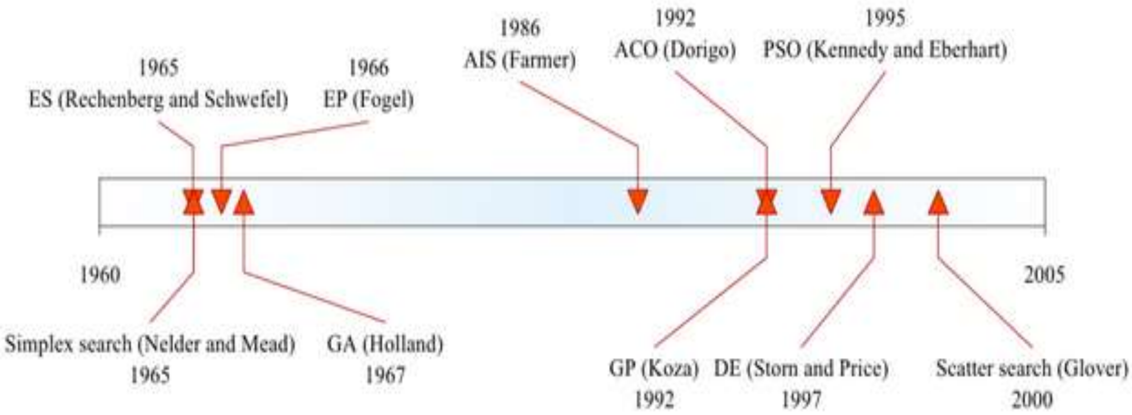


**Figure 2. 1: Timeline of various evolutionary algorithm(s)**
**Source: Yu and Gen (2010)**

The evolutionary algorithms have been employed as optimisation techniques or methods for optimising various machine learning models' parameters so that forecasting performance and accuracy can be improved. For instance; Liu et al. (2001) proposed improved naïve Bayes classifier technique and investigated the use of genetic algorithm (GA) for selection of a subset of input features in classification problem. The study employed specially designed genetic algorithm to obtain the optimal model parameters of a proposed adaptive radial basis function (Jareanpon et al., 2004). Vesterstrom and Thomsen (2004) compared differential evolution, PSO and other evolutionary algorithms on 34 suites widely used to benchmark problems and the results revealed that differential evolution outperformed other algorithms. However, both differential evolution and particle swarm optimisation were both outperformed by the EA on two noisy functions.

Behnamian and Ghomi (2010) found that a hybrid ensemble model of nonlinear regression-based particle swarm optimisation (PSO) and simulated annealing (SA) produces lower error in forecasting the data sets. Furthermore, Adhikari and Agrawal (2011) contributed to the literature by proposing a hybrid ensemble model of neural networks based particle swarm optimisation (PSO) of Treleal and Trelea2 for seasonal time series forecasting. The results obtained revealed that neural network (NN) based on PSO-Treleal and PSO-Trelea2 has higher forecasting accuracy than standard back-propagation algorithm. Jin et al. (2012) applied particle swarm optimisation (PSO) and neural network in a hybrid model for predicting monthly mean rainfall.

The study predicted rainfall employing a proposed model incorporating Immune evolutionary algorithm based back-propagation networks. Results of the study demonstrated that not only is the algorithm higher in accuracy but also has a better stability (Du et al., 2012). Wang et al. (2015a) applied a hybrid model of back-propagation neural network based on adaptive differential evolution (ADE) called ADE-BPNN for forecasting time series data. Results showed that the proposed model of ADE-BPNN improves forecasting accuracy than standard BPNN, ARIMA and other hybrid models. The study proposed an ensemble hybrid model of nu-SVR and epsilon-SVR optimised by differential evolution for forecasting building energy consumption (Zhang et al., 2016a). It was revealed that the proposed model shows a higher accuracy for forecasting building energy consumption.

Boyd et al. (2007) defines decomposition as a “general approach to solving a problem by breaking it up into smaller ones and solving each of the smaller ones separately, either in parallel or sequentially”. Moreover, sequentially breaking up a problem into smaller ones has an advantage from a fact that problem complexity grows more than linearly. Decomposition is a method that has been employed for so many years in time series modelling and forecasting.

Decomposition methods have been used in literature in an ensemble hybrid model with machine learning models, for instance (Yu et al., 2008b, Xiang et al., 2018) , time series / econometrics models (Abadan et al., 2015, Büyükşahin and Ertekin, 2019) and they have shown to improve forecasting performance of these models. These decomposition methods amongst others include empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), wavelet decomposition (WD), discreet wavelet transformation (DWT) and signal decomposition.

Khandelwal et al. (2015) employed Discreet wavelet transformation based ARIMA-ANN hybrid model for forecasting four real world time series data and compared the employed method with ARIMA, ANN, and Zhang’s hybrid model in terms of forecasting accuracy. Results demonstrated that the proposed model had achieved the best forecasting accuracy than the listing models. Zhang et al. (2017a) employed wavelet decomposition for time series decomposition.

Annual runoff time series was predicted in the study by a hybrid model of EEMD-ARIMA. EEMD was employed to decompose the series into several IMFs and one residue. Three annual runoff series from Biuliuhe reservoir, Dahuofang reservoir and Mopanshan reservoir in China, were employed as time series data. It was demonstrated that EEMD method can significantly improve forecasting accuracy of ARIMA model (Wang et al., 2015b).

Annual runoff time series was predicted in the study by a hybrid model of EEMD-RBFN. EEMD was employed to decompose the series into several IMFs and one residue. Results revealed that EEMD can improve forecasting performance of RBFN and the hybrid model has shown superior forecasting model in comparison to ARIMA and BPNN (Ma and Li, 2016).

Zhang et al. (2017b) proposed a new two stage methodology combining ensemble EMD with multidimensional k-nearest neighbour model for predicting closing price and high price of stock simultaneously and compared its forecasting accuracy to those of EMD-KNN, KNN and ARIMA models simultaneously. EEMD was employed to decompose the series into IMFs. Based on the results, it was found that the proposed model outperformed other listing models.

Qiu et al. (2017) employed a hybrid model consisting of empirical mode decomposition and Deep Belief Network (DBN) with two restricted Boltzmann machine for forecasting load demand time series. EMD was employed to decompose load demand time series into several IMF and the models used to model each IMF so that they can be accurately forecasted. Load demand time series data from Australian Energy Market Operator was used to demonstrate the ability of the proposed model. The results reveal that the proposed model is superior to the other nine forecasting models in terms of forecasting load demand series.

### 2.3.1 Artificial neural network (ANN) model

The study by McCulloch and Pitts (1943) titled "A logical calculus of the ideas immanent in nervous activity" was the first model of artificial neural networks developed. Ever since the development of the first ANNs model, there has been numerous models of artificial neural networks (ANNs) developed. Artificial neural networks (ANNs) models are both for classification and function approximation. Zhang and Hu (1998) states that neural networks are data driven and self-adaptive methods which are non-parametric in nature.

In the numerous models of ANNs developed, the mostly approved and model applied is the feed-forward multilayer network or the multilayer perceptron (MLP). This model has been adopted and applied for forecasting purposes. This model is composed of several layers which are the input layer, output layer and in between is one or several hidden layers. Except in the input layer, the other layers have an activation function which helps in the calculation of the weights of the model.

In recent years, artificial neural networks (ANNs) have been applied in modelling and forecasting economic and financial time series data, particularly stock market returns. White (1988) used neural network modelling and learning techniques to search for and decode nonlinear regularities in asset price movements. Moreover, Egeli et al. (2003) showed that ANN (MLP and GFFN) is better in forecasting Istanbul Stock Exchange than moving averages models. Enke and Thawornwong (2005) extended the study of stock returns by introducing information gain technique in machine learning. It was found that this trading strategy is better than buy and hold strategy as it gives more risk adjusted profits. It was found in the study that the previous day return, other than the first three days' input, influences the predictive power of neural network (Thenmozhi, 2006).

Over the recent past, linear models, particularly ARIMA, were used to forecast exchange rate (Tseng et al., 2001, Weisang and Awazu, 2008, Newaz, 2008, Babu and Reddy, 2015). Due to the flexibility and exceptional properties of artificial neural network model, particularly multilayer perceptron, the model started to be used for forecasting exchange rate time series. Franses and Van Homelen (1998) stated that ANN does not yield good forecasting performance in the sample data using exchange rate. However, Zhang and Hu (1998) and Panda and Narasimhan (2007) showed successfully that ANN model outperformed linear autoregressive and random walk models in forecasting exchange rate. Furthermore, Zhang and Hu (1998) state that the number of input nodes than hidden nodes has a greater impact on the forecasting performance of ANN. Moreover, Kamruzzaman and Sarker (2003) and Kadilar et al. (2009) successfully applied neural network for exchange rate forecasting.

Witt and Witt (1995) state that many possible forecasting methods were listed in the previous reviews of tourism forecasting. Furthermore, they state that the majority of research in tourism forecasting employed econometric models, spatial models and time series models. Amongst most tourism research time series models, ARIMA model has been used in the past to forecast tourism demand (Chu, 1998, Chu, 2009) and this study, amongst others, reported promising results obtained by ARIMA model. However, in the 1990s and early 2000s and before, ANN model was used to forecast tourism demand. For instance, Pattie and Snyder (1996) found out that Census II decomposition and neural network are the best models to forecast tourism 12 months ahead, Law (1998) revealed that for forecasting room occupancy rates, it is better with neural network as it outperforms both multiple regression and naïve extrapolation. Burger et al. (2001) reiterated that neural network forecasts tourism better.

Neural network has also been applied in time series data. This has led to new developments as new findings were highlighted regarding how to make sure that neural networks perform to their desirable ability. Nelson et al. (1999), Virili and Freisleben (2000) and Zhang and Kline (2007) have stated that deseasonalising time series data improves the forecasting accuracy of artificial neural network. On the other hand, Zhang et al. (2001) also state that for nonlinear time series, neural network is the best model and that input nodes are crucial for neural network model building for forecasting.

Although Neural network seems to be a great model in time series and a promising alternative in other disciplines, it has shortcomings. In the study, Palmer et al. (2006) recommended that future research must be undertaken and aimed at conquering shortcomings of ANN model, that is, selection of a proper ANN architecture and effect of input variables. Eğrioğlu et al. (2008), according to the results obtained in the study, recommended that weighted information criteria is reliable for selecting a proper ANN architecture. However, Aladag (2011), based on the results, recommended that tabu search is the best method for selecting the best ANN architecture. Bal et al. (2016) state that for volatility time series, measures that should be used for selecting the best

ANN model should be based on the median (MdAE, MdAPE, SMdAPE and MdRAE) and not to use measures based on penalising (AIC, BIC).

Lin et al. (2011) contribute to the literature by trying to build a forecasting model using ARIMA, artificial neural network, multivariate adaptive regression splines (MARS) of visitors to Taiwan. A monthly dataset of visitors to Taiwan was used for illustrative purposes. The results revealed that the ARIMA, based on RMSE, MAD and MAPE, had outperformed ANN and MARS and it gives an effective alternative to forecast tourism demand.

Claveria and Torra (2014) used the models which were used in the study of Lin et al. (2011), with the exception of MARS, to investigate the predicting performance of neural net relative to time series models in forecasting tourism demand in the regional level. The data employed is from 2001 to 2009 of the overnight stays and tourism arrivals from all various countries of origin to Catalonia. Results shows that in a shorter horizon ARIMA model has outperformed self-exciting threshold auto-regression and artificial neural networks. Furthermore, the study suggests that pre-processing can increase neural network forecasting accuracy which is more suitable for nonlinear data.

Bing et al. (2012) in their study predicted Shanghai Stock Exchange Composite Index by employing back-propagation neural network model. The results demonstrate that neural network can successfully predict daily highest, lowest and closing value of the Shanghai Stock Exchange Composite Index in the short run, however it cannot predict the return rate of the Stock Exchange Composite Index. Patel and Yalamalle (2014) also predicted stock price of companies listed under LIX 15 index of National Stock Exchange by employing a feed-forward MLP technique. The results show that MLP neural network gives satisfactory results based on median normalised error, median correct direction, and median standard deviation.

Dhamija and Bhalla (2011) employed MLP and RBF, which are neural network models, in evaluating their predictive accuracy for forecasting exchange rate and various structures of the models used. It was found that neural network is effective in forecasting exchange rate and RBF outperformed MLP. Erdogan and Goksu (2014), (Dhamija and Bhalla, 2011) considered the predictive accuracy of ANN using historical data of exchange rate for the Euro and Turkish Lira with a normalised back-propagation over the period from 2010 to 2013. The study altered several factors such as number of neurons, transfer functions and learning algorithms in order to achieve higher performance. Based on the results, it was shown that ANN can closely predict Euro/TRY exchange rates.

In a study by Rapoo and Xaba (2017), the forecasting performance of ARIMA and ANN was investigated using ZAR/USD exchange rate series. The study employed MSE and MAE as error metrics to compare the forecasting performance of the models. Based on the results, it was demonstrated that the ARIMA model had a superior performance to that of ANN in forecasting ZAR/USD exchange rate.

Sigo et al. (2018) did a study similar to White (1988). Sigo et al. (2018) in their study, neural network was employed to analyse nonlinear movement patterns of the most volatile top three stocks listed in the Bombay Stock exchange (BSE) in India which are Reliance Industries Limited (Virili and Freisleben), Tata Consultancy Services (TCS) limited and HDFC Bank limited for the period from 2008 to 2017. The results showed that ANN is the better model than time series models.

Barapatre et al. (2018) applied in their study multilayer feed-forward network trained with back-propagation as this model was also applied in the study of Bing et al. (2012) to predict stock market price. The results demonstrated that multilayer feed-forward network trained with back-propagation algorithm predicts the direction of changes of the value of stock better.

Orjuela-Canon et al. (2022) applied in their study three neural network models which are nonlinear autoregressive model, recurrent neural network and radial basis functions for time series forecasting, which were trained with information on reported cases obtained from the national vigilance institution in Columbia. The results, based on mean average percentage error, showed that the model based on traditional method shows better performance as compared to connectionists models, which are known as neural network models, are a class of computational models inspired by the structure and functioning of the human brain.

### 2.3.2 Artificial neural network based evolutionary algorithm hybrid model(s)

A global search technique known as genetic algorithm (GA) was developed by Holland (1975). A GA attempts to find a good (or best) solution to a problem by genetically breeding a population of individuals over a series of generations (Holland, 1975). Genetic algorithm is usually employed to enhance the forecasting performance of artificial intelligence techniques. For instance, in the ANN model, GA is used for the optimal selection of the neural network topology, for selecting the optimal number of hidden layers and processing elements (Kim, 2006). Since one of the shortcomings of ANN is to fall into a local optimum, there has been a suggestion in the literature that there might be a prevention of ANN falling into local optimum given by global search techniques including the genetic algorithm (GA) (Kim, 2006) cited from (Gupta and Sexton, 1999, Sexton et al., 1998, Kim and Han, 2000).

Artificial neural networks (ANNs) have their own gradient search techniques, i.e. back-propagation amongst others. However, there are two significant shortcomings in the back-propagation algorithm, i.e. slowness in convergence and an inability to escape local optima. Therefore, global search engines, such as GA have been proposed to overcome the local convergence problem for nonlinear optimisation problems (Qiu et al., 2016).

In the recent past and of late, there has been quite a lot of research on the application of an artificial neural network-based genetic algorithm hybrid model to time series data. Amongst other researches disciplines; GA and artificial neural network has been applied to structural reliability analysis (Yan-gang and Jin-ren, 1995, Shao and Murotsu, 1999, Cheng, 2007). A neural network model using back-propagation BP) and neural network using genetic algorithm (GA) was studied in the paper of Sexton et al. (1998). It was found that results obtained by GA dominated those found by BP.

The studies by Kim and Han (2000); Kuo et al. (2001) and Kim (2006) employed artificial neural network based genetic algorithm to stock price forecasting successfully. Moreover, Nag and Mitra (2002) and Alvarez-Diaz and Alvarez (2003) applied ANN based GA successfully to foreign exchange rate forecasting. In their study, they found that back-propagation neural network based genetic algorithm outperformed VAR in making ex-post forecast of U.S oil price movement (Mirmirani and Li, 2004). Doganis et al. (2006) successfully applied a hybrid model of RBF neural network and genetic algorithm to sales data of fresh milk provided.

Pacelli et al. (2011), predicted the trend of exchange rate Euro/USD up to three days ahead using optimal topology of multilayer perceptron (MLP) designed by genetic algorithm multi-objective Pareto-Based. The results shows that the developed ANN can forecast the trend of Euro/USD exchange rate to three days ahead.

Natural gas consumption in the city of Yasouj, Iran was predicted by the developed hybrid model of artificial neural network-based genetic algorithm. The results shows that the developed hybrid is good for predicting natural gas consumption as it shows only a deviation of 2.19% and a high correlation (R = 0.998%). The study further recommended that the model be used to forecast natural gas consumption and it can assist authorities in developing policies for natural gas distribution (Karimi and Dastranj, 2014).

In the conducted study, the researchers focused on predicting the closing prices of the Indian stock market. Two neural network models and two neuro-genetic hybrid models and the impact of data pre-processing was examined. Models considered are ANN trained with gradient descended (ANN-GD), genetic algorithm (ANN-GA), FLANN-GD, and FLANN-GA. The results demonstrate that these models are promising to be effective in forecasting Indian Stock market and data pre-processing plays a significant role in the forecasting performance of the models

(Nayak et al., 2014). Moreover, Ebadati and Mortazavi (2018) applied artificial neural network based genetic algorithm to predicting stock price and time series and the results that the applied model reached improved in SSE of 99.99% and time improvement of 90.66%.

Qiu et al. (2016) predicted the return of Japanese Nikkei 225 index by using artificial neural network. Genetic algorithm and simulated annealing were employed to improve the forecasting accuracy of ANN model and to deal with the short coming of local convergence present in back-propagation. It was found that hybrid models based on GA and SA improve forecasting accuracy and outperform the traditional BP training algorithm.

Huang and Hou (2017) tried to establish an air ticket sales revenue prediction model by applying a hybrid model of artificial neural network based genetic algorithm. Genetic algorithm was employed to determine optimal number of input and hidden nodes of feed-forward neural network. Results found that predictive capacity of the model was good and may be used as it will give reliable and highly efficient analysis of the data.

However, it is also noted that in some research, ANN based GA does not give reliable results and is outperformed by other models and evolutionary algorithms. For instance, in the study of Radityo et al. (2017), a comparative analysis of four neural network models; i.e. back-propagation neural network (BPNN), genetic algorithm neural network (GANN), genetic algorithm back-propagation neural network (GABPNN) and neuro-evolution of augmenting topologies (NEAT) in predicting the close value of Bitcoin in the next day was conducted. Based on the results, it was found that BPNN outperformed the other models based on MAPE of 1.998 $\pm 0.038\%$ and training time of 347 $\pm 63$ seconds.

Furthermore, Ghasemiyeh et al. (2017) contributed in the literature by employing metaheuristic algorithms including particle swarm optimisation, genetic algorithm, improved cuckoo search, cuckoo search and improved cuckoo search genetic algorithm in a hybrid system with artificial neural network to predict prices on stock exchange. It was found that particle swarm optimisation is the best algorithm to predict stock price.

Empirical mode function (EMD) was firstly proposed by Huang et al. (1998) as a form of adaptive time series decomposition technique employing the Hilbert-Huang transformation (HHT) for non-linear and non-stationary time series. Furthermore, Chen et al. (2012) and Wei and Chen (2012) defined EMD as an empirical, intuitive, direct and adaptive data processing method developed especially for dealing with non-linear and non-stationary data.

Empirical mode function (EMD) method uses sifting process to decompose an original time series data into several intrinsic mode functions and residue based on the local characteristic time scale (Huang et al., 1998). Therefore, according to the scale, each mode physical feature can be identified. There are several successful applications of EMD on various research fields, for instance, non-stationary sea wave data (Huang et al., 1999), earthquake signal and structural analysis, bridge and construction state monitoring (Vincent et al., 1999), ocean waves (Veltcheva and Soares, 2004), biomedical engineering (Liang et al., 2005, Jiang and Yan, 2008), signal processing (Li and Meng, 2006, Blanco-Velasco et al., 2008), wind engineering (Li and Wu, 2007), earthquake engineering (Yinfeng et al., 2008), and financial time series (Zhang et al., 2008).

Empirical mode decomposition (EMD) has also been used to analyse time series data in a hybrid model with artificial neural networks (ANNs). Research has shown that the model has been successfully applied in the recent past, such as for forecasting tourism demand (Chen et al., 2012), Wei and Chen (2012) employed the hybrid model of EMD-ANN to forecast short-term metro passenger flow and Kisi et al. (2014) used EMD-ANN to predict the following month's stream flows.

A novel forecasting model is proposed in the study based on empirical mode decomposition (EMD) and back-propagation neural network (BPNN) to predict tourism demand (that is, the number of arrivals), and the forecasting performance of the proposed model is compared with that of single back-propagation neural network (BPNN) and traditional ARIMA models. EMD is employed to decompose a time series of tourism demand into a finite set of intrinsic mode functions (IMF) and residue, which has a simpler frequency and components and higher correlations. Based on the experimental results, it was found that the proposed model outperforms both single BPNN and traditional ARIMA model respectively (Chen et al., 2012).

Yahya et al. (2017) contributes to the literature by extending the normal EMD method to a modified EMD method. In this case, a proposed model in their study was based on modified empirical mode decomposition (EMD) and artificial neural network (ANN) model used for forecasting tourism demand. Data used in the study covered monthly tourism arrivals from Singapore and Indonesia in the period from 2000 to 2013. The results showed that the proposed model outperformed both individual ANN and EMD-ANN models.

The study proposed a hybrid model named EMD-ANN for wind speed prediction based on empirical mode decomposition and artificial neural networks and the proposed model was compared to ARIMA model and individual ANN. Based on the results, it was found that there is a high satisfactory forecasting performance achieved by the proposed model. Furthermore, the model is able to deal robustly with the jumping sampling in non-stationary wind series Liu et al. (2012).

Shabri and Samsudin (2014) aimed to improve the accuracy of water demand forecasting by applying the proposed hybrid ensemble learning paradigm integrated ensemble empirical mode decomposition and several single forecasting methods such as autoregressive integrated moving average and artificial neural network. The study used monthly water demand record data from Batu Pahat city in Johor of Peninsular Malaysia. The results demonstrate that the proposed EMD-ANN model yields better forecasting results as compared to ANN, ARIMA and EMD-ARIMA models respectively.

The study developed a model based on EMD and ANN for Baltic Dry Index (BDI) forecasting and the results have shown that the developed model indeed outperforms both the ANN and VAR models in forecasting Baltic Dry Index (BDI). It was thus stated that, the EMD-based method seems to be a useful technique for dry bulk market analysis and forecasting (Zeng et al., 2016). The paper applied a hybrid model of EMD with stochastic time strength neural network (STNN) in an attempt to improve forecasting accuracy of stock price fluctuations. The results reveal that stock market fluctuations are well forecast using the proposed model as the model displays a good forecasting performance.

Abdullah et al. (2019) investigated the effectiveness of integrating EMD into ANN model to forecast paddy prices and compared the model to a simple ANN model. The data applied was that of paddy prices form February 1999 to May 2018. The study found out that the forecasting capabilities are improved when integrating EMD into ANN model. Furthermore, the forecasting error was reduced significantly when EMD integrated with ANN, as RMSE was reduced by 0.012, MAE by 0.0002 and MPE by 0.448.

The applied four intelligent methods for maximum power point tracking (MPPT), which represents one of the significant challenges for designing photovoltaic (PV) systems. The study used both ambient temperature and irradiance as the systems' input, and maximum power was employed as output. The models were assessed on speed, stability and complexity. Based on the RMSE and MAE, it was found that fuzzy systems provide faster, more accurate and more stable performance than the other methods. Based on hybrid models, it was found that ANN-ICA is faster but complicated in implementation than ANN-PSO and ANN-GA (Fathi and Parian, 2021).

The study applied genetic algorithm (GA) and emotional artificial neural network (EANN) for modelling of two watersheds with totally distinct climatic conditions. GA has been implemented to choose the important feature's candidate as EANN input and automatically diagnose the optimal number of hidden nodes and hormones simultaneously. The results show that a proposed hybrid GA-EANN model enhances the better results compared to ANN and EANN models up to 19% and 35% in terms of testing sustainability criteria for Aji Chai and Murrumbidgee catchments, respectively (Molajou et al., 2021).

Sharma et al. (2022) applied ANN-GA hybrid model for forecasting two US stock markets, DOW30 and NASDAQ100 respectively. The study partitioned the series into training, testing and validation sets. The validation of the model was done on the stock data of COVID 19. The results revealed that the performance of the hybrid model was superior to the single ANN model for both DOW30 and NASDAQ100 series in the short and long run.

### 2.3.3 Support vector regression (SVR) model

Support vector machine (SVM) is a novel neural network algorithm developed by Vapnik and his co-workers (Vapnik, 1995). Support vector machine (SVM) applies the principle of structural risk minimisation, unlike most traditional neural network models which implement empirical risk minimisation principle. Structural risk minimisation principle seeks to minimise an upper bound of the generalisation error than minimise the training error (Tay and Cao, 2001, Cao and Tay, 2003).

This induction principle is based on the fact that the generalisation error is bounded by the sum of the training error and a confidence interval term that depends on the Vapnik-Chervonenkis (VC) dimension. This principle enables SVM to achieve an optimal network structure by striking a right balance between the empirical error and the VC-confidence interval. Thus, this affords SVM to have a better generalisation than other neural network models. Training SVM is equivalent to solving a linear constrained quadratic programming. This then means that SVM solution is unique, optimal and absent from local minima, unlike other neural network trainings which require non-linear optimisation thus having a danger of getting stuck in a local minimum.

Originally, support vector machine was developed to solve problems concerned with pattern recognition. Early studies which applied SVM for pattern recognition problems are amongst others (Schiilkop et al., 1995, Joachims, 1998). Thereafter, there was an introduction of Vapnik - insensitive loss function. This enabled SVM to be extended and applied to non-linear regression problems and time series forecasting and it was therefore named support vector regression (SVR). SVM has shown to exhibit excellent forecasting performance (Müller et al., 1997, Mukherjee et al., 1997, Vapnik et al., 1997).

Since its development, support vector machine (SVM) or support vector regression (SVR) has been applied and made quite a successful breakthrough and obtained plausible performance particularly for time series forecasting; such as for forecasting financial time series (Tay and Cao, 2001, Yang et al., 2002, Cao and Tay, 2003); forecasting of electricity price (Sansom et al., 2002) and estimation of power consumption (Chen and Chang, 2004).

Furthermore, Kim (2003) contributed to the literature by predicting stock price index using SVM. Results demonstrated that SVM provides a promising alternative to stock price prediction. Wu et al. (2004) applied SVR for travel time prediction and compared SVR's results with baseline travel time prediction methods. It was found that SVR is applicable and performs well for traffic data analysis. Huang et al. (2005) used SVM for investigating the predictability of financial movement direction by forecasting weekly movement direction of NIKKEI 225 index. It was found that SVM outperformed classification methods. Lai and Liu (2010) compared neural network and support vector regression (SVR) for financial market prediction. Results demonstrated that SVR in the short-term forecasts performs well.

In the study, stock market price and stock market trend were predicted by support vector regression (SVR) using various window functions. The model was applied on Dhaka Stock Exchange (DSE), named ACI group of company limited from 2009 to 2012 time series data. It was found that SVR based rectangular window and SVR based flatten window are good predictors of stock price for 1 day, 5 days and 22 days ahead (Meesad and Rasel, 2013).

Okasha (2014) compared the forecasting performance of SVR, ANN and ARIMA by fitting them to Al-Quds Index of the Palestinian Stock Exchange market time series data and two months future points were forecast. The study employed root mean squared error (RMSE) as a forecasting performance measure on the models used. Results concluded that SVM's results give more accurate results and provide a suitable forecasting technique for financial data than both ANN and ARIMA respectively.

The study contributed to the literature by forecasting RMB exchange rate series by using a build nonlinear combination model of autoregressive fractionally integrated moving average (ARFIMA), support vector regression (SVR) and back-propagation neural network (BPNN). The series used was exchange rate of RMB/USD and RMB/EURO. Results showed that the predictive performance of NCM is better than those of the single models and the LCM. Furthermore, it was concluded that NCM is suitable for predicting a special series such as the RMB exchange rate series (Xie et al., 2015).

The study developed a building energy consumption time series model by applying support vector regression algorithm. Experimental results have shown that the predictive accuracy of SVR algorithm is better than traditional time series analysis model (Liu et al., 2015). Notwithstanding the great success of SVR, there are some instances where other machine learning models outperform SVR, for instance; the study performed a comparative analysis of a new methodology using three machine learning techniques namely, support vector regression, random forest and ANN for investigating the dynamic interactions between four commodity prices and the exchange rate for an emerging economy in Malaysia. Results shows that random forest has a better comparative accuracy and performance than support vector regression and neural networks. Furthermore, it was shown that Malaysian exchange rate is influenced by dynamic parameters of specific commodity prices – crude oil, palm oil, rubber and gold (Ramakrishnan et al., 2017).

Rapoo et al. (2020) analysed the forecasting efficiency of support vector regression, artificial neural network and VAR for in sample forecasting of portfolio inflows. Models were applied on portfolio inflows as dependent variables and exchange rate, inflation linked bonds and GDP as independent variables series covering the period from 1st March 2004 to 1st Feb 2016. It was found that SVR has a better forecasting performance than ANN models respectively. Furthermore, in determining factors that affect allocation of PIs into South Africa based on SVAR, 69% of the variation was explained by pull factors while 9% was explained by push factors.

### 2.3.4 Support vector regression based evolutionary algorithm(s) hybrid model(s)

Support vector regression (SVR) has demonstrated its successful application to time series forecasting. Although this is the case, SVR model forecasting performance depends largely on the right selection of its hyper-parameters' values. Lin (2001) and (Duan et al., 2003) state that in order to construct the SVR model efficiently, SVR's parameters must be set carefully. Various evolutionary algorithms have been developed in the recent past to aid in selection of artificial intelligent models or machine learning models' hyper-parameters.

Stability of any support vector regression (SVR) model and its accuracy strongly depends on the kernel, its parameters and time penalisation term (Wu et al., 2009). A highly effective and representative SVR model can only be built after these parameters are carefully selected. Therefore, genetic algorithm has been employed successfully in choosing the optimal values of the hyper-parameters of support vector regression (SVR) model. That is because, genetic algorithms are able to search for global optimum values in a non-linear search space without prior information and with a modest computational cost.

There are several successful applications of genetic algorithm (GA) to various types of optimisation problems in recent years, for instance; (Golberg, 1989, Fogel, 1994, Cao and Wu, 1999, McCall and Petrovski, 1999, Alba and Dorronsoro, 2005, Aurnhammer and Tonnies, 2005, Venkatraman and Yen, 2005, McCall, 2005, Hokey et al., 2006). Furthermore, successful application of GA-SVR hybrid model has been reported in the recent past, for example; Cheng (2007) examined the feasibility of GA-SVR model in predicting system reliability by comparing it with neural network model and ARIMA model. Results showed that the model outperforms NN and ARIMA models based on normalised RMSE and MAPE. Chen and Wang (2007) examined the feasibility of applying GA-SVR model to forecasting tourism demand by comparing it with back-propagation NN and ARIMA. It was found that GA-SVR model has a higher forecasting accuracy than BPNN and ARIMA. The study forecast Taiwan Stock Exchange Market weighted index (TAIEX) by applying GA-SVR model and comparing the model to ANN and random walk model (Chen and Ho, 2005). Results demonstrate the superiority of GA-SVR model over ANN and RW models respectively. Wu et al. (2009) found that the proposed model of hybrid GA-SVR outperformed previous models in forecasting electricity load.

Liu et al. (2013) proposed a prediction model based on support vector regression (SVR) model which can solve the problem of aqua water quality. Furthermore, for the search of SVR parameters, the study used genetic algorithm technique. The data used was that of aquaculture water quality collected from YiXing in China. The results show that both the traditional SVR and neural networks models are outperformed by the random genetic algorithm-based support vector regression (RGA-SVR) based on the accuracy measurements of root mean square error and mean percentage square error.

Based on selecting indicators method for stock price forecasting, the study proposed a novel GA-SVR model. Furthermore, multivariate adaptive regression splines and stepwise regression were employed to select important indicators and GA was used to optimise the parameters of the forecasting model. The study used stock price data of Chunghwa Telecom from 2003 to 2012 and RMSE was an evaluation criterion employed by the study. Results showed that SR-GASVR has a better forecasting performance in both the short and long run periods and performs better than other models. Furthermore, MARS-GASVR has a good forecasting performance of smooth fluctuations of stock price irrespective of the short or long run periods (Cheng and Shiu, 2014).

Santamaría-Bonfil et al. (2015) proposed a new hybrid support vector regression based genetic algorithm method for forecasting financial volatility and compared the model with GARCH model. Two ASEAN and two Latin-American market indexes were employed as time series data. MAPE and directional accuracy are used as evaluation criterions. It was demonstrated by the results that SVR based genetic algorithm outperformed both GARCH (1, 1) model and SVR based grid search model.

Jiang and Wu (2016) employed a hybrid technique of genetic algorithm and support vector regression to forecast CNY exchange rates series using back propagation neural network as a benchmark model. Various functions are examined and presented which are linear, radial basis, polynomial and sigmoid. Based on the results, it is shown that the hybrid model in studying CNY exchange rates series is effective.

The damping ratio of cantilever beam with particle damper is predicted by an optimised support vector regression. The study compared a cross validation combined with support vector regression and genetic algorithm-support vector regression. The optimised parameters of the models were adopted to construct a new model. Based on the experimental results, it shows that genetic algorithm-support vector regression is better in predicting the ratio of the damping cantilever beam with particle damper (Xia et al., 2017).

The study optimised a nonlinear support vector regression model for its application to stock index prediction. The optimisation parameters were selected by grid search (Lee et al.), particle swarm optimization (PSO) and genetic algorithm (GA). The results demonstrate that the methods were able to show that the time dependent trend of the stock index and the prediction accuracy of the models were significant (Chen et al., 2017).

Fu et al. (2019) forecasted four typical RMB exchange rates (CNY against USD, EUR, JPY and GBP) by employing the developed two evolutionary support vector regression models (PSO & GA) and in assessing the out-of-sample performance of the developed models; they employed four evaluation criteria (MAE, RMSE, MAPE and DS). Evolutionary algorithms were employed to optimise parameters of SVR model by balancing the search between the global and local optima. Results demonstrated that the proposed evolutionary support vector regression model outperformed all other benchmark models based on the evaluation criteria used. It shows that for RMB exchange rate forecasting, the proposed model is the best model to employ.

Stationary time series were decomposed as time went on by the widely utilised Fourier analysis and associated spectrum analysis created by (Wiener, 1949). However, the principles of these methods involve analysing time series in the frequency domain, which leads to loss of important information regarding instantaneous frequency (Boashash, 1992). In contrast, EMD is based on the theory of local scale separation and does not need a predetermined basis functions, which is totally different from wavelet analysis and almost all the other decomposition techniques (Meng et al., 2019) cited from (Lee and Ouarda, 2010, Kim and Oh, 2013).

Empirical mode decomposition (EMD) is employed as a pre-processing method of the original time series to improve prediction accuracy. Furthermore, EMD exhibits a strong generality in dealing with non-stationary data. Support vector regression (SVR) has shown to be successful in predicting time series data over the years. Even though SVR is an effective prediction method, non-stationary time series have a great impact on its prediction accuracy (Fan et al., 2016).

Empirical mode decomposition-based support vector regression hybrid models have been applied in the recent past for analysing and modelling time series data. For instance, Zhu et al. (2007) employed a hybrid model based on empirical mode decomposition and support vector regression for electric power load forecasting. Results demonstrated that the employed hybrid model has a higher precision and greater generalisation ability. The study predicted FMS frequency spectrum series using a novel prediction algorithm of EMD-SVR compared to AR model and a traditional SVR model. Results showed that the EMD-SVR algorithm outperformed both AR model and traditional SVR model on predicting FMS frequency spectrum series (Yu et al., 2008a). Fu (2010) attempted to propose an exchange rate ensemble learning paradigm called EMD-SVR. It was found that the proposed model has a superior forecasting accuracy than the normal SVR model. Fan and Tang (2013) built a model for non-stationary time series prediction using empirical mode decomposition and support vector regression.

Cheng and Wei (2014) forecast stock price for Taiwan stock exchange capitalisation weighted stock index using a proposed hybrid time series model of support vector regression based empirical mode decomposition while comparing the model with autoregressive (AR) and SVR model. Results based on root mean squared error (RMSE) showed that the proposed model has a superior forecasting ability than the other two listed models.

The study compared the developed hybrid model of empirical mode decomposition (EMD) support vector regression (SVR) with AR model, single SVR model, and EMD AR model for non-linear and non-stationary wave prediction. They furthermore compared Wavelet based SVR and EMD based SVR to investigate the performance of WV and EMD techniques (Duan et al., 2016). Significant wave height data obtained from National Data Buoy Centre was used as experiment in the study. Results found out that EMD-SVR shows a good model performance and it can be an alternative model for short term prediction of nonlinear and non-stationary waves.

Yaslan and Bican (2017) employed a hybrid model incorporating empirical mode decomposition (EMD) and support vector regression (SVR) algorithm for predicting electricity load demand. The hybrid model was compared with SVR model and EMD was employed as a denoising step on the training data. Three electricity load datasets of various countries were employed in the study. The results demonstrated that the proposed model outperforms the SVR model and non-feature employed denoised-SVR on forecasting electricity load.

Duan et al. (2015) employed a hybrid model of autoregressive (AR) based EMD-SVR (AR-EMD-SVR) for forecasting short-term of non-linear and non-stationary ship motion. The employed models' forecasting performance was compared with those of AR model, non-linear SVR and a hybrid EMD-SVR models respectively. Results shows that the difficulty of non-stationarity to SVR was overcome by AR-EMD technique, moreover, AR-EMD-SVR has obtained better predictions than the models employed to forecast non-linear and non-stationary ship motion.

Meng et al. (2019) contribute to the literature by proposing a modified EMD based SVR model for improving monthly stream-flow prediction accuracy. Furthermore, for quantifying the prediction performance of the proposed model, the model was compared to the forecasting performance of ANN, SVM, WA-SVM, EMD-SVM and performance measurements used were RMSE, MAE, MAPE and Nash-Sutcliffe efficiency coefficient. Monthly stream-flow series with various non-stationary levels were employed as data set. The findings of the study generally showed that the proposed model (Nie et al., 2020) depicted more accurate prediction of non-stationary stream-flow data than all the models employed in the study.

Nie et al. (2020) extend the study of Duan et al. (2015) by offering the hybrid model of mirror symmetry EMD-SVR (MSEMD-SVR) for short term prediction of ship motion employing mirror symmetry and SVR algorithm to eliminate EMD boundary effect. Results demonstrate that the offered model of MSEMD-SVR is feasible and more reliable than EMD-SVR for predicting short term of ship motion. This is because EMD-SVR does not deal with the boundary effect, it only deals with it by applying mirror symmetry.

Jiang and Liang (2021) applied a hybrid model of support vector regression based genetic algorithm (SVR-GA) for shear strength capacity of medium- to ultra-high strength concrete beams with longitudinal reinforcement and vertical stirrups prediction. The proposed model was compared to traditional SVR, multiple linear regression model, and ACI-318. Based on coefficient of determination, root mean squared error and mean absolute error, it was established that the proposed model performed better than the other models in the study.

Xu et al. (2021) employed a hybrid AI model of genetic algorithm-based support vector regression for forecasting electric energy consumption for crude oil pipelines. The data used was of three crude oil pipelines in China. The results show the proposed genetic algorithm-based support vector regression to be the best model as it improves the prediction accuracy compared to the state-of-the-art models.

### 2.3.5 Least square support vector regression (LSSVR) model

Least square support vector machine for regression (LSSVR) is another version of support vector machine (SVM). Least square support vector regression (LSSVR) is normally and widely employed in complex system studies for modelling, regression and parameter prediction. LSSVR uses just least square loss function to obtain a set of linear equations in contrast to SVR algorithm that requires Epsilon insensitive loss function to obtain convex quadratic programming in the feature space (Suykens, 2000), so that the learning rate is faster and the complexity of calculation in convex programming in SVR is relaxed.

Moreover, LSSVR avoids shortcomings faced by SVR which has three trade-off parameters selection, rather LSSVR only has two parameters while training the model. Since LSSVR possesses all these great qualities, the model has been applied successfully to regression problems, classification, and time series forecasting. For instance, Van Gestel et al. (2001) applied Bayesian evidence framework to LSSVR for predicting time varying volatility of DAX 30 index. Based on MSE and MAE it was found that a better forecasting performance was obtained by the proposed model than by GARCH (1, 1) and AR (10) models respectively.

Mei-Ying and Xiao-Dong (2004) and Espinoza et al. (2005) applied least square SVR for chaotic time series prediction. Baylar et al. (2009) forecast entrainment rate and aeration efficiency of weirs using least square SVR compared with multi nonlinear and linear regression model. Results reveal that LSSVR can be used successfully in predicting air entrainment rate and aeration efficiency weirs. Three ASEAN stock markets' financial volatility were forecasted by comparing the performance of semi-parametric model LSSVM, with classical GARCH family models Ou and Wang (2010). Their study found that employing hybrid models improves forecasting performance on the leverage effect volatility.

The study explored the potential of employing least squares SVR for streamflow prediction using hydroclimatic inputs. The input variables used in the model were rainfall, maximum temperature, minimum temperature and streamflow. It is shown based on correlation coefficient, root mean square error and Nash-Sutcliffe that LSSVR is the best model to be applied (Bhagwat and Maity, 2013).

Ahmadi et al. (2016) predicted output power, shaft torque and brake specific fuel consumption using least square support vector regression. Results demonstrated that LSSVR has a higher accuracy for predicting output power, shaft torque and brake specific fuel consumption. Shaghaghi et al. (2019) aimed at gaining new insight and develop methods which are accurate for assessment of longitudinal slope (S), water surface width (W) and mean water depth (D) of rivers in regime state by employing and comparing the performance of M5Tree, MARS and LSSVR models respectively. According to the values of correlation coefficient and mean absolute relative error, it was found that M5Tree model has outperformed the other models in the study.

The study added to the existing literature by conducting a comparative analysis of predicting the water quality index using support vector regression and least square SVR. Moreover, the study used different functions to determine the best model to be employed for SVR. Results obtained showed that the best function to employ for SVR is a polynomial kernel function. Moreover, on the overall analysis, it was demonstrated that LSSVR model has a superior accuracy compared to SVR (Leong et al., 2021).

### 2.3.6 Least square support vector regression based evolutionary algorithm(s) hybrid model(s)

Though it has shown and still showing a great deal of successful application to time series forecasting, and can well fit intrinsic nonlinear relation between historical data and future data, the optimal selection of least square support vector regression (LSSVR) parameters plays a vital role in its forecasting accuracy. In the recent past, various evolutionary algorithms have been employed to find optimal parameters of the LSSVR model such as the following evolutionary algorithm, ant colony optimisation algorithm (Hai-yan and Ze-jun, 2013); fruit fly optimisation algorithm (Cong et al., 2016); artificial bee colony algorithm (Sulaiman et al., 2012) and particle swarm optimisation (Liao et al., 2011, Kennedy, 2011) amongst others.

These algorithms have only one objective; to optimize parameters of machine learning models by various ways of intelligence. However, PSO has advantages over the other algorithms which are fast searching speed, simple structure, easy implementation and the property of memory. Therefore, it is more often that PSO is applied. However, there are some disadvantages with PSO; it is easy to cause particle aggregation and fall into local optimal. These shortcomings are addressed by genetic algorithm (GA). There are various accurate applications of GA; for instance, Goodarzi et al. (2010) applied GA based LSSVR for pKa modelling and prediction of pH indicators series; Wang and Wang (2012) found out that GA based LSSVR outperformed traditional LSSVR and traditional SVR for material demand forecasting. GA was also applied by Zhang et al. (2016b) to forecast shop flow and Tian et al. (2019) improved GA by including quantum algorithm to select the route of urban traffic management.

The application of LSSVR based GA to tourism demand forecasting was extended by including fuzzy e-means with logarithmic LSSVR and GA. The data employed in the study was tourism arrivals to Taiwan and Hong Kong. Results shows that higher accuracy of forecasting tourism demand is achieved by the developed model as it has a higher forecasting performance than the other models (Pai et al., 2014).

Yu et al. (2016) modelled crude oil by employing a hybrid paradigm incorporating LSSVR with grid search method and genetic algorithm. The model was applied to West Texas Intermediate and the Brent markets time series data. It was found that the proposed grid-GA based LSSVR model has outperformed other benchmark models (for example, LSSVR) in terms of accuracy and computational efficiency. The models also seem to be a very effective model for volatility and irregularity.

Particle swarm optimisation (PSO) was incorporated with genetic algorithm (GA) into LSSVR for predicting short-term traffic flow. The proposed algorithms (PSO and GA) were used to search for optimal parameters of LSSVR model. It was evident, based on the results, that the proposed model yielded more accurate results than other models and showed computational efficiency (Luo et al., 2019).

The study added to the existing literature by using LSSVR based GA model to train and forecast future gold prices. Furthermore, opinion score was used as an input predictor for the model. Results demonstrate that predictive accuracy of the model is improved through incorporating input predictor and score opinion in the model as shown by MAPE (Yuan et al., 2020).

As literature has demonstrated, least square SVM for regression has been successfully applied to various fields of research. Nevertheless, often time series data are non-linear and non-stationary. If these issues are ignored in the modelling and forecasting process, they will lead to error in forecasting.

Empirical mode decomposition (EMD) has shown to be the successful method to handle non-linear and non-stationary in time series data. This shows that EMD offers solutions to non-linearity and non-stationary issues, as EMD is a time frequency resolution approach which gives a suitable way in which non-linear and non-stationary time series can be decomposed into a series made of valuable independent time resolutions. Furthermore, EMD also reveals the hidden patterns and trends of time series applications. Various scholars have employed hybrid model EMD coupled with LSSVM for regression (LSSVR) for time series forecasting.

Lin et al. (2012) forecast foreign exchange rate by the proposed hybrid model of empirical mode decomposition based LSSVR. Foreign exchange rate series is decomposed into several intrinsic mode functions and a residue by the EMD algorithm. It was found that the proposed model has a higher forecasting accuracy as compared to EMD based ARIMA, single LSSVR and ARIMA.

Ismail et al. (2015) investigate the ability of EMD based LSSVM in terms of improving the forecasting accuracy of river flow. The model was compared with a standard LSSVM model. The model was divided into three segments which are decomposition, component identification and forecasting stages. Mean monthly river flow dataset of Bernam River from January 1996 to December 2008 was used as the time series data in the study. It was found, based on MAE, RMSE, correlation coefficient and correlation of efficiency, that EMD based LSSVM model has a higher accuracy for river flow forecasting than the standard LSSVM model.

The study proposed a hybrid model made of EMD and LSSVM for forecasting water demand. The proposed model was compared with standard LSSVM, ANN and EMD-ANN. EMD was employed in the study for decomposing time series into IMD and a residue. Monthly water demand from Batu Pahat city in Johor of Peninsular, Malaysia was employed as a dataset. Proposed models have shown that it does outperform single LSSVM, ANN and EMD-ANN for water demand forecasting. It further shows that it can be an alternative model for forecasting water demand (Shabri et al., 2015).

The study extended the application of the hybrid model of EMD based LSSVM by including optimisation algorithm and employed EMD based LSSVM as a proposed model for analysing CSI 300 index and compared the model with WD-LSSVM as the benchmark model. Furthermore, simplex, grid search, PSO and GA were used to select optimal parameters of the model. Results show that EMD-LSSVM with grid search (GS) outperformed other models in analysing CSI 300 index series (Chai et al., 2015).

The study enriched the body of knowledge by incorporating a hybrid model of maximal wavelet decomposition, non-dominated sorting genetic algorithm II and least squares support vector machine with three kernels for predicting short term wind power. The proposed hybrid model was compared with empirical mode decomposition-least squares support vector machine and wavelet-decomposition-least squares support vector machine models. The findings of the study demonstrated that the proposed model has a much better performance than the benchmark models (Ding et al., 2021).

Zhu et al. (2022) forecasted asset prices by a proposed model of a new multi-objective least squares support vector machines with mixture kernels. Furthermore, particle swarm optimisation was employed to synchronously search the optimal model selection of mixture kernels of the least squares support vector machine. The results showed that the proposed model can achieve higher forecasting accuracy and higher trading performance as compared to popular models.

The study sought to establish load forecasting model by employing a proposed short-term power load forecasting technique based on AI algorithm known as IGA-LSSVM. The input data used in the prediction model are temperature, load, weather state, working and holiday days and the output was load value. The data used was sourced form a city in Yunnan province. The study found out that the proposed model was appropriate for short term power load prediction as compared to others (Bao-De et al., 2021).

The study incorporating the multi-objective sine cosine optimisation algorithm with least squares support vector machine for forecasting electricity demand. The findings of the study showed that the proposed model ensures high accuracy and strong stability in the case of Australia (Li et al., 2021).

### 2.3.7 Neural network autoregressive with exogenous inputs (NNARX) model

Neural network models are classified into dynamic networks and static networks (Karbasi et al., 2009, Nor and Gharleghi, 2011). Neural network autoregressive with exogenous inputs (NNARX) is a recurrent dynamic neural network model. In dynamic network inputs, previous inputs, outputs and the state of the network plays a significant role in the output of the model.

Dynamic neural networks are generally more powerful than static networks except for somewhat being more difficult to train. Dynamic networks can be trained to learn sequential or time varying patterns because they have a memory (Racine, 2001). There are various applications of NNARX model to time series forecasting; for instance, Frosini and Petrecca (2000) reported that NNARX model gives better results than a linear ARX model in the short-term prediction of the thermal energy consumption of a hospital. Karbasi et al. (2009) conducted a comparative study for forecasting (1, 2 and 4 weeks ahead) poultry retail prices using NNARX as a dynamic nonlinear NN model, ANN as a static nonlinear NN model and ARIMA as a linear model. The study found that NNARX model and ANN have outperformed ARIMA model, and overall NNARX model outperformed ANN model in all three forecasting horizons. A comparative study using NNARX, ANN, GARCH and ARIMA models was undertaken for forecasting Singapore dollar/ US dollar. Results demonstrated that NNARX model has a superior forecasting performance than the other models (Nor and Gharleghi, 2011).

Georgescu and Dinucă (2011) presumed that global interaction effects and extraneous influences from mature markets tremendously affect Bucharest Stock Exchange (BSE). In order to test these assumptions, the study employed NNAR to capture only the endogenous dynamics of BSE index and NNARX was used to capture exogenous influences made by global market indices. Results demonstrate that more accurate findings are found when NNARX model is used.

Abounoori et al. (2013) combined the long memory feature and dynamic neural network model. Furthermore, fractal markets hypothesis was investigated in daily data of the Tehran Stock Exchange (TSE). Results confirmed that fractal market hypothesis was confirmed for TSE series, meaning that fractal structure is present in TSE series. Moreover, it was found that dynamic NN model outperforms ARFIMA model and the recommendation of using a hybrid model to improve forecasting performance was made.

Ruslan and Adnan (2016) employed NNARX model structure to propose a 3-hour flood water level prediction. The series used was obtained from Department of Irrigation and Drainage of Malaysia. The study successfully reported that NNARX model has managed predict a 3 hours ahead flood water level.

Noor et al. (2017) stated that 24 hours (one day ahead) of river flow can be modelled successfully using NNARX model. Furthermore, the study stated that Bayesian regularisation backpropagation produces more accurate results than Levenberg-Marquardt (LM). The study examined volatility and forecasted foreign exchange rate using a comparative approach between GARCH, ARIMA, NNARX and KPCA-SVR. Furthermore, NNARX model applied different training functions to examine the best model to be used, Levenberg-Marquardt (LM), Bayesian regularisation and scaled conjugate gradient (SCG). Daily nominal spot rate data from 2014 to 2018 for Euro/MUR, GBP/MUR, CAD/MUR, and AUD/MUR were employed as series. Results found that for volatility clustering and prediction, GARCH model outperformed ARIMA model. Moreover, it was found that for exchange rate forecasting, LM and Bayesian regularisation are the best training functions to be employed. Overall, it was found that NNARX model is the best model as it outperformed all the models included in the study (Amelot et al., 2020).

As default risk is deemed to be one of the most significant risks, credit default swap (CDS) is used as a financial instrument to cover this risk. Therefore, Beytollahi and Zeinali (2020) employed Merton model, ANFIS model, NNARX model, AdaBoost model and SVM regression in a comparative study to forecast the price of CDS contracts. Data employed was that of 125 companies which were sampled from 2008 to 2015. The study reported that the NNARX average predictive power was higher than that of the model used in the study.

Kajan et al. (2022), in their comparative study, employed NNARX, LSTM and gated recurrent unit (GRU) for modelling dynamic changes in the microgrid. The analysis of the models was carried out in MATLAB. The results show that in the short term, in the open loop, the NNARX model is the best model, however, in the close loop, LSTM is the best model.

The study employed in a comparative study autoregressive neural network with grey-box and black-box linear models to simulate indoor temperature. The actual experimental data was obtained from a building in Montreal, Canada. The findings of the study shows that the neural network models to be the best model in simulation of indoor temperature as compared to grey-box and black-box models (Delcroix et al., 2021).

### 2.3.8 Neural network autoregressive with exogenous inputs based evolutionary algorithm(s) hybrid model(s)

Literature has shown a successful application of NNARX model to time series data. As it is the case with neural network models, NNARX also needs to be optimised in order for the model to have a superior forecasting accuracy. Han et al. (2019) state that in neural network modelling, it is crucial to decide the best structure of a neural network model. Input variables, numbers of hidden layers and neurons, and the internal structure between the input layer and the output layer specifies neural network models. To optimally select appropriate parameters of the NNARX model, an optimisation or evolutionary algorithm should be employed to select those optimal hyper-parameters.

Genetic algorithm is one evolutionary algorithm to aid in achieving an optimised NNARX model. Genetic algorithm has been successfully employed in NNARX modelling; for instance, in minimal structure detection (Mao and Billings, 1997), simulation model order and parameter estimation (Chen et al., 2007). Moreover, Ghosh and Maka (2011) employed a hybrid model of NARX neural network with genetic algorithm for deriving an index of insulin sensitivity. The study found that there was a higher accuracy fit reported by the proposed hybrid model.

Jawad et al. (2018) compared a proposed hybrid model of genetic algorithm based non-linear auto-regressive with exogenous inputs neural network (GA-NARX-NN) with five traditional techniques for forecasting electric load and wind speed. The study employed mean absolute percentage error, root mean square error and variance to validate the forecasting performance of the models. The results demonstrate that the proposed hybrid model is more accurate and provides more reliable results than the five traditional models included in the study.

The study compared a hybrid model of genetic algorithm based NNARX model with feed-forward neural network model for suggesting daily bitcoin return model. Furthermore, the methods used to select NNARX architecture were also compared which includes GA, AIC and SIC. The results show that GA is the best method to optimise NNARX model and the proposed hybrid model shows higher accuracy for daily bitcoin return than feed-forward neural network model (Han et al., 2019).

In most cases, time series data is non-stationary and non-linear. Therefore, for appropriate modelling of the series, a more suitable forecasting model(s) should be applied to model such time series data. Empirical mode decomposition (EMD) method is much suitable to take care of the property of non-stationarity in the series. Furthermore, Jun et al. (2018) stated that EMD based models have been recently successfully applied in various fields of research such as for wind speed assessment, electricity load, crude oil pricing, energy consumption, foreign exchange rate, and tourism arrivals and they have depicted a very good forecasting performance for both non-linear and non-stationary time series.

Zhang et al. (2016c) added to the body of knowledge by proposing a hybrid model of EMD-NARX suitable for non-stationary exchange rate series and for its application for exchange rate forecasting. Five minutes and daily US dollar against Japanese yen were employed as an exchange rate series. Results reveal that when the time interval is shorter, the forecasting precision of the model is higher. Furthermore, it was shown that the reform of exchange rate does not affect the EMD-NARX as its stability is still satisfactory.

The study extended the application of EMD by introducing a weighted recombination model. In the study, Jun et al. (2018) proposed a weighted recombination model for one-step ahead forward prediction of non-linear and non-stationary time series based on NARX and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN). The proposed model was compared to 12 commonly employed non-decomposition and decomposition models which were applied as benchmark models. The study applied two non-linear, non-stationary time series data to validate the performance of the models. The results showed that as compared to the 12 models, the proposed model noticeably improved forecasting accuracy.

The study proposed a prediction model for foundation pit deformation based on a hybrid model of NARX based empirical wavelet transformation (EWT). The forecasting ability of the proposed model was compared with the hybrid models of NAR, EMD-NAR, EWT-NAR, NARX, and EMD-NARX. According to the results, the proposed model had a superior forecasting performance than the other models included in the study. Moreover, the results demonstrated that EWT is the best method as compared to EMD in decomposing time series data (KD25) (Ma et al., 2020).

The study adopted EMD as a data augmentation technique. The method aids to obtain a surrogate data series after the resulting intrinsic mode functions were recombined with various weights, the data was used to train NARX model for prediction. M4 data series was employed for validating the proposed model. Results show forecasting accuracy improvement when using EMD as an augmentation technique over baseline method (Abayomi-Alli et al., 2020).

Medi and Asadbeigi (2021) employed genetic algorithm based neural network (NNARX) controller for modelling chemical and biochemical processes. Furthermore, the proposed controller was compared to the conventional proportional integral (PI) controller. The results indicated that the GA based NNARX controller had better performance as compared to the conventional proportional integral (PI) controller.

### 2.3.9 Adaptive neuro-fuzzy inference systems (ANFIS) model

Adaptive neuro-fuzzy inference system (ANFIS) was first developed by (Jang, 1993) and the model has been applied to various research problems. ANFIS is made up of integrating fuzzy logic and neural networks known as neuro-fuzzy, and it is another powerful technique for modelling nonlinear systems. In neuro fuzzy approach, the capability of fuzzy ruled based systems in handling uncertain and noisy data and the learning capability of neural networks is combined to form better estimators (Talebizadeh and Moridnejad, 2011). ANFIS is a particular form of neuro fuzzy system that has shown significant results in modelling non-linear functions (Jang et al., 1997).

ANFIS model has been successfully applied in the recent past in various fields of research, for instance; in hydrological time series modelling (Chang and Chang, 2001, Bazartseren et al., 2003, Nayak et al., 2005), non-linear function modelling (Jang, 1991), time series prediction (Jang and Sun, 1993) and parameter identification for control system (Masugi and Takuma, 2007).

Sfetsos (2000) applied ANFIS model in comparison with ARIMA, feed forward model, recurrent neural networks and neural logic networks for forecasting mean hourly wind speed data. Keskin et al. (2006) found that extension of input and output data in the training phase plays a significant role in improving the forecasting accuracy of ANFIS model. In the study, internet traffic series was forecasted by applying ANFIS model (Chabaa et al., 2009). Wang et al. (2009) employed ARIMA, ANN, ANFIS, genetic programming (GP), SVM for examining their forecasting performance using the long term observations of monthly river flow discharge. Talebizadeh and Moridnejad (2011) compared ANN model with ANFIS model for lake level fluctuations in Lake Urmia in northwest of Iran forecasting. It was found that the ANFIS has the highest forecasting accuracy than ANN model.

The study forecasts weather of Goztepe, Istanbul, Turkey by applying in a comparative study ARIMA model and ANFIS model respectively. Daily average temperature, air pressure and wind speed covering the period from 2000 to 2008 was employed as a time series data. Based on MAE, RMSE - (root mean squared error) it was found that ANFIS model yields better results (Tektaş, 2010). In contrast to the results of the study conducted by Tektaş (2010), Rahman et al. (2013) forecasted the weather conditions in Dhaka, Bangladesh by comparing ARIMA model and ANFIS model respectively. Maximum temperature, minimum temperature, humidity and air pressure covering the period of ten years from 2000 to 2009 were used in the study as a time series data. Results showed that based on MAE, RMSE, r-squared and SSE, ARIMA model had a superior forecasting performance compared to ANFIS model.

Abounoori et al. (2012) conducted a comparative study between static and dynamic neural network models for forecasting the return of Tehran Stock Exchange (TSE) and with the aim of finding the best model for forecasting the series. The study employed daily series from 25 March 2009 to 22 October 2011. ANFIS model showed that it had the superior forecasting performance compared to other models based on mean square error and root mean square error.

Attendance rates at soccer games were forecast by developing and applying NN approach and ANFIS model in a comparative study. Based on mean absolute deviation and mean absolute percentage error, it was demonstrated that NN approach outperformed ANFIS model (Şahin and Erol, 2017).

Huang et al. (2019) added to the body of knowledge by undertaking a comparative study in investigating and comparing feed-forward NN and ANFIS on stock prediction employing fundamental financial ratios. The experimental results revealed that both architectures possess great abilities in separating winners and losers from a sample universe in stocks, and the benchmark were outperformed by selected portfolios. Moreover, the study had an argument that ANFIS model was outperformed by FFNN.

The study used ANFIS, ANN and RMS models in a comparative study for the purpose of modelling and optimising of the substrate treatment process in biogas energy production from yam peel substrates via anaerobic digestion. The results showed ANFIS model to be marginally better in simulating and modelling the anaerobic digestion (Onu et al., 2022).

Ani and Agu (2022) employed both ANFIS and ANN for optimising the experimental response of total petroleum hydrocarbon (TPH) contaminated soil. The study used RMSE, MSE and MAE for evaluating model performance and for comparison purposes. The results demonstrated that ANFIS model had a better model performance as compared to ANN model. Furthermore, it shows that ANFIS prediction capacity was considerably better than the ANN model.

### 2.3.10 Adaptive neuro-fuzzy inference system based evolutionary algorithm(s) hybrid model

Wei (2013) enhanced the existing knowledge by forecasting the Taiwan stock index employing a hybrid model of GA based ANFIS and using fluctuations of other national stock markets as forecasting factors. The proposed model was compared with Chen's model, Yu's model, Huarng's model and the ANFIS model. Based on the results, it was evident that the proposed model outperformed all the listing models based on root mean squared error (RMSE).

The study aimed at developing a robust model (FFNN and GA based ANFIS) for short-term prediction of Photovoltaics (PV) generation. In order to investigate fully the influence of exogenous variables on the PV predicted time series, various combination of inputs is investigated. Results successfully show that the developed model can be implemented in the decision-making process of retailers, distribution system operators, and prosumers and others, in order to fully exploit the generation capacity of grid connected PV systems in day-ahead electricity markets. On the inputs, it is demonstrated that lower prediction errors can be achieved when the inputs combination that involves the panels temperature and historical PV power values are used (Panapakidis and Christoforidis, 2017).

The study predicted day ahead natural gas demand with the aim of testing the robustness of the novel hybrid computational intelligence model of wavelet transform, genetic algorithm, adaptive neuro-fuzzy inference system and feed-forward neural network. Wavelet transform is employed as a method to decompose the original series and the genetic algorithm is used to optimise ANFIS model. Then ANFIS output is used as an input to FFNN model so that it can refine the forecasts of the ANFIS model and increase overall forecasting accuracy. The study concluded that the proposed model is characterised by high flexibility, comprehensive operation and low requirements for computational resources (Panapakidis and Dagoumas, 2017).

Khosravi et al. (2018b) developed machine learning models in a comparative study which are MLFFNN, SVR, FIS, ANFIS, GMDH, ANFIS based PSO and ANFIS based GA for time series wind speed data prediction. Wind speed data considered is in 5-min, 10-min, 15-min and 30-min intervals which the developed models will be validated on. Results illustrated that GMDH has a higher forecasting accuracy for all the time intervals. Furthermore, it was shown that PSO and GA in combination with ANFIS surely increases the prediction accuracy of the ANFIS model for all the time intervals.

The study forecast aggregate and long-term energy demand using a proposed novel ensemble methodology combining an auto regressive integrated moving average, artificial neural network, fuzzy inference system model, adaptive neuro-fuzzy inference system, support vector regression, extreme learning machine, and genetic algorithm. The proposed model is compared with several benchmark models by loss function mean squared error and mean absolute percentage error. It is found that the proposed model improves forecasting accuracy, as mean squared error and mean absolute percentage error decreased by 22.3% and 33.1% respectively. It was also found that hybrid models improve accuracy, and the study stated that improving forecasting techniques is extremely significant in areas where there are upper bound constraints on supply and lower bound on minimum operation levels especially (Prado et al., 2020).

Okolobah and Ismail (2013) enhanced the existing knowledge by forecasting peak load using a hybrid model of empirical mode decomposing (EMD based adaptive neuro-fuzzy inference system (ANFIS). The proposed model was compared with ANN and EMD-ANN models respectively. EMD is employed to decompose the series into several intrinsic mode functions and a residue. Time series data obtained from Power Holding Company based in Nigeria was employed to evaluate the forecasting performance of the models. It was found that the proposed model EMD-ANFIS yielded better results when compared to ANN and EMD-ANN. There was a recorded improvement of 2.76% and 50.05% by EMD-ANFIS model over ANN and EMD-ANN model.

Wei (2015) forecasts stock price for Taiwan stock exchange capitalisation weighted stock index (TAIEX) using a hybrid model of adaptive network-based fuzzy inference systems (ANFIS) based empirical mode decomposition. The proposed model was compared with autoregressive (AR) model, ANFIS model, and SVR model. Based on the root mean squared error, it was found that the proposed model is the best model for forecasting stock price for Taiwan stock exchange capitalisation weighted stock index.

Stock prices in the Taiwan Stock Exchange Capitalisation Weighted Stock Index (TAIEX) and Hang Seng Stock Index (HIS) were forecast using a hybrid time series model of adaptive network-based fuzzy inference system (ANFIS) coupled with empirical model decomposition. The model was compared for its forecasting validation against Chen's model, Yu's model, autoregressive (AR) model, the ANFIS, and the support vector regression (SVR) model. Results illustrate that the proposed hybrid model has a higher prediction accuracy than other models based on the root mean squared error (RMSE) (Wei, 2016).

Semero et al. (2019) extended the application of hybrid EMD based ANFIS for forecasting short term load in microgrids by including PSO employed to optimise ANFIS model. EMD is used to decompose time series into IMF and then PSO is employed to optimise ANFIS model. Load demand dataset in microgrid from Beijing is employed as a time series dataset. Results demonstrated the superior forecasting performance of the proposed model in forecasting short-term microgrid load demand as compared to other models.

Mahrooghi and Lakzian (2021) employed a hybrid model of ANFIS-GA-CFD for optimisation purposes of several variables which are (the blade tip thickness, the blade form, and the input-output angles of guide vanes) simultaneously. The results shows that the optimal blade increased the upper limit of efficiency up to 18%, torque coefficient up to 8%, and reduced the pressure drop coefficient up to almost 3.45% compared to the best used previous models.

## 2.4 Concluding remarks

This chapter has outlined both the advantages and disadvantages of machine learning (ML) models. It is well documented in the literature that optimisation of machine learning models' hyper-parameters improves forecasting performance of ML models, hence over the recent past, there has been development of various evolutionary algorithms with their unique advantages and shortcomings. There have been a lot of comparative studies of machine learning based evolutionary algorithms hybrid models in modelling and forecasting time series data. However, to the knowledge of the researcher, there has been few or none in which all the three ML models together with evolutionary algorithm have been compared in terms of forecasting performance. Furthermore, this chapter also has presented a gap in terms of modelling and forecasting of FDI series using machine learning models. It is shown in the chapter that conventional models are the ones employed to determine that the determinants of FDI and ML models are never adopted and employed in modelling and forecasting FDI. Therefore, the current study will seek to contribute to the literature by comparing these hybrid models and using traditional machine learning models as benchmark models for modelling and forecasting foreign direct investment (FDI). Furthermore, many determinants of FDI have been listed in the literature based or depending on the receiving economy, the current study will adopt and use exchange rate, gross domestic product (GDP) and inflation rate as the determinants of FDI.

# CHAPTER 3

# METHODOLOGY

## 3.1 Introduction

Over the recent past, various methodologies have been employed to model foreign direct investments (FDIs). The models predominantly employed were econometric models and were employed to study both developing and developed economies' determinants of foreign direct investments (FDIs). The success of these models is well documented in the literature and has given insights into what drives FDI investors to invest into an economy. Nevertheless, there are various time series (TS) and machine learning (ML) models that can be adopted and used to model and forecast FDIs. Machine learning (ML) models have, over the recent past, demonstrated their forecasting ability, thereby the current study adopts three nonlinear machine learning models, and one evolutionary algorithm (EA) to model and forecast FDI. Although there are several ML models, the current study is limited to using only the following methods: artificial neural network (ANN), support vector regression (SVR), least square support vector regression (LSSVR) together with genetic algorithm (GA) coupled with an ensemble hybrid model with the specified ML models.

## 3.2 Data description and sources

For addressing the study aim(s) and objectives, the current study, which is quantitative in nature, uses a historical monthly time series data of foreign direct investment as a dependent variable, gross domestic product, inflation rate and exchange rate as explanatory variables covering the period from January 1970 to June 2019; sourced from World Bank website and South African Reserve Bank website respectively. The choice of the variables used in the study was based on the literature of determinants of foreign direct investment into Africa/or South Africa. The stated explanatory variables time series data was downloaded from the World Bank website and South African Reserve Bank website for usage because of availability. The current variables used are those of foreign direct investment into South Africa, and the explanatory variables used are those of the South Africa context. Therefore, the study is predominantly focusing on the South African context. The following sections under this chapter will outline the methodology of tests and models adopted and used in the current study.

### 3.3 Preliminary data analysis

Prior to the main analysis, as suggested by Kline (2005a), the current study looks to address issues of normality of the actual time series data. Furthermore, other descriptive statistics such as median, mean and standard deviation of time series variables are also discussed. Moreover. For the purpose of assessing normality of the data, skewness-kurtoses measures will be estimated following the work of Xaba (2014).

### 3.4 Multicollinearity analysis of the data

The current study uses variance inflation factor (VIF) to test for multicollinearity in the data. Multicollinearity exists when there is a correlation between explanatory variables in a multiple regression model and thus multicollinearity can adversely affect the model and the results of the model. Thus, the variance inflation factor can estimate how much the variance of a regression coefficient is inflated due to multicollinearity. Multicollinearity reduces the statistical significance of the explanatory variables.

While there might be no specific universally agreed-upon benchmark for multicollinearity, some common metrics and guidelines can help identify its presence. A variance inflation factor (VIF) value greater than 5 or 10 is often considered an indication of problematic multicollinearity, while any value less than 5 indicates low or little concern of multicollinearity. The variance inflation factor (VIF) statistic is obtained as following:

$$VIF\left(k\right)=\frac{1}{1-R_k^2}, k=2,...,p, \qquad 3.1$$

Given that $R_k^2$ is the coefficient of determination of the regression of explanatory variables $X_k$ and $k$ represents number of explanatory variables.

### 3.5 Nonlinearity analysis of the data

Various nonlinear machine learning models are employed in the study to address the set research aim(s) and objectives. Furthermore, several nonlinearity tests are adopted and applied to investigate the appropriateness of employing nonlinear models. Ismail and Isa (2007) stated and advised that in testing for nonlinearity, it is wise to employ various tests of nonlinearity as nonlinearity in time series occurs in several ways. Therefore, the study used the following nonlinearity tests to investigate whether the data is nonlinear in nature: Brock, Dechert, and Scheinkman (BDS) test developed by Broock et al. (1996), cumulative sum (CUSUM) test for stability developed by Brown et al. (1975), White neural network test developed by (White, 1989), Terasvirta neural network test developed by Teräsvirta et al. (1993), McLeod-Li test developed by McLeod and Li (1983), Tsay test developed by Tsay (1986), and Keenan test developed by Keenan (1985).

Furthermore, as Ismail and Isa (2007) already stated, nonlinearity can occur in different forms hence it is wise to employ various nonlinearity tests as they have different advantages and shortcomings. Luo et al. (2020) state both the advantages and shortcomings of BDS test and according to Luo et al. (2020), BDS is a nonparametric test meaning that it does not need to have any knowledge about the random sequence. Moreover, the sequence of higher moments of random sequence is not required when the BDS test is applied in practice which gives the BDS test an advantage over other independence tests. Furthermore, BDS test can also be employed to examine the goodness of model fitting in time series analysis. However, there are several shortcoming also involved or present in BDS test, this includes the fact that the BDS's performance is somewhat sensitive to the choice of other parameters which are embedding dimension $m$, dimensional distance $\varepsilon$, and sample size $T$. Another discovery made by Broock et al. (1996) is that BDS test has another shortcoming of over-rejecting the null hypothesis irrespective of how large the sample size is, even if $m$ and $\varepsilon$ are well specified.

The cumulative sum (CUSUM) family of tests, especially CUSUM test, are adopted and employed in conjunction with BDS to test the stability of a time series data. These tests have an advantage over the Chow (1960) tests of not making an assumption of the point at which the hypothesised structural break takes place. However, the shortcoming of CUSUM test is that when the structural change involves a slope coefficient or the variance of the error term, the test has no power as cited from (Turner, 2010). Whereas Keenan test was motivated by the similarities of Volterra expansions to polynomials, it has the advantage of being extremely simple both conceptually and computationally. However, it has the shortcoming of having a low power. Moreover, Tsay test is a generalisation of Keenan test and uses the cross product of lagged variables in the second step. This makes the Tsay test of nonlinearity more powerful. Furthermore, Prabowo et al. (2020) stated that teresvirta test has a greater power in detecting nonlinearity, however, it is sensitive to the presence of outliers.

### 3.5.1 BDS test

Time series data has an element of nonlinearity. Broock et al. (1996) developed the BDS test of nonlinearity based on the idea that arises in the theory of chaos. Furthermore, chaos theory depends on the assumption that the underlying system is a nonlinear procedure, and the underlying system is in the deterministic framework. The BDS test is dependent on the series integral correlation. Several studies in the past have employed BDS test; empirical studies of time series, amongst others, are the studies of Xaba et al. (2019). The statistic of the BDS test is given by:

$$BDS_{m,M}(r)=\sqrt{M}\,\frac{c_m(r)-c_1^T(r)}{\sigma_{m.M}(r)} \qquad 3.2$$

where M is the embedded point of the space with m dimension; $r$ is the radius of a sphere centred on $X_i, C$ is the constant and $\sigma_{m.M}(r)$ is the standard deviation of $\sqrt{M}c_m(r)-c_1^T(r)$ Moreover, the cumulative sum (CUSUM) test is used in accordance with the BDS test to assess the stability of foreign direct investment, gross domestic product, exchange rate and inflation rate time series data employed in the current study. Brock et al. (1996) showed that when the residuals under the null hypothesis are independent and identically distributed (IID) or the white noise process is being followed,

$$BDS_{m,n} \sim N(0,1) \qquad 3.3$$

In the instance that the null hypothesis is rejected (linearity), the test statistic of the BDS will exceed the critical value at a given significance level or if the probability value of $BDS_{m,n}$ is lower than $\alpha$. If the latter is to happen, it means that there is presence of nonlinearity in the series.

### 3.5.2 CUSUM test

Stability is another part of nonlinearity in data. CUSUM test looks at data stability by testing for conceivable structural change in the series. Several studies have employed the test in the past. Amongst others are the studies of (Xaba, 2014, Shao and Zhang, 2010, Halicioglu and Mehmet, 2005). The other test that can be used to test stability is CUSUMSQ. In the condition that there is stability in the model, then there will be no change over time in the coefficient $(\beta)$ and $\sigma_\varepsilon{}^2$. Then, if the latter is achieved, the matrix can be used to obtain the coefficients (Brown et al., 1975):

$$\hat{\beta}=\left(Y'_{t-1}Y_{t-1}\right)^{-1}Y'_{t-1}Y_t \qquad 3.4$$

dependent variable is given by $Y_t$ and $Y_{t-1} = \left(1Y_{t-1}, Y_{t-1}, .., Y_{t-p}\right)'$ and $\varepsilon_t \sim iid\left(0, \sigma_\varepsilon^{\ 2}\right)$. Given that the model is unstable, then $(\beta)$ coefficient and the $\sigma_\varepsilon^{\ 2}$ possibly change over time. In such a case, then coefficient $(\beta)$ is replaced by $b_t$ and so

$$b_t = \left(Y'_{t-1} Y_{t-1}\right)^{-1} Y'_{t-1} Y_{t-1} \qquad 3.5$$

Where $Y_{t-1} = \left(1Y_{t-1}Y_{t-2}, ..., Y_{t-p}\right)'$ and $\varepsilon_t \sim iid\left(0, \sigma_\varepsilon^{\ 2}\right)$. Should the $AR(p)$ be stable, then there will be consistency in the parameters over time, this will show that there is no structural change in the series. Then hypotheses put forth to be tested are given as:

$$H_o : \beta_1 = \beta_2 = ... = \beta_p = \beta \qquad 3.6$$

Or

$$H_o : \sigma^2_{\ 1,\varepsilon} = \sigma^2_{\ 2,\varepsilon} = ... = \sigma^2_{\ n,\varepsilon} = \sigma^2_{\ \varepsilon} \qquad 3.7$$

Then, the calculation of the $\sigma_\varepsilon^{\ 2}$ is as follows:

$$\operatorname{var}\left(\varepsilon_t\right) = E\left(\varepsilon_t^{'}\varepsilon_t\right) = \operatorname{var}\left\{\varepsilon_t - Y^{'}_{\ t}\left(b_{t-1} - \beta\right)\right\} = \sigma_\varepsilon^{\ 2} + Y^{'}_{\ t}\sigma_\varepsilon^{\ 2}\left(Y^{'}_{\ t-1}Y_{t-1}\right)^{-1} Y_t = \sigma_\varepsilon^{\ 2} + \left\{1 + Y^{'}_{\ t}\left(Y^{'}_{\ t-1}Y_{t-1}\right)^{-1} Y_t\right\} \qquad 3.8$$

The scaled recursive residuals, $\omega$, may be defined as

$$\omega_t = \frac{\varepsilon_t}{\sqrt{\left\{1 + Y'_{\ t}\left(Y'_{\ t-1}Y\right)^{-1} Y_t\right\}}} = \frac{\varepsilon_t}{\operatorname{var}\left(\varepsilon_t\right)} = \frac{\varepsilon_t}{s.e\left(\varepsilon_t\right)} = \frac{\varepsilon_t}{\sigma_\omega}, \qquad 3.9$$

Under constant parameters, $\omega \sim iid\left(0, \sigma_\omega^{\ 2}\right)$. Therefore, the CUSUM statistic will be given by:

$$W_t = \sum_{j=l+1}^{t} \omega_j, t = l+1, l+2, ..., n. \qquad 3.10$$

The statistic of CUSUMSQ (cumulative sum of squares) is given by mathematically:

$$S_t = \sum_{j=l+1}^{t} \omega_j \Big/ \sum_{j=l+1}^{n} \omega_j, t = l+1, l+2, ..., n. \qquad 3.11$$

The application of the two tests is by plotting $W_t$ and $S_t$ against time $t$. The confidence bounds are found through plotting the two lines that join the points $\left[l, \pm\alpha\sqrt{n-1}\right]$ and $\left[n, \pm 3\alpha l\right]$. At 5% level (95% confidence interval), $a = 0.0948$ and at 1% level (99% confidence interval), $a = 1.143$ . The existence of structural break can be seen when the test statistic is outside the interval at a specific significance level, together with non-consistency in the parameters which leads to instability in the series. Furthermore, this will lead to the rejection of the null hypothesis of stability.

### 3.5.3 White neural network test

White (1989) developed a white neural network test for testing for a neglected nonlinearity present in a time series data and the test is computed by first letting the pragmatic training set that is attained as a haphazard sequence to be $Z^n = \{Z_t, t = 1, 2, ..., n\}, Z_t = (Y_t, X_t)$ Given a scalar quantity as $Y_t$, and row vector for finite dimension as $X_t$ then, White (1989) alleged that $Z_t$ is neither identical nor independent from the past values be distributed within $t = 1, 2, ..., n$. In this case, the conditional $E((Y_t \mid X_t))$ is the one governing the studying of the relationship between $X_t$ and $Y_t$. To be more direct, the above expression can be extended to the following regression function:

$$g(X_t) = E((Y_t \mid X_t)) \tag{3.12}$$

where the identical distribution assumption $g$ does not depend on $t$. Here, the goal line is to investigate the competence of a given multilayer feed-forward network as a representation of the unknown mapping function of $g$ by considering a class of feed-forward networks in which the network output scalar is determined by some given input $x$. Therefore, the network was demonstrated by White as follows:

$$f(x, \theta) = \tilde{x}\gamma_0 + \sum_{j=1}^{q} \beta_j \varphi(\tilde{x}\gamma_1) \tag{3.13}$$

Where $\tilde{x}\gamma_0$ is a linear component, $\sum_{j=1}^{q} \beta_j \delta(\tilde{x}\gamma_1)$ is a nonlinear component, $\beta_j$ is the weight of the neural network model from the hidden layer to the output layer, $\gamma_1$ is the weight of the neural network model from the input layer to the hidden layer, and $\varphi$ is a sigmoid activation function.

When the value of $q = 0$, there are only 2 layers, namely the input layer and the output layer only. And the model is linear. The null hypothesis in the White test is as follows:

$$H_0 : P\left[\left(g\left(X_t\right) = f\left(X_t, \theta^*\right)\right)\right] = 1 \text{ for some } \theta^* \qquad 3.14$$

This establishes a detailed statement of the null hypothesis of interest. Henceforth, the alternative is that the network is unable to exact the representation of $g$ and it is formulated as:

$$H_a : P\left[\left(g\left(X_t\right) = f\left(X_t, \theta^*\right)\right)\right] < 1 \text{ for all } \theta \qquad 3.15$$

A Lagrange multiplier test of Engle (1982) is a statistical test of the null hypothesis and the test is based on the linear regression:

$$\hat{\varepsilon}_t^{\,2} = \eta_0 + \sum_{i=0}^{m} \eta_i \varepsilon^2_{\;t-i} + \varepsilon_t \qquad 3.16$$

Where parameter estimates are given by $\eta_0, ..., \eta_m, v_t \sim i.i.d\left(\mu = 0, \sigma_\varepsilon^{\;2} = 1\right)$ and the F-statistic is used as the test statistic.

$$F^* = \frac{\dfrac{\left(\dfrac{RSS}{TSS}\right)}{m}}{\dfrac{1 - \left(\dfrac{RSS}{TSS}\right)}{n - m - 1}} \sim F_{\alpha,(m,n-m-1)} \qquad 3.17$$

RSS and TSS are the regression sum of squares and total sum of squares of equation (3.16) respectively. In general, the quantity $\frac{RSS}{TSS}$ is the mathematical computation of the coefficient of determination also known as $R^2$ The null hypothesis is rejected if the $F$ value is greater than critical value of $F_{\alpha,(m,n-m-1)}$ and it can be concluded that foreign direct investment, gross domestic product, exchange rate and inflation rate are non-linear in nature.

### 3.5.4 Teresvirta neural network test

One of the nonlinearity test is Terasvirta test developed by (Teräsvirta et al., 1993). This test is almost similar to the White test in that they are both using neural network model. However, the

White test and Terasvirta test differ in determining neural network model parameter values: White uses random sampling, while Terasvirta employs Taylor expansion. (Prabowo et al., 2020). The following shows an example of a nonlinear model:

$$y_t = \varphi(\gamma' w_t) + \beta' w_t + u_t \tag{3.18}$$

Where a linear component is given by $\beta' w_t$, $\varphi(\gamma' w_t)$ is nonlinear component, $\gamma'$ is the weight of the neural network model from the input layer to the hidden layer for nonlinear components, $\beta'$ is the weight of the neural network model from the input layer to the output layer for linear components and a sigmoid threshold function is given by $\varphi$. Equation (3.18) can be mathematically rewritten as follows:

$$y_t = \beta' w_t + \sum_{i=1}^{q} \theta_{0i} \left\{ \varphi(\gamma' w_t) - \frac{1}{2} \right\} + u_t \tag{3.19}$$

Where the weight of a neural network model from a nonlinear component from the hidden layer to the output layer is given by $\theta_{0i}$, the data will have a linear relationship if the nonlinear component is $0$. Therefore, the hypothesis will be as follows:

$$H_0 : \theta_{01} = \ldots = \theta_{0q} = 0 \tag{3.20}$$

Or can be written as:

$$H_0 : \gamma_1 = \ldots = \gamma_q = 0 \tag{3.21}$$

The value of neural network parameters in the Teravista test uses Taylor expansion so that a new model is obtained:

$$y_t = \beta' w_t + \sum_{i=1}^{p} \sum_{j=1}^{p} \delta_{ij} y_{t-i} y_{t-j} + \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} \delta_{ijk} y_{t-i} y_{t-j} y_{t-k} + u_t \tag{3.22}$$

If the quadratic and cubic component terms are $0$, then fail to reject $H_0$, we get a linear model. Terasvirta test can be tested using the chi-square distribution and F distribution, it similar to White test. Using Terasvirta test with chi-squared distribution follows the following procedure:

1. Regress $y_t$ with $1, X_1, \ldots, X_p$ and calculate the residual value $\hat{u}_i = Y_i - \hat{Y}_i$.

2. Regress $\hat{u}_i$ with $1, X_1, ..., X_p$ and $m$ additional predictor, then calculate the coefficient of determination from regression $R^2$, $m$ predictor is the quadratic and cubic component values of Taylor expansion.
3. Calculate $X^2 = nR^2$ where $n$ is the number of observations. If $nR^2 > X^2_{(m)}$, then reject $H_0$.

Whereas the testing procedure with F distribution is as follows:

1. Regress $y_t$ with $1, X_1, ..., X_p$ and calculate the residual value $\hat{u}_i = Y_i - \hat{Y}_i$ and calculate the value of $SSR_0 = \sum \hat{u}_i^2$.
2. Regress $\hat{u}_i$ with $1, X_1, ..., X_p$ and $m$ additional predictor. Then, calculate the residual value $\hat{v}_i = \hat{u}_i - \hat{\hat{u}}_i$ and calculate the value of $SSR_1 = \sum \hat{v}_i^2$, $m$ predictor is the quadratic and cubic component values of Taylor expansion.
3. Then compute

$$F^* = \frac{\frac{(SSR_0 - SSR_1)}{m}}{\frac{SSR_1}{(n-p-m-1)}} \qquad 3.23$$

Where $m$ is the number of additional predictors, $n$ is the number of observations and $p$ is the number of initial model predictors. Then $H_0$ is rejected if $F_{calculated} > F_{(\alpha, m, n-p-m-1)}$.

### 3.5.5 McLeod-Li test

The McLeod and Li test developed by McLeod and Li (1983) is a Portmanteau test for nonlinear dependence and it is carried out by examining the Ljung-Box Q statistic of the squared residuals from an Autoregressive Moving Average (ARMA) representation. Employing the $k$ autocorrelation coefficients $\rho$, the squared residuals $e_t^2$ are examined. The test is employed to detect the presence of serial correlation as suggested by Granger and Newbold (1986). That is, if

$$\rho_e(K) = [\rho_e(K)]^2 \qquad 3.24$$

then the series are linear. However, there is an implication of nonlinearity when the expression in equation (3.24) is not equal. The Ljung-Box Q-statistic to test the null hypothesis of independence is calculated as follows:

$$Q_{ee}\left(K\right)=n\left(n+2\right)\sum_{k=1}^{m}\left(n-k\right)^{-1}\rho^{2}{}_{ee}\left(K\right) \tag{3.25}$$

Where the number of observations is given by $n$ and $Q_{ee}\left(K\right)$ is the autocorrelation coefficient of the squared series. The critical value $\tau$ obtained from the chi-square distribution at a given significance level, $\chi^{2}\left(K\right)$ distributed with $K$ degree of freedom, is compared to $Q_{ee}\left(K\right)$. The null hypothesis of independence will be rejected if $Q_{ee}\left(K\right)>\tau$, suggesting that the residuals time series exhibit possible conditional heteroscedasticity, that is, they are nonlinearly dependent.

### 3.5.6 Keenan test

Another method of testing linearity/(nonlinearity) of a time series data was developed by Keenan (1985), which essentially tests

$$H_0 : X_t = \mu + \sum_{u=0}^{M} a_u e_{t-M} \tag{3.26}$$

against

$$H_a : X_t = \mu + \sum_{u=0}^{M} a_u e_{t-u} + \sum_{u=0}^{M'}\sum_{v=0}^{M''} a_{uv} e_{t-u} e_{t-v}, \tag{3.27}$$

where the sequence of zero mean i.i.d. are given by $\{e_t\}$ and $M, M', M''$ are presumably sufficiently large. The motivation seems to stem from regarding the following functional expansion as the most general form of non-linear stationary time-series models:

$$X_t = \mu + \sum_{u=-\infty}^{\infty} a_u e_{t-u} + \sum_{u,v=-\infty}^{\infty} a_{uv} e_{t-u} e_{t-v} + \sum_{u,v,w=-\infty}^{\infty} a_{uvw} e_{t-u} e_{t-v} e_{t-w} + ..., \tag{3.28}$$

where a sequence of zero mean i.i.d. random variables is given by $\{e_t\}$. The mechanics of the test are similar to Tukey's one-degree-of-freedom test for non-additivity and runs as follows, using the observations $X_1, X_2, ..., X_n$:

(i) Fit model in equation (3.28) to the data and calculate the estimated values $\{\hat{X}_t\}$ and the residuals, $\hat{e}_t$, for $t = M+1,...,n,$ and the residuals sum of squares, $\langle \hat{e}\hat{e} \rangle = \sum \hat{e}_t^{\,2}$ .

(ii) Regress $\hat{X}_t^{\,2}$ on $\{1, X_{t-1}, X_{t-2},..., X_{t-M}\}$ and calculate the residuals $\{\hat{\varphi}_t\}$, for $t = M+1,...,n,$

(iii) Regress $\hat{e} = (\hat{e}_{M+1},...,\hat{e}_n)$ on $\hat{\varphi} = (\hat{\varphi}_{M+1},...,\hat{\varphi}_n)$ and obtain $\hat{\eta}$ and $\hat{F}$ via

$$\hat{\eta} = \hat{\eta}_0 \left( \sum_{t=M+1}^{n} \hat{\varphi}_t^2 \right)^{\frac{1}{2}} \qquad 3.29$$

where $\hat{\eta}_0$ is the regression coefficient, and

$$\hat{F} = \frac{\hat{\eta}^2 (n-2M-2)}{\langle \hat{e}, \hat{e} \rangle - \hat{\eta}^2}. \qquad 3.30$$

And under $H_0, \hat{F} \sim F_{1,n-2M-2}$ . The choice of $M$ does not seem to be too critical. Keenan tests are easy and quick to implement. This is the advantage of the Keenan test. This test involves little subjective choice of parameters ($M$ because it's the only one) and seems to provide generally stable results. However, even if we admit that functional expansions are the most common form of nonlinear time series models, it is not immediately clear why. Keenan's test accepts the functional expansions as linear. A more fundamental point concerns the generality of extensions. It can be assumed that not all currently favoured nonlinear time series models have truncated enhancements; for instance, (Tong, 1986). The following simple example is considered:

$$X_t = \begin{cases} a + e_t, & if \quad X_{t-1} \le 0, \\ b + e_t, & if \quad X_{t-1} > 0, \end{cases} \qquad 3.31$$

where a sequence of zero mean i.i.d. random variables is given by $\{e_t\}$, and $\alpha$ and $b$ are unequal real constants. Does it possess a truncated functional expansion? A natural question to follow concerns the performance of Keenan’s test for cases where no valid truncated functional expansions exist.

### 3.5.7 Tsay test

Tsay (1986) test of nonlinearity is a generalisation of the Keenan (1985) test. Tsay test explicitly looks in the data for quadratic serial dependence. Let's let $k = k(k-1)/2$ column vectors $V_1,...,V_k$ contain all of the possible cross-products of the form $e_{t-i}e_{t-j}$, given that $i \in [1,k]$ and $j \in [i,k]$ Thus, $v_{t,1} = e^2{}_{t-1}, v_{t,2} = e_{t-1}e_{t-2}, v_{t-3} = e_{t-1}e_{t-3}, v_{t,k+1} = e_{t-2}e_{t-3}, v_{t,k+2} = e_{t-2}e_{t-4},...,v_{t,k} = e^2{}_{t-k}$. And let $\hat{v}_{t,j}$ denote the projection of $v_{t,i}$ on the orthogonal subspace $e_{t-1},...,e_{t-k}$, (that is, the residual from a regression of $v_{t,j}$ on $e_{t-1},...,e_{t-k}$.

The parameters $\gamma_1,...,\gamma_k$ are then estimated by applying OLS to the regression equation.

$$e_t = \gamma_0 + \sum_{i=1}^{k} \gamma_i \hat{v}_{t,j} + \eta_t \tag{3.32}$$

Given that the $jth$ regressor in this equation is $\hat{v}_{t,j}$, the period $t$ fitting error from a regression of $v_{t,j}$ on $e_{t-1},...,e_{t-k}$. As long as $p$ exceeds $K$, this projection is unnecessary for the dependent variable $\{e_t\}$ if it is pre-whitened using an $AR(p)$ model. The Tsay test statistic then is just the usual F statistic for testing the null hypothesis that $\gamma_1,...,\gamma_k$ are all zero.

## 3.6 Modelling and forecasting methods

The following sub-sections present the methodology of three nonlinear machine learning models employed in the study to model and forecast foreign direct investment. These models, in a systematic way, are ANNs, SVR, and LSSVR respectively.

### 3.6.1 Artificial neural networks (ANNs)

Computational intelligence systems, particularly the artificial neural network (ANNs) which are basically free of dynamics, have been used in approximation function and forecasting. Khashei and Bijari (2010), indicate that the ANN model has one significant advantage over other nonlinear models and, that is, ANNs are universal approximators that can approximate a large class of functions with a high class of accuracy. Their power comes from the parallel processing of the information from the data.

Zhang (2003) states that ANN can be a promising methodology in research to the traditional linear models (for example, multiple regression models and ARIMA). ANNs have found extensive use as new computational tools in solving many complex real-world problems (Basheer and Hajmeer, 2000) and “can be defined as information processing systems whose structure and functioning is inspired by biological neural networks” (Palmer et al., 2006).

According to Basheer and Hajmeer (2000) and Palmer et al. (2006), ANNs’ attractiveness stems from their remarkable information processing characteristics pertinent mainly to nonlinearity, high parallelism, fault and noise tolerance, and learning and generalisation. As stated by the study of Constantino et al. (2016) cited from (Law and Au, 1999), the architecture of the neural feed-forward is composed of three distinct layers, which are input layer, one or several hidden layers and an output layer, each of these layers contains nodes, and they are connected to nodes on the adjacent layer. For them, each node of a neural network is a processing unit that contains a weight and a sum function.

A mathematical value is returned by a weight (w) for the relative strength of connections to transfer data from one layer to another layer, whereas a weighted sum of the inputs elements entering a processing unit is computed by sum function (y). Independent problem variables are represented by the nodes in the input layer, the hidden layer is employed to add an internal representation of handling nonlinear data and the output of a neural network is the solution to a problem (Law and Au, 1999). The relationship between the output $(Y)$ and the inputs $(X_1, X_2, ..., X_k)$ has the following mathematical representation (Constantino et al., 2016).

$$y_k = b_k + \sum_{j=1}^{n} w_j f\left(\sum_{i=1}^{m} W_{ji} X_i + \beta_{oj}\right) \tag{3.33}$$

where $y_k$ $(k = 1, 2, 3, ..., p)$ representing the output variable, $W_{ji}\left(i = 0, 1, 2, 3..., m; j = 1, 2, 3, ...n\right)$ and $w_j\left(j = 1, 2, 3, ..., n\right)$ are model parameters often called connections weights, number of inputs nodes is given by $m$, and number of hidden nodes is given by $n$. Time series data enters the network through input layer, moves through hidden layer and exits through the output layer. Moreover, $f$ represents sigmoid activation function (equation also indicates the use of a linear activation function in the output layer). The $b_k$ and $\beta_{oj}$ indicates the deviations of the independent terms (bias) associated with each output layer node and hidden layer node respectively (figure 3.1).

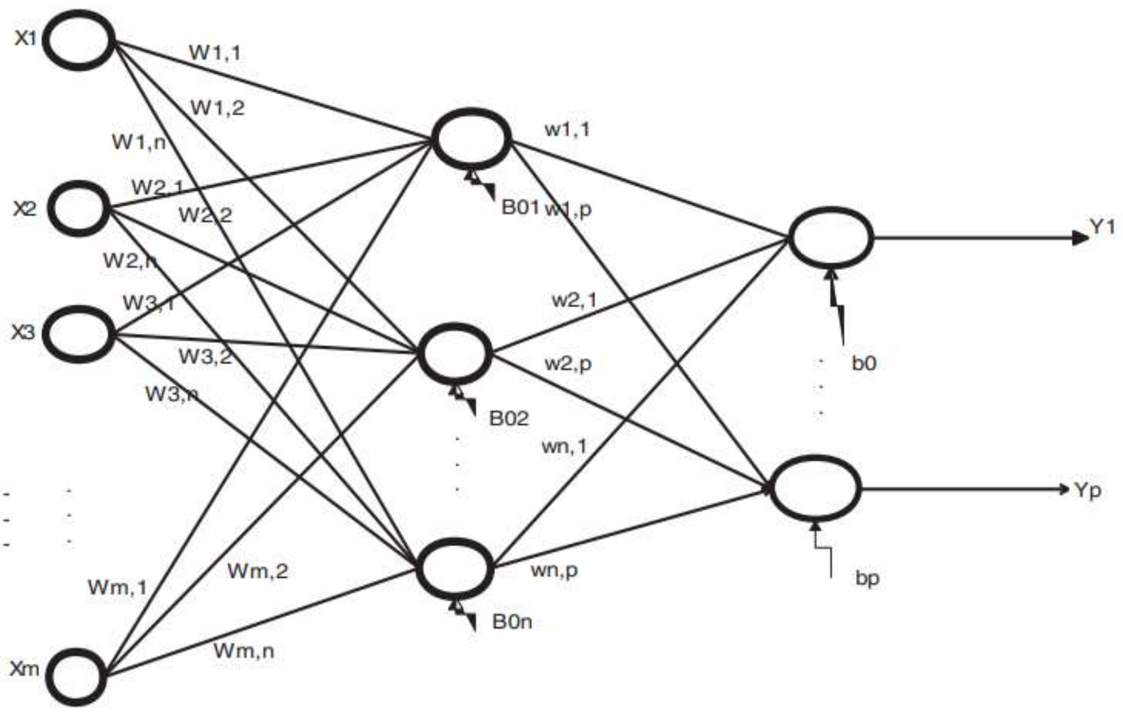


**Figure 3. 1: Feed-forward neural network architecture**
**Source: Constantino et al. (2016)**

Sigmoid logarithmic function is the most often used activation function in the hidden layer in the construction of artificial neural networks models (Zhang et al., 1998, Hykin, 1999). The logarithmic sigmoidal activation function is given in equation (3.33) and in figure 3.2a. The logarithmic sigmoidal activation function extends from $0$ to $+1$. This function is used to transform output so that it falls within an acceptable zone and defined as a strictly increasing function which exhibits an appropriate balance between linear and nonlinear behaviour (Fernandes, 2005). The transformation is done before the output reaches the next level and the purpose of this function is to prevent the output value to be large, as the value must be scaled to be between $0$ and $+1$.

$$f(x) = \frac{1}{1+e^{-x}} \quad 3.34$$

Hyperbolic tangent function or what can also be called tangent sigmoidal can also be used in the hidden layer as an activation function. Equation (3.35) and figure 3.2a shows the hyperbolic tangent function. The mathematical representation of hyperbolic tangent function is as follows:

$$f(x) = \frac{\left(e^{x} - e^{-x}\right)}{\left(e^{x} - e^{-x}\right)} = 1 \quad 3.35$$

Purelin is the most used activation function in the output layer, and it is a linear activation function. Purelin activation function can be represented mathematically as follows:

$$f(x) = x \tag{3.36}$$

Figure 3.2 shows three activation functions that are adoptable to be used in the hidden and output layer. These activation functions are logarithmic functions, tangent sigmoidal and linear function known as purelin.

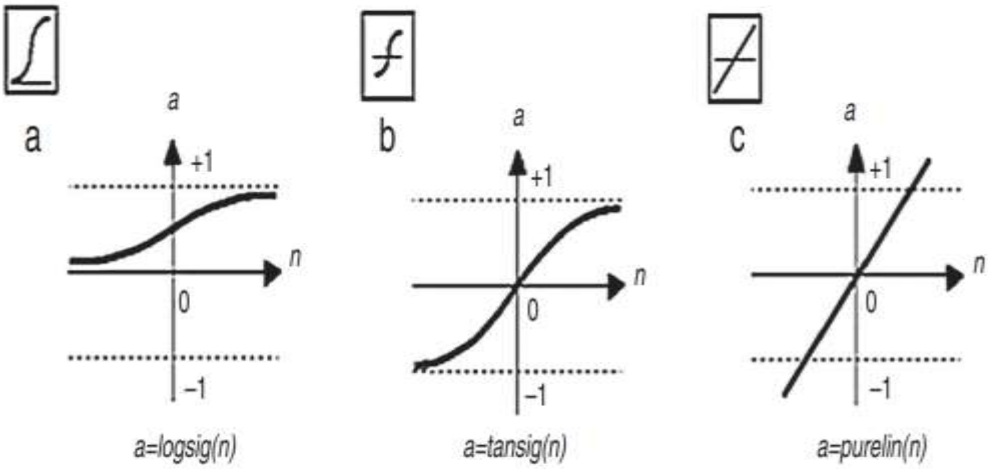


**Figure 3. 2: Activation function chart**
**Source: Constantino et al. (2016)**

In artificial neural networks (ANNs) modelling, supervised learning is by far the mostly employed technique. In this type of learning, the main condition is the presence of a teacher who can provide precise corrections to the network outputs when errors occur or relate them to the environment of input units and free network outputs. From training data, a supervised feed-forward neural network learns to reveal patterns in data representing input and output variables. Law and Au (1999) state that the learning process of a feed-forward neural network involves the following steps:

- Random numbers to assign the weights;
- Each element in the training set (a set of observations of the sample used to develop a pattern or relationship between observations) adjust the weights W and w in the error back propagation process (back propagation algorithm);
- Compare the computed output with the observed value. The process stops when the error between the output and the target drops below the current value, or when another validation set of errors has not fallen for a given number of iterations. This latter process, called cross-validation, prevents the model from overfitting the training data. This guarantees the neural network's ability to penalisation.

### 3.6.2 Support vector regression (SVR) model

There is a recent machine learning theory-based classification paradigm which is known as support vector machine (SVM), whereby support vector regression (SVR) is an adaptation of this support vector machine. In SVR concept, a typical regression problem is formulated. Let's consider a set of data, given that is a vector of the model inputs, is actual value and represents total number of data patterns.

Regression analysis has an objective of determining a function, as so to accurately predict desired output. Given that objective, the typical regression function can be formulated as, where is a random error with distribution of. There are two types of regression problem(s) namely, linear and nonlinear respectively. Nonlinear regression problems are somehow difficult to deal with. SVR was mainly developed for tackling the nonlinear regression problems.

In SVR, solving nonlinear regression problem, the inputs are firstly nonlinear mapped into a high dimensional feature space, wherein they are correlated linearly with the outputs. The SVR formalism considers the following linear estimation function (Lu et al., 2009).

$$f(x) = \left(v \bullet \varphi(x)\right) + b \tag{3.37}$$

given $v$ is weight vector, $b$ is a constant, $\varphi(x)$ denotes a mapping function in the feature space, and $\left(v \bullet \varphi(x)\right)$ describes the dot production in the feature space $F$. In SVR, the problem of nonlinear regression in the lower dimension input space $(x)$ is transformed into a linear regression problem in a high dimension feature space $(F)$. This means, the original optimisation problem involving a nonlinear regression is recast as searching the flatter function in the feature space and not in the input space.

There are several cost functions that can be used in the SVR formulation. These functions include Laplacian, Huber's Gaussian and -insensitive loss function $l_\varepsilon$ given below is the most commonly adopted (Lu et al., 2009).

$$l_\varepsilon\left(f(x), q\right) = \begin{Bmatrix} \left|f(x) - q\right| - \varepsilon & if \left|f(x) - q\right| \geq \varepsilon \\ 0 & otherwise \end{Bmatrix} \tag{3.38}$$

where $\varepsilon$ is a precision parameter representing the radius of the tube located around the regression function $f(x)$. Figure 3.3 represents a schematic representation of the SVR using the -insensitive loss function. In figure 3.3, the region enclosed by the tube is known as "-insensitive" since the loss function assumes a zero value in this region and does not penalise the prediction errors with magnitudes smaller than  (Lu et al., 2009).

According to (Lu et al., 2009) the weight $(v)$ and constant $(b)$ in equation (3.37) can be estimated by minimising the following regularised risk function.

$$R(C)=C\frac{1}{n}\sum_{i=1}^{n}L_e\left(f\left(x_i\right),q\right)+\frac{1}{2}|w|^2 \qquad 3.39$$

given that $L_e\left(f\left(x_i\right),q\right)$ is $\varepsilon$-insensitive loss function in equation (3.38); $\frac{1}{2}|w|^2$ is the regularisation term which controls the trade-off between the complexity and the approximation accuracy of the regression model to ensure that the model possesses an improved generalised performance; C is the regularisation constant used to specify the trade-off between empirical risk and regularisation term. Both C and $\varepsilon$ are the user-determined parameters.

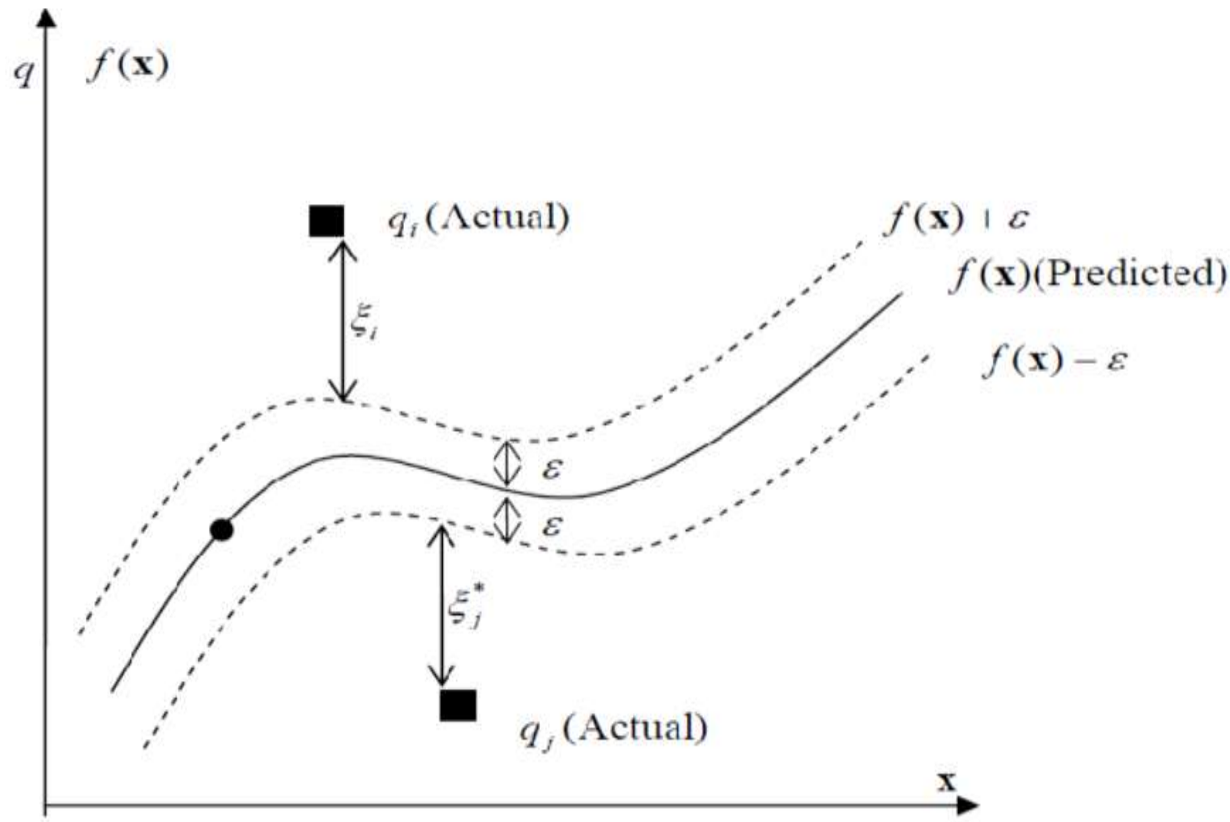


**Figure 3. 3: A schematic representation of the SVR using $\varepsilon$ -insensitive loss function**
**Source : Khashei and Bijari (2010)**

Two positive slack variables, $\check{c}_i$ and $\check{c}_i^*$, $i = 1, 2, ..., n$ can be used to measure the deviation $\left(q_i - f\left(x_i\right)\right)$ from the boundaries of the  zone. This means they represent the distance from the actual values to the corresponding boundary value of $\varepsilon$ insensitive zone as this is shown in figure 3.3. Equation (3.39) is transformed by using the slack variables into the following constrained form:

Minimize: $R_{reg}\left(f\right) = \frac{1}{2}\left|w\right|^2 + C\sum_{i=1}^{n}\left(\breve{c}_i + \breve{c}_i^*\right)$ 3.40

subjected to:

$$\begin{Bmatrix} q_i - \left(w \cdot \vartheta\left(x_i\right)\right) - b \le \varepsilon + \breve{c}_i \\ \left(w \cdot \vartheta\left(x_i\right)\right) + b - q_i \le \varepsilon + \breve{c}_i^* \\ \breve{c}_i, \breve{c}_i^* \ge 0,\ for i = 1, ..., n \end{Bmatrix} \quad 3.41$$

By using Lagrangian multipliers and Karush-Kuhn-Tucker conditions to equation (3.40), this will yield the following dual Lagrangian form Lu, Maximize:

$$L_d\left(\alpha, \alpha^*\right) = -\varepsilon\sum_{i=1}^{n}\left(\alpha_i^* + \alpha_i\right) + \sum_{i=1}^{n}\left(\alpha_i^* - \alpha_i\right)q_i - \frac{1}{2}\sum_{i}\sum_{j=1}^{n}\left(\alpha_i^* + \alpha_i\right)\left(\alpha_j^* + \alpha_j\right)k\left(x_i, x_j\right) \quad 3.42$$

subject to the constraints:

$$\begin{Bmatrix} \sum_{i=1}^{n}\left(\alpha_i^* - \alpha_i\right) = 0 \\ 0 \le \alpha_i \le C, i = 1, .., n \\ 0 \le \alpha_i^* \le C, i = 1, .., n \end{Bmatrix} \quad 3.43$$

The Lagrangian multipliers in equation (3.42) satisfy the equality $\alpha_i\alpha_i^* = 0$. The Lagrangian multipliers, $\alpha_i$ and $\alpha_i^*$ are calculated and an optimal desired weight vector of the regression hyperplan is;

$v^* = \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)k\left(x, x_j\right)$. SVR based regression function general form can be written as (Lu et al., 2009, Chen and Wang, 2007).

$$f\left(x,v\right)=f\left(x,\alpha,\alpha^{*}\right)=\sum_{i=1}^{n}\left(\alpha_{i}-\alpha_{i}^{*}\right)k\left(x,x_{i}\right)+b \qquad 3.44$$

given that $k\left(x,x_{i}\right)$ is called kernel function. The value of the kernel equals the inner product of two vectors, $x_i$ and $x_i$ in the feature space $\varphi\left(x_i\right)$ and $\varphi\left(x_j\right)$; that is, $k\left(x,x_i\right)=\varphi\left(x_i\right)\varphi\left(x_j\right)$. If any function can satisfy the Mercer's condition, it can be employed as a kernel function (Chen and Wang, 2007). The widely used kernel function is the radial basis function, although several functions are available (Lu et al., 2009). The radial basis function is defined as follows:

$$k\left(x_{i},x_{j}\right)=\exp\left(\frac{-\left\|x_{i}-x_{j}\right\|^{2}}{2\delta^{2}}\right) \qquad 3.45$$

where $\sigma$ denotes the width of the RBF. Therefore, the current study will employ the RBF as it was employed in the study of (Lu et al., 2009).

### 3.6.3 Least square support vector regression (LSSVR) model

Kumar and Kar, (2009) and Kisi, (2016) state that Suykens and Vandewalle (1999) were the first to propose or develop LSSVR method for solving nonlinear problems and function estimation. The nonlinear LSSVR function is written as:

$$f\left(X\right)=w^{T}f\left(x\right)+b \qquad 3.46$$

given that weight vector with $m$ dimension is given by $w$, function of mapping is denoted by $f$ and term of bias is given by $b$ (Cao et al. 2008). By considering error of function estimation, the regression problem can be solved by the minimisation of standard relation as:

$$\min J\left(w,\ e\right)=\frac{w^{T}w}{2}+\frac{\Upsilon}{2}\sum_{i=1}^{m}{}_{i}^{2} \qquad 3.47$$

subject to these following constraints as:

$$y_{t}=w^{T}f\left(x_{i}\right)+b+e_{i}\left(i=1,2,3,\ldots,m\right) \qquad 3.48$$

given that $\Upsilon$ is the regularisation constant and $e_i$ accounts for corresponding training error to $x_i$. The optimal programming technique of Lagrange multiple is applied to obtain coefficients of $w$

and $e$ (to solve equation (3.47). The objective is determined by changing the constraint problem into an unconstraint problem. The Lagrange function can be written as follows:

$$L(w,b,e,\alpha) = J(w,\ \mathrm{b},\ e) - \sum_{i=1}^{m} \alpha_i \left\{ w^T f(x_i) + b + e_i - y_i \right\} \tag{3.49}$$

where $\alpha_i$ is given as Lagrange multipliers and $\sum_{i=1}^{m} \alpha_i \left\{ w^T f(x_i) + b + e_i - y_i \right\}$ is used to minimize the training error. When the Karush-Kuhn-Tucker conditions are assumed (Fletcher, 1987), the best conditions can be obtained by taking the partial derivatives of (3.49) and considering $w$, $b$ , $e$ and $\alpha$ represents Lagrange multipliers as follows:

$$\begin{cases} w = \sum_{i=1}^{m} \alpha_i \varphi(x_i) \\ \sum_{i=1}^{m} \alpha_i = 0 \\ \alpha_i = \gamma e_i \\ w^T f(x_i) + b + e_i - y_i = 0 \end{cases} \tag{3.50}$$

After removing $e_i$ and $w$, the linear equations are obtained as:

$$\begin{bmatrix} 0 & -Y^T \\ Y & ZZ^T + I/Y \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{3.51}$$

given that $Y = y_I ... y_m$, $z = f(x_1)^T y_1, ..., f(x_m)^T y_m$, $I = [1,...,1]$, $\alpha = [\alpha_1, ..., \alpha_1]$. The kernel function is written as $K(x, x_i) = f(x)^T f(x_i), i = 1,...,m$, that is satisfied with Mercer's condition. In this condition, the LSSVR is expressed as (Shaghaghi et al, 2018):

$$f(x) = \sum_{i=1}^{m} \alpha_i K(x, x_i) + b \tag{3.52}$$

The study employs the RBF kernel function which is written in the following equation.

$$K(x, x_i) = \exp\left( \frac{-x - x_i^{\ 2}}{2\sigma^2} \right) \tag{3.53}$$

### 3.6.4 Genetic algorithm (GA)

Global search technique known as genetic algorithm (GA) was developed by Holland (1975). "Genetic algorithm is a meta-search algorithm based on natural evolutionary selection to solve complex optimisation and high dimensional problems either with or without restrictions" (Prado et al., 2020). GA has two types, real and binary coded GAs respectively. The study will adopt real value GA to optimise the hyper-parameters of the machine learning models because it is straight forward, fast and efficient. In genetic algorithm (GA), a candidate's goodness is determined by a fitness function, which must be defined based on the problem being solved. A process of genetic algorithm can be illustrated with pseudocode (Prado et al., 2020):

1. The population must be initialised;
2. Evaluate the goodness of each candidate;
3. If goodness is not achieved, or generations do not reach their maximum, then;
4. Genetic crossing and mutation operators must be applied on the candidates;
5. Create a new population with these new candidates;
6. Evaluate the goodness of each candidate;
7. Increase the generation by one;
8. Present the best individual reached.

## 3.7 Machine learning based genetic algorithm hybrid models

This section of methodology presents a hybrid machine learning based genetic algorithm model for modelling and forecasting foreign direct investment. The genetic algorithm is employed to optimise hyper-parameters of the following machine learning models which are ANNs and SVR to achieve optimal forecasting performance and accuracy.

### 3.7.1 Artificial neural network based genetic algorithm

Although artificial neural networks (ANNs) are one of the efficient and performing machine learning model, they have several shortcomings. These shortcomings include long training time, easily falling into local minima, global search capabilities are poor, and many parameters. These shortcomings affect the model accuracy of ANN. To address these shortcomings of ANN, attempts have been made to use it in hybrid methodologies with other algorithms, aiming to overcome these limitations. Over the recent past, ANNs have been employed in the hybrid model with genetic algorithm (GA). For example, the following studies have employed ANN based GA and achieved desirable results (Whitley et al., 1990, Kim, 2006, Tian and Gao, 2009, Ahmad et al., 2010, Karimi and Yousefi, 2012, Sangwan et al., 2015).

Genetic algorithm (GA) has advantages that can be able to aid in case of shortcomings that are present in ANN model. These advantages of GA include its ability to search over the whole solution space, it obtains the global optimal solution easily, and does not require the objective function continuously, it is differentiable, and does not even require an explicit function of the objective function, it only requires problem computable. It's not often trapped in local minima especially when the error function is not differentiable or there are no gradient solutions.

The problem of finding the optimal set of parameters for training an ANN can be viewed as a search problem in the space of all possible parameters. The parameters include weight sets between input hidden layers, weight sets between output hidden layers, bias values, and number of neurons in hidden layers. Genetic algorithms can be applied as search algorithms using exploitation and exploration. GA has the following key issues: The parameters of the network are contained in the chromosome, i.e. the solution space, how these parameters are encoded in the chromosome, i.e. the coding scheme, and finally how the individual chromosomes are evaluated and transformed into a fitness function. The architecture of the ANN-GA model is presented in figure 3.5. The chromosomes of GA represent the weight and bias values for a set of ANN models. Input data (gross domestic product, inflation rate and exchange rate) along with the chromosomes values are fed to the set of ANN models. The absolute difference between target $y$ and the estimated output $\hat{y}$ gives the fitness. The less the fitness value of an individual, GA considers it as better fit. The overall steps of genetic algorithm are described as follows:

**Pseudo code for Genetic Algorithm**

```
Choose an initial population of chromosomes:
   While termination conditions not satisfied do
   Repeat
        If crossover condition satisfied, then
          {
       Select parent chromosomes:
       Choose crossover parameters:
       Perform crossover:
          }
        If mutation conditions satisfied then
          {
         Choose mutation point:
         Perform mutation;
          }
     Evaluate fitness of offspring
     Until sufficient offspring created
     Select new population
 End while.
```

**Figure 3. 4: Overall steps of genetic algorithm**
**Source: Nayak et al. (2014)**

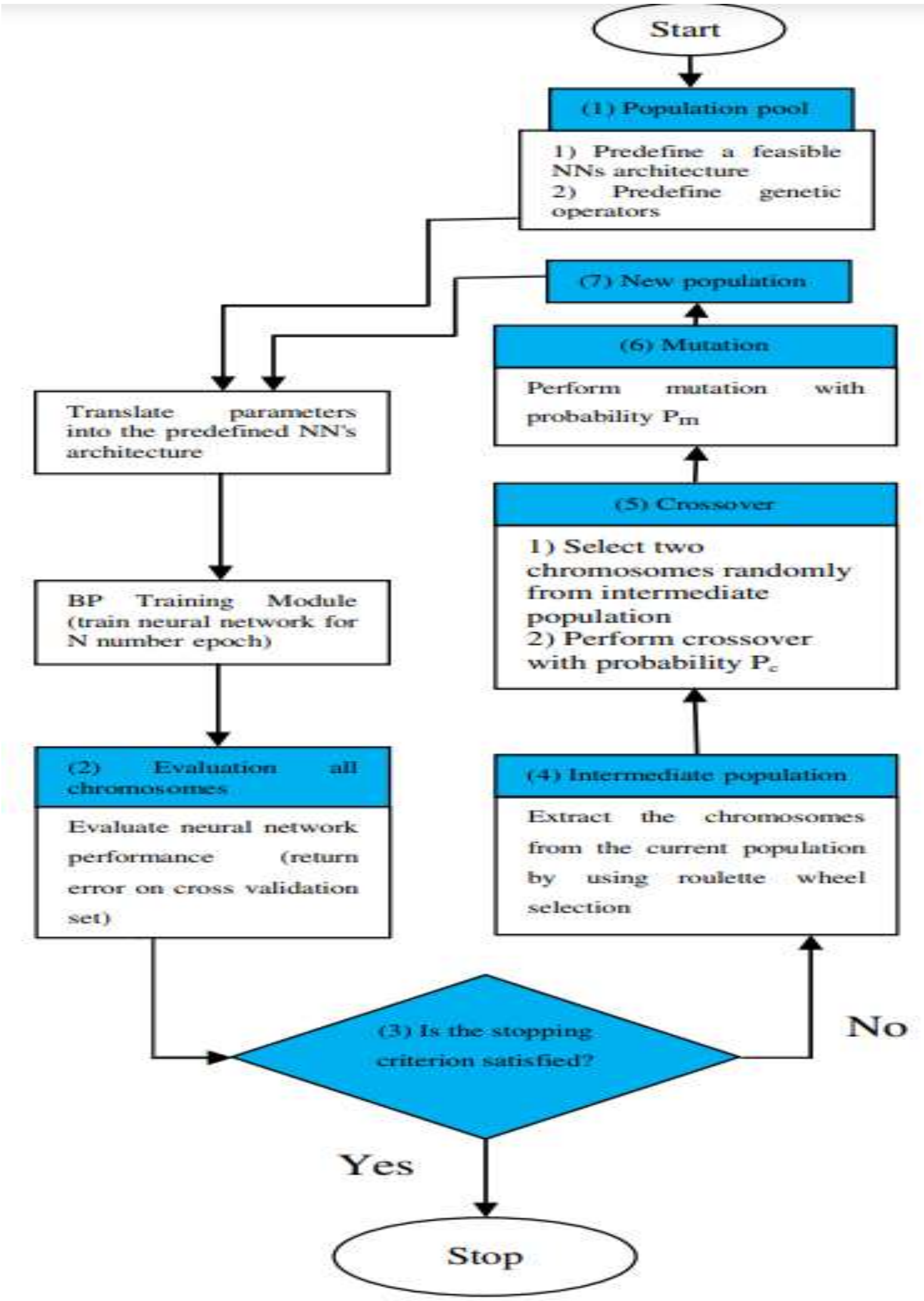


**Figure 3. 5: Using GA to optimise ANN hyper-parameter**
**Source: Karimi and Dastranj (2014)**

### 3.7.2 Support vector regression based genetic algorithm

Incorrect setting of SVR's hyper-parameters affects the generalisation performance (estimation accuracy) in both the industry and academia and efficiency of SVR model depends largely on its hyper-parameters $\left(C, \varepsilon, \sigma^2\right)$ which have correctly been set, where $\sigma^2$ is a kernel parameter. Xia et al. (2017) state that the inappropriate parameters in SVR lead to over-fitting. Even though this is the case, there has never been a guideline employed to correctly select optimal values for SVR's hyper-parameters. This then led researchers to depend on the trial-and-error method using grid algorithm, where several models would be estimated based on various parameter set first, then testing them on the validation set to obtain optimal parameters. However, this procedure is time consuming and inconvenient for use. Therefore, the current study follows the studies of Liu et al. (2013) and Xia et al. (2017) and adopts and uses a real value genetic algorithm to seek optimal hyper-parameters of SVR model in order to improve the efficiency of forecasting. In the proposed SVR-GA hybrid model, the values of the SVR parameters $(C, \varepsilon, \sigma)$ are directly coded in the chromosome with real value data, this study will optimise SVR and the values of SVR hyper-parameters through real value genetic algorithm (GA) process, and use those acquired optimal parameters values to construct an optimised SVR model in order to proceed with the forecasting. Figure 3.10 presents the framework of optimising SVR hyper-parameters with a real value genetic algorithm, and that framework can also be summarised as follows.

*Step 1 code the chromosome:* The parameters of SVR model $(C, \varepsilon, \sigma)$ are directly coded to form a chromosome. Here the range of $C$ is defined as $[1,100]$, the range of $\varepsilon$ is defined as $[0.0001, 0.01]$ and the range of $\sigma$ is defined as $[0,1]$. Randomly generate an initial population of chromosomes which represents the values of hyper-parameters in SVR. The population size is 30.

*Step 2 evaluate fitness:* The fitness of training data is easy to calculate but prone to over-fitting. Cross validation and genetic algorithm (GA) can be used to handle the problem of over-fitting. The current study adopts and uses fivefold cross validation (CV) technique to overcome the over-fitting problem. Five-fold CV randomly splits the training data into five mutually exclusive subsets of approximately equal size.. The four subsets are employed as training data to build a regression function with given set of parameters $(C, \varepsilon, \sigma)$. The performance of the set of hyper-parameters $(C, \varepsilon, \sigma)$ is measured by the root mean squared error (RMSE) on the last subset. The above steps are repeated 5 times. That is, each subset is used once for validation. The fitness function is defined as the RMSE CV of the training dataset, given by (3.55).

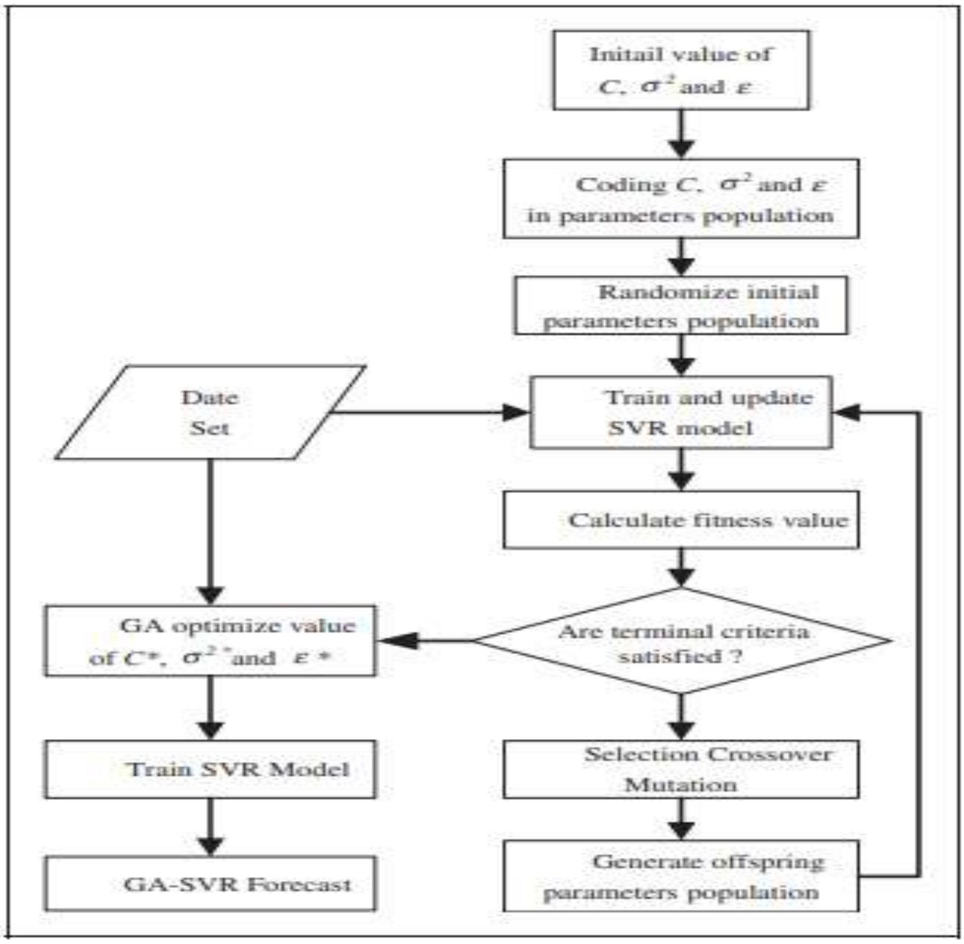


**Figure 3. 6: Optimising SVR parameters using real genetic algorithm**
**Source: Xia et al. (2017)**

$$Minf = RMSE_{CV} = \frac{1}{5}\sum_{i=1}^{n}\sqrt{\frac{5}{m}\sum_{j=1}^{+k/5}\left(a_j - p_j\right)^2} \quad 3.54$$

given that the actual and the predicted values are given by $a_j$ and $p_j$ respectively.

*Step 3 GA operators:* In the operator's standard roulette, wheel is performed to select excellent chromosomes to reproduce. Single point cross over is randomly adopted to exchange genes between two chromosomes and the probability of creating new chromosomes in each pair is set to 0.5. The mutation operation follows the crossover operation and determines whether a chromosome should be mutated in the next generation. Each chromosome in the new population is subject to mutation with a probability of 0.02.

*Step 4 stopping criteria*: If the new population does not meet the termination criteria, steps 3 through 4 are repeated until $C$ , $\varepsilon$ and $\sigma$ are satisfied with minimised model error. The optimal $C$ , $\varepsilon$ and $\sigma$ returned according to the value of the optimal fitness function.

The converged solution is mainly affected by parameter settings. The selection of other parameters such as mutation probabilities and population size is based on numerous experiments as these values are minimized in $RMSE_{CV}$ $RSE_{POS}$ in the training data set.

### 3.8 Comparison of model performance

Comparative analysis research to date still has the attention of researchers worldwide. This is due to the fact and assumption that various models will have different forecasting performance and forecasting accuracy. These models differ in terms of performance, accuracy, and reliability. This being the case, various performance metrics (or performance measures) have been developed and proposed in the literature for the purpose of selecting the model that shows more ability in terms of forecasting performance and forecasting accuracy. This study employs the following performance metrics (or performance measure) for evaluating forecasting performance and accuracy namely, mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and mean absolute scaled error (Abayomi-Alli et al.) for the purpose of selecting the most accurate model for forecasting foreign direct investment time series data. MSE, RMSE, MAE, MAPE and MASE measures the deviation between the actual data and the predicted data. The closer the values of the predicted time series data to the actual time series data, the smaller the values of MSE, RMSE, MAE, and MAPE. Each of the stated performance measures (MSE, RMSE, MAE, MAPE and MASE) are the function of the actual value $X_t$ and forecasted values $\hat{X}_t$ of the time series data, $\varepsilon_t = X_t - \hat{X}_t$ and $n$ being the sample size. The following formulae shows the mathematical representation of these performance metrics (or performance measures):

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^{\,2} \qquad 3.55$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^{\,2}} \qquad 3.56$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|\varepsilon_t\right| \qquad 3.57$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{X_t - \hat{X}_t}{X_t}\right| \qquad 3.58$$

$$MASE = \frac{\frac{1}{j}\sum_j \left|X_t - \hat{X}_t\right|}{\frac{1}{T-1}\sum_{t=2}^{T}\left|X_t - X_{t-1}\right|} \qquad 3.59$$

## 3.9 Comparing model accuracy

Statistical models exhibit promising forecasting performance when quantifying their performance using performance measures. Comparing these models in terms of forecasting accuracy is the way to tell if there is any difference of forecasting performance. Therefore, this current study employs Diebold-Mariano test of forecasting accuracy to see if there is any difference in forecasting performance of the models included in the study. The following sub section gives the methodology of Diebold-Mariano test.

### 3.9.1 Diebold-Mariano (DM) test

In empirical applications, we often have more than one time series model available to predict a particular variable of interest. Many statistical methods show construction performance in terms of forecasting ability. Comparing these methods or models is a way to see the differences these models have in terms of forecasting performance. The standard is that the smaller the error, the better the model, but this is not always the case because the difference in error measurements of competing models is not much different from zero. The test used to compare the predictive performance of competing models is the Diebold Mariano (DM) test, which tests whether a pair of competing models built on given data have equal forecasting accuracy. If the pair of forecast errors $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ have the same prediction accuracy, the expected squared difference (mean of the squared differences),

$$\bar{d} = E\left(d_t\right) = E\left({\varepsilon^2}_{1,t} - {\varepsilon^2}_{2,t}\right) \qquad 3.60$$

should be approximately zero, that is $\bar{d} = \frac{1}{n}\sum_{t=1}^{n} d_t = 0$. The following hypotheses are constructed for Diebold-Mariano test:

$H_0 : \bar{d} = 0$ (equal forecasting accuracy of pair of models)

$H_1 : \bar{d} \neq 0$ (no equality of forecasting accuracy of pair of models)

In order to test the null hypothesis that the two forecasts have the equal forecasting accuracy, Diebold-Mariano utilised the following statistic:

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi\hat{\tau}_d(0)}{T}}} \qquad 3.61$$

given that $\hat{\tau}_d(0)$ is a consistent estimate of $\tau_d(0)$ defined by:

$$\hat{\tau}_d(0) = \frac{1}{2\pi} \sum_{k=-(T-1)}^{T-1} I\left(\frac{K}{h-1}\right) \hat{\gamma}_d(k) \qquad 3.62$$

where:

$$\hat{\gamma}_d(k) = \frac{1}{T} \sum_{t=|k|+1}^{T} \left(d_t - \bar{d}\right)\left(d_{t-|k|} - \bar{d}\right) \qquad 3.63$$

and

$$I\left(\frac{K}{h-1}\right) = \begin{cases} 1 & for \left|\frac{k}{h-1}\right| \leq 1 \\ 0 & otherwise \end{cases} \qquad 3.64$$

The null hypothesis of Model 1 having the same predictive accuracy as Model 2 is rejected if the test statistic DM is greater than the critical value or the p-values of DM are less than 1%, 5%, and 10%. If the null hypothesis is rejected, a higher positive t-statistic for Model 1 than for Model 2 indicates that Model 2 is more accurate than Model 1.

### 3.10 Concluding remarks

This current chapter outlined various tests and methods/models which will be employed to analyse and address the aim(s) and objectives of this study. Also, what was discussed was the measurement metrics which are used to evaluate forecasting performance or forecasting accuracy of the models included in the study. This chapter will be followed systematically in the following chapter and tests and models will be fully described in terms of statistical analysis in the subsequent chapter.

# CHAPTER 4

# EMPIRICAL ANALYSIS AND RESULTS

## 4.1 Introduction

This chapter addresses the research aim(s) and objectives of the study. The results in this chapter are presented in figures and graphs. Moreover, traditional machine learning models and hybrid machine learning models will be estimated, and their forecasting performance and accuracy will be analysed in this chapter. R studio, Eviews, Matlab and Python are the main statistical software or programming languages employed for executing the analysis of this study.

## 4.2 Preliminary data analysis results

This research section is crucial as a preliminary analysis in chapter 4. It outlines the key characteristics of the variables in a clear and important manner in question which are foreign direct investment (FDI), gross domestic product (GDP), inflation rate, and exchange rate. These characteristics mentioned include, but not limited to, the type of data used, graphical methods employed for analysing the data, and descriptive statistics. These descriptive statistics include, but not limited to, the mean of the variables mentioned, median, standard deviation, skewness, and kurtosis. Furthermore, the study will also test for multicollinearity and nonlinearity of the variables using the following tests: variance inflation factor; and BDS, CUSUM, white neural network test of neglected nonlinearity, Terasvirta neural network test of neglected nonlinearity, McLeod-Li test, Tsay test, and Keenan test respectively.

**Table 4.1: Descriptive statistics**

| Variable(s) | Foreign direct investment | Gross domestic product | Inflation rate | Exchange rate |
|---|---|---|---|---|
| Mean | -631,000,000 | 168,000,000,000 | 9.032817 | 4.95067 |
| Median | 1,101,768 | 134,000,000,000 | 8.558710 | 3.57265 |
| Std. dev. | 2,960,000,000 | 119,000,000,000 | 4.317522 | 4.08431 |
| Skewness | -1.617092 | 0.642650 | 0.196221 | 0.78562 |
| Kurtosis | 6.749146 | 2.025802 | 2.170742 | 2.60835 |
| Jarqua Bera | 606.7720 | 64.37617 | 20.83158 | 64.89826 |
| Probability value | 0.000000 | 0.000000 | 0.000030 | 0.00000 |

Descriptive statistics, as a part of explanatory data analysis, is one section of empirical research that is significant. As noted from the study of Xaba (2014), Schumacker and Lomax (2004) and Kline (2005b) proposed that before execution of the actual analysis of a study, significant issues must be addressed and analysed such as descriptive statistics. These statistics include, amongst others but not limited to, mean, standard deviation and the normality of the original time series data. Table 4.1 reports the results of descriptive statistics of foreign direct investment, gross domestic product, inflation rate and exchange rate.

Based on table 4.1 it is shown that on average, South Africa from January 1970 to June 2019, recorded a negative mean of -631,000,000 implying that, on average, $631,000,000.00 is leaving the country than entering it through foreign direct investment (FDI). This might have been because of investors not having a solid confidence for investing in South Africa. Moreover, GDP of South Africa during the same period on average recorded a figure of 168 billion US dollars and the inflation rate from month to month in the same period recorded on average 9.033 % which was above the targeted values of 3% to 6%. Furthermore, the exchange rate shows that between Jan 1970 and June 2019, one US dollar was R4.95.

Table 4.1 further reveals, based on the median, that foreign direct investment received by South Africa was having a central value of $1,101,768 dollars, quantifying that some of the investments received in the same period might have been less or greater than the median value. Moreover, South African GDP recorded a median value of $134 billion from January 1970 to June 2019 which also implies that some of the recorded figures might be less or greater than the reported 134 billion US dollars, a reported inflation rate of 8.56% for South Africa means that the general price level of goods and services in the country increased by 8.56% over the period of January 1970 to June 2019. Furthermore, R3.57 was the median value for one US dollar in exchange to South African rand.

A standard deviation of FDI is recorded as $2,960,000,000.00 which indicates a significant spread or variability in the data points from the mean. It means that the values in the dataset are dispersed, and they deviate from the mean by an average of $2,960,000,000. Moreover, GDP recorded standard deviation of $119 000 000 000.00 and it implies that the values in the dataset are dispersed significantly from the mean by an average of $119 000 000 000.00. While inflation rate shows standard deviation of 4.32% which means that, on average, the inflation rate deviates from the mean (average) by approximately 4.32% percentage points. Furthermore, it means that, on average, the exchange rate deviates from the mean value by approximately R4.08 units.

Table 4.1 also shows the statistic that can be used for analysing normality of the actual data. Skewness and Kurtosis are used based on their theoretical values to measure the normality of time series data. Based on table 4.1, it is reported that foreign direct investment recorded a negative value of skewness which implies that the data is negatively skewed and deviates from the normality assumption. Whereas, GDP, inflation rate and exchange rate recorded a positive skewness value, it indicates that the time series data of these variables is positively skewed. Kurtosis is also one significant measure, for all the variables shown in table 4.1 has a positive value of kurtosis which shows that the distribution of these variables deviates from normal distribution since they have a heavier tail and a sharper peak than the normal distribution.

Both measures of skewness and kurtosis are also used to measure normality. A kurtosis value of 0 (or close to 0) indicates a distribution with tails and a peak like that of the normal distribution and since for all the variables the value is not close to zero it means they're not normality distributed. Furthermore, A skewness value of 0 (or close to 0) indicates a distribution with equal symmetry on both sides of the mean and since the values of skewness for all the variables is not close to zero it means normality assumption is violated. The results of skewness and kurtosis were supported by the test of Jarqua Bera and the probability value of the test for all the included variables is 0.000, which rejects the null hypothesis of normal distribution at all levels of significance - 1%, 5% and 10% for all the included series.

Time series plot of foreign direct investment, gross domestic product, inflation rate and exchange rate is in figure 4.1. The main aim of plotting time series data is that various characteristics of the data can be identified. Furthermore, figure 4.1 shows the plot of FDI, GDP, inflation rate and exchange rate time series data from January 1970 to June 2019.

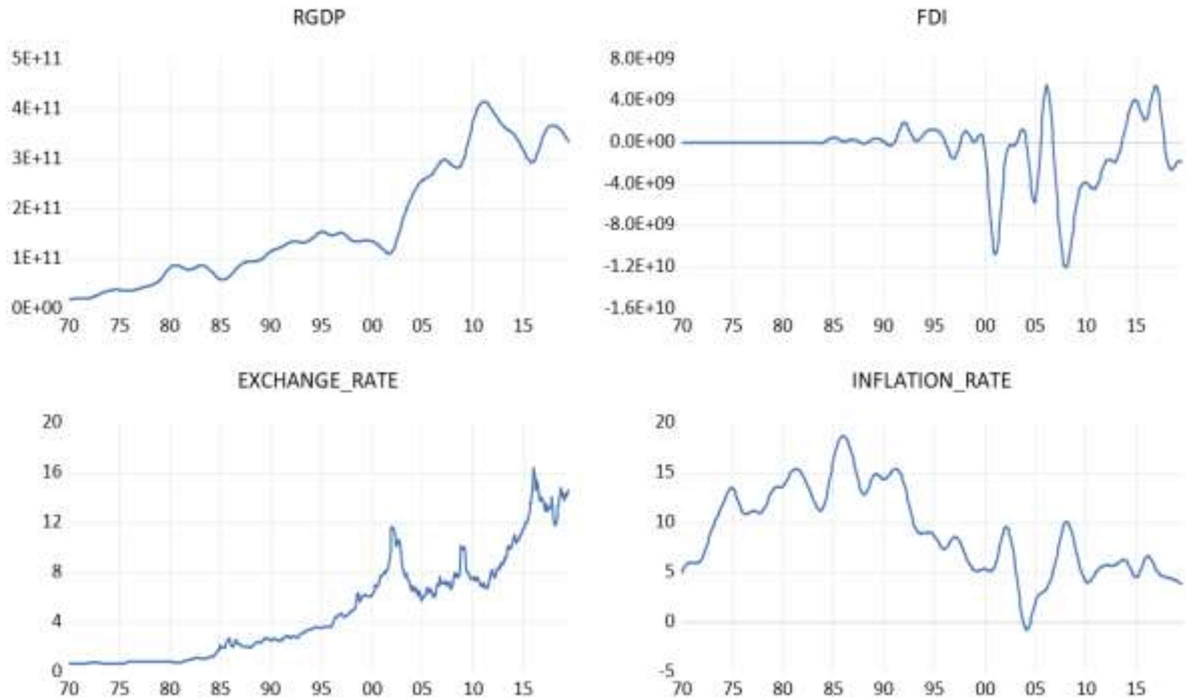


**Figure 4. 1: Time series plot of FDI, GDP, inflation rate and exchange rate from January 1970 to June 2019**

Time series plot of these four time series variables shows that FDI from January 1970 to June 2019 has been having a fairly equilibrium investment into South Africa. At some instances, it was having a negative investment into the country, meaning investors during this time were not investing in the form of FDI investment into South Africa. Furthermore, the plot of FDI is also characterised by irregular movements. Moreover, inflation rate has also been constant on the month-to-month basis, with the aim of keeping it around the range of 3% and 6%. Inflation rate has almost similar irregular movement around the zero line of the time series plot. South African gross domestic product, on the other hand, has been steadily showing an upward trend which then becomes evident that there has been a good output from the country. This is shown by the upward trend of the GDP plot. ZAR/USD exchange rate presents a weakening rand against US dollar, as the plot of the series shows an upward trend. This means that in the period considered in the current study, ZAR has been losing strength against the US dollar. The following section will now analyse whether these series considered in the study are linear or nonlinear. The study employs seven portmanteau nonlinearity tests.

### 4.3 Multicollinearity test results

The study applied variance inflation factor (VIF) to assess multicollinearity between the explanatory variables. Multicollinearity happens when the explanatory variables are collinear. The results are presented in the following table:

**Table 4.2: Variance inflation factor results**

| VIF | Explanatory variables | | |
|---|---|---|---|
| | Inflation rate | Exchange rate | GDP |
| VIF statistics | 1.78491 | 3.57449 | 3.46786 |

## 4.4 Nonlinearity tests results

Nonlinearity tests are employed to test whether a time series variable is linear or nonlinear in nature, which in turn helps in determining the proper model to model and forecast that variable. The current study employs seven portmanteau tests of nonlinearity to investigate the linearity/nonlinearity characteristics of foreign direct investment, gross domestic product, inflation rate and exchange rate. These tests include BDS test, CUSUM tests of stability, white neural network nonlinearity test, Terasvirta neural network nonlinearity test, McLeod-Li test, Tsay test and Keenan test. The results of these tests are reported in table 4.2, 4.3, 4.4, 4.5 and 4.6 and in figure 4.2, 4.3, 4.4, 4.5 and 4.6 respectively.

**Table 4.3: BDS nonlinearity test results**

| Variable(s) | Dimension | BDS Statistic | Std. Error | Z-statistic | P-value |
|---|---|---|---|---|---|
| Foreign direct investment | 2 | 0.191778 | 0.005332 | 35.96496 | 0.00000 |
| | 3 | 0.321925 | 0.008517 | 37.79769 | 0.00000 |
| | 4 | 0.408483 | 0.010201 | 40.04432 | 0.00000 |
| | 5 | 0.464305 | 0.010697 | 43.40352 | 0.00000 |
| | 6 | 0.498727 | 0.010382 | 48.03585 | 0.00000 |
| Gross domestic product | 2 | 0.205089 | 0.002417 | 84.84073 | 0.00000 |
| | 3 | 0.348056 | 0.003829 | 90.88977 | 0.00000 |
| | 4 | 0.447612 | 0.004544 | 98.51277 | 0.00000 |
| | 5 | 0.516930 | 0.004718 | 109.5676 | 0.00000 |
| | 6 | 0.565142 | 0.004532 | 124.6985 | 0.00000 |
| Inflation rate | 2 | 0.200676 | 0.001947 | 103.0536 | 0.00000 |
| | 3 | 0.338888 | 0.003075 | 110.1977 | 0.00000 |
| | 4 | 0.433388 | 0.003637 | 119.1624 | 0.00000 |
| | 5 | 0.497306 | 0.003764 | 132.1296 | 0.00000 |
| | 6 | 0.539815 | 0.003603 | 149.8123 | 0.00000 |
| Exchange rate | 2 | 0.195667 | 0.002705 | 72.33988 | 0.00000 |
| | 3 | 0.331106 | 0.004278 | 77.39813 | 0.00000 |
| | 4 | 0.424577 | 0.005068 | 83.77098 | 0.00000 |

| | | | | | |
|---|---|---|---|---|---|
| | 5 | 0.488858 | 0.005255 | 93.02681 | 0.00000 |
| | 6 | 0.533693 | 0.005041 | 105.8727 | 0.00000 |

Based on the results in table 4.2, it is found that the null hypothesis of linearity is rejected for foreign direct investment, gross domestic product, inflation rate and exchange rate in favour of nonlinearity, this is because for all the dimensions of BDS tests the probability values associated with the test are all 0.0000 for all the mentioned variables. This implies that FDI, GDP, inflation rate and exchange rate are all nonlinear in nature. Moreover, CUSUM test of stability is used to support the results of BDS test of nonlinearity. Based on CUSUM stability test results in figure 4.2 the series is shown to be unstable as the blue line on the graph is outside the critical red line on the graph, this then confirms the results of BDS test that the series of FDI, GDP, inflation rate, and exchange rate are inherently nonlinear.

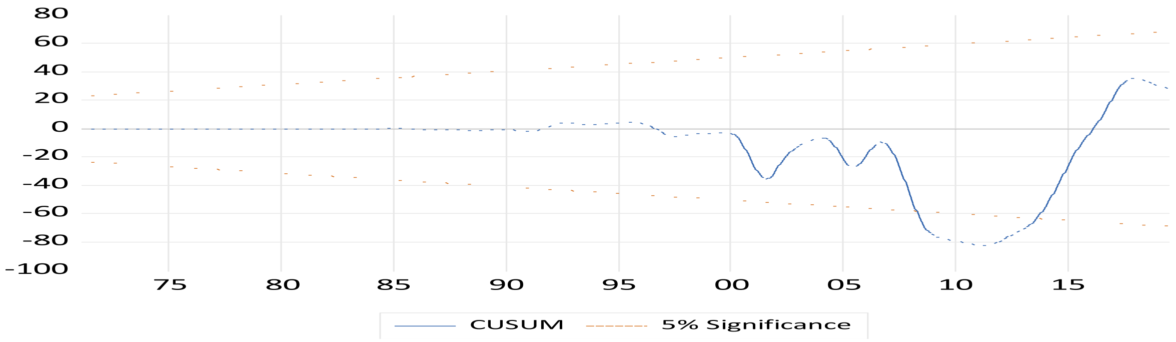


**Figure 4. 2: CUSUM test of stability results**

**Table 4.4: White neural network nonlinearity test results**

| Variable(s) | $F_{calculated}$ | $DF_1$ | $DF_2$ | $P_{value}$ |
|---|---|---|---|---|
| Foreign direct investment | 4.1907 | 10 | 564 | 0.00001 |
| Gross domestic product | 2.858 | 10 | 564 | 0.00176 |
| Inflation rate | 2.5835 | 10 | 564 | 0.00457 |
| Exchange rate | 2.821 | 10 | 564 | 0.00200 |

Another nonlinearity test employed to test for nonlinearity of the variables included in the study is white neural network test of nonlinearity. Table 4.4 presents the results of the test. The white neural network test of nonlinearity employed was used with F distribution. The calculated F distribution probability values of white neural network test of nonlinearity for foreign direct investment is (0.00001), gross domestic product is (0.00176), inflation rate is (0.00457) and exchange rate is (0.00200). Since the probability values are less than the default significant value of 5%, it means the null hypothesis of linearity for all the four time series variables is rejected. Furthermore, this means that the series of all these variables are inherently nonlinear.

**Table 4.5: Terasvirta neural network nonlinearity test results**

| Variable(s) | $F_{calculated}$ | $DF_1$ | $DF_2$ | $P_{value}$ |
|---|---|---|---|---|
| Foreign direct investment | 12.128 | 7 | 585 | 0.00000 |
| Gross domestic product | 4.1225 | 7 | 585 | 0.00019 |
| Inflation rate | 14.071 | 7 | 585 | 0.00000 |
| Exchange rate | 4.783 | 7 | 585 | 0.00003 |

Terasvirta neural network nonlinearity test is another test employed in the study to test for linearity/nonlinearity of the series. Table 4.5 presents the results of the test. The Terasvirta neural network test of nonlinearity employed was used with F distribution. The calculated F distribution probability values of Terasvirta neural network test of nonlinearity for foreign direct investment is (0.00000), gross domestic product is (0.00019), inflation rate is (0.00000) and exchange rate is (0.00003). Since the probability values are less than the default significant value of 5%, it means the null hypothesis of linearity for all the four time series variables is rejected. Furthermore, this means that the series of all these variables are inherently nonlinear. The results also show that Terasvirta neural network test of nonlinearity has more power than the white neural network test of nonlinearity.

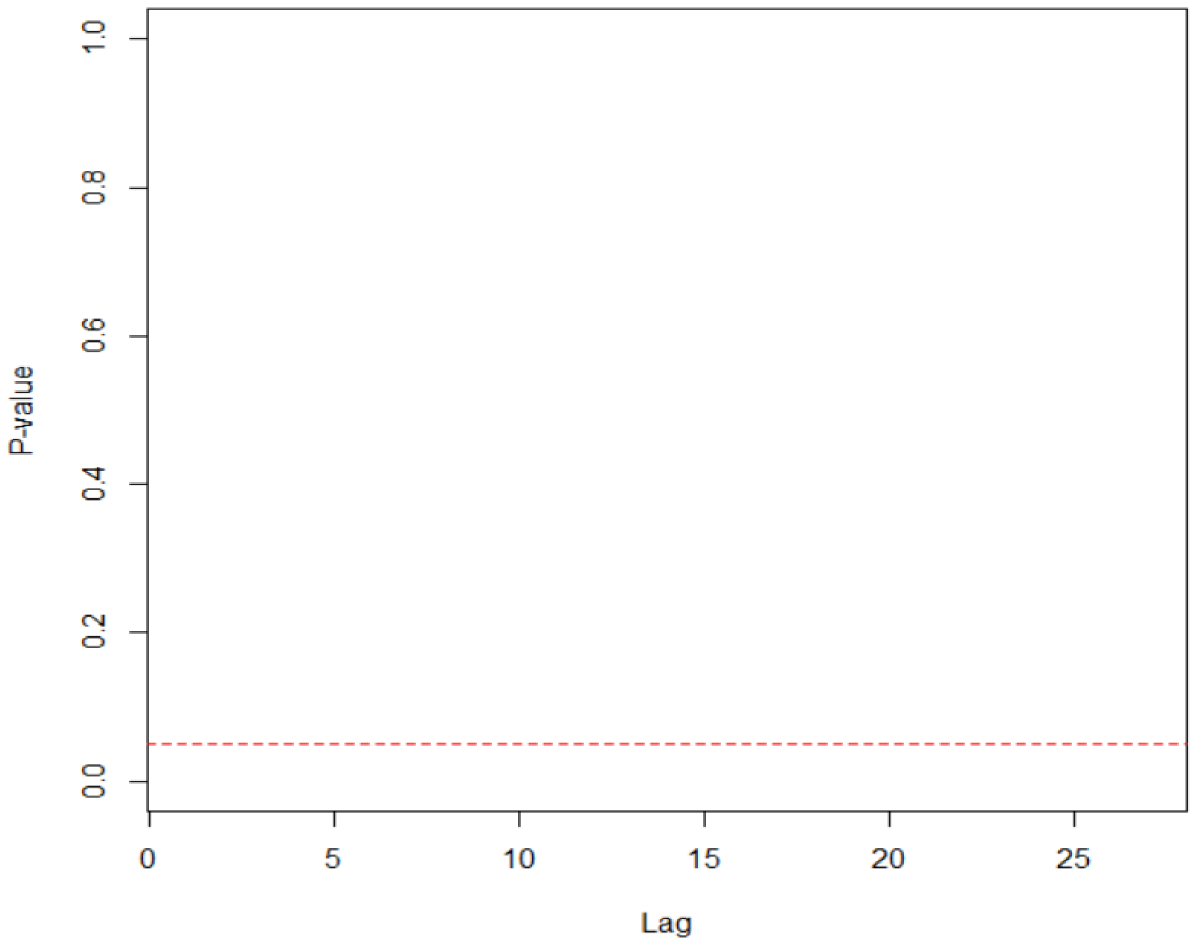


**Figure 4. 3: McLeod-Li nonlinearity test results for foreign direct investment time series data**

Figure 4. 4: McLeod-Li nonlinearity test results for foreign direct investment time series data

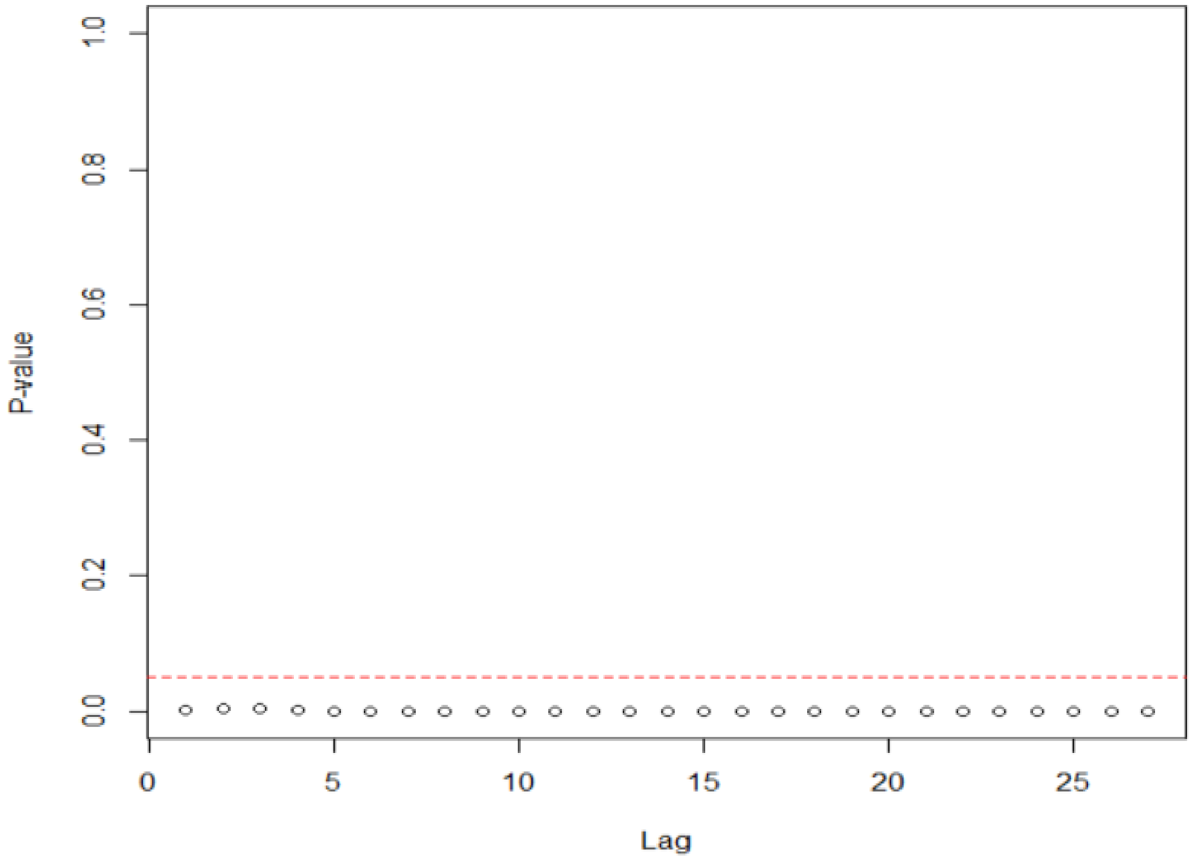


Figure 4. 5: McLeod-Li nonlinearity test results for gross domestic product time series data

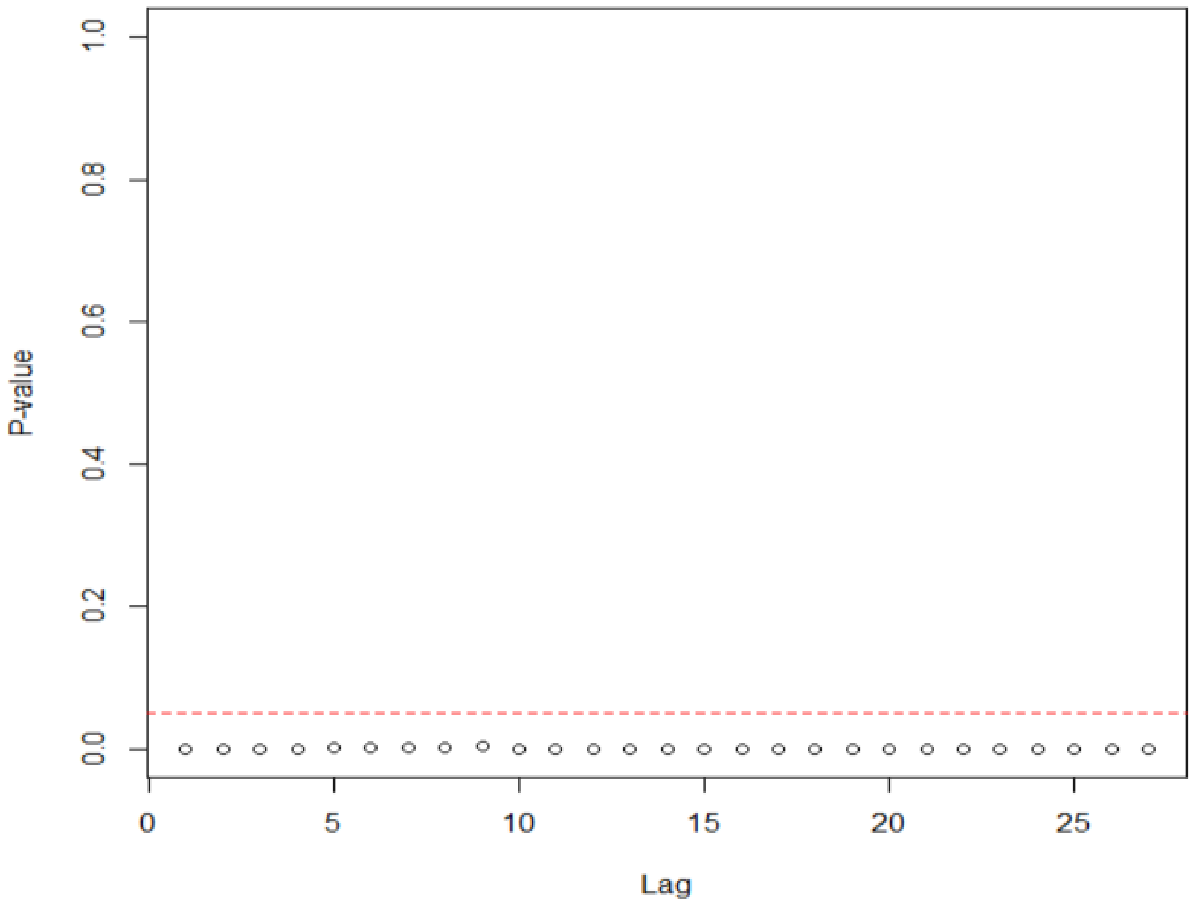

**Figure 4. 6: McLeod-Li nonlinearity test results for inflation rate time series data**

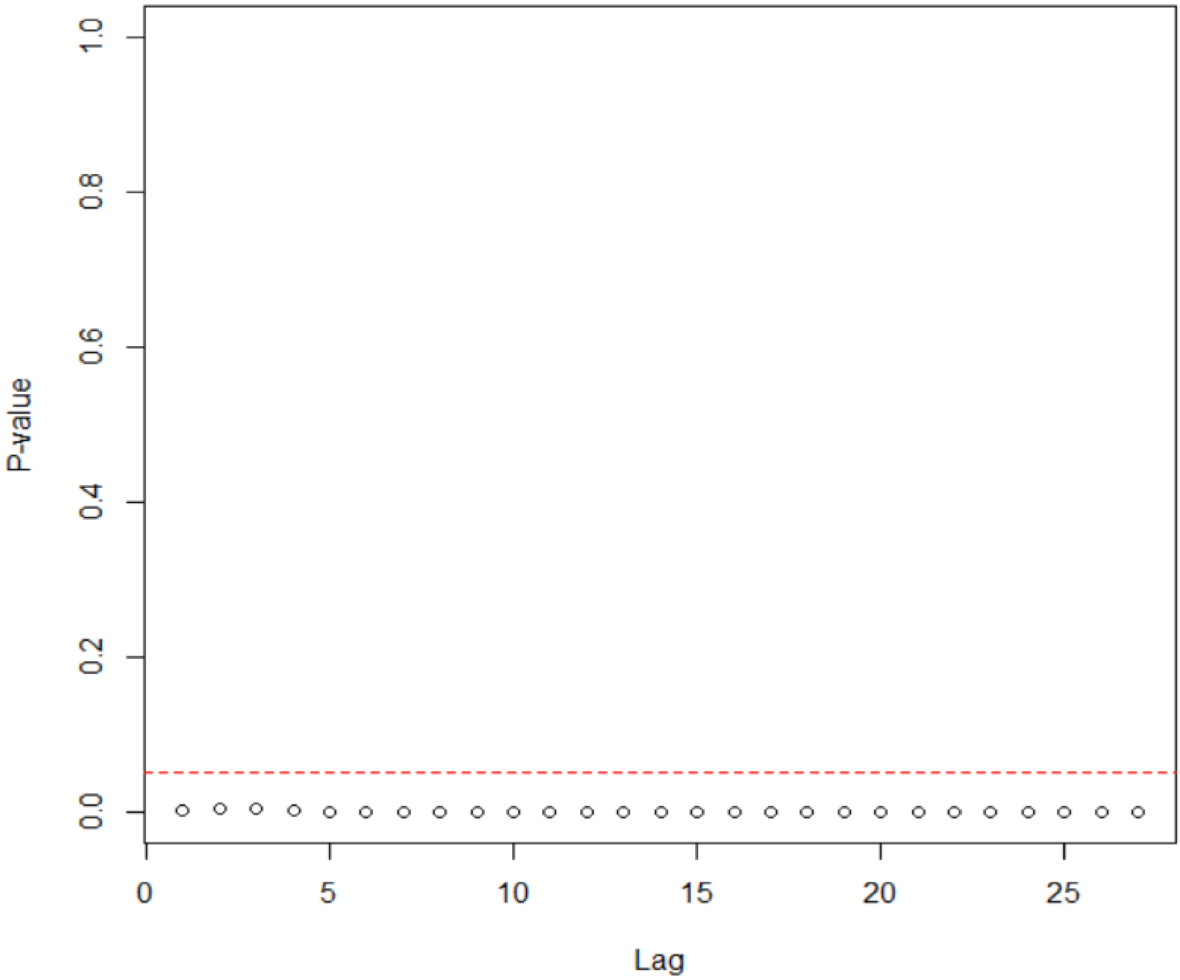


**Figure 4. 7: McLeod-Li nonlinearity test results for exchange rate time series data**

Figures 4.3, 4.4, 4.5 and 4.6 report the results of McLeod-Li test of nonlinearity for foreign direct investment, gross domestic product, inflation rate and exchange rate respectively. These figures show on the vertical line the observed probability values associated with the test and on the horizontal line there is a k number of lags. The red horizontal line shows the critical value associated with the McLeod-Li test. Furthermore, based on the figures 4.3 – 4.6, it is evident that the observed probability values lie beneath the red critical horizontal line, implying that the observed probability values are significant and that the time series variables of foreign direct investment, gross domestic product, inflation rate and exchange rate exhibit the presence of strong conditional heteroscedasticity, which is the evidence of nonlinearity in all the four mentioned time series variables.

**Table 4.6: Tsay nonlinearity test results**

| Variable(s) | Test statistic | Order | $P_{value}$ |
|---|---|---|---|
| Foreign direct investment | NA | NA | NA |
| Gross domestic product | 3.045 | 1 | 0.08152 |
| Inflation rate | NA | NA | NA |
| Exchange rate | 1.957 | 1 | 0.16230 |

Another test utilised to test for nonlinearity in the variables included in the study is the test of (Tsay, 1986). Table 4.6 above reports the results of Tsay test and according to the results presented, foreign direct investment and inflation rate have not been able to be analysed as the variables contain negative values in their observation. Tsay test of nonlinearity calculates using a square root thereby making it difficult to test whether FDI and inflation rate are nonlinear or not. However, Tsay test of nonlinearity shows that both gross domestic product and exchange rate are linear in nature as their reported observed probability values are greater than the default significant level.

**Table 4.7: Keenan nonlinearity test results**

| Variable(s) | Test statistic | Order | $P_{value}$ |
|---|---|---|---|
| Foreign direct investment | NA | NA | NA |
| Gross domestic product | 23.07532 | 1 | 0.00000 |
| Inflation rate | NA | NA | NA |
| Exchange rate | 1.849049 | 1 | 0.17441 |

Another test utilised to test for nonlinearity in the variables included in the study is the test of (Keenan, 1985) presented results in table 4.7. As it was the case with Tsay test of nonlinearity, Keenan test of nonlinearity also takes the square root of the original time series observations resulting in foreign direct investment and inflation rate not being successfully analysed for nonlinearity since they have negative values in their observations. However, gross domestic product and exchange rate, according to the results of Keenan, shows that the former variable is nonlinear in nature as its observed probability value is less than the standard significant level of 0.05, indicating that gross domestic product is nonlinear in nature. In contrary, exchange rate has an observed probability value greater than the significant level of 0.05 indicating that the variable is linear in nature.

The study employed seven portmanteau nonlinearity tests for testing whether the time series variables of foreign direct investment, gross domestic product, inflation rate and exchange rate are linear/nonlinear in nature. Out of the seven employed nonlinearity tests, five have indicated that all the variables are inherently nonlinear and the other two have shown that some of the variables are linear. The consensus is that the variables are nonlinear and can be modelled using nonlinear models since their characteristics are nonlinear in nature.

## 4.5 Model estimation and results

This section presents the estimation and results of traditional machine learning models and the machine learning based evolutionary algorithms hybrid models. These models have a great success in terms of the modelling and forecasting in various disciplines of research as outlined in the literature review chapter, thereby leading to the current study adopting and employing these models for modelling and forecasting foreign direct investment (FDI).

## 4.6 Traditional machine learning models

“Machine learning is one of the fastest growing areas of computer science, with far-reaching applications” (Shalev-Shwartz and Ben-David, 2014). Machine learning models or algorithms have been widely and extensively used in various research disciplines. The current study adopts and employs these machine learning models for modelling and forecasting foreign direct investment. These models include artificial neural network (ANN), support vector regression (SVR), and least squares support vector regression (LS-SVR). The following subsections presents the estimation and results of the above-mentioned machine learning (ML) models.

### 4.6.1 Artificial neural network (ANNS) model

As chapter 2 of the current study has shown, Artificial Neural Network (ANN) has been widely used for time series forecasting. This model has shown a good generalisation ability and forecasting ability to various research disciplines. The current study adopted ANN for modelling and forecasting a time series data of foreign direct investment. *R studio* is the main employed programming language for the coding of ANN and analysing of foreign direct investment.

In coding and employing Artificial Neural Network in *R studio* for modelling and forecasting a time series data of foreign direct investment various packages are available. However, for the purpose of the current study, a package in *R studio software* called *neuralnet* package was the one installed and called for the purpose of analysing foreign direct investment using artificial neural network (ANN). The *neuralnet* package was introduced by Fritsch et al. (2016) and it is a very flexible and easy to use package.

In training the artificial neural network (ANN) model, the data was scaled to the interval of [0,1] so that the values which are large cannot affect the values which are small, and the interval of the scaled values were in line with the interval of the employed nonlinear activation function. Furthermore, the activation function employed in the hidden layer of the ANN model is called logistic activation function, and the study also employed an algorithm of resilient backpropagation. Since the current research is quantitative in nature, the sum of squared error (SSE) is employed as the measure of error in the training of the model and the study adopted and employed the default threshold value of 0.01.

In modelling and forecasting foreign direct investment (FDI), twenty-four (24) models of artificial neural networks (ANNs) were trained in the current study. All twenty-four (24) trained models were having various hyper-parameter settings, such as the number of hidden layers, number of hidden nodes on each hidden layer etc. Furthermore, this showed that when the number of hidden layers in connection with their respective hidden nodes are increased, the more forecasting accuracy of ANN is improved. However, this was only evident until the architecture of the artificial neural network (ANN) model reached three hidden layers with their respective hidden nodes, in this case the error of the model started to increase. The current study chose a model of artificial neural network (ANN) with the least error out of the twenty-four trained models. The following figure, 4.7, shows the architecture of ANN chosen for modelling and forecasting foreign direct investment in the study.

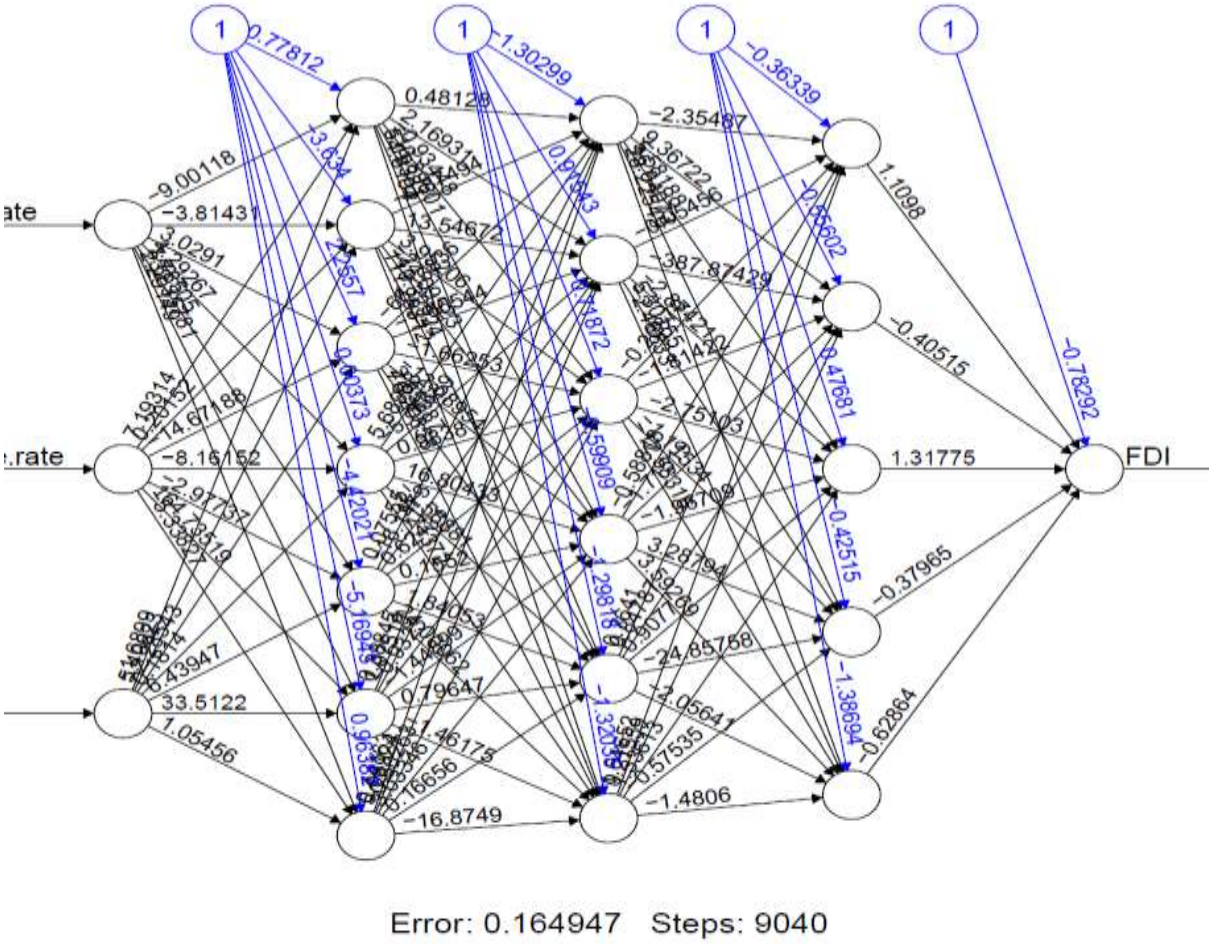


**Figure 4. 8: Artificial neural network architecture for modelling and forecasting foreign direct investment (FDI)**

The architecture of the best performing ANN model shows that the artificial neural network (ANN) model employed in the study is a three-layered feed-forward artificial neural network (ANN) model. The chosen model of ANN with the above architecture converged and the above architecture shows that it took the training process of the model *9040* steps for the model to converge and obtain a minimum error value of 0.164947 which was the smallest error amongst the twenty-four trained artificial neural network (ANN) models. The ANN model is a three-layered neural network model, the model has a three-input layer since the study only considered three explanatory time series variables. Furthermore, the input layer is made up of each input node, this means that per explanatory variable, there is one input node. Since the model is a feed-forward artificial neural network (ANN), the input layer is connected to hidden layer(s). Moreover, the model in the study has three hidden layers with logistic function employed in the study as an activation function. The three hidden layers employed have seven, six and five hidden nodes respectively and the hidden nodes are interconnected by the various weights of the model. Moreover, the model is also made up of four constants which are also connected to the hidden nodes of the first layer through to the output nodes. The output layer is made up of the response variable which is foreign direct investment (FDI) and has only one output node since the study only considered one time series response variable.

### 4.6.2 Support vector regression (SVR) model

Support vector regression is one of the widely employed machine learning (ML) models for modelling and forecasting time series data, this model has shown great success in terms of modelling and forecasting time series data. The current study adopts this machine learning (ML) model and employs it for modelling and forecasting foreign direct investment (FDI). *R studio* is the main programming language used for the analysis of FDI using SVR as the model for modelling FDI time series data.

For building and training support vector regression model, the *R* package of *e1071* is employed for the purpose of using SVR for modelling and forecasting time series data of foreign direct investment (FDI). The *e1071 R* package was proposed by (Meyer et al., 2015). The original time series data was scaled to the interval of $[0,1]$ so that the large values present in the data do not affect the small values thereby leading to the biased modelling of the data. Furthermore, as it is a pre-requisite in machine learning discipline, the original data was partitioned to both the training set and testing set. This was for the purpose that the training set (80%) can be used to train the SVR model, and the testing set (20%) can be used for accuracy of the best selected SVR model in terms of forecasting the unseen test data.

The study trained the initial model of SVR machine learning model and after training the model, root mean squared error (RMSE) was used as the measure of accuracy for the trained model and it was shown that the initial trained model had *0.1090777* as the value of RMSE. Moreover, in the study, the optimal search of hyper-parameters of SVR model was undertaken to ensure an improvement of accuracy of the initial trained SVR model.

In improving the accuracy of the initially trained model, the SVR model was tuned. This allowed the study to set an interval of tuned hyper-parameters of the study which are epsilon $(\varepsilon)$ and cost $(C)$ to $seq(0,1,0.1)$ and $(1:100)$ respectively. Thus, the optimal hyper-parameter values of both epsilon and cost chosen are 0.4 and 100 respectively with a best performance of the model at *0.0050630*.

From the chosen tuned model of SVR, the study trained 1100 models of SVR and the best model of SVR which was chosen had a root mean squared error (RMSE) of *0.0449601* which shows an improvement of accuracy from the initially trained SVR model which had an RMSE value of *0.1090777*. The overall chosen hyper-parameter values of epsilon, cost, and gamma for the chosen best model of SVR are 0.4, 100 and 0.333 respectively, with radial basis function used as a kernel function and the model having *99* support vectors.

The other significant model parameters are the coefficients of the SVR model shown in chapter 3 in equation 3.37. These parameters are estimated to be the coefficients of equation 3.37 in chapter 3. Since the study had three explanatory time series variables, this study will have three $(w)$ coefficients with a constant value of $(b)$. Equation 4.1 shows the estimated coefficients of the SVR formalism linear estimation function:

$$w = [1,1:3]\begin{Bmatrix} InflationRate & ExchangeRate & GDP \\ 10.51 & -19.36 & -2.87 \end{Bmatrix} \quad b = 1.6275008 \qquad 4.1$$

The estimated coefficients of the SVR considered formalism linear estimation shows that there is a positive relationship between inflation rate and FDI and a negative relationship between exchange rate, GDP to FDI, with a constant show of a positive sign to the FDI. Furthermore, the following figure, 4.8, shows the performance of SVR with its respective parameters of epsilon and cost hyper-parameters. The best performance of the model in terms of the chosen hyper-parameters is shown in the darker blue area where the epsilon value is given as 0.4 and cost value is given as 100 respectively.

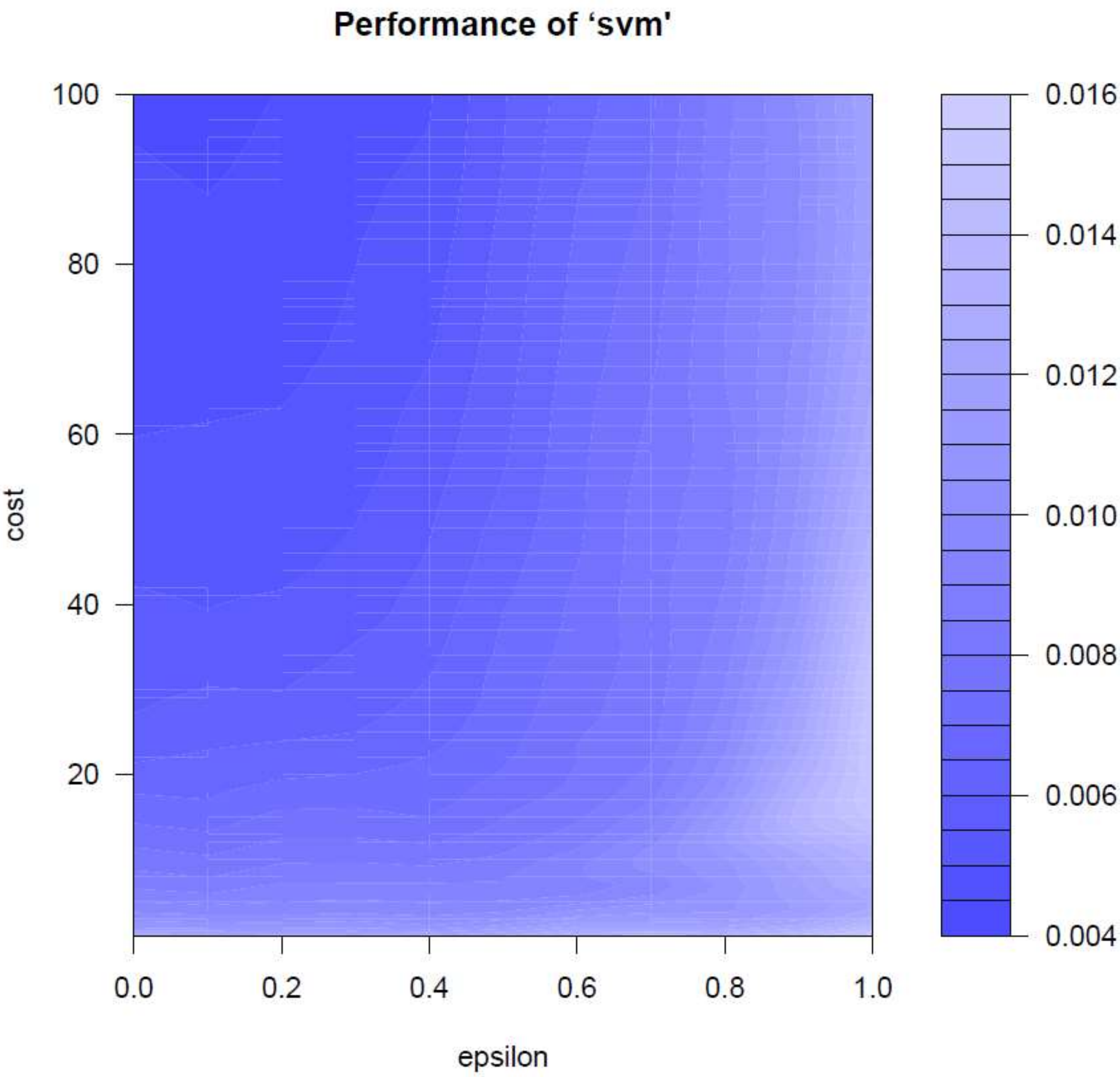


**Figure 4. 9:SVR model performance for modelling and forecasting FDI**

### 4.6.3 Least squares support vector regression (LSSVR) model

Over the recent years, in-time series modelling and forecasting least square support vector regression (LSSVR) has received a lot of attention and is being employed successfully for time series forecasting and in other disciplines. The literature in the current study has shown that this machine learning (ML) model has had great success in terms of modelling and forecasting time series data. The current study adopted and employed this ML model for modelling and forecasting foreign direct investment time series data. The programming language or software employed for the analysis, modelling and forecasting of FDI is Python using libraries of tensor-flow and keras.

The original time series data was scaled to the interval of $[0,1]$ so that the large values present in the data does not affect the small values in the data thereby leading to the biased modelling of the data. Furthermore, as it is a pre-requisite in machine learning discipline, the original data was partitioned to both the training set and testing set respectively. This was for the purpose that the training set (80%) can be used to train the LSSVR model, and the testing set (20%) can be used for testing the accuracy of the best selected LSSVR model in terms of forecasting the unseen test data.

In the context of the current study, LSSVR model is employed to model and forecast foreign direct investment (FDI) using the radial basis function as a kernel function. The LSSVR model was trained using the method of trial and error with different values for the gamma hyper-parameter. This process was done so that it could be clear that which value of the gamma hyper-parameter is best for accurate forecasting of the LSSVR model.

The performance of the LSSVR model significantly improved in terms of forecasting performance when the gamma value was set to be *0.20* instead of the *0.01* default value. This chosen model of LSSVR was chosen from a total of *63* LSSVR trained models and the model with hyper-parameter value of gamma which was *0.20* was the best trained model. The model showed a significant improvement in terms of generalisation. The forecasting results in terms of error measurements of the model are reported in table 4.8.

## 4.7 Machine learning based genetic algorithm hybrid models

Various machine learning models employ trial and error method which is time consuming to select the proper, relevant and optimised hyper-parameters of the models. This method does not however guarantee a selection of optimal hyper-parameters. Therefore, search algorithms have been developed in the past to select optimised hyper-parameters which enable these models to have more accurate forecasting performance. In this study, both artificial neural networks (ANN) and support vector regression (SVR) are optimised employing genetic algorithm (GA). The limitation of optimising LSSVR model came as a result of the lack of programme syntax.

### 4.7.1 Artificial neural network based genetic algorithm

Optimisation of various machine learning models, specifically artificial neural network (ANN), can be done by employing evolutionary algorithms. The current study employed genetic algorithm (GA) for search and optimisation of hyper-parameters of artificial neural network (ANN). The appropriate selection of hyper-parameters is important for the optimal forecasting performance and accuracy of the model, failing to do that can lead to poor forecasting performance and forecasting accuracy.

The appropriate hyper-parameter optimisation of ANN model using GA was based on back-propagation algorithm. The hybrid model of genetic algorithm (GA) based artificial neural network (ANN) was executed using the TPOT in python environment. The optimal solution of the model hyper-parameters using GA was obtained after 50 iterations. The optimal hyper-parameters chosen by genetic algorithm are three hidden layer sizes with 50, 100, 50 as neurons per layer; hyperbolic tangent as activation function; alpha with the optimal value of 0.0001; and learning rate used is adaptive.

After the genetic algorithm (GA) was used to search for the above optimal parameters of the model ANN, the ANN optimal model was built and trained on the train data. The $RMSE$ of train data used in the study was at its minimum value which was found to be 0.1001. Furthermore, the study tested the hybrid model of GA-ANN against the unseen data of test data. The error measurements found on the test data showed that the model had a good forecasting and generalisation of the test data. This shows that the forecasting performance of the GA-ANN model is fair excellent. The results made by the GA-ANN hybrid model are reported in table 4.8.

### 4.7.2 Support vector regression based genetic algorithm

In machine learning algorithms or models, to be precise, SVR, there is no standard procedure to determine the hyper-parameters of the model(s). The procedure employed often is the trial-and-error method which is time consuming. For the same reason, the current study optimised the hyper-parameters of SVR model using genetic algorithm (GA). As it was stated in the study of Xia et al. (2017), “the inappropriate parameters in SVR lead to over-fitting”. The first step involved employing 5-fold cross validation technique and later genetic algorithm was then used to search for optimal hyper-parameters of SVR model.

The study, in terms of the process of searching for optimal hyper-parameters, was operated with 50 population size and the overall optimal fitness was reached at generations 10. Thus, the individuals at generation 10 produced the following optimal hyper-parameters of the model SVR, which are as follows; cost is $C = 5.807201$, gamma is $\sigma^2 = 1.660981$ and epsilon is $\varepsilon = 0.1818941$. The other parameters of GA are crossover probability value of 0.8, elitism value of 2 and mutation probability value of 0.1. The maximum iterations for which the genetic algorithm was trained is 10 with a fitness function value of -0.00158.

After the genetic algorithm (GA) was used to search for the above optimal parameters of the model SVR, the SVR optimal model was built and trained on the train data. The $RMSE$ of train data used in the study was at its minimum value which was found to be 0.034445. Furthermore, the study tested the hybrid model of GA-SVR against the unseen data of test data. The error measurements found on the test data showed that the model had a good forecasting and generalisation of the test data. This shows that the forecasting performance of the GA-SVR model is excellent. The results made by the GA-SVR hybrid model are reported in table 4.7. The results will further be compared with other results obtained by other models included in the study in section 4.5.

### 4.8 Comparison of model performance results

The current study employed five machine learning models; of which three of the five machine learning models are traditional machine learning models and the two out of five are evolutionary algorithm-based machine learning hybrid models. These machine learning models, to the knowledge of the researcher, have not been employed to model and forecast foreign direct investment into South Africa. Therefore, the current study sought to determine whether these models could be of any use to modelling and forecasting financial time series of foreign direct investment. In terms of measuring the forecasting performance of these machine learning models; the study employed the following error measurements: mean squared error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and mean absolute scaled error. These error measurements were discussed in chapter 3. Table 4.8 summarises the results of the error measurements of these machine learning model for modelling and forecasting foreign direct investments.

**Table 4.8: Forecasting performance of traditional machine learning models and hybrid machine learning models**

| Models | Performance criterion | | | | |
|---|---|---|---|---|---|
| | MSE | RMSE | MAE | MAPE | MASE |
| ANN | 0.00180 | 0.04248 | 0.02817 | 5.67016 | 0.73075 |
| SVR | 0.00225 | 0.04743 | 0.03199 | 9.57313 | 0.082976 |
| LSSVR | 0.02527 | 0.15898 | 0.10125 | 15.66068 | 2.62555 |
| Machine learning based genetic algorithm hybrid models | | | | | |
| Hybrid models | Performance criterion | | | | |
| | MSE | RMSE | MAE | MAPE | MASE |
| ANN-GA | 0.86663 | 0.93093 | 0.69106 | 392.35030 | 17.92044 |
| SVR-GA | 0.01437 | 0.11987 | 0.07797 | 42.07455 | 2.02197 |

From the results reported in table 4.8 particularly for traditional machine learning models; five error measurements are used to measure the forecasting performance of the models included in the study. These error measurements are MSE, RMSE, MAE, MAPE and MASE. The results show that ANN model is the best performing model in forecasting the foreign direct investment (FDI) as the model has the least statistic for all the five error measurements. Closely following the ANN model is the SVR model which has the second-best forecasting performance, and it should also be noted that the error measurements of these two models are closely related. Moreover, the LSSVR model in the traditional machine learning models is the worst performing model in forecasting foreign direct investment (FDI) as the statistic for the error measurements is almost twice the ones to the counter parts models. Even though LSSVR model is the worst performing model, it still shows a great generalisation ability and shows great capabilities in forecasting FDI.

Genetic algorithm (GA) was employed so that it can be seen whether it improves forecasting performance of the two traditional machine learning models. Based on the two-hybrid models employed in the study for modelling and forecasting foreign direct investment, it is evident that genetic algorithm (GA) based support vector regression (SVR-GA) hybrid model has the superior forecasting performance/accuracy over the genetic algorithm based artificial neural network (ANN-GA). This is the case as the five employed error measurement chooses SVR-GA as the model that has superior forecasting performance. On the overall, it is clearly shown that ANN model has the best forecasting accuracy and performance against all the models included in the study. Furthermore, it is evident that the GA based machine learning models (ANN-GA and SVR-GA) have performed worse than their traditional counter parts (ANN and SVR). This shows that other evolutionary algorithms can also be employed in the future for optimisation purposes. Therefore, it can be concluded that genetic algorithm (GA) did not do such a great work for improving the forecasting performance/accuracy of the included models. The following section will deal with assessing the forecasting accuracy of the models.

## 4.9 Comparing model accuracy results

According to Diebold (2015), the question that the Diebold-Mariano (DM) test answers is whether one model is really better than another, or is it pure luck? A smaller error matrix does not necessarily mean that one model has better predictive power than another, because differences in the error matrices may not be statistically significant (Xaba, 2014). The need to assess the predictive accuracy of models is evident, and in the current study the Diebold-Mariano test was used to measure the predictive accuracy of models. A summary of the test results is shown in Table 4.9.

**Table 4.9: Diebold-Mariano test of forecasting accuracy results**

| Variable | Hypothesis | t-statistic | P-value |
|---|---|---|---|
| FDI | SVR vs. ANN | 0.76165 | 0.44780 |
| FDI | SVR vs. LSSVR | -3.9638 | 0.00013 |
| FDI | LSSVR vs. ANN | 4.0327 | 0.00010 |
| FDI | ANN-GA vs. ANN | 4.50130 | 0.00002 |
| FDI | ANN vs. SVR-GA | -3.50580 | 0.00064 |
| FDI | ANN-GA vs. SVR | 4.49890 | 0.00002 |
| FDI | SVR vs. SVR-GA | -3.39990 | 0.00092 |
| FDI | ANN-GA vs. LSSVR | 4.37420 | 0.00003 |
| FDI | SVR-GA vs. LSSVR | -3.9953 | 0.00011 |
| FDI | SVR-GA vs. ANN-GA | 3.5058 | 0.00064 |

According to the results presented in the table 4.9 above, the null hypothesis of equal forecasting accuracy between SVR and ANN is accepted since the calculated p-value (0.44780) is greater than the 5% significance level. In that case, it can be concluded that SVR and ANN models have equal accuracy in forecasting FDI time series data.

According to the results, the null hypothesis that ANN-GA and LSSVR models have equal forecasting accuracy is rejected since the calculated p-value (0.00003) is less than the default significance level of 5%. In that case, since the t-statistic (4.37420) is positive, it can be concluded that the estimated LSSVR model is more accurate in forecasting FDI than ANN-GA model.

According to the results, the null hypothesis that LSSVR and ANN models have equal forecasting accuracy is rejected since the calculated p-value (0.00010) is less than the default significance level of 5%. In that case, since the t-statistic (4.03270) is positive, it can be concluded that the estimated ANN model is more accurate in forecasting FDI than LSSVR model. Furthermore, the null hypothesis that ANN-GA and ANN models have equal forecasting accuracy is rejected since the calculated p-value (0.00002) is less than the default significance level of 5%. In that case, since the t-statistic (4.50130) is positive, it can be concluded that the estimated ANN model is more accurate in forecasting FDI than the hybrid model of ANN-GA.

According to the results, the null hypothesis that ANN and SVR-GA models have equal forecasting accuracy is rejected since the calculated p-value (0.00064) is less than the default significance level of 5%. In that case, since the t-statistic (-3.50580) is negative, it can be concluded that the estimated ANN model is more accurate in forecasting FDI than SVR-GA hybrid model. Furthermore, the null hypothesis that SVR-GA and LSSVR models have equal forecasting accuracy is rejected since the calculated p-value (0.00011) is less than the default significance level of 5%. In that case, since the t-statistic (-3.9953) is negative, it can be concluded that the estimated SVR-GA model is more accurate in forecasting FDI than LSSVR hybrid model.

According to the results, the null hypothesis that SVR-GA and ANN-GA hybrid models have the same forecasting accuracy is rejected since the calculated p-value (0.00064) is less than the standard significance level of 5%. In that case, since the t-statistic (3.5058) is positive, it can be concluded that the estimated ANN-GA hybrid model is more accurate in predicting FDI than SVR-GA hybrid model. Moreover, the null hypothesis that ANN-GA and SVR models have the same forecasting accuracy is rejected since the calculated p-value (0.00002) is less than the standard significance level of 5%. In that case, since the t-statistic (4.49890) is positive, it can be concluded that the estimated SVR model is more accurate in forecasting FDI than ANN-GA hybrid model.

According to the results, the null hypothesis that the SVR and SVR-GA models have the same predictive accuracy is rejected because the calculated p-value (0.00092) is less than the standard 5% significance level. In this case, the t-statistic (-3.39990) is negative, so it can be concluded that the estimated SVR model is more accurate than the hybrid SVR-GA model in predicting FDI. Furthermore, the null hypothesis that the SVR and LSSVR models have the same predictive accuracy is rejected because the calculated p-value (0.00013) is less than the default 5% significance level. In this case, since the t-statistic (-3.9638) is negative, it can be concluded that the estimated SVR model is more accurate than the LSSVR model in predicting FDI.

### 4.10 Concluding remarks

In this chapter the empirical analysis of foreign direct investment is presented. The researcher also did exploratory data analysis to understand the underlying characteristics of the financial time series data used in the study. The study objectives outlined in chapter one is addressed by the analysis. Furthermore, the best performing model(s) in the study in terms of modelling and forecasting foreign direct investment was provided. A summary of findings and recommendations for future research will be presented in the next chapter.

# CHAPTER 5

# CONCLUSION AND RECOMMENDATIONS

## 5.1 Introduction

The primary objective/aim of the current thesis has been to seek to assess whether genetic algorithm (GA) can improve forecasting efficiency and performance of the included machine learning models in a hybrid system in modelling and forecasting foreign direct investment. This is the case because the optimal performance of machine learning models depends largely on the correct setting of hyper-parameter values. This chapter presents a summary of key findings as well as recommendations for future research.

## 5.2 Research objectives

The current study had an aim and research objectives that were supposed to be achieved. The following are the objectives of the study:

- To statistically test the linearity/nonlinearity nature of foreign direct investment time series.
- To statistically test the forecasting efficiency and performance of traditional machine learning models on foreign direct investment time series.
- To statistically test whether genetic algorithm improves forecasting accuracy of machine learning models on foreign direct investment.
- To compare the efficiency and performance of traditional machine learning models and hybrid machine learning models on foreign direct investment.
- To compare the predictive power of both the traditional machine learning models and the hybrid machine learning models on foreign direct investment.

## 5.3 Summary of research findings

As a first step in analysing data, preliminary data analysis was conducted in the current study to investigate the nature of the data employed in the study. Descriptive analysis was conducted, and it showed that GDP had the highest mean and median, with FDI recording the lowest mean. The first objective was to determine whether FDI is linear/(or) nonlinear in nature and to address this objective various nonlinear test were conducted and it was found out that FDI is nonlinear in nature. Furthermore, five nonlinear supervised machine learning models were employed to model and forecast FDI series. The five models were made of three traditional machine learning models (ANN, SVR and LSSVR) and two machine learning based evolutionary algorithm hybrid models (ANN-GA and SVR-GA). All the five models were estimated as discussed in chapter 4 and for comparing forecasting performance, the study employed the following error measurements: MSE, RMSE, MAE, MAPE and MASE. It was demonstrated in the results that in the overall picture, ANN model is the best model for modelling and forecasting foreign direct investment time series data. Furthermore, the study found that genetic algorithm did not improve the forecasting accuracy of the two optimised machine learning models (ANN and SVR).

The last objective sought to compare the traditional machine learning models and machine learning based evolutionary hybrid models in their predictive powers using the notion that a lower error measurement value does not necessarily translate to a superior forecasting ability. To address this objective, the study employed Diebold-Mariano test and it was found that in the overall, the traditional machine learning models had a superior predictive power in comparison to the hybrid models.

## 5.4 Achievement(s) of research objectives

The study had objectives that needed to be achieved and in doing so, various tests and models were employed in the study to aid in the achievement of the objectives put forth in the study. The following are the findings per research objective(s) of the study.

- The study established that the foreign direct investment series are nonlinear in nature as this was achieved using the seven nonlinearity tests outlined in chapter 3. The tests employed include: BDS test, CUSUM test, White neural network test of nonlinearity, Teresvirta neural network test of nonlinearity, Tsay test, Keenan test and McLeod-Li test.

- Three traditional machine learning models were estimated for modelling and forecasting foreign direct investment and it was found that amongst the three models; ANN model was the model with the best forecasting efficiency and performance followed closely by SVR model. And the least performing model was the LSSVR model which had a poor forecasting efficiency and performance on foreign direct investment.

- Based on the error measurements which tests the forecasting performance of a particular model, it is found that in the current study, genetic algorithm (GA) did not improve the forecasting accuracy of both ANN and SVR respectively. This is so because the traditional machine learning models still have better forecasting performance and accuracy than the hybrid models based on the results of the error measurements.

- The overall comparison of the forecasting performance of the five models in the study shows that traditional machine learning models have better forecasting performance compared to hybrid models. The ANN model is the best performing model for modelling and forecasting foreign direct investment time series data.

- This objective is based on the results of Diebold-Mariano test and based on the results of the mentioned test, it is found that traditional machine learning models (ANN and SVR) have better predictive power in the study than hybrid models in exception to the predictive power shown by the SVR-GA hybrid model. However, in the overall picture, it is evident that the predictive power is with the traditional machine learning models.

## 5.5 Discussions of key findings

The findings of the current study show that foreign direct investment is nonlinear in nature and this can then be employed as a recommendation that nonlinear models need to be adopted more and used for studying foreign direct investment into South Africa. As the existing knowledge has documented, those conventional linear models have a limitation in modelling a series that is nonlinear in nature. Furthermore, the study showed artificial neural network to be the best performing model with high forecasting performance and accuracy as compared to the other models. This signifies the ability of ANN in modelling time series data and the ability of the model in learning the characteristics of the data while being able to generalise the unseen data. For several years since the development of ANN model, it has been doing well in forecasting time series data but not in the setting of foreign direct investment as it is not documented before that ANN has ever been employed to model the variable in question. Nevertheless, it shows that ANN can model and perform well in time series modelling and forecasting.

Genetic algorithm (GA) was employed to improve forecasting performance and accuracy of the two supervised nonlinear machine learning models (ANN and SVR). The findings of the performance of these hybrid models have shown that, based on the error measurements, there was no improvement in the forecasting performance and accuracy of the models. This therefore shows that GA was not able to improve the forecasting accuracy of the models. These findings contradict the existing knowledge as it is documented that optimising of the hyper-parameters of the models will improve its forecasting accuracy. Furthermore, while genetic algorithms have been successful in other domains, their application to foreign direct investment forecasting, in this case, did not lead to better results. This finding may encourage further research to explore alternative approaches or evolutionary algorithms that could potentially provide better results in FDI forecasting. The study also shows that the predictive powers of machine learning models are more superior to the ones of hybrid models. These findings of the study offer valuable insights into the effective use of machine learning models for FDI forecasting and emphasize the importance of selecting appropriate modelling techniques based on the data's characteristics and the forecasting goals. Further research could explore alternative hybrid approaches or other machine learning models to potentially enhance FDI forecasting accuracy.

## 5.6 Strengths and limitations

The machine learning models used in the study are commendable in terms of being non-parametric and nonlinear in nature. This allows the research to not take too much attention in the underlying function of the time series data employed in the study. Furthermore, the other strengths of the models are that they have a great generalisation ability, with high forecasting accuracy and this allows these models to perform better in forecasting problems. However, with all the strengths of the models highlighted, there are some limitations attached to the models. These models are prone to noise in the data which will hinder their forecasting ability and they have the problem of overfitting or underfitting a time series data modelled and forecast. Furthermore, they also have a greater computational burden. Therefore, it would be advisable to consider that these models are faced with limitations and that these limitations might be the reason why the models did not perform to the desired ability in the current study.

The other limitations that the study have are in terms of not using other machine learning models due to the lack of programming syntax. The other limitation is also the inability of the study to employ correct uptrend, current downtrend, and directional symmetry in measuring the performance of the models in the study. Furthermore, the study was limited to only employing the five models of machine learning and one evolutionary algorithm due to the unavailability of proper and understandable programming syntax.

## 5.7 Limitations and delimitation(s) of the study

- Only three supervised nonlinear machine learning models (ANN, SVR, and LSSVR) together with one evolutionary technique (GA) will be employed to analyse foreign direct investment series.

- Not all machine learning models parameters included in the study will be optimised due to lack of appropriate software syntax.

- For assessing forecasting performance of the models, the following measurements are utilised: MSE, RMSE, MAE, MAPE, and MASE.

- The study will use only *R version 3.6.1, Python, Eviews and Matlab R2018b* as programming languages.

- The study used a dependent variable of foreign direct investment and explanatory variables of gross domestic product (GDP), inflation rate, and exchange rate.

### 5.8 Research contribution

The current study firstly demonstrated that both machine learning models and machine learning based evolutionary hybrid models can be employed to model and forecast foreign direct investment. A time series data that is predominantly, according to the knowledge of the researcher, being modelled using econometric models. The results of the study also contributed by showing that the nature of the series is nonlinear therefore advocating for a better choice of models in the future. The study showed that artificial neural network (ANN) is still a model that can be employed to model not only FDI but other time series data. Furthermore, the contribution of the study was also in the application of evolutionary algorithm although it was revealed that in this case, it did not improve the forecasting accuracy. However, it can be applied or maybe other EA methods can also be used for improving forecasting accuracy.

### 5.9 Recommendations of the study

The included nonlinear machine learning models in the study demonstrated to be good, there is still room for improvement especially in hybrid models. Perhaps more evolutionary algorithms can also be used in a hybrid setting to improve the accuracy of the included machine learning models other than just using the genetic algorithm (GA). There are also more machine learning models that can also be considered for modelling and forecasting foreign direct investment. Furthermore, the study recommends for more machine learning models to be employed in modelling foreign direct investment.

### 5.10 Chapter summary

This chapter has given a steady close to the entirety of the study. The chapter gives the important findings of the study in relation to the study objectives and presented the limitations encountered by the study. Moreover, it was shown that machine learning models are appropriate for modelling and forecasting foreign direct investment and that EA method was not capable of improving the forecasting accuracy of the models. The recommendations for future research are also presented.

## 5.11 Acknowledgements

Major thanks to my entire family for the support, love, and encouragement throughout this journey of research. I would like to extend my sincere gratitude to all who stood by me through this journey.

# BIBLIOGRAPHY


ABADAN, S. S., SHABRI, A. & ISMAIL, S. Hybrid empirical mode decomposition-ARIMA for forecasting exchange rates. AIP Conference Proceedings, 2015. American Institute of Physics, 256-263.

ABAYOMI-ALLI, O. O., SIDEKERSKIENĖ, T., DAMAŠEVIČIUS, R., SIŁKA, J. & POŁAP, D. Empirical Mode Decomposition Based Data Augmentation for Time Series Prediction Using NARX Network. International Conference on Artificial Intelligence and Soft Computing, 2020. Springer, 702-711.

ABDULLAH, S. N. S., SHABRI, A. & SAMSUDIN, R. 2019. Use of Empirical Mode Decomposition in Improving Neural Network Forecasting of Paddy Price. *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics,* 35**,** 53-64.

ABOUNOORI, A. A., MOHAMMADALI, H., GANDALI ALIKHANI, N. & NADERI, E. 2012. Comparative study of static and dynamic neural network models for nonlinear time series forecasting.

ABOUNOORI, A. A., NADERI, E., GANDALI ALIKHANI, N. & AMIRI, A. 2013. Financial Time Series Forecasting by Developing a Hybrid Intelligent System.

ABRAHAM, A., PHILIP, N. S. & SARATCHANDRAN, P. 2004. Modeling chaotic behavior of stock indices using intelligent paradigms. *arXiv preprint cs/0405018*.

ADHIKARI, R. & AGRAWAL, R. K. Effectiveness of PSO Based Neural Network for Seasonal Time Series Forecasting. IICAI, 2011. 231-244.

AHMAD, F., MAT-ISA, N. A., HUSSAIN, Z., BOUDVILLE, R. & OSMAN, M. K. Genetic Algorithm-Artificial Neural Network (GA-ANN) hybrid intelligence for cancer diagnosis. 2010 2nd international conference on computational intelligence, communication systems and networks, 2010. IEEE, 78-83.

AHMADI, M. H., AHMADI, M. A., ASHOURI, M., ASTARAEI, F. R., GHASEMPOUR, R. & ALOUI, F. 2016. Prediction of performance of Stirling engine using least squares support machine technique. *Mechanics & Industry,* 17**,** 506.

AHMED, N. K., ATIYA, A. F., GAYAR, N. E. & EL-SHISHINY, H. 2010. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews,* 29**,** 594-621.

AJIDE, K. B. & IBRAHIM, R. L. 2022. Bayesian model averaging approach of the determinants of foreign direct investment in Africa. *International Economics,* 172**,** 91-105.

AKINSETE, F. M. 2016. *Perceived determinants of foreign direct investment into South Africa.*

ALADAG, C. H. 2011. A new architecture selection method based on tabu search for artificial neural networks. *Expart Systems with Applications,* 38**,** 3287-3293.

ALBA, E. & DORRONSORO, B. 2005. The exploration/exploitation tradeoff in dynamic cellular genetic algorithms. *IEEE transactions on evolutionary computation,* 9**,** 126-142.

ALFARO, L. 2003. Foreign direct investment and growth: Does the sector matter. *Harvard Business School,* 2003**,** 1-31.

ALVAREZ-DIAZ, M. & ALVAREZ, A. 2003. Forecasting exchange rates using genetic algorithms. *Applied Economics Letters,* 10**,** 319-322.

AMELOT, L. M. M., AGATHEE, U. S. & SUNECHER, Y. 2020. Time series modelling, NARX neural network and hybrid KPCA–SVR approach to forecast the foreign exchange market in Mauritius. *African Journal of Economic and Management Studies*.

ANI, K. A. & AGU, C. M. 2022. Predictive comparison and assessment of ANFIS and ANN, as efficient tools in modeling degradation of total petroleum hydrocarbon (TPH). *Cleaner Waste Systems***,** 100052.

ASAFO-AGYEI, G. & KODONGO, O. 2022. Foreign direct investment and economic growth in Sub-Saharan Africa: A nonlinear analysis. *Economic Systems, Forthcoming*.

AURNHAMMER, M. & TONNIES, K. 2005. A genetic algorithm for automated horizon correlation across faults in seismic images. *IEEE Transactions on evolutionary computation,* 9**,** 201-210.

BABU, A. & REDDY, S. 2015. Exchange rate forecasting using ARIMA. *Neural Network and Fuzzy Neuron, Journal of Stock & Forex Trading,* 4**,** 01-05.

BAL, C., DEMIR, S. & ALADAG, C. H. 2016. A comparison of different model selection criteria for forecasting EURO/USD exchange rates by feed forward neural network. *Int. Journal of Computing, Communications and Instrumentation Engineering,* 3**,** 271-275.

BAO-DE, L., XIN-YANG, Z., MEI, Z., HUI, L. & GUANG-QIAN, L. 2021. Improved genetic algorithm-based research on optimization of least square support vector machines: an application of load forecasting. *Soft Computing,* 25**,** 11997-12005.

BARAPATRE, O., TETE, E., SAHU, C. L., KUMAR, D. & KSHATRIYA, H. 2018. Stock price prediction using artificial neural network. *Int. J. Adv. Res. Ideas Innov. Technol,* 4**,** 916-922.

BASHEER, I. A. & HAJMEER, M. 2000. Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods,* 43**,** 3-31.

BAYLAR, A., HANBAY, D. & BATAN, M. 2009. Application of least square support vector machines in the prediction of aeration performance of plunging overfall jets from weirs. *Expert Systems with Applications,* 36**,** 8368-8374.

BAZARTSEREN, B., HILDEBRANDT, G. & HOLZ, K.-P. 2003. Short-term water level prediction using neural networks and neuro-fuzzy approach. *Neurocomputing,* 55**,** 439-450.

BEHNAMIAN, J. & GHOMI, S. F. 2010. Development of a PSO–SA hybrid metaheuristic for a new comprehensive regression model to time-series forecasting. *Expert Systems with applications,* 37**,** 974-984.

BEYTOLLAHI, A. & ZEINALI, H. 2020. Comparing Prediction Power of Artificial Neural Networks Compound Models in Predicting Credit Default Swap Prices through Black–Scholes–Merton Model. *Iranian Journal of Management Studies,* 13**,** 69-93.

BHAGWAT, P. P. & MAITY, R. 2013. Hydroclimatic streamflow prediction using least square-support vector regression. *ISH Journal of Hydraulic Engineering,* 19**,** 320-328.

BING, Y., HAO, J. K. & ZHANG, S. C. Stock market prediction using artificial neural networks. Advanced Engineering Forum, 2012. Trans Tech Publ, 1055-1060.

BLANCO-VELASCO, M., WENG, B. & BARNER, K. E. 2008. ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Computers in biology and medicine,* 38**,** 1-13.

BLONIGEN, B. A. 2005. A review of the empirical literature on FDI determinants. *Atlantic economic journal,* 33**,** 383-403.

BOASHASH, B. 1992. Estimating and interpreting the instantaneous frequency of a signal. II. Algorithms and applications. *Proceedings of the IEEE,* 80**,** 540-568.

BORENSZTEIN, E., DE GREGORIO, J. & LEE, J.-W. 1998. How does foreign direct investment affect economic growth? *Journal of international Economics,* 45**,** 115-135.

BOYD, S., XIAO, L., MUTAPCIC, A. & MATTINGLEY, J. 2007. Notes on decomposition methods. *Notes for EE364B, Stanford University,* 635**,** 1-36.

BROOCK, W. A., SCHEINKMAN, J. A., DECHERT, W. D. & LEBARON, B. 1996. A test for independence based on the correlation dimension. *Econometric reviews,* 15**,** 197-235.

BROWN, R. L., DURBIN, J. & EVANS, J. M. 1975. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)***,** 149-192.

BURGER, C., DOHNAL, M., KATHRADA, M. & LAW, R. 2001. A practitioners guide to time-series methods for tourism demand forecasting—a case study of Durban, South Africa. *Tourism Management,* 22**,** 403-409.

BÜYÜKŞAHIN, Ü. Ç. & ERTEKIN, Ş. 2019. Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. *Neurocomputing,* 361**,** 151-163.

CAO, L.-J. & TAY, F. E. H. 2003. Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on neural networks,* 14**,** 1506-1518.
CAO, Y. & WU, Q. 1999. Optimization of control parameters in genetic algorithms: a stochastic approach. *International journal of systems science,* 30**,** 551-559.
CHABAA, S., ZEROUAL, A. & ANTARI, J. ANFIS method for forecasting internet traffic time series. 2009 Mediterrannean microwave symposium (mms), 2009. IEEE, 1-4.
CHAI, J., DU, J., LAI, K. K. & LEE, Y. P. 2015. A hybrid least square support vector machine model with parameters optimization for stock forecasting. *Mathematical Problems in Engineering,* 2015.
CHANG, L. C. & CHANG, F. J. 2001. Intelligent control for modelling of real-time reservoir operation. *Hydrological processes,* 15**,** 1621-1634.
CHARANIYA, N. & DUDUL, S. Focused time delay neural network model for rainfall prediction using indian ocean dipole index. 2012 Fourth International Conference on Computational Intelligence and Communication Networks, 2012. IEEE, 851-855.
CHATTERJEE, S. & BANDOPADHYAY, S. 2012. Reliability estimation using a genetic algorithm-based artificial neural network: An application to a load-haul-dump machine. *Expart Systems with Applications,* 39**,** 10943-10951.
CHEN, B.-J. & CHANG, M.-W. 2004. Load forecasting using support vector machines: A study on EUNITE competition 2001. *IEEE transactions on power systems,* 19**,** 1821-1830.
CHEN, C.-F., LAI, M.-C. & YEH, C.-C. 2012. Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowledge-Based Systems,* 26**,** 281-287.
CHEN, J., CHEN, H., HUO, Y. & GAO, W. 2017. Application of SVR Models in Stock Index Forecast Based on Different Parameter Search Methods. *Open Journal of Statistics,* 7**,** 194-202.
CHEN, K.-Y. & HO, C.-H. An improved support vector regression modeling for Taiwan Stock Exchange market weighted index forecasting. 2005 International Conference on Neural Networks and Brain, 2005. IEEE, nil15-1638.
CHEN, K.-Y. & WANG, C.-H. 2007. Support vector regression with genetic algorithms in forecasting tourism demand. *Tourism Management,* 28**,** 215-226.
CHEN, Q., WORDEN, K., PENG, P. & LEUNG, A. 2007. Genetic algorithm with an improved fitness function for (N) ARX modelling. *Mechanical Systems and Signal Processing,* 21**,** 994-1007.
CHENG, C.-H. & SHIU, H.-Y. A novel GA-SVR time series model based on selected indicators method for forecasting stock price. 2014 International Conference on Information Science, Electronics and Electrical Engineering, 2014. IEEE, 395-399.
CHENG, C.-H. & WEI, L.-Y. 2014. A novel time-series model based on empirical mode decomposition for forecasting TAIEX. *Economic Modelling,* 36**,** 136-141.
CHENG, J. 2007. Hybrid genetic algorithms for structural reliability analysis. *Computers & Structures,* 85**,** 1524-1533.
CHENG, J. 2010. An artificial neural network based genetic algorithm for estimating the reliability of long span suspension bridges. *Finite Elements in Analysis Design,* 46**,** 658-667.
CHOW, G. C. 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society***,** 591-605.
CHU, F.-L. 1998. Forecasting tourism demand in Asian-Pacific countries. *Annals of Tourism Research,* 25**,** 597-615.
CHU, F.-L. 2009. Forecasting tourism demand with ARMA-based methods. *Tourism Management,* 30**,** 740-751.
CLAVERIA, O. & TORRA, S. 2014. Forecasting tourism demand to Catalonia: Neural networks vs. time series models. *Economic Modelling,* 36**,** 220-228.
CONG, Y., WANG, J. & LI, X. 2016. Traffic flow forecasting by a least squares support vector machine with a fruit fly optimization algorithm. *Procedia Engineering,* 137**,** 59-68.
CONSTANTINO, H., FERNANDES, P. O. & TEIXEIRA, J. P. 2016. Tourism demand modelling and forecasting with artificial neural network models: The Mozambique case study. *Tékhne,* 14**,** 113-124.

DEB, C., ZHANG, F., YANG, J., LEE, S. E. & SHAH, K. W. 2017. A review on time series forecasting techniques for building energy consumption. *Renewable Sustainable Energy Reviews,* 74**,** 902-924.

DELCROIX, B., NY, J. L., BERNIER, M., AZAM, M., QU, B. & VENNE, J.-S. Autoregressive neural networks with exogenous variables for indoor temperature prediction in buildings. Building Simulation, 2021. Springer, 165-178.

DHAMIJA, A. K. & BHALLA, V. 2011. Exchange rate forecasting: comparison of various architectures of neural networks. *Neural Computing and Applications,* 20**,** 355-363.

DIEBOLD, F. X. 2015. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. *Journal of Business & Economic Statistics,* 33**,** 1-1.

DING, M., ZHOU, H., XIE, H., WU, M., LIU, K.-Z., NAKANISHI, Y. & YOKOYAMA, R. 2021. A time series model based on hybrid-kernel least-squares support vector machine for short-term wind power forecasting. *ISA transactions,* 108**,** 58-68.

DOGANIS, P., ALEXANDRIDIS, A., PATRINOS, P. & SARIMVEIS, H. 2006. Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *Journal of Food Engineering,* 75**,** 196-204.

DOMINGOS, S. D. O., DE OLIVEIRA, J. F. & DE MATTOS NETO, P. S. J. K.-B. S. 2019. An intelligent hybridization of ARIMA with machine learning models for time series forecasting. 175**,** 72-86.

DRUCKER, H., BURGES, C. J., KAUFMAN, L., SMOLA, A. J. & VAPNIK, V. Support vector regression machines. Advances in neural information processing systems, 1997. 155-161.

DU, J., ZHAO, B. & MIAO, S. An Application on the Immune Evolutionary Algorithm Based on Back Propagation in the Rainfall Prediction. 2012 International Conference on Computer Science and Electronics Engineering, 2012. IEEE, 313-317.

DUAN, K., KEERTHI, S. S. & POO, A. N. 2003. Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing,* 51**,** 41-59.

DUAN, W.-Y., HUANG, L.-M., HAN, Y., ZHANG, Y.-H. & HUANG, S. 2015. A hybrid AR-EMD-SVR model for the short-term prediction of nonlinear and non-stationary ship motion. *Journal of Zhejiang University-SCIENCE A,* 16**,** 562-576.

DUAN, W., HAN, Y., HUANG, L., ZHAO, B. & WANG, M. 2016. A hybrid EMD-SVR model for the short-term prediction of significant wave height. *Ocean Engineering,* 124**,** 54-73.

EBADATI, O. & MORTAZAVI, M. 2018. An efficient hybrid machine learning method for time series stock market forecasting. *Neural Network World,* 28**,** 41-55.

EGELI, B., OZTURAN, M. & BADUR, B. 2003. Stock market prediction using artificial neural networks. *Decision Support Systems,* 22**,** 171-185.

EĞRIOĞLU, E., ALADAĞ, Ç. H. & GÜNAY, S. 2008. A new model selection strategy in artificial neural networks. *Applied Mathematics and Computation,* 195**,** 591-597.

ENGLE, R. F. 1982. A general approach to Lagrange multiplier model diagnostics. *Journal of Econometrics,* 20**,** 83-104.

ENKE, D. & THAWORNWONG, S. 2005. The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with applications,* 29**,** 927-940.

ERDOGAN, O. & GOKSU, A. 2014. Forecasting Euro and Turkish Lira Exchange Rates with Artificial Neural Networks (ANN). *International Journal of Academic Research in Accounting, Finance and Management Sciences,* 4**,** 307-316.

ESPINOZA, M., SUYKENS, J. A. & DE MOOR, B. Short term chaotic time series prediction using symmetric LS-SVM regression. Proc. of the 2005 International Symposium on Nonlinear Theory and Applications (NOLTA), 2005. 606-609.

FAN, G.-F., PENG, L.-L., HONG, W.-C. & SUN, F. 2016. Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression. *Neurocomputing,* 173**,** 958-970.

FAN, J. & TANG, Y. An EMD-SVR method for non-stationary time series prediction. 2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE), 2013. IEEE, 1765-1770.

FATHI, M. & PARIAN, J. A. 2021. Intelligent MPPT for photovoltaic panels using a novel fuzzy logic and artificial neural networks based on evolutionary algorithms. *Energy Reports,* 7**,** 1338-1348.

FAULINA, R., LUSIA, D. A., OTOK, B. W. & KUSWANTO, H. Ensemble method based on anfis-arima for rainfall prediction. 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), 2012. IEEE, 1-4.

FEI, S.-W., WANG, M.-J., MIAO, Y.-B., TU, J. & LIU, C.-L. 2009. Particle swarm optimization-based support vector machine for forecasting dissolved gases content in power transformer oil. *Energy Conversion Management,* 50**,** 1604-1609.

FERNANDES, P. O. 2005. Modelling, Prediction and Behaviour Analysis of Tourism Demand in the North of Portugal. *PhD, Valladolid University, Valladolid*.

FOGEL, D. B. 1994. An introduction to simulated evolutionary optimization. *IEEE transactions on neural networks,* 5**,** 3-14.

FRANSES, P. H. & VAN HOMELEN, P. 1998. On forecasting exchange rates using neural networks. *Applied Financial Economics,* 8**,** 589-596.

FRITSCH, S., GUENTHER, F., SULING, M. & MUELLER, S. 2016. Package “neuralnet.” The Comprehensive R Archive Network.

FROSINI, L. & PETRECCA, G. 2000. Linear and neural dynamical models for energy flows prediction in facility systems. *Soft Computing in Industrial Applications.* Springer.

FU, C. Forecasting exchange rate with EMD-based support vector regression. 2010 International Conference on Management and Service Science, 2010. IEEE, 1-4.

FU, S., LI, Y., SUN, S. & LI, H. 2019. Evolutionary support vector machine for RMB exchange rate forecasting. *Physica A: Statistical Mechanics and its Applications,* 521**,** 692-704.

GEORGESCU, V. & DINUCĂ, E. C. Evidence of improvement in neural-network based predictability of stock market indexes through co-movement entries. Recent Advances in Applied & Biomedical Informatics and Computational Engineering in Systems Applications, 11th WSEAS International Conference on Applied Informatics and Communications, Florence, Italy, 2011. 412-417.

GHASEMIYEH, R., MOGHDANI, R. & SANA, S. S. 2017. A hybrid artificial neural network with metaheuristic algorithms for predicting stock price. *Cybernetics and Systems,* 48**,** 365-392.

GHOSH, S. & MAKA, S. 2011. Genetic algorithm based NARX model identification for evaluation of insulin sensitivity. *Applied Soft Computing,* 11**,** 221-226.

GOLBERG, D. E. 1989. Genetic algorithms in search, optimization, and machine learning. *Addion wesley,* 1989**,** 36.

GOODARZI, M., FREITAS, M. P., WU, C. H. & DUCHOWICZ, P. R. 2010. pKa modeling and prediction of a series of pH indicators through genetic algorithm-least square support vector regression. *Chemometrics and Intelligent Laboratory Systems,* 101**,** 102-109.

GUPTA, J. N. & SEXTON, R. S. 1999. Comparing backpropagation with a genetic algorithm for neural network training. *Omega,* 27**,** 679-684.

HAI-YAN, L. & ZE-JUN, J. 2013. Research on failure prediction technology based on time series analysis and aco-lssvm. *Computer and Modernization,* 1**,** 219.

HALICIOGLU, F. & MEHMET, U. 2005. On Stability of the Demand for Money in a Developing OECD. University Library of Munich, Germany.

HAN, J.-B., KIM, S.-H., JANG, M.-H. & RI, K.-S. 2019. Using genetic algorithm and NARX neural network to forecast daily bitcoin price. *Computational Economics***,** 1-17.

HOKEY, M., HYUN, J. & CHANG, S. 2006. A genetic algorithm approach to developing the multi-echelon reverse logistics network for product returns. *Omega,* 34**,** 56-69.

HOLLAND, J. H. 1975. *Adaptations in Natural Artificial Systems*, University of Michigan Press, Michigan.

HONG, W.-C. 2008. Rainfall forecasting by technological machine learning models. *Applied Mathematics Computation,* 200**,** 41-57.
HONG, W.-C. 2009. Hybrid evolutionary algorithms in a SVR-based electric load forecasting model. *International Journal of Electrical Power Energy Systems,* 31**,** 409-417.
HUANG, H.-C. & HOU, C.-I. 2017. Tourism Demand Forecasting Model Using Neural Network. *International Journal of Computer Science & Information Technology (IJCSIT),* 9**,** 19-29.
HUANG, N. E., SHEN, Z. & LONG, S. R. 1999. A new view of nonlinear water waves: the Hilbert spectrum. *Annual review of fluid mechanics,* 31**,** 417-457.
HUANG, N. E., SHEN, Z., LONG, S. R., WU, M. C., SHIH, H. H., ZHENG, Q., YEN, N.-C., TUNG, C. C. & LIU, H. H. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences,* 454**,** 903-995.
HUANG, W., NAKAMORI, Y. & WANG, S.-Y. 2005. Forecasting stock market movement direction with support vector machine. *Computers & operations research,* 32**,** 2513-2522.
HUANG, Y., CAPRETZ, L. F. & HO, D. Neural network models for stock selection based on fundamental analysis. 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), 2019. IEEE, 1-4.
HYKIN, S. 1999. Neural networks: A comprehensive introduction. *NJ: PHI*.
ISMAIL, M. T. & ISA, Z. 2007. Detecting regime shifts in Malaysian exchange rates. *Malaysian Journal of Fundamental and Applied Sciences,* 3.
ISMAIL, S., SHABRI, A. & ABADAN, S. S. Empirical mode decomposition coupled with least square support vector machine for river flow forecasting. AIP Conference Proceedings, 2015. American Institute of Physics, 232-241.
JANG, J.-S. 1993. ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man,*

*cybernetics,* 23**,** 665-685.
JANG, J.-S. & SUN, C.-T. Predicting chaotic time series with fuzzy if-then rules. [Proceedings 1993] Second IEEE International Conference on Fuzzy Systems, 1993. IEEE, 1079-1084.
JANG, J.-S. R. Fuzzy modeling using generalized neural networks and kalman filter algorithm. AAAI, 1991. 762-767.
JANG, J.-S. R., SUN, C.-T. & MIZUTANI, E. 1997. Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]. *IEEE Transactions on automatic control,* 42**,** 1482-1484.
JAREANPON, C., PENSUWON, W., FRANK, R. & DAVEY, N. An adaptive RBF network optimised using a genetic algorithm applied to rainfall forecasting. IEEE International Symposium on Communications and Information Technology, 2004. ISCIT 2004., 2004. IEEE, 1005-1010.
JAWAD, M., ALI, S. M., KHAN, B., MEHMOOD, C. A., FARID, U., ULLAH, Z., USMAN, S., FAYYAZ, A., JADOON, J. & TAREEN, N. 2018. Genetic algorithm-based non-linear auto-regressive with exogenous inputs neural network short-term and medium-term uncertainty modelling and prediction for electrical load and wind speed. *The Journal of Engineering,* 2018**,** 721-729.
JIANG, C.-S. & LIANG, G.-Q. 2021. Modeling shear strength of medium-to ultra-high-strength concrete beams with stirrups using SVR and genetic algorithm. *Soft Computing,* 25**,** 10661-10675.
JIANG, F. & WU, W. 2016. Hybrid Genetic Algorithm and Support Vector Regression Performance in CNY Exchange Rate

Prediction. *International Conference on Engineering Science and Management (ESM 2016)*.
JIANG, R. & YAN, H. 2008. Studies of spectral properties of short genes using the wavelet subspace Hilbert–Huang transform (WSHHT). *Physica A: Statistical Mechanics and its Applications,* 387**,** 4223-4247.

JIN, L., HUANG, Y. & ZHAO, H.-S. Ensemble prediction of monthly mean rainfall with a Particle Swarm Optimization-neural network model. 2012 IEEE 13th International Conference on Information Reuse & Integration (IRI), 2012. IEEE, 287-294.

JOACHIMS, T. Text categorization with support vector machines: Learning with many relevant features. European conference on machine learning, 1998. Springer, 137-142.

JOHNSON, A. 2006. The effects of FDI inflows on host country economic growth. *The Royal Institute of technology. Centre of Excellence for studies in Science and Innovation http://www. infra. kth. se/cesis/research/publications/working papers*.

JUN, W., YUYAN, L., LINGYU, T. & PENG, G. 2018. A new weighted CEEMDAN-based prediction model: An experimental investigation of decomposition and non-decomposition approaches. *Knowledge-Based Systems,* 160**,** 188-199.

KADILAR, C., ŞIMŞEK, M. & ALADAĞ, A. G. Ç. H. 2009. Forecasting the exchange rate series with ANN: the case of Turkey. *Ekonometri ve İstatistik e-Dergisi***,** 17-29.

KAJAN, S., MÁCSIK, P., KÖRÖSI, L., GOGA, J. & PAVLOVIČOVÁ, J. Comparison of Neural Models for Modeling Dynamic Changes in Microgrids. 2022 Cybernetics & Informatics (K&I), 2022. IEEE, 1-5.

KAJORNRIT, J., WONG, K. W. & FUNG, C. C. Rainfall prediction in the northeast region of Thailand using modular fuzzy inference system. 2012 IEEE International Conference on Fuzzy Systems, 2012. IEEE, 1-6.

KAMRUZZAMAN, J. & SARKER, R. A. Forecasting of currency exchange rates using ANN: A case study. Proceedings of the 2003 international conference on neural networks and signal processing, 2003. IEEE.

KARBASI, A. R., LASKUKALAYEH, S. S. & FAHIMIFARD, S. M. 2009. Comparison of NNARX, ANN and ARIMA Techniques to Poultry Retail Price Forecasting.

KARIMI, H. & DASTRANJ, J. 2014. Artificial neural network-based genetic algorithm to predict natural gas consumption. *Energy Systems,* 5**,** 571-581.

KARIMI, H. & YOUSEFI, F. 2012. Application of artificial neural network–genetic algorithm (ANN–GA) to correlation of density in nanofluids. *Fluid Phase Equilibria,* 336**,** 79-83.

KEENAN, D. M. 1985. A Tukey nonadditivity-type test for time series nonlinearity. *Biometrika,* 72**,** 39-44.

KENNEDY, J. 2011. Encyclopedia of machine learning. *Particle Swarm Optimization (ed)***,** 760-766.

KESKIN, M. E., TAYLAN, D. & TERZI, O. 2006. Adaptive neural-based fuzzy inference system (ANFIS) approach for modelling hydrological time series. *Hydrological sciences journal,* 51**,** 588-598.

KHANDELWAL, I., ADHIKARI, R. & VERMA, G. J. P. C. S. 2015. Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. 48**,** 173-179. .

KHASHEI, M. & BIJARI, M. 2010. An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with applications,* 37**,** 479-489.

KHOSRAVI, A., KOURY, R., MACHADO, L. & PABON, J. 2018a. Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system. *Sustainable Energy Technologies Assessments,* 25**,** 146-160.

KHOSRAVI, A., MACHADO, L. & NUNES, R. 2018b. Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil. *Applied Energy,* 224**,** 550-566.

KIM, D. & OH, H.-S. 2013. EMD: Empirical mode decomposition and Hilbert spectral analysis. *R package version,* 1.

KIM, K.-J. 2003. Financial time series forecasting using support vector machines. *Neurocomputing,* 55**,** 307-319.

KIM, K.-J. 2006. Artificial neural networks with evolutionary instance selection for financial forecasting. *Expert Systems with Applications,* 30**,** 519-526.

KIM, K.-J. & HAN, I. 2000. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert systems with Applications,* 19**,** 125-132.
KISI, O., LATIFOĞLU, L. & LATIFOĞLU, F. 2014. Investigation of empirical mode decomposition in forecasting of hydrological time series. *Water resources management,* 28**,** 4045-4057.
KLINE, R. B. 2005a. Methodology in the social sciences. Principles and practice of structural equation modeling.
KLINE, R. B. 2005b. Methodology in the social sciences. Principles and practice of structural equation modeling. Guilford press New York.
KROLLNER, B., VANSTONE, B. J. & FINNIE, G. R. Financial time series forecasting with machine learning techniques: a survey. ESANN, 2010.
KUO, R. J., CHEN, C. & HWANG, Y. 2001. An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. *Fuzzy sets and systems,* 118**,** 21-45.
LAI, L. K. & LIU, J. N. Stock forecasting using support vector machine. 2010 International Conference on Machine Learning and Cybernetics, 2010. IEEE, 1607-1614.
LAW, R. 1998. Room occupancy rate forecasting: a neural network approach. *International Journal of Contemporary Hospitality Management*.
LAW, R. & AU, N. 1999. A neural network model to forecast Japanese demand for travel to Hong Kong. *Tourism Management,* 20**,** 89-97.
LEE, J., KIM, G., PARK, S. & JANG, D. 2016. Hybrid Corporate Performance Prediction Model Considering Technical Capability.
LEE, T. & OUARDA, T. B. 2010. Long-term prediction of precipitation and hydrologic extremes with nonstationary oscillation processes. *Journal of Geophysical Research: Atmospheres,* 115.
LEONG, W. C., BAHADORI, A., ZHANG, J. & AHMAD, Z. 2021. Prediction of water quality index (WQI) using support vector machine (SVM) and least square-support vector machine (LS-SVM). *International Journal of River Basin Management,* 19**,** 149-156.
LEVINE, R. & RENELT, D. 1992. A sensitivity analysis of cross-country growth regressions. *The American economic review***,** 942-963.
LI, H. & MENG, G. 2006. Detection of harmonic signals from chaotic interference by empirical mode decomposition. *Chaos, Solitons & Fractals,* 30**,** 930-935.
LI, Q. & RESNICK, A. 2003. Reversal of fortunes: Democratic institutions and foreign direct investment inflows to developing countries. *International organization,* 57**,** 175-212.
LI, Q. & WU, J. 2007. Time–frequency analysis of typhoon effects on a 79-storey tall building. *Journal of Wind Engineering and Industrial Aerodynamics,* 95**,** 1648-1666.
LI, R., CHEN, X., BALEZENTIS, T., STREIMIKIENE, D. & NIU, Z. 2021. Multi-step least squares support vector machine modeling approach for forecasting short-term electricity demand with application. *Neural computing and applications,* 33**,** 301-320.
LIANG, H., BRESSLER, S. L., DESIMONE, R. & FRIES, P. 2005. Empirical mode decomposition: a method for analyzing neural data. *Neurocomputing,* 65**,** 801-807.
LIAO, R., ZHENG, H., GRZYBOWSKI, S. & YANG, L. 2011. Particle swarm optimization-least squares support vector regression based forecasting model on dissolved gases in oil-filled power transformers. *Electric Power Systems Research,* 81**,** 2074-2080.
LIN, C.-J., CHEN, H.-F. & LEE, T.-S. 2011. Forecasting tourism demand using time series, artificial neural networks and multivariate adaptive regression splines: Evidence from Taiwan. *International Journal of Business Administration,* 2**,** 14-24.
LIN, C.-S., CHIU, S.-H. & LIN, T.-Y. 2012. Empirical mode decomposition–based least squares support vector regression for foreign exchange rate forecasting. *Economic Modelling,* 29**,** 2583-2590.
LIN, P. 2001. Support vector regression: Systematic design and performance analysis. *Unpublished Doctoral Dissertation, Department of Electronic Engineering, National Taiwan University*.

LIU, D., CHEN, Q. & MORI, K. Time series forecasting method of building energy consumption using support vector regression. 2015 IEEE international conference on information and automation, 2015. IEEE, 1628-1632.

LIU, H.-H., CHANG, L.-C., LI, C.-W. & YANG, C.-H. 2018. Particle swarm optimization-based support vector regression for tourist arrivals forecasting. *Computational Intelligence Neuroscience,* 2018.

LIU, H., CHEN, C., TIAN, H.-Q. & LI, Y.-F. 2012. A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks. *Renewable energy,* 48**,** 545-556.

LIU, J. N., LI, B. N. & DILLON, T. S. 2001. An improved naive Bayesian classifier technique coupled with a novel input solution method [rainfall prediction]. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* 31**,** 249-256.

LIU, S., TAI, H., DING, Q., LI, D., XU, L. & WEI, Y. 2013. A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. *Mathematical and Computer Modelling,* 58**,** 458-465.

LU, C.-J., LEE, T.-S. & CHIU, C.-C. 2009. Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems,* 47**,** 115-125.

LUO, C., HUANG, C., CAO, J., LU, J., HUANG, W., GUO, J. & WEI, Y. 2019. Short-term traffic flow prediction based on least square support vector machine with hybrid optimization algorithm. *Neural processing letters,* 50**,** 2305-2322.

LUO, W., BAI, Z., ZHENG, S. & HUI, Y. 2020. A modified BDS test. *Statistics & Probability Letters,* 164**,** 108794.

MA, C. & LI, Y. 2016. Improving forecasting accuracy of annual runoff time series using RBFN based on EEMD decomposition. *DEStech Transactions on Engineering and Technology Research*.

MA, Q., LIU, S., FAN, X., CHAI, C., WANG, Y. & YANG, K. 2020. A Time Series Prediction Model of Foundation Pit Deformation Based on Empirical Wavelet Transform and NARX Network. *Mathematics,* 8**,** 1535.

MABULE, R. 2012. *The determinants of foreign direct investment on the South African economic growth.*

MAHROOGHI, A. & LAKZIAN, E. 2021. Optimization of Wells turbine performance using a hybrid artificial neural fuzzy inference system (ANFIS)-Genetic algorithm (GA). *Ocean Engineering,* 226**,** 108861.

MAJAVU, A. & KAPINGURA 2016a. The Determinants of Foreign Direct Investment Inflows in South Africa: An Application of the Johansen Co-integration Test and VECM. *Journal of Economics,* 7**,** 130-143.

MAJAVU, A. & KAPINGURA, F. 2016b. The Determinants of Foreign Direct Investment Inflows in South Africa: An Application of the Johansen Co-integration Test and VECM. *Journal of Economics* 7**,** 130-143.

MAJAVU, A. & KAPINGURA, F. 2016c. The Determinants of Foreign Direct Investment Inflows in South Africa: An Application of the Johansen Co-integration Test and VECM. *Journal of Economics,* 7**,** 130-143.

MAO, K. & BILLINGS, S. 1997. Algorithms for minimal model structure detection in nonlinear dynamic system identification. *International journal of control,* 68**,** 311-330.

MASUGI, M. & TAKUMA, T. 2007. Using a Volterra system model to analyze nonlinear response in video-packet transmission over IP networks. *Communications in Nonlinear Science and Numerical Simulation,* 12**,** 411-421.

MATHUR, A. & SINGH, K. 2013. Foreign direct investment, corruption and democracy. *Applied Economics,* 45**,** 991-1002.

MCCALL, J. 2005. Genetic algorithms for modelling and optimisation. *Journal of computational and Applied Mathematics,* 184**,** 205-222.

MCCALL, J. & PETROVSKI, A. A decision support system for cancer chemotherapy using genetic algorithms. Proceedings of the international conference on computational intelligence for modeling, control and automation, 1999. 65-70.

MCCULLOCH, W. S. & PITTS, W. 1943. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics,* 5**,** 115-133.

MCLEOD, A. I. & LI, W. K. 1983. Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of time series analysis,* 4**,** 269-273.

MEDI, B. & ASADBEIGI, A. 2021. Application of a GA-Optimized NNARX controller to nonlinear chemical and biochemical processes. *Heliyon,* 7**,** e07846.

MEESAD, P. & RASEL, R. I. Predicting stock market price using support vector regression. 2013 International Conference on Informatics, Electronics and Vision (ICIEV), 2013. IEEE, 1-6.

MEI-YING, Y. & XIAO-DONG, W. 2004. Chaotic time series prediction using least squares support vector machines. *Chinese physics,* 13**,** 454.

MENG, E., HUANG, S., HUANG, Q., FANG, W., WU, L. & WANG, L. 2019. A robust method for non-stationary streamflow prediction based on improved EMD-SVM model. *Journal of hydrology,* 568**,** 462-478.

MEYER, D., DIMITRIADOU, E., HORNIK, K., WEINGESSEL, A., LEISCH, F., CHANG, C. & LIN, C. 2015. Misc functions of the Department of Statistics. *Probability Theory Group (Formerly: E1071), TU Wien*.

MIRMIRANI, S. & LI, H. C. 2004. A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil. *Applications of artificial intelligence in finance and economics.* Emerald Group Publishing Limited.

MOLAJOU, A., NOURANI, V., AFSHAR, A., KHOSRAVI, M. & BRYSIEWICZ, A. 2021. Optimal design and feature selection by genetic algorithm for emotional artificial neural network (EANN) in rainfall-runoff modeling. *Water Resources Management,* 35**,** 2369-2384.

MÜHLENBEIN, H. Evolutionary algorithms: Theory and applications. Local search in combinatorial optimization, 1993. Citeseer.

MUKHERJEE, S., OSUNA, E. & GIROSI, F. Nonlinear prediction of chaotic time series using support vector machines. Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop, 1997. IEEE, 511-520.

MÜLLER, K.-R., SMOLA, A. J., RÄTSCH, G., SCHÖLKOPF, B., KOHLMORGEN, J. & VAPNIK, V. Predicting time series with support vector machines. International Conference on Artificial Neural Networks, 1997. Springer, 999-1004.

NAG, A. K. & MITRA, A. 2002. Forecasting daily foreign exchange rates using genetically optimized neural networks. *Journal of Forecasting,* 21**,** 501-511.

NAYAK, P., SUDHEER, K., RANGAN, D. & RAMASASTRI, K. 2005. Short-term flood forecasting with a neurofuzzy model. *Water Resources Research,* 41.

NAYAK, S., MISRA, B. B. & BEHERA, H. S. 2014. Impact of data normalization on stock index forecasting. *International Journal of Computer Information Systems and Industrial Management Applications,* 6**,** 257-269.

NELSON, M., HILL, T., REMUS, W. & O'CONNOR, M. 1999. Time series forecasting using neural networks: Should the data be deseasonalized first? *Journal of forecasting,* 18**,** 359-367.

NEWAZ, M. 2008. Comparing the performance of time series models for forecasting exchange rate.

NIE, Z., SHEN, F., XU, D. & LI, Q. 2020. An EMD-SVR model for short-term prediction of ship motion using mirror symmetry and SVR algorithms to eliminate EMD boundary effect. *Ocean Engineering,* 217**,** 107927.

NONNENBERG, M. J. B. & MENDONCA, M. J. C. 2004. The determinants of direct foreign investment in developing countries. *Available at SSRN 525462*.

NOOR, H. M., NDZI, D., YANG, G. & SAFAR, N. Z. M. Rainfall-based river flow prediction using NARX in Malaysia. 2017 IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA), 2017. IEEE, 67-72.

NOR, A. H. S. M. & GHARLEGHI, B. 2011. Application of dynamic models for exchange rate prediction. *International Journal of Innovation, Management and Technology,* 2.

OKASHA, M. K. 2014. Using support vector machines in financial time series forecasting. *International Journal of Statistics and Applications,* 4**,** 28-39.
OKE, B. O., EZIKE, J. E. & OJOGBO, S. O. 2012. Locational determinants of foreign direct investments in Nigeria. *International Business Research,* 5**,** 103-111.
OKOLOBAH, V. A. & ISMAIL, Z. 2013. New approach to peak load forecasting based on EMD and ANFIS. *Indian Journal of Science and Technology,* 6**,** 5600-6.
ONGSRITRAKUL, P. & SOONTHORNPHISAJ, N. Apply decision tree and support vector regression to predict the gold price. Proceedings of the International Joint Conference on Neural Networks, 2003., 2003. IEEE, 2488-2492.
ONU, C. E., NWEKE, C. N. & NWABANNE, J. T. 2022. Modeling of thermo-chemical pretreatment of yam peel substrate for biogas energy production: RSM, ANN, and ANFIS comparative approach. *Applied Surface Science Advances,* 11**,** 100299.
ONYEIWU, S. & SHRESTHA, H. 2004. Determinants of foreign direct investment in Africa. *Journal of Developing Societies,* 20**,** 89-106.
ORJUELA-CANON, A. D., JUTINICO, A. L., GONZALEZ, M. E. D., GARCÍA, C. E. A., VERGARA, E. & PALENCIA, M. A. 2022. Time series forecasting for tuberculosis incidence employing neural network models. *Heliyon,* 8.
ORO, O. U. & ALAGIDEDE, I. P. 2021. Does petroleum resources or market size drive foreign direct investment in Africa? New evidence from time-series analysis. *Resources Policy,* 71**,** 101992.
OU, P. & WANG, H. 2010. Financial volatility forecasting by least square support vector machine based on GARCH, EGARCH and GJR models: evidence from ASEAN stock markets. *International Journal of Economics and Finance,* 2**,** 51-64.
PACELLI, V., BEVILACQUA, V. & AZZOLLINI, M. 2011. An artificial neural network model to forecast exchange rates. *Journal of Intelligent Learning Systems and Applications,* 3**,** 57.
PAI, P.-F., HUNG, K.-C. & LIN, K.-P. 2014. Tourism demand forecasting using novel hybrid system. *Expert Systems with applications,* 41**,** 3691-3702.
PALMER, A., MONTANO, J. J. & SESÉ, A. 2006. Designing an artificial neural network for forecasting tourism time series. *Tourism Management,* 27**,** 781-790.
PANAPAKIDIS, I. P. & CHRISTOFORIDIS, G. C. A hybrid ANN/GA/ANFIS model for very short-term PV power forecasting. 2017 11th IEEE International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG), 2017. IEEE, 412-417.
PANAPAKIDIS, I. P. & DAGOUMAS, A. S. 2017. Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model. *Energy,* 118**,** 231-245.
PANDA, C. & NARASIMHAN, V. 2007. Forecasting exchange rate better with artificial neural network. *Journal of Policy Modeling,* 29**,** 227-236.
PANIGRAHI, S., KARALI, Y. & BEHERA, H. 2013. Time series forecasting using evolutionary neural network. *International Journal of Computer Applications,* 75.
PATEL, M. B. & YALAMALLE, S. R. 2014. Stock price prediction using artificial neural network. *International Journal of Innovative Research in Science, Engineering and Technology,* 3**,** 13755-13762.
PATTIE, D. C. & SNYDER, J. 1996. Using a neural network to forecast visitor behavior. *Annals of Tourism Research,* 23**,** 151-164.
PHALADISAILOED, T. & NUMNONDA, T. Machine learning models comparison for bitcoin price prediction. 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE), 2018. IEEE, 506-511.
PRABOWO, H., SUHARTONO, S. & PRASTYO, D. D. 2020. The Performance of Ramsey Test, White Test and Terasvirta Test in Detecting Nonlinearity. *Inferensi,* 3**,** 1-12.
PRADO, F., MINUTOLO, M. C. & KRISTJANPOLLER, W. 2020. Forecasting based on an ensemble autoregressive moving average-adaptive neuro-fuzzy inference system–neural network-genetic algorithm framework. *Energy,* 197**,** 117159.

QIU, M., SONG, Y. & AKAGI, F. 2016. Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market. *Chaos, Solitons & Fractals,* 85**,** 1-7.

QIU, X., REN, Y., SUGANTHAN, P. N. & AMARATUNGA, G. A. 2017. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. *Applied Soft Computing,* 54**,** 246-255.

RACINE, J. 2001. On the nonlinear predictability of stock returns using financial and economic variables. *Journal of Business & Economic Statistics,* 19**,** 380-382.

RADITYO, A., MUNAJAT, Q. & BUDI, I. Prediction of Bitcoin exchange rate to American dollar using artificial neural network methods. 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2017. IEEE, 433-438.

RAHMAN, M., ISLAM, A. S., NADVI, S. Y. M. & RAHMAN, R. M. Comparative study of ANFIS and ARIMA model for weather forecasting in Dhaka. 2013 International Conference on Informatics, Electronics and Vision (ICIEV), 2013. IEEE, 1-6.

RAMAKRISHNAN, S., BUTT, S., CHOHAN, M. A. & AHMAD, H. Forecasting Malaysian exchange rate using machine learning techniques based on commodities prices. 2017 International Conference on Research and Innovation in Information Systems (ICRIIS), 2017. IEEE, 1-5.

RAPOO, M. I., MUNAPO, E., CHANZA, M. M. & EWEMOOJE, O. S. 2020. Modelling and Forecasting Portfolio Inflows: A Comparative Study of Support Vector Regression, Artificial Neural Networks, and Structural VAR Models. *Handbook of Research on Smart Technology Models for Business and Industry.* IGI Global.

RAPOO, M. I. & XABA, D. 2017. A Comparative Analysis of Artificial Neural Network and Autoregressive Integrated Moving Average Model on Modeling and Forecasting Exchange Rate. *International Journal of Economics and Management Engineering,* 11**,** 2669-2672.

REMESAN, R., SHAMIM, M. A., HAN, D. & MATHEW, J. ANFIS and NNARX based rainfall-runoff modeling. 2008 IEEE International Conference on Systems, Man and Cybernetics, 2008. IEEE, 1454-1459.

RESNICK, A. L. 2001. Investors, turbulence, and transition: Democratic transition and foreign direct investment in nineteen developing countries. *International Interactions,* 27**,** 381-398.

RUSLAN, F. A. & ADNAN, R. 3 Hours flood water level prediction using NNARX structure: Case study Kuala Lumpur. 2016 6th International Conference on System Engineering and Technology (ICSET), 2016. IEEE, 53-56.

ŞAHIN, M. & EROL, R. 2017. A comparative study of neural networks and ANFIS for forecasting attendance rate of soccer games. *Mathematical and Computational Applications,* 22**,** 43.

SANGWAN, K. S., SAXENA, S. & KANT, G. 2015. Optimization of machining parameters to minimize surface roughness using integrated ANN-GA approach. *Procedia Cirp,* 29**,** 305-310.

SANSOM, D., DOWNS, T. & SAHA, T. 2002. Evaluation of support vector machine based forecasting tool in electricity price forecasting for australian national electricity market participants.

SANTAMARÍA-BONFIL, G., FRAUSTO-SOLÍS, J. & VÁZQUEZ-RODARTE, I. 2015. Volatility forecasting using support vector regression and a hybrid genetic algorithm. *Computational Economics,* 45**,** 111-133.

SAPANKEVYCH, N. I. & SANKAR, R. 2009. Time series prediction using support vector machines: a survey. *IEEE Computational Intelligence Magazine,* 4**,** 24-38.

SARKHEYLI, A., ZAIN, A. M. & SHARIF, S. 2015. Robust optimization of ANFIS based on a new modified GA. *Neurocomputing,* 166**,** 357-366.

SCHIILKOP, P., BURGEST, C. & VAPNIK, V. Extracting support data for a given task. Proceedings, First International Conference on Knowledge Discovery & Data Mining. AAAI Press, Menlo Park, CA, 1995. 252-257.

SCHUMACKER, R. E. & LOMAX, R. G. 2004. *A beginner's guide to structural equation modeling*, psychology press.

SEMERO, Y. K., ZHANG, J. & ZHENG, D. 2019. EMD–PSO–ANFIS-based hybrid approach for short-term load forecasting in microgrids. *IET Generation, Transmission & Distribution,* 14**,** 470-475.
SEXTON, R. S., DORSEY, R. E. & JOHNSON, J. D. 1998. Toward global optimization of neural networks: A comparison of the genetic algorithm and backpropagation. *Decision Support Systems,* 22**,** 171-185.
SFETSOS, A. 2000. A comparison of various forecasting techniques applied to mean hourly wind speed time series. *Renewable energy,* 21**,** 23-35.
SHABRI, A. & SAMSUDIN, R. A new approach for water demand forecasting based on empirical mode decomposition. 2014 8th. Malaysian Software Engineering Conference (MySEC), 2014. IEEE, 284-288.
SHABRI, A., SAMSUDIN, R. & TEKNOLOGI, U. 2015. Empirical mode decomposition–least squares support vector machine based for water demand forecasting. *Int. J. Adv. Soft Comput. Its Appl,* 7.
SHAGHAGHI, S., BONAKDARI, H., GHOLAMI, A., KISI, O., BINNS, A. & GHARABAGHI, B. 2019. Predicting the geometry of regime rivers using M5 model tree, multivariate adaptive regression splines and least square support vector regression methods. *International Journal of River Basin Management,* 17**,** 333-352.
SHALEV-SHWARTZ, S. & BEN-DAVID, S. 2014. *Understanding machine learning: From theory to algorithms*, Cambridge university press.
SHAO, S. & MUROTSU, Y. 1999. Approach to failure mode analysis of large structures. *Probabilistic Engineering Mechanics,* 14**,** 169-177.
SHAO, X. & ZHANG, X. 2010. Testing for change points in time series. *Journal of the American Statistical Association,* 105**,** 1228-1240.
SHARMA, D. K., HOTA, H., BROWN, K. & HANDA, R. 2022. Integration of genetic algorithm with artificial neural network for stock market forecasting. *International Journal of System Assurance Engineering and Management,* 13**,** 828-841.
SIGO, M. O., SELVAM, M., MANIAM, B., KANNAIAH, D., KATHIRAVAN, C. & VADIVEL, T. 2018. Big data analytics-application of artificial neural network in forecasting stock price trends in India. *Academy of Accounting and Financial Studies,* 22.
SMOLA, A. J. & SCHÖLKOPF, B. 1998. *Learning with kernels*, Citeseer.
SULAIMAN, M. H., MUSTAFA, M. W., SHAREEF, H. & KHALID, S. N. A. 2012. An application of artificial bee colony algorithm with least squares support vector machine for real and reactive power tracing in deregulated power system. *International Journal of Electrical Power & Energy Systems,* 37**,** 67-77.
SUYKENS, J. 2000. Least squares support vector machines for classification and nonlinear modelling. *Neural Network World,* 10**,** 29-48.
SUYKENS, J., VANDEWALLE, J. & EDITORS 1998. Nonlinear Modeling: Advanced Black-Box Techniques. 209-239.
SUYKENS, J. A., VAN GESTEL, T., DE BRABANTER, J., DE MOOR, B. & VANDEWALLE, J. P. 2002. *Least squares support vector machines*, World scientific.
SYLVAIRE, D. D. Y., QING, W. H., RAN, C. H., KASSAI, D. L., VINCENT, N., DOUCE, D. A. C., FRANK, O.-K., NICAISE, N. P., TRAORE, F. & BORIS, A. F. 2022. The impact of China's foreign direct investment on Africa's inclusive development. *Social Sciences & Humanities Open,* 6**,** 100276.
SZKUTA, B. R., SANABRIA, L. A. & DILLON, T. S. 1999. Electricity price short-term forecasting using artificial neural networks. *IEEE transactions on power systems,* 14**,** 851-857.
TALEBIZADEH, M. & MORIDNEJAD, A. 2011. Uncertainty analysis for the forecast of lake level fluctuations using ensembles of ANN and ANFIS models. *Expert Systems with applications,* 38**,** 4126-4135.
TAY, F. E. & CAO, L. 2001. Application of support vector machines in financial time series forecasting. *omega,* 29**,** 309-317.
TEKTAŞ, M. 2010. Weather forecasting using ANFIS and ARIMA models. *Environmental Research, Engineering and Management,* 51**,** 5-10.

TERÄSVIRTA, T., LIN, C. F. & GRANGER, C. W. 1993. Power of the neural network linearity test. *Journal of time series analysis,* 14**,** 209-220.
THENMOZHI, M. 2006. Forecasting stock index returns using neural networks. *Delhi Business Review,* 7**,** 59-69.
TIAN, J. & GAO, M. Network intrusion detection method based on high speed and precise genetic algorithm neural network. 2009 International Conference on Networks Security, Wireless Communications and Trusted Computing, 2009. IEEE, 619-622.
TIAN, Y., HU, W., DU, B., HU, S., NIE, C. & ZHANG, C. 2019. IQGA: A route selection method based on quantum genetic algorithm-toward urban traffic management under big data environment. *World Wide Web,* 22**,** 2129-2151.
TONG, H. 1986. "on-linear time series models of regularly sampled data: a review', an invited paper presented at the First World Congress of the Bernoulli Society for Mathematical Statistics and Probability.
TRAFALIS, T. B. & INCE, H. Support vector machine for regression and applications to financial forecasting. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, 2000. IEEE, 348-353.
TSAY, R. S. 1986. Nonlinearity tests for time series. *Biometrika,* 73**,** 461-466.
TSENG, F.-M., TZENG, G.-H., YU, H.-C. & YUAN, B. J. 2001. Fuzzy ARIMA model for forecasting the foreign exchange market. *Fuzzy sets systems,* 118**,** 9-19.
TURNER, P. 2010. Power properties of the CUSUM and CUSUMSQ tests for parameter instability. *Applied Economics Letters,* 17**,** 1049-1053.
VAN GESTEL, T., SUYKENS, J. A., BAESTAENS, D.-E., LAMBRECHTS, A., LANCKRIET, G., VANDAELE, B., DE MOOR, B. & VANDEWALLE, J. 2001. Financial time series prediction using least squares support vector machines within the evidence framework. *IEEE Transactions on neural networks,* 12**,** 809-821.
VAPNIK, V. 2013. *The nature of statistical learning theory*, Springer science & business media.
VAPNIK, V., GOLOWICH, S. E. & SMOLA, A. 1997. Support vector method for function approximation, regression estimation, and signal processing. *Advances in neural information processing systems***,** 281-287.
VAPNIK, V. N. 1995. The nature of statistical learning theory. *New York: Springer*.
VELTCHEVA, A. & SOARES, C. G. 2004. Identification of the components of wave spectra by the Hilbert Huang transform method. *Applied Ocean Research,* 26**,** 1-12.
VENKATRAMAN, S. & YEN, G. G. 2005. A generic framework for constrained optimization using genetic algorithms. *IEEE Transactions on Evolutionary Computation,* 9**,** 424-435.
VESTERSTROM, J. & THOMSEN, R. A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems. Proceedings of the 2004 congress on evolutionary computation (IEEE Cat. No. 04TH8753), 2004. IEEE, 1980-1987.
VINCENT, H., HU, S. & HOU, Z. Damage detection using empirical mode decomposition method and a comparison with wavelet analysis. Proceedings of the 2nd International Workshop on Structural Health Monitoring, 1999. Stanford Univeristy Standford, 891-900.
VIRILI, F. & FREISLEBEN, B. Nonstationarity and data preprocessing for neural network predictions of an economic time series. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, 2000. IEEE, 129-134.
WAFURE, O. G. & NURUDEEN, A. 2010. Determinants of foreign direct investment in Nigeria: An empirical analysis. *Global journal of human social science,* 10**,** 26-34.
WANG, L., ZENG, Y. & CHEN, T. 2015a. Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. *Expert Systems with Applications,* 42**,** 855-863.
WANG, S. & WANG, Q. 2012. Prediction and dispatching of workshop material demand based on least squares support vector regression with genetic algorithm. *International Information Institute (Tokyo). Information,* 15**,** 213.

WANG, W.-C., CHAU, K.-W., CHENG, C.-T. & QIU, L. 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of hydrology,* 374**,** 294-306.

WANG, W.-C., CHAU, K.-W., XU, D.-M. & CHEN, X.-Y. 2015b. Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resources Management,* 29**,** 2655-2675.

WANG, Z.-L. & SHENG, H.-H. Rainfall prediction using generalized regression neural network: case study Zhengzhou. 2010 International conference on computational and information sciences, 2010. IEEE, 1265-1268.

WEI, L.-Y. 2013. A GA-weighted ANFIS model based on multiple stock market volatility causality for TAIEX forecasting. *Applied Soft Computing,* 13**,** 911-920.

WEI, L.-Y. 2015. A Hybrid Model Based on ANFIS and Empirical Mode Decomposition for Stock Forecasting. *Journal of Economics, Business and Management,* 3.

WEI, L.-Y. 2016. A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting. *Applied Soft Computing,* 42**,** 368-376.

WEI, Y. & CHEN, M.-C. 2012. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C: Emerging Technologies,* 21**,** 148-162.

WEISANG, G. & AWAZU, Y. 2008. Vagaries of the Euro: an Introduction to ARIMA Modeling. *Case Studies in Business, Industry, and Government Statistics,* 2**,** 45-55.

WHITE, H. Economic prediction using neural networks: The case of IBM daily stock returns. ICNN, 1988. 451-458.

WHITE, H. An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks. Proceedings of the international joint conference on neural networks, 1989. Washington, DC, 451-455.

WHITLEY, D., STARKWEATHER, T. & BOGART, C. 1990. Genetic algorithms and neural networks: Optimizing connections and connectivity. *Parallel computing,* 14**,** 347-361.

WIENER, N. 1949. The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications. *Wiley, New York*.

WITT, S. F. & WITT, C. A. 1995. Forecasting tourism demand: A review of empirical research. *International Journal of forecasting,* 11**,** 447-475.

WU, C.-H., HO, J.-M. & LEE, D.-T. 2004. Travel-time prediction with support vector regression. *IEEE transactions on intelligent transportation systems,* 5**,** 276-281.

WU, C.-H., TZENG, G.-H. & LIN, R.-H. 2009. A novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression. *Expert Systems with Applications,* 36**,** 4725-4735.

XABA, D., MOROKE, N. D. & RAPOO, I. 2019. Modeling stock market returns of BRICS with a Markov-switching dynamic regression Model. *Journal of Economics and Behavioral Studies,* 11**,** 10-22.

XABA, L. D. 2014. *A comparative study of threshold autoregressive, smooth transition regression and markov switching autoregressive models on stock price behaviour.*

XIA, Z., MAO, K., WEI, S., WANG, X., FANG, Y. & YANG, S. 2017. Application of genetic algorithmsupport vector regression model to predict damping of cantilever beam with particle damper. *Journal of low frequency, noise, vibration and active control,* 36**,** 138–147.

XIANG, Y., GOU, L., HE, L., XIA, S. & WANG, W. 2018. A SVR–ANN combined model based on ensemble EMD for rainfall prediction. *Applied Soft Computing,* 73**,** 874-883.

XIE, C., MAO, Z. & WANG, G.-J. 2015. Forecasting RMB exchange rate based on a nonlinear combination model of ARFIMA, SVM, and BPNN. *Mathematical Problems in Engineering,* 2015.

XU, L., HOU, L., ZHU, Z., LI, Y., LIU, J., LEI, T. & WU, X. 2021. Mid-term prediction of electrical energy consumption for crude oil pipelines using a hybrid algorithm of support vector machine and genetic algorithm. *Energy,* 222**,** 119955.

YAHYA, N. A., SAMSUDIN, R. & SHABRI, A. 2017. Tourism forecasting using hybrid modified empirical mode decomposition and neural network. *Int. J. Advance Soft Compu. Appl,* 9.

YAN-GANG, Z. & JIN-REN, J. 1995. A structural reliability analysis method based on genetic algorithm. *Earthquake Engineering and Engineering Vibration,* 15**,** 47-58.

YANG, H., CHAN, L. & KING, I. Support vector machine regression for volatile stock market prediction. International Conference on Intelligent Data Engineering and Automated Learning, 2002. Springer, 391-396.

YASLAN, Y. & BICAN, B. 2017. Empirical mode decomposition based denoising method with support vector regression for time series prediction: A case study for electricity load forecasting. *Measurement,* 103**,** 52-61.

YINFENG, D., YINGMIN, L., MINGKUI, X. & MING, L. 2008. Analysis of earthquake ground motions using an improved Hilbert–Huang transform. *Soil Dynamics and Earthquake Engineering,* 28**,** 7-19.

YU, C.-J., HE, Y.-Y. & QUAN, T.-F. Frequency spectrum prediction method based on EMD and SVR. 2008 Eighth International Conference on Intelligent Systems Design and Applications, 2008a. IEEE, 39-44.

YU, L., DAI, W., TANG, L. & WU, J. 2016. A hybrid grid-GA-based LSSVR learning paradigm for crude oil price forecasting. *Neural computing and applications,* 27**,** 2193-2215.

YU, L., WANG, S. & LAI, K. K. 2008b. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics,* 30**,** 2623-2635.

YU, X. & GEN, M. 2010. *Introduction to evolutionary algorithms*, Springer Science & Business Media.

YUAN, F.-C., LEE, C.-H. & CHIU, C. 2020. Using market sentiment analysis and genetic algorithm-based least squares support vector regression to predict gold prices. *International Journal of Computational Intelligence Systems,* 13**,** 234-246.

ZENG, Q., QU, C., NG, A. K. & ZHAO, X. 2016. A new approach for Baltic Dry Index forecasting based on empirical mode decomposition and neural networks. *Maritime Economics & Logistics,* 18**,** 192-210.

ZHANG, F., DEB, C., LEE, S. E., YANG, J. & SHAH, K. W. 2016a. Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique. *Energy and Buildings,* 126**,** 94-103.

ZHANG, G. & HU, M. Y. 1998. Neural network forecasting of the British pound/US dollar exchange rate. *omega,* 26**,** 495-506.

ZHANG, G., PATUWO, B. E. & HU, M. Y. 1998. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting,* 14**,** 35-62.

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

ZHANG, G. P. & KLINE, D. M. 2007. Quarterly time-series forecasting with neural networks. *IEEE transactions on neural networks,* 18**,** 1800-1814.

ZHANG, G. P., PATUWO, B. E. & HU, M. Y. J. C. 2001. A simulation study of artificial neural networks for nonlinear time-series forecasting. *Computers & Operations Research,* 28**,** 381-396.

ZHANG, H., XIAO, Y., BAI, X. & CHEN, L. 2016b. GA-support vector regression based ship traffic flow prediction. *International Journal of Control and Automation,* 9**,** 219-228.

ZHANG, K., GENÇAY, R. & YAZGAN, M. E. 2017a. Application of wavelet decomposition in time-series forecasting. *Economics Letters,* 158**,** 41-46.

ZHANG, K. H. 2021. How does South-South FDI affect host economies? Evidence from China-Africa in 2003–2018. *International Review of Economics & Finance,* 75**,** 690-703.

ZHANG, N., LIN, A. & SHANG, P. 2017b. Multidimensional k-nearest neighbor model based on EEMD for financial time series forecasting. *Physica A: Statistical Mechanics and its Applications,* 477**,** 161-173.

ZHANG, X., LAI, K. K. & WANG, S.-Y. 2008. A new approach for crude oil price analysis based on empirical mode decomposition. *Energy economics,* 30**,** 905-918.

ZHANG, Z., OUYANG, Z., MA, C., XU, J., LI, L. & WEN, Q. 2016c. A Novel Combination Forecast Method on Non-Stationary Time Series and its Application to Exchange Rate Forecasting. *Int. J. Simulation—Systems, Sci. Technol.,* 17.

ZHANG, Z., ZHENG, H., LIU, Y., HE, G. & DING, D. Study on the PSA-ANFIS approach for inverse design in slope engineering. 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, 2012. IEEE, 170-174.

ZHU, B., YE, S., WANG, P., CHEVALLIER, J. & WEI, Y. M. 2022. Forecasting carbon price using a multi-objective least squares support vector machine with mixture kernels. *Journal of Forecasting,* 41**,** 100-117.

ZHU, Z., SUN, Y. & LI, H. Hybrid of EMD and SVMs for short-term load forecasting. 2007 IEEE International Conference on Control and Automation, 2007. IEEE, 1044-1047.

# Annexures

## Annexure 1: Multicollinearity R codes/syntax

```
lm_model <- lm(FDI ~ Inflation.rate + Exchange.rate
               + gdp, data = NewPhDTSponki)
vif_values <- car::vif(lm_model)
print(vif_values)
```

## Annexure 2: Nonlinearity tests R codes/syntax

```
Tsay.test(sqrt(FDI))
Tsay.test(sqrt(gdp))
Tsay.test(sqrt(Exchange.rate))
Tsay.test(sqrt(Inflation.rate))


McLeod.Li.test(y=r.crefgdp)
McLeod.Li.test(y=r.crefExRate)
McLeod.Li.test(y=r.crefInRate)
McLeod.Li.test(y=r.creffdi)


terasvirta.test(Exchange.rate, lag = 1, type = c("Chisq"),
                scale = TRUE)
terasvirta.test(gdp, lag = 1, type = c("Chisq"),
                scale = TRUE)
terasvirta.test(Inflation.rate, lag = 1, type = c("Chisq"),
                scale = TRUE)
terasvirta.test(FDI, lag = 1, type = c("Chisq"),
                scale = TRUE)


white.test(Inflation.rate, lag = 1, qstar = 2, q = 10, range = 4,
           type = c("Chisq"), scale = TRUE)
white.test(Exchange.rate, lag = 1, qstar = 2, q = 10, range = 4,
           type = c("Chisq"), scale = TRUE)
white.test(gdp, lag = 1, qstar = 2, q = 10, range = 4,
           type = c("Chisq"), scale = TRUE)
white.test(FDI, lag = 1, qstar = 2, q = 10, range = 4,
           type = c("Chisq"), scale = TRUE)


Keenan.test(Exchange.rate, order = 1)
Keenan.test(gdp, order = 1)
Keenan.test(Inflation.rate, order = 1)
Keenan.test(FDI, order = 1)
```

## Annexure 3: Traditional machine learning R/Python codes/syntax

## Artificial neural network (ANN)

```
#Installation of packages
install.packages("neuralnet")
require(neuralnet)
```

```
install.packages("forecast")
require(forecast)
install.packages("Metrics")
require(Metrics)

#Uploading of data to R global environment
PhDTSdata_original_5   <-   read.csv("C:/Users/RAPOOM/Desktop/PhDTSdata_original_5.csv",
sep=";")
PhDTSdata_original1ts6 <- data.frame(PhDTSdata_original_5)
head(PhDTSdata_original1ts6, n = 20)
PhDPonkiFDI<-as.numeric(PhDTSdata_original1ts6$FDI)
PhDPonkiInflation<-as.numeric(PhDTSdata_original1ts6$Inflation)
PhDPonkiExchangerate<-as.numeric(PhDTSdata_original1ts6$Exchange_rate)
PhDPonkigdp<-as.numeric(Datammm12$gdp)
Datammm121TS
data.frame(PhDPonkiFDI,PhDPonkiInflation,PhDPonkiExchangerate,PhDPonkigdp)
FDI <- PhDPonkiFDI
Inflation.rate <- PhDPonkiInflation
Exchange.rate <- PhDPonkiExchangerate
gdp <- PhDPonkigdp
head(gdp, n =5)

#Plotting of time series data
NewPhDTS <- data.frame(FDI, Inflation.rate, Exchange.rate, gdp)
plot(NewPhDTS)
NewPhDTSponki <- ts(NewPhDTS, frequency = 12, start = c(1970,1))
str(NewPhDTSponki)
head(NewPhDTSponki, n =20)
plot(NewPhDTSponki)

#Checking for missing values
apply(NewPhDTSponki,2,function(x) sum(is.na(x)))

#Scaling and spliting the data before fitting NN model
maxs <- apply(NewPhDTSponki, 2, max)
maxs
mins <- apply(NewPhDTSponki, 2, min)
mins
scaled <- as.data.frame(scale(NewPhDTSponki, center = mins, scale = maxs - mins))
maxscaled <- apply(scaled, 2, max)
maxscaled
minscaled <- apply(scaled, 2, min)
minscaled
index = sample(1:nrow(NewPhDTSponki),round(0.80*nrow(NewPhDTSponki)))
train_dataPonki <- as.data.frame(scaled[index,])
train_dataPonki
test_dataPonki <- as.data.frame(scaled[-index,])


#Training NN model
n <- names(train_dataPonki)
f <- as.formula(paste("FDI ~", paste(n[!n %in% "FDI"], collapse = " + ")))
nn <- neuralnet(f,data=train_dataPonki,hidden=c(7,6,5),linear.output=TRUE)
plot(nn)

#Predicting FDI
temp_test  <-  subset(test_dataPonki,  select  =  c("Inflation.rate",  "Exchange.rate",
"gdp"))
head(temp_test)
nn.results <- compute(nn, temp_test)
results <- data.frame(actual = test_dataPonki$FDI, prediction = nn.results$net.result)

#Forecasting performance
actualAnn <- results$actual
preditedAnn <- results$prediction
etAnn <- actualAnn - preditedAnn
mse = mean((actualAnn-preditedAnn)^2)
```

```
rmse <- sqrt(mse)
```

### Support vector regression

```
#Installation of packages
install.packages('e1071')
require('e1071')
install.packages('Metrics')
require('Metrics')

#Scalling and spliting the data before fitting SVR model
maxs <- apply(NewPhDTSponki, 2, max)
maxs
mins <- apply(NewPhDTSponki, 2, min)
mins
scaled <- as.data.frame(scale(NewPhDTSponki, center = mins, scale = maxs - mins))
maxscaled <- apply(scaled, 2, max)
maxscaled
minscaled <- apply(scaled, 2, min)
minscaled
index = sample(1:nrow(NewPhDTSponki),round(0.80*nrow(NewPhDTSponki)))
train_dataPonki <- as.data.frame(scaled[index,])
train_dataPonki
test_dataPonki <- as.data.frame(scaled[-index,])
test_dataPonki

#Regression with SVM
modelsvr <- svm(FDI~. ,train_dataPonki)

#Predict using SVM regression
predFDI = predict(modelsvr, train_dataPonki)
predFDI

#Calculate parameters of the SVR model
W3FDI = t(modelsvr$coefs)
b3FDI = modelsvr$rho

#RMSE for SVR Model
#Calculate RMSE
RMSEsvr =rmse(predFDI,train_dataPonki$FDI)

#Tuning SVR model by varying values of maximum allowable error and cost parameter
#Tune the SVM model
OptModelsvmMy_leader_1=tune(svm, FDI~., data=train_dataPonki,
                            ranges=list(epsilon=seq(0,1,0.1), cost=1:100))

#Print optimum value of parameters
print(OptModelsvmMy_leader_1)

#Plot the perfrormance of SVM Regression model
plot(OptModelsvmMy_leader_1)

#Select the best model out of 1100 trained models and compute RMSE
#Find out the best model
tunedModel <- OptModelsvmMy_leader_1$best.model

#Predict Y using best model
PredFDIBst=predict(tunedModel,test_dataPonki)
head(PredFDIBst)

#Calculate RMSE of the best model
RMSLEBst=rmsle(PredFDIBst,test_dataPonki$FDI)

#Calculate parameters of the Best SVR model
```

```
#Find value of W
W = t(tunedModel$coefs)

#Find value of b
b = tunedModel$rho
```

### Least square support vector regression

```
#Define Path to PhD Analysis Folder

PATH = " C:/Users/Jabulani/Desktop/2022-Life/4. Experiments/PhD Analysis/"


import numpy as np

import pandas as pd


import sys

sys.path.append(PATH)

from PHd_Functions import LSSVR, Focasting_Errors


#Set Class version

__all__ = ['LSSVR']

__version__ = '0.0.1'


# # Step2: Data Import

df = pd.read_excel(PATH + 'Research_data5.csv')

#Convert Pandas df to numpy array

X = np.array(X)#.to_numpy()

y = np.array(y)#y.to_numpy()


#Data Spliting

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

print('Train Sets ', X_train.shape, y_train.shape,)

print('Test Sets ', X_test.shape, y_test.shape)


#Fit LSSVR Regressor

model = LSSVR(kernel='rbf', gamma=0.01)

model.fit(X_train, y_train)
```

```
model


Preds =
pd.concat([pd.DataFrame(y_test).rename(columns={0:'Actual'}).reset_index(drop=True),
pd.DataFrame(model.predict(X_test)).rename(columns={0:'Predictions'})],axis=1)

Preds


# # Step5: Model Selection

Preds ['Residual']= [Preds['Actual'][i]-Preds['Predictions'][i] for i in
range(len(Preds))]

Preds


Focasting_Errors(y_true = Preds['Actual'], y_pred = Preds['Predictions'], Model_Name =
model)
```

## Annexure 4: Machine learning based evolutionary algorithm hybrid model Python codes/syntax

### ANN-GA

```
pip install tpot


import pandas as pd

from google.colab import files

data_to_load = files.upload()

import io

train = pd.read_excel(io.BytesIO(data_to_load['Research_data5-csv.xlsx']))

train


X, y = train.drop(['FDI'],axis=1), train[['FDI']]

X

Y


from sklearn.preprocessing import StandardScaler

scale = StandardScaler()

x_scaled = scale.fit_transform(X)

x_scaled

y_scaled = scale.fit_transform(y)

y_scaled
```

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(x_scaled, y_scaled, test_size=0.20
)

print(X_train.shape, X_test.shape)

print(y_train.shape, y_test.shape)


#Base model


from sklearn.neural_network import MLPRegressor

regressor = MLPRegressor()

regressor.fit(X_train, y_train)


regressor.score(X_train, y_train)

regressor.score(X_test,y_test)



from tpot import TPOTRegressor


param_grid = {'hidden_layer_sizes': [(50,50,50), (50,100,50), (100,1)],

          'activation': ['relu','tanh','logistic'],

          'alpha': [0.0001, 0.05],

          'learning_rate': ['constant','adaptive'],

          'solver': ['adam']

          }


model = TPOTRegressor(generations = 5,

                      population_size = 24,

                      offspring_size = 12,

                      verbosity = 2,

                      early_stop = 5,

                      config_dict=

                      {'sklearn.neural_network.MLPRegressor':param_grid},

                      cv = 5,

                      scoring = 'r2'

                      )
```

```
model.fit(X_train, y_train)


model.score(X_test, y_test)


model.score(X_train, y_train)
model.score(X_test, y_test)


pred = pd.DataFrame(regressor.predict(X_test))
pred


from google.colab import files, drive
drive.mount('drive')
pred.to_csv('ANNGApred.csv', encoding = 'utf-8-sig', index=False)
files.download('ANNGApred.csv')
```

## SVR-GA

```
#Installation of GA package
install.packages("GA")
library(GA)


#Setup the data for cross-validation
#5-fold cross-validation
K = 5
fold_inds <- sample(1:K, nrow(NewPhDTSponki), replace = TRUE)
lst_CV_data <- lapply(1:K, function(i) list(
  train_dataPonki = NewPhDTSponki[fold_inds != i, , drop = FALSE],
  test_dataPonki = NewPhDTSponki[fold_inds == i, , drop = FALSE]))


#Given the values of parameters 'cost', 'gamma' and 'epsilon', return the rmse of the
model over the test data


evalParams <- function(train_dataPonki, test_dataPonki, cost, gamma, epsilon) {
  # Train
  model <- svm(FDI ~ ., data = train_dataPonki, cost = cost, gamma = gamma, epsilon =
epsilon, type = "eps-regression", kernel = "radial")
  # Test
  rmse <- mean((predict(model, newdata = test_dataPonki) - test_dataPonki) ^ 2)
  return (rmse)
}


#Fitness function (to be maximized)
```

```
#Parameter vector x is: (cost, gamma, epsilon)

fitnessFunc <- function(x, Lst_CV_Data) {
#Retrieve the SVM parameters
  cost_val <- x[1]
  gamma_val <- x[2]
  epsilon_val <- x[3]

#Use cross-validation to estimate the RMSE for each split of the dataset
  rmse_vals <- sapply(Lst_CV_Data, function(in_data) with(in_data,
                                                          evalParams(train_dataPonki,
test_dataPonki, cost_val, gamma_val, epsilon_val)))

#As fitness measure, return minus the average rmse (over the cross-validation folds),
#so that by maximizing fitness we are minimizing the rmse
  return (-mean(rmse_vals))
}

#Range of the parameter values to be tested
#Parameters are: (cost, gamma, epsilon)
theta_min <- c(cost = 1e-4, gamma = 1e-3, epsilon = 1e-2)
theta_max <- c(cost = 10, gamma = 2, epsilon = 2)

#Run the genetic algorithm
results <- ga(type = "real-valued", fitness = fitnessFunc, lst_CV_data,
              names = names(theta_min),
              lower = theta_min, upper =  theta_max,
              popSize = 50, maxiter = 10)
summary(results)

#Substitute the optimal values to the model: ost = 2.33213, gamma = 0.09883, and epsilon
= 0.37549
modelGAtuned <- svm(FDI ~., data = train_dataPonki, cost = 2.33213,
             gamma   =   0.09883,   epsilon   =   0.37549,   type   =   "eps-regression",
kernel="radial")

#Predicted values of SVR-GA
PredFDIGA=predict(modelGAtuned,test_dataPonki)
head(PredFDIGA)
RMSLEGA=rmsle(PredFDIGA,test_dataPonki$FDI)
```

## Annexure 5: Model accuracy (Diebold-Mariano) R codes/syntax

```
dm.test(etANN, etSvrga)
dm.test(etLSSVR, etANN)
dm.test(etAnnGA, etANN)
dm.test(etAnnGA, etLSSVR)
dm.test(etSvrga, etLSSVR)
dm.test(etSvrga, etANN)
dm.test(etAnnGA, etsvr)
dm.test(etsvr, etSvrga)
dm.test(etsvr, etLSSVR)
dm.test(etsvr, etANN)
```

**Appendix A**

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT

Private Bag X6001, Potchefstroom
South Africa 2520 Tel: 018
299-1111/2222
Web: http://www.nwu.ac.za

Economic and Management Sciences Rese
Ethics Committee (EMS-REC)

25 November 2020

Profs E Munapo and Dr M Chanza
*Per e-mail*

Dear Prof Munapo and Dr Chanza,

## EMS-REC FEEDBACK: 20112020

## Student: Rapoo, MI (23809213)(NWU-00928-20-A4)

## Applicant: Prof E Munapo / Dr M Chanza – PhD in Statistics

Your ethics application on, ***Modelling and forecasting foreign direct investment: A comparative application of machine learning based evolutionary algorithms hybrid models**,* which served on the EMS-REC meeting of 20 November 2020, refers.

### Outcome:

Approved as a minimal risk study. A number **NWU-00928-20-A4** is given for one year of ethics clearance.

Due to the Covid-19 lock down ethics clearance for applications that involve data collection or any form of contact with participants are subject to the restrictions imposed by the South African government.

Kind regards,

**Prof Mark Rathbone**
**Chairperson: Economic and Management Sciences Research Ethics Committee (EMS-REC)**

## Appendix B

*MBIRAZ INVESTMENTS PROOFREADERS AND EDITORS*

# CERTIFICATE OF PROOFREADING

THIS IS TO ACKNOWLEDGE THAT THE THESIS ENTITLED:

## Modelling and Forecasting Foreign Direct Investment: A Comparative Application of Machine Learning Based Evolutionary Algorithms Hybrid Models

WRITTEN BY:

MOGARI ISHMAEL RAPOO

Has been proofread and returned to the principal owner on 02 December 2021. This thesis has been edited for English language usage, grammar, spelling, and punctuation by one or more qualified editors at Mbiraz Investments (PTY) LTD. Research content and the author's intentions were not altered in any way.

---

*Victor Mambiravana*
*Chief Editor*
*Mbiraz Investments*



*This certificate may be verified by emailing us at info@mbirazinvestments.co.za*

**Appendix C**

23809213:Modelling and Forecasting Foreign Direct Investment....pdf

ORIGINALITY REPORT

| | | | |
|---|---|---|---|
| 13% | 14% | 10% | 6% |
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Submitted to North West University<br>Student Paper | 5% |
| 2 | dspace.nwu.ac.za<br>Internet Source | 2% |
| 3 | journals.sagepub.com<br>Internet Source | 1% |
| 4 | subasish.github.io<br>Internet Source | 1% |
| 5 | www.elsevier.es<br>Internet Source | 1% |

| | | |
|---|---|---|
| 6 | Roscoe Bertrum van Wyk, Bianca Flavia van Wyk, Katleho Daniel Makatjane. "Trends in foreign agricultural trade and its impact on households in South Africa", Outlook on Agriculture, 2020<br>Publication | 1% |
| 7 | W. S. Chan. "On tests for non-linearity in time series analysis", Journal of Forecasting, 10/1986<br>Publication | 1% |
| 8 | www.mirlabs.org<br>Internet Source | 1% |
| 9 | iptek.its.ac.id<br>Internet Source | 1% |
| 10 | dse.unibo.it<br>Internet Source | 1% |

| | | | |
|---|---|---|---|
| Exclude quotes | On | Exclude matches | < 1% |
| Exclude bibliography | On | | |