TIME SERIES ANALYSIS AND IMPROVED DEEP LEARNING MODEL FOR ELECTRICITY PRICE FORECASTING

MD RASHED IQBAL

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2022

TIME SERIES ANALYSIS AND IMPROVED DEEP LEARNING MODEL FOR ELECTRICITY PRICE FORECASTING

MD RASHED IQBAL

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ENGINEERING SCIENCE

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2022

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Md Rashed Iqbal

Matric No: 17221223/1

Name of Degree: Master of Engineering Science

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

TIME SERIES ANALYSIS AND IMPROVED DEEP LEARNING MODEL FOR ELECTRICITY PRICE FORECASTING

Field of Study: Power System

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

TIME SERIES ANALYSIS AND IMPROVED DEEP LEARNING MODEL FOR ELECTRICITY PRICE FORECASTING

ABSTRACT

Accurate electricity price forecasting (EPF) is important for the purpose of bidding strategies and minimizing the risk for market participants in the competitive electricity market. However, accurate prediction is very challenging due to complex nonlinearity in electricity prices. Therefore, forecasting accuracy highly depends on the nature of time series. An improved deep-learning framework is proposed for short and mid-term EPF which consists of four modules: time-series data pre-processing, the deep learning-based prediction methodology, spike prediction module and reliability checking of prediction model. The feature pre-processing module is based on linear trend of the correlated features of electricity price series and test time series for unit root by augmented dickey fuller (ADF). In addition, the time series data is transformed with box-cox transformation method for better training process. Firstly, the prediction module combines linear scaled hyperbolic tangent (LISHT) with the long short-term memory (LSTM) and compared with bidirectional long short-term memory (BiLSTM) which is a recurrent neural network (RNN) to adjust complex nonlinear features and improve the precision of day ahead prediction. The residual autocorrelation determined in the reliability check section. Secondly, an optimized gated recurrent unit (GRU) which incorporates bagged tree ensemble (BTE) is developed in the recurrent neural network (RNN) architecture for the mid-term EPF. A tanh layer is employed to optimize the hyperparameters of the heterogeneous GRU with the aim to improve the model's performance, error reduction and predict the spikes. This study is performed based on the Australian price, load and renewable energy supply data from five major economical states New South Wales (NSW), Queensland (QLD), South Australia (SA), Tasmania (TAS), Victoria (VIC). The experimental results obtained show that the proposed EPF framework performed better compared to previous techniques.

Keywords: LSTM, deep learning, time series, electricity price forecasting, smart grid.

ANALISIS SIRI MASA DAN MODEL PEMBELAJARAN MENDALAM YANG DIPERBAIKI UNTUK RAMALAN HARGA ELEKTRIK

ABSTRAK

Ramalan harga elektrik (EPF) yang tepat adalah penting untuk tujuan strategi pembidaan dan untuk meminimumkan risiko peserta pasaran dalam pasaran elektrik yang kompetitif. Walau bagaimanapun, ramalan yang tepat adalah sangat mencabar disebabkan oleh ketidaklinearan yang kompleks dalam harga elektrik. Oleh itu, ketepatan ramalan sangat bergantung pada sifat siri masa. Rangka kerja pembelajaran mendalam yang dipertingkatkan dicadangkan untuk EPF jangka pendek dan sederhana yang terdiri daripada empat modul: pra-pemprosesan data siri masa, metodologi ramalan berasaskan pembelajaran mendalam, modul ramalan lonjakan dan semakan kebolehpercayaan model ramalan. Modul pra-pemprosesan ciri adalah berdasarkan aliran linear ciri berkorelasi siri harga elektrik dan siri masa ujian untuk sumber unit oleh augmented dickey fuller (ADF). Selain itu, data siri masa telah diubah suai dengan kaedah transformasi box-cox untuk proses latihan yang lebih baik. Pertama, modul ramalan menggabungkan tangen hiperbolik skala linear (LISHT) dengan ingatan jangka pendek (LSTM) jangka panjang dan membandingkannya dengan BiLSTM yang merupakan rangkaian saraf berulang (RNN) untuk melaraskan ciri tak linear yang kompleks dan meningkatkan ketepatan ramalan masa hadapan. Autokorelasi sisa ditentukan dalam bahagian semakan kebolehpercayaan. Kedua, unit berulang berpagar (GRU) yang dioptimumkan yang menggabungkan bagged tree ensemble (BTE) telah dibangunkan dalam seni bina rangkaian neural nerulang (RNN) untuk EPF jangka sederhana. Lapisan tanh telah digunakan untuk mengoptimumkan hiperparameter GRU heterogen dengan tujuan untuk meningkatkan prestasi model, pengurangan ralat dan meramalkan pancang. Kajian ini dijalankan berdasarkan harga Australia, beban dan data bekalan tenaga boleh diperbaharui daripada lima negeri ekonomi utama New South Wales (NSW), Queensland

(QLD), South Australia (SA), Tasmania (TAS), Victoria (VIC). Keputusan eksperimen yang diperolehi menunjukkan rangka kerja EPF yang dicadangkan menghasilkan prestasi yang lebih baik daripada teknik sebelumnya.

Kata kunci: LSTM, pembelajaran mendalam, siri masa, ramalan harga elektrik, grid pintar.

ACKNOWLEDGEMENTS

All deepest thanks to almighty Allah who is kind, great merciful and source of all knowledge.

First, I would like to thank my honorable supervisors Professor Ir. Dr. Hazlie Bin Mokhlis and Dr. Anis Salwa Binti Mohd Khairuddin. It has been an honor to be one of them masters' students under their guidance. From the beginning to end of the research their inspiration and contributions in terms of ideas, time management, productive way of working and simulation were outstanding. I want to thank them again for having trust on me and co-operating in every possible way which made the research successful and effective. They have been very friendly and as guardian for me.

Personal thanks to my senior colleagues specially Dr. Munir Azam Mohammad, Syafiqah Ismail for supporting in this research project. There are few lab members in power system lab specially Dr. Abdullah, Hamza, Dr. Naveed, Dr. Sohail, Dr. Shameem and Dr. Ali have contributed immensely to my personal and professional matter while discussed some issues related to work.

Lastly, I want to thank my family members who supported me throughout the whole journey of two years by supporting financially and mentally.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

ACF	:	Autocorrelation function
ADF	:	Augmented dickey fuller
AI	:	Artificial intelligence
ANN	:	Artificial neural network
AR	:	Auto regressive
ARIMA	:	Auto regressive integrated moving average
AUD	:	Australian dollar
EPF	:	Electricity price forecast
ESOO	:	Electricity statement of opportunities
GRU	:	Gated recurrent unit
LISHT	:	Linear scaled hyperbolic tangent
LSTM	:	Long short-term memory
LSSVM	:	Least square support vector machine
MAPE	:	Mean average percent error
ML	:	Machine Learning
MWH		Megawatt hour
NSW	:	New south Wales
PSO	:	Particle swarm optimization
QLD	:	Queensland
RNN	:	Recurrent neural network
SA	:	South Australia
SARIMA	:	Seasonal auto regressive integrated moving average
SFE	:	Supply function equilibrium
SVM	:	Support vector machine

SVR	:	Support vector regression
TAS	:	Tasmania
VIC	:	Victoria

CHAPTER 1: INTRODUCTON

1.1 Introduction

Smart grids (SG) are being utilized to enhance the traditional grid's performance and efficiency. With the privatization of the energy business, electricity price forecasting has become vital for optimal power system planning and operation. Market players will be able to adapt their bidding methods and production/consumption schedules in order to optimize their earnings in the electricity market with the help of accurate price forecasts (Almeshaiei & Soltan, 2011; Ozer, Efe, & Ozbay, 2021). In several countries, deregulations of the electricity sector have been developed to enhance congestion control, facilitate renewable energy, and maximize the resource allocation of the power system (Pourdaryaei, Mokhlis, Illias, Kaboli, & Ahmad, 2019). Electricity markets depend largely on electricity prices. Precise electricity price forecasting (EPF) provides vital information to all stakeholders in the power sector marketplace (Alazab et al., 2020). In the electricity distribution sector, the accuracy of EPF influence the performance and rational analysis of energy resource optimization. Accurate price forecasting can boost profitability in day-ahead trades and energy management by improving commercial electricity pricing and production (Zhou, Zhou, Mao, Tai, & Wan, 2019). Artificial intelligence (AI) approaches such as shallow learning models and deep learning models have demonstrated to be feasible in electricity price forecasting due to the nonlinearity and high volatility of the features in EPF (Huaizhi Wang et al., 2020). Electricity price prediction forecasts are an important aspect of regulators', governments', and financial market participants' expectations. EPF has been a critical tool for managing the competitive electricity market as well as complex renewable energy and emission policy goals. This is because, accurate and consistent price forecasting reduces risk, maximises profit in the day-to-day market and improves bidding and production measures (Bunn, 2000). Besides, accurate EPF can improve wholesale electricity price bidding strategy and production which can increase the profits in day-ahead trading and energy management. Usually, power portfolio managers are interested with short-term and mid-term price forecasts. Short-term projections (intra-day) are crucial in day-to-day market operations, particularly when bidding on a power exchange or implementing effective demand adjustment. Meanwhile, mid-term forecasting is applied for planning objectives such as refining mid-term plans and resource allocation, risk management, and the valuation of exchange traded futures and bilateral contracts. This is due to high saturation of intermittent technologies and the evolving concerns related to resource adequacy in the longer-term. These forecasting operations will affect the baseload electricity price, such as the peak load price, the mean tariff for the 24 hours of the day or the baseload price (Maciejowska & Weron, 2015).

1.2 Problem Statement

Autocorrelation is important in forecasting to evaluate the correlation between the current value of a variable and any previous values that would have direct exposure in time series for performing regression analysis. One of the assumptions of regression analysis is that the data is devoid of autocorrelation. This can be troublesome since attempting to perform a regression analysis on data with autocorrelation will result in false results (Chenyu, Ruihua, & Yuandong, 2018). There are several sectors' where the autocorrelation of the residuals is analysed for time series data of S&P500, Brent, bitcoin price (Muglia, Santabarbara, & Grassi, 2019). In EPF, time series data analysis is very significant to get accurate regression outcome. To evaluate the performance of the forecasting in electricity data, test of autocorrelation is obvious. There should not exist autocorrelation of residuals in time series data for any improved prediction model. The Durbin-Watson test has been used widely to evaluate autocorrelation in residuals of regression analysis for time series EPF (Kabaila, Farchione, Alhelli, & Bragg, 2021). However, the Durbin-Watson test has the disadvantage of not being applicable to models that have autoregressive impacts. Further, it cannot not be used to assess higher-order serial correlation or other type of autocorrelations (Palomino, Reyes, Núñez, Valencia, & Herrera Acosta, 2020). Thus, in order to develop better prediction model for EPF, improved method of autocorrelation test of time series data needs to propose.

There are many statistical techniques have been proposed to forecast electricity price in literature. The existing statistical techniques (Hong, Taylar, & Fajardo, 2020; Rafal Weron, 2007) tried to reveal the specific pattern of historic power price utilizing curve fitting. Application of statistical models had shown to be challenging when predicting multi-dimensional nonlinear price of electricity since they are mainly based on linear equations. In the field of short and mid-term EPF, machine learning (ML) approaches like Support Vector Machine (SVM), hybrid support vector regression (SVR) and ANN, hybrid ANFIS and Backtracking Search algorithm have been applied to predict ranges of nonlinear quantities and perform feature selection (Lv, Liu, Yu, Zheng, & Lv, 2020; Pourdaryaei, Mokhlis, Illias, Kaboli, & Ahmad, 2019). Feature selection, over-fitting and gradient disappearance have been found the common challenges in these models. It can also be seen that ML techniques seemed to be less feasible for day ahead EPF due to limited compatibility with big data and perplexing nonlinear problems. Alternatively, deep learning algorithms such as bidirectional gated recurrent unit (BiGRU), deep belief network (DBN), LSTM, RNN, and convolutional neural network (CNN) have increasingly become popular in the field of EPF due to their ability to generate efficient classification approximations from a huge volume of input data and extract the data's underlying properties (Lv et al., 2020; Zhang, Li, & Ma, 2020). However, while processing EPF data with the mentioned deep learning models, there have found unusual price spike and gradient vanishing problem. In light of this, an optimized deep learning framework needs to be proposed for short and midterm EPF.

The accuracy of the EPF method needs to be validated to ensure its capability to produce highly accurate results. Therefore, there is a need for comprehensive statistical analysis for obtaining the accurate results in a volatile electricity market. There have been several techniques proposed to obtain good accuracy for EPF in terms of RMSE and MAPE. However, comprehensive statistical analysis has not been considered in those methods to obtain highly accurate forecasting outcomes, which means the overall time series data need to be analyze and prepare utilizing statistical tool to ensure accuracy of the prediction model. Hence, it is important to validate the performance of the proposed deep leaning models for accurate EPF by conducting comprehensive statistical and deep learning analyses to ensure the proposed forecasting models have high accuracy for practical application.

1.3 Research Objective

Therefore, the prime focus of the research is to utilize artificial intelligence (AI) based on deep learning method for prediction future days electricity price in the competitive electricity market. The contributions of this study are as follow:

1. To analyze the autocorrelation of the time series electricity data for the electricity price forecasting model.
2. To propose optimized RNN-based algorithm for short and mid-term Electricity Price Forecast.
3. To validate and compare the proposed deep learning electricity price forecasting models performance through past statistical and deep learning model.

1.4 Scope of Study

Electricity price prediction is split into three types: short term, medium term, and long-term projections. The pattern of electricity consumption, on the other hand, is dependent on the seasons and the peak, off-peak hours of the day. Though short-term forecasting is the most effective, mid-term and long-term forecasts have a considerable impact on competitive power market bidding strategies. Therefore, this study aimed at short and mid-term electricity price forecasting using machine learning and deep learning technique.

Time series analysis and data transformation were used to assess the accuracy and dependability of electricity price forecasts. As a case study, five of Australia's most economically significant zones were chosen. The Australian Energy Market Operator (AEMO) provides a variety of planning and forecasting trends to help decision-makers, such as the Electricity Statement of Opportunities (ESOO) and the integrated system plan (ISP). The fact that the Australian Energy Market Operator (AEMO) provides the most classified electricity data, including information on renewable and non-renewable energy sources, to determine electricity prices every five minutes, is one of the primary reasons to choose the Australian electricity market. Australia also has five distinct states, each with a wide range of meteorological conditions, economic conditions, and energy resources that include distinctive electrical data patterns. The National Electricity Market (NEM) is a single electricity market that allows generators, retailers, and interconnected areas like New South Wales (NSW), Queensland (QLD), South Australia (SA), Tasmania (TAS), and Victoria (VIC) to buy and sell electricity.

In this research, a different method applied for developing effective and trustworthy deep learning forecasting model, as well as a new strategy. The suggested approach ensures the forecasting accuracy of the deep learning model's projections regardless of the time-series data employed. This is accomplished through the use of a sequence of transformations that guarantee a time-series meets the stationarity criteria and is acceptable for building a deep learning method. Additionally experimental validation and resilience of the suggested method provided with theoretical advantages. Precisely, significant experiment has done with time series from Australia's five most vital economic zones. The models were tested on their capacity to anticipate time-series pricing (regression) and the accuracy of their predictions by looking for autocorrelation in the errors. This research indicates that the suggested technique significantly increased the accuracy and dependability of a deep learning model's forecasting performance. It can be claimed that the proposed method can be employed for any deep learning framework, further optimized and reconfigured deep learning methods can have performed more better.

1.5 Thesis outline

There are five chapters in this thesis, each chapter elaborated the main topic as follow.

Chapter 1 Following the context and rationale for the prospective study, the problem statement is presented. The study's objectives are provided first, followed by the research's scopes. Finally, the research strategy and structure for the research report are presented.

Chapter 2 focuses on a detailed information of the literature on time series data analysis and electricity price forecasting. The phenomenon of existing EPF technique is explored, as well as their technological problems. The latest developments and developments on short-term and mid-term electricity price projection for various regions, seasons, forecasting periods, and precision growth using various AI-based algorithms are reviewed at the end of this chapter.

Chapter 3 describes the concept of time series analysis and its transformation to make the time series effective for forecasting along with this machine learning and deep learning optimization approach utilized to increase the performance of forecasting. A supervised short-term deep learning structure of Recurrent Neural Network (RNN) based Long Short-Term Memory (LSTM) and optimized the prediction using Linear scaled hyperbolic Tangent (LiSHT) which helps to adjust biases of the neural network again where there are sudden high price rising or lowest falling of price. Besides for mid-term forecasting structure bagged tree ensemble (BTE) model is developed which is optimized by gated recurrent unit (GRU). In the time series data analysis Augmented Dickey Fuller (ADF) test performed to detect unit root and removed the unit root by transformation when there exists any unit root otherwise data trained with deep learning and check the Residual autocorrelation occurs. Therefore, if there is no autocorrelation the data is fine to train through deep learning prediction module.

Chapter 4 represents the result of the formulated of supervised short-term deep learning structure of Recurrent Neural Network (RNN) based Long Short-Term Memory (LSTM) and optimized the prediction using Linear scaled hyperbolic Tangent (LISHT) which helps to adjust biases of the neural network again where there are sudden high price rising or lowest falling of price. Besides for mid-term forecasting structure BTE model is developed which is optimized by GRU. Moreover, Australia's most economically significant regions take into account as every five minutes interval and compared with other deep learning approach for the proposed method. In this section regression analysis done for finding the residuals of the observed and predicted values and statistical test to show acceptability and robustness of the proposed LSTM model.

Chapter 5 elaborates the concluding statement and what are the improvements that can be done in future. A detail references listed at the end of the thesis.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

Electricity price forecasting (EPF) is one of the most important fields of energy forecasting, and it is primarily concerned with predicting spot and forward prices in large-scale electricity markets. Throughout the last 15 years, electricity price projections became an unavoidable concern for energy providers. The electrical industry has been reshaping its outmoded monopolistic structure, which has been influenced by the authority in the power industry, since the early 1990s. With deregulation and the emergence of competing electrical markets, this restructuring was completed. It should be emphasized that electricity is a unique commodity which cannot be economically stored, so the power system's reliability depends on a continuous equilibrium between production and customer demand. Furthermore, meteorological conditions (temperature, wind speed, precipitation), daily activities (on-peak vs off-peak hours, weekdays vs weekends, holidays), and market intensity all influence power consumption. Electricity prices are dynamic because of these involving variables, which is not seen in other industries or services. In general, time series forecasting is analysing time series data via modelling and statistics approach to aid strategic decision-making process. Time series forecasting incorporates information related to historical values and associated patterns to foresee future activity (Li, Wu, & Wang, 2020; Westland, Mou, & Yin, 2019). Nowadays, researchers are capable of employing and extracting complex information from time series data to solve various problems such as wind speed prediction (Westland et al., 2019), stock market prediction (Yin et al., 2020) and forecasting electricity prices (Amjady, 2001; Pappas et al., 2008). Time series models have been widely used to forecast electricity prices, although they can be challenging due to large variations (Ciarreta, Muniain, & Zarraga, 2017; G.-F. Fan, Wei, Li, & Hong, 2020; Heydari, Keynia, Garcia, & De Santoli, 2018; Hu et al., 2020; Karabiber & Xydis, 2019; Kraft, Keles, &

Fichtner, 2020; Pourdaryaei, Mokhlis, Illias, Kaboli, Ahmad, et al., 2019; Wu, Cattani, Song, & Zio, 2020; Yan, Song, & Chowdhury, 2016; Zhou et al., 2019). Existing statistical techniques tried to reveal the specific pattern of historic power price utilizing curve fitting. For instance, German electricity market has tested a k-factor Guégan Introduced Generalized Autoregressive Conditionally Heteroskedastic (GIGARCH) for forecasting electricity price (G.-F. Fan et al., 2020; Pourdaryaei, Mokhlis, Illias, Kaboli, Ahmad, et al., 2019). An iterative neural network methodology is also adopted along with this combinatorial neural network-based prediction technique to forecast upcoming electricity price. The advantages of this method include good precision, model functionality, and reliability. Meanwhile, Auto-regressive Integrated Moving Average (ARIMA) was proposed for electricity and power load forecasting (Hu et al., 2020; Karabiber & Xydis, 2019; Kraft et al., 2020; Wu et al., 2020). However, application of statistical models had shown to be challenging when predicting multi-dimensional nonlinear price of electricity since they are mainly based on linear equations. Moreover, statistical methods are inadequate for solving nonlinear multi-dimensional data for prediction purpose, as Its more suitable in handling linear data (Singh, Mohanty, & Shukla, 2017). Finally, from the literature review it is observed that, there are several limitations in statistical method and machine learning method which is described in the following sections.

2.2 Time-series Analysis in forecasting model

A time series is a set of observations made in a particular order over a period of time. Sales of a specific product over the course of a month, the temperature at noon in a specific region over the course of a day and electricity usage in a specific area over the course of a one-hour period all are examples of time series displayed in the Fig 2.1.

Time series forecasting has a variety of applications:

- Financial planning
- Commercial forecasting
- Stock market control
- Production and storage planning
- Alternative strategies for financial evaluation
- Budgeting
- Economic risk management

The majority of the applications on this list are self-explanatory. For instance, accurate estimates of future sales will make production planning much easier. A time series model will foresee a fresh future of observation that can be compared to what has already been observed(Chatfield, 2000).

Time-series are used in a wide range of real-world purposes, including banking (Liu, Zhang, & Ma, 2017) and commodities (Livieris, Pintelas, & Pintelas, 2020), as well as healthcare (Urtnasan, Park, & Lee, 2020) and pollution control(Gocheva-Ilieva, Voynikova, Stoimenova, Ivanov, & Iliev, 2019). Time-series data is made up of discrete data points collected at evenly spaced intervals in time. Time-series data are distinguished from other types of data by their primary qualities and characteristics. More specifically, they frequently contain a lot of noise, have a lot of volatility, and have a lot of extreme directional moves, with a chance of reversing these movements in the coming days. Time-series forecasting is widely regarded as one of the most difficult tasks in data mining due to these fundamental qualities. As a result, time series data analysis has been a hot topic in academia for years.

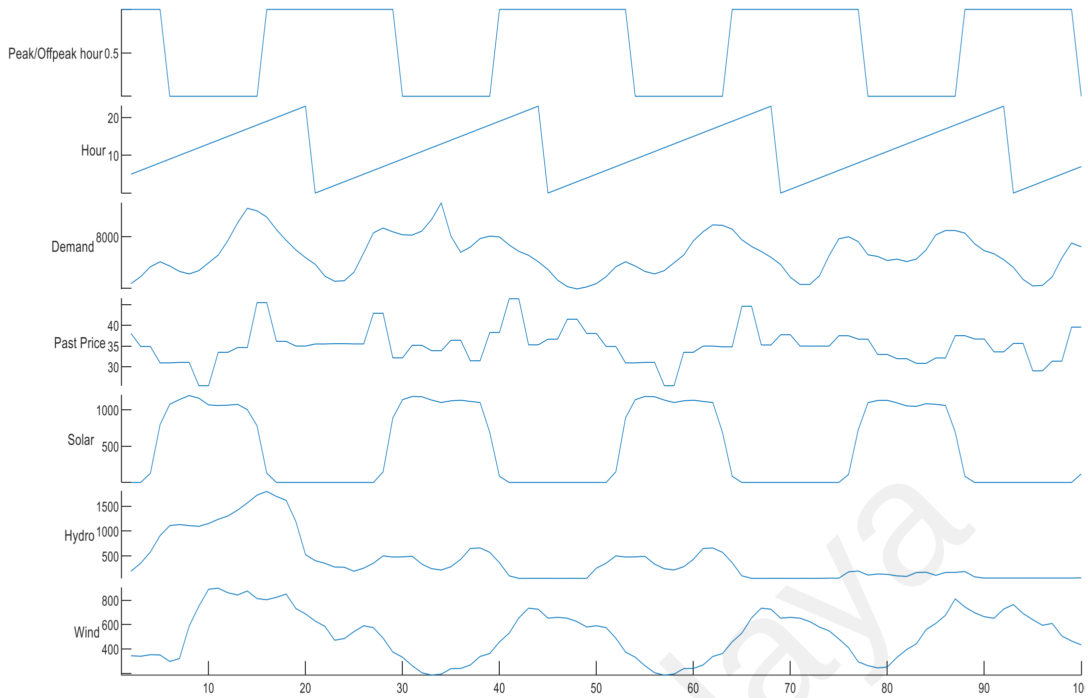


Figure 2.1: Different time series for forecasting

2.3 Features Impacting Electricity Price Forecasting

Price volatility is one of most typical price behaviors in deregulated power markets. The fundamental cause of these variations is a combination of economic and technical variables. In some work together on projects, scholars used either historical pricing data or both price and demand data to forecast the spot price, but they ignored other elements such as weather, generating reserve, and fuel cost.

2.3.1 Electricity Demand

The overall demand of the consumers influences the electricity price as an important aspect. According to analyses, a rise in power demand causes the spot price to rise. While demand is higher than normal, electricity supply will not enough. Because of this insufficient supply and high demand there need to balance the situation, which is usually done by increasing spot price of the electricity. The price of electricity depends on the demand of the customers which also have impact in fluctuations of various energy sources which may act as substitutes for electricity, consumer income changes and so on. Such as, high electricity price increases the industrial production cost and living cost. To overcome these instability of energy price an appropriate electricity tariff need to design (Andruszkiewicz, Lorenc, & Weychan, 2019). EPF can help to design and set the feasible energy price for the stakeholders.

2.3.2 Seasonality

The state of the climate has a direct impact on electricity usage, particularly the seasonal temperature, air flow and hydro flow. As a result, weather variations have a significant impact on spot prices. In the same season the temperature and wind flow may not vary much yet, in different season it will certainly change. For short term prediction it may have less effect but in terms of mid and long term the seasonality will play vital role in overall electricity price. At certain periods of the year, electricity is more expensive to deliver than at others. For example, in summer season sun light is available long period of time but in winter sunlight is very less, in different seasons air flow is not same then the production of electricity highly depending on fossil fuel which is expensive. Higher operating costs are spent during peak hours or peak seasons, and additional capacity must be available. It may be preferable to go from non-time differentiated to time-of-use pricing, especially if it is based on marginal costs. We provide welfare calculations based on two different measures: one computed for an uncompensated Marshallian demand curve and the other based on the theoretically more desirable compensating variation(Lillard & Acton, 1981). Uncompensated demand is function which maximize the product price. When it is considered products or services that account for a major portion of costs, Marshallian demand becomes understanding. The income effect is significant in this case. For minor purchases, people are often prepared to pay whatever feel is reasonable as long as there is benefit. The compensated demand curve indicates the amount of a commodity that a client will purchase if he were receiving income compensation for a change in the good's price. To put it another way, the compensated demand curve for a commodity is a graph that indicates how much would be purchased by the client at the altered price if the income effect were to be removed.

2.3.3 Peak and Off-peak Hours

There are some high consumer demands in a particular period of time in a day, mostly when there is less demand price is lower than regular price. So, these hours have significant impact on peak and off-peak hours. Off-peak hours electricity price, in contrast to peak hours lower (Huang, Hong, & Li, 2017). This is usually the case since there are minimal individuals attempting to access the grid around these hours, resulting in lower total demand and no need to pay extra for each KWh used.

2.3.4 Supply of Renewable Energy

Normally in a smart grid there should have one or more renewable energy source, the integration of these source can fulfil extra demand of power when necessary. The higher the supply then the lower power generation need from other expensive source. This renewable source has huge impact on pricing. Electricity production is becoming more climate sensitive as wind, solar, and hydro capacity grows. The supply curve shifts to the right when the wind blows or the sun shines, because these kinds of power generation have relatively low marginal costs. As a result, the price of energy may become more closely linked to weather conditions than the marginal costs of fossil-fuel facilities. Furthermore, when Combined Heat & Power installations (CHP) capacity grows, the potential influence of outdoor temperature on power supply grows as well, because these plants are primarily dispatched to produce heat. As a result, climate variables may begin to play a larger role in the electricity market. Increased temperatures, for example, can cause more cooling difficulties for thermal power plants, especially if they are unable to use river water for cooling. As the temperature rises, so does the demand for cooling in the home(Mulder & Scholtens, 2013).

2.3.5 Fossil Fuel Cost

Another significant influence on the power spot price is the cost of fuel, which accounts for the majority of generation costs. Since fossil fuels (oil, natural gas and coal) account for a large portion of the energy produced, a recent body of research has looked into the relationship between crude oil, natural gas, coal, and electricity costs. The significance of alternative energy prices like oil, natural gas, and coal on electricity pricing is investigated in this study.

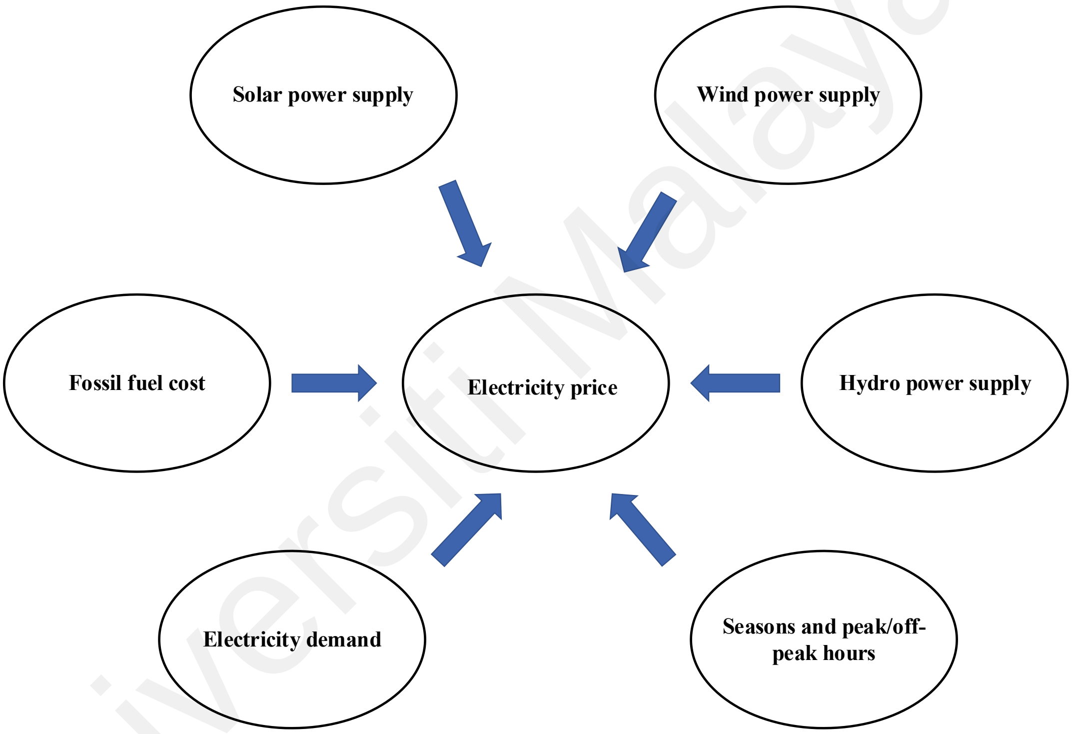


Figure 2.2: Factors that impacting Electricity Price

2.4 Different Models of Electricity Price Forecasting

Since the advent of privatized energy, the key emphasis has been on maximizing the profits of multiple market players. When it comes to predicting, the electricity price and demand are inherently associated and have a collaborative respect because they rely on each other and a mistake in one would result in inefficiency or a problem in the other. The key concerns that determine the price are non-storability, seasonal behavior and electricity transportability.

Electricity price predictions is generally divided into three classes shown in the Figure 2.3: short-term, medium-term, and long-term(Rafał Weron, 2014). Yet, there is no clear borderline insane in the research to distinguish them.

- **Short-term:** It is one of the most crucial segments for day-to-day trading operations, with predicting periods ranging from a few minutes to a few days.
- **Medium-term:** Medium-term forecasting is important for balancing sheet adjustments, risk assessment, and futures pricing. It can range from just few days to a few weeks. When it comes to energy price predicting, the majority of the time, the assessment is done on the basis of prices over a specific future time period rather than exact point projections.
- **Long-term:** This forecasting is primarily focused on investment efficiency analysis and preparation, and it predicts for months, quarters, or even years in advance. The insight from this forecasting is useful in selecting future locations for power station sources of fuel.

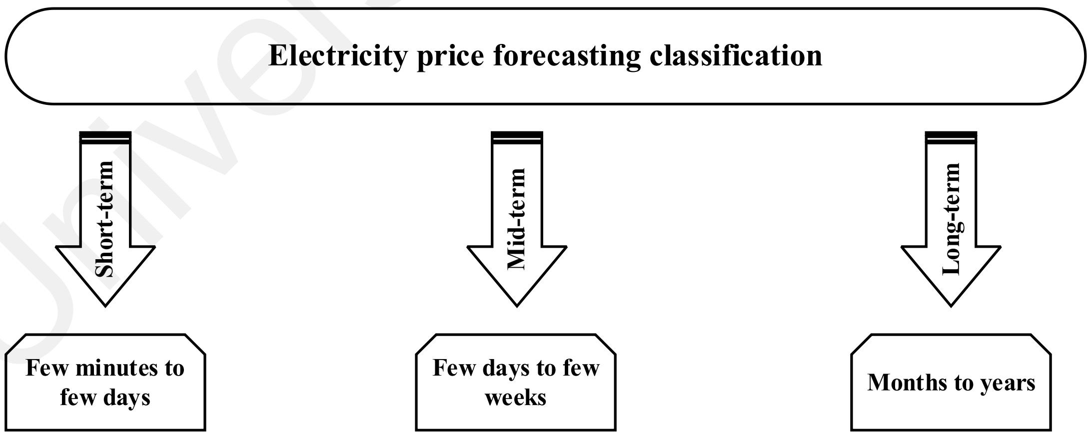


Figure 2.3: Classification of electricity price forecasting

Several strategies and procedures for EPF have been designed or launched during the last decade in the field of machine learning and deep learning method, with varying degrees of success. Statistical method of forecasting had to utilize to predict before the development of AI. EPF can be divided into six categories. The graphical classifications of the models are given below at Figure 2.4 (Pourdaryaei, Mokhlis, Illias, Kaboli, & Ahmad, 2019).

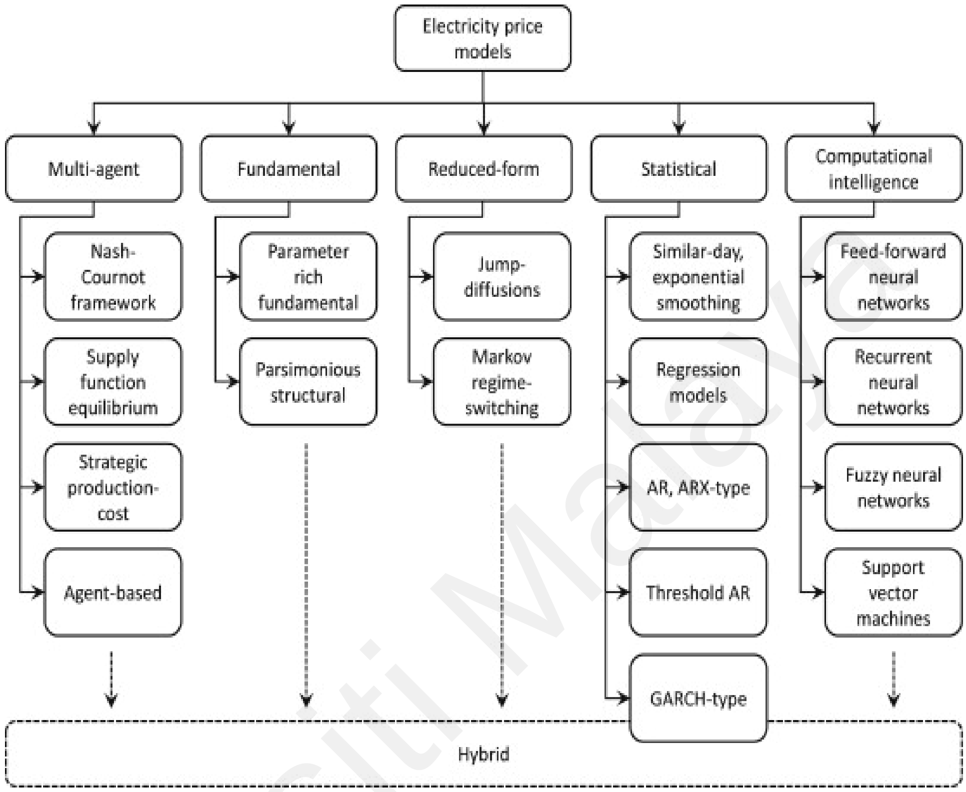


Figure 2.4: General approaches of EPF

2.4.1 Multi-agent Models

Multi agent approaches, like multi - agent systems simulation, cognitive science, and stability, can evaluate the system operation of heterogeneous agents, including generating units and corporations, and obtain price processes by ensure that there are adequate and demand in the market. Cost - oriented modeling, supply function equilibrium (SFE), homeostasis or tradeoff models like the Nash-Cournot architecture, strategic production cost models (SPCM), and agent-based models are the most common methods under this category. Multi agent approaches are mostly concerned with qualitative issues rather than quantitative outcomes(Zimmermann, Neuneier, & Grothmann, 2001). They are able to provide insight into whether a pricing will be beyond marginal cost or not, as well as how this may affect market outcomes. This model has certain downsides; for example, when more quantitative findings are required, this model is not the best option, particularly when prices must be forecast with a high level of precision.

2.4.2 Fundamental models

The underlying physical and economic relationships that are accessible in the generation and trade of electricity are captured in the design of primary models. It is hypothesized that basic correlations exist between basic drives such as loads, system characteristics, meteorological circumstances, and others. Furthermore, essential inputs are typically modelled and predicted individually using statistics, which is a simplified kind of AI. The access to data and the inclusion of stochastic variations in the underlying drivers are two major hurdles in the practical application of these models(Eydeland & Wolyniec, 2002). During the model's development, specific assumptions are anticipated to be made. As a result, the model's predicted price is extremely vulnerable to changes in the hypotheses.

2.4.3 Reduced form model

The major aim of this process is to reproduce the basic characteristics of daily power prices instead of delivering high-precision hourly price projections. It is relevant to residuals at future time periods, price dynamics, and commodity price correlations(Rafal Weron, 2007). The retrieved findings from the model are less reliable if the chosen price does not fit the main features of power prices. However, if the model is extremely complex, the computing overhead will prevent it from being implemented in the trading section online. There are two groups in this model, and both have some disadvantages, The biggest drawback is determining the risk premium that connects spot and future prices. Another drawback is that they miss data, which is required for validation, and they are reluctant extract the spot price attributes from forward curve assessment.

2.4.4 Statistical models

In statistical techniques, such as multiple regression and fundamental analysis, a numerical mixture of prior prices and/or previous or present values of exogenous elements is utilized to predict the existing electricity price(Rafal Weron, 2007). Two popular subclasses of this concept are linear and non-linear models. The projected price in a linear model is the sum of several components, but in a non-linear model, the projected price is the product of the number of elements. The projected price in an additive model is the sum of a number of components, whereas the estimated price in a non-linear model is the result of the several factors. The linear model is significantly more prevalent, although both of the models are closely connected because a non-linear model for log-price can be converted to a linear model. The appeal of statistical models derives from the fact that some practical interpretations of their elements exist, enabling system administrators and developers to comprehend their role. However, when it comes to modelling the nonlinear behavior of electricity pricing and other fundamental variables, they have some drawbacks. Their efficacy, on the other hand, is comparable to that of nonlinear computational intelligence techniques. The best winning contestants in the load forecasting way of the rivalry in nation's energy projection for example, used regression type models across several of rivals.

Statistical frameworks are a broad category that includes:

- Similar -day and exponential smoothing approaches
- Regression Models
- Time series models without (ARIMA, AR, Seasonal ARIMA-SARIMA) and with exogenous variables (ARX, ARMAX, ARIMAX, AR-GARCH)

2.4.4.1 Autoregressive Integrated Moving Average

Among various statistical forecasting methods, the autoregressive integrated moving average (ARIMA) is a very reliable nonstructural approach for time series prediction modelling. Since there are no intervals in the dataset, an ARIMA design acts as a 'filter,' attempting to separate the time series from the noise, following that, the time series is extended in order to either anticipate future points in the series or acquire a wider range of data, which is particularly useful in time series analysis. The letters (p,d,q) represent quasi ARIMA method, where q, d, and p are all positive integers. The parameter p specifies the number of time delays for the autoregressive model, d the number of differences needed to make the series stationary, and q the moving average order (Yuan, Liu, & Fang, 2016).

The ARIMA framework with order p has an autoregressive component of the type below Eq. (2.1):

$$y_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \ldots\ldots + \emptyset_p y_{t-p} + e_t + c \tag{2.1}$$

The random variables $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ are time delayed readings of the prediction variable, and the modeling parameter are $\emptyset_1, \emptyset_2 \ldots . \emptyset_p$. This model is called autoregression since the forecasts are entirely based on observed data from past time periods.

The previous errors serve as the explanatory variable in the ARIMA model's moving average (MA) component. A moving average framework in Eq. (2.2) having order of q takes the following form:

$$y_t = \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots\ldots + \theta_q e_{t-q} + e_t + c \tag{2.2}$$

Where, $\theta_1, \theta_2, \ldots\ldots, \theta_q$ are the model's elements and $e_t, e_{t-1}, e_{t-2}, \ldots\ldots, e_{t-q}$ are wide range frequency error terms.

For stationarity, an autoregressive approach is used to create an ARMA (p,q) (autoregressive moving average) framework. Differentiation is introduced to the ARMA model for non-stationary variables. Differencing is a technique for stabilising a series' mean, removing seasonality, and making the series stationary. To mathematically difference the data, the initial difference between consecutive observations is obtained using Eq. (2.3).

$$\triangle y_t = y_t - y_{t-1} \tag{2.3}$$

Finally, Eq. (2.4) expresses the general form of the ARIMA model, It necessitates at least p+d prior specimens in order to initialize the time series.

$$\begin{gathered}\triangle_\alpha y_t = c + \phi_1 \triangle_\alpha y_{t-1} + \phi_2 \triangle_2 y_{t-2} + \ldots\ldots\ldots + \phi_p \triangle_p y_{t-p} + e_t \\ - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots\ldots - \theta_q e_{t-q}\end{gathered} \tag{2.4}$$

2.4.4.2 Generalized autoregressive conditional heteroscedasticity models

Another model is the generalized autoregressive conditional heteroscedasticity (GARCH), which is designed to capture price volatility instead of the changing price as in the ARMA Model. The instant of a time series as a variation is taken into account in this model. In other words, unlike an ARIMA process, the error term, which is actual value minus prediction value, does not have a zero mean and constant variance. The error term can then be assumed to be connected and represented using an AR process. As a result, a GARCH can be used to measure the indirect volatility of a time series as a result of price spikes. Considering a time series with the variable x_t and a constant mean offset, a GARCH model can be established, then:

$$x_t = \mu + \varepsilon_t \tag{2.5}$$

Where μ denotes offset and $\epsilon_t = \sigma_t Z_t$

$$\varepsilon_t^2 = c + \sum\nolimits_{i=1}^{q} \varphi_i {\varepsilon_{t-i}}^2 + \sum\nolimits_{i=1}^{p} \psi_i \varepsilon_{t-i}^2 \tag{2.6}$$

Where p presenting the sequence of GARCCH factor σ^2 and q presenting the sequence of ARCH factor ε^2. It is obvious that Eq. (2.6) with p=0, the GARCH (0, q) transform to an ARCH model. This should be highlighted that GARCH model applies exclusively to the stationary time series. As a result, before using this model the following equation must be provided for stationary time series:

$$\sum_{i=1}^{q} \varphi_i + \sum_{i=1}^{p} \psi_i < 1 \tag{2.7}$$

2.4.5 Computational intelligence model

Artificial machine learning, intelligence-based, non-parametric, and non-linear analysis tools are examples of computational techniques that combine aspects of learning, evolution, and fuzziness to provide solutions that can adapt to complex systems. In this respect, they could be considered "intelligent." Artificial neural networks are one of the subclasses of computational intelligence (Keles, Scelle, Paraschiv, & Fichtner, 2016), support vector machine (SVM)(Yan & Chowdhury, 2013) and fuzzy systems(Rodriguez & Anders, 2004) which are widely used in EPF. These models can solve and handle exceedingly complicated and non-linear problems. In general, computational intelligence outperforms statistical methodologies when it comes to simulating the characteristics of electricity costs. Flexibility, on the other hand isn't usually a sign of strength. In fact, flexibility to nonlinear and spiky behaviors isn't usually a strength for these models and better results or probabilistic projections aren't always the result. Figure 2.5 shows many forms of artificial neural networks that can be used for various forecasting applications.

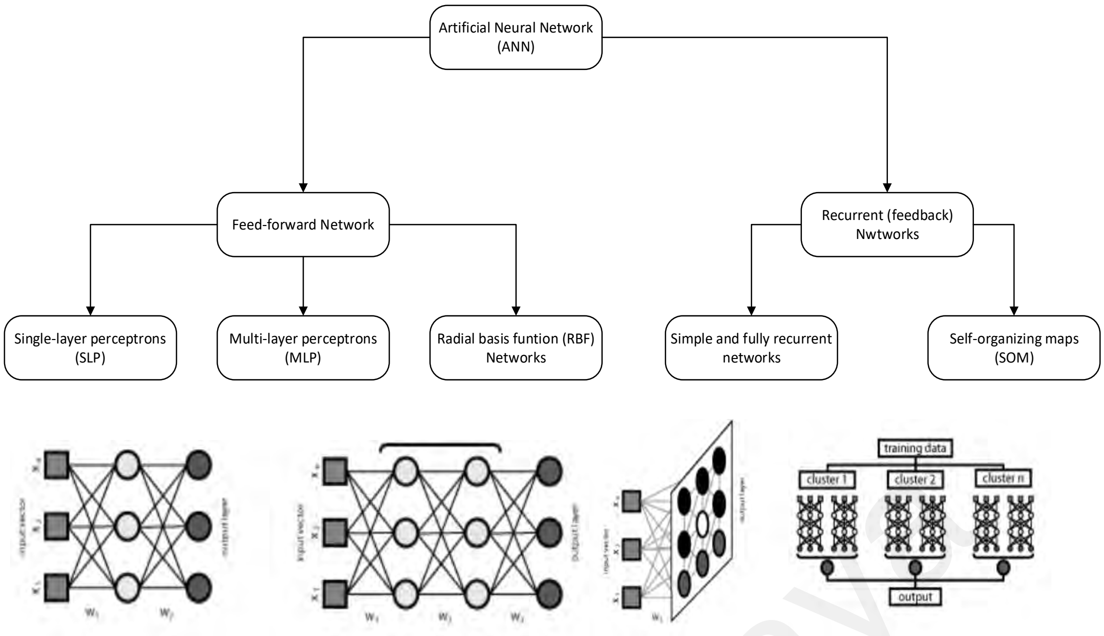


Figure 2.5: Computational approaches architecture

2.4.5.1 Artificial neural network

Artificial neural networks (ANNs) are artificial networks that are inspired by the brain and are thought to be human initiatives to understand what happens in a biological neural network. ANNs replicate the nervous system's learning process in the hopes of capturing the power of biological networks. Collectively, they do jobs that even supercomputers with high-level computational capacity have not been able to process (Abiodun et al., 2018).

The ANN is similar to a biological neural network, which is made up of a densely connected network of relatively simple processors called neurons. Weighted interconnections connect neurons, allowing them to communicate by transferring signals from one neuron to another, with the intensity of the weighted connections indicating the value of each neuron's input. A transfer function is assigned to each neuron and specifies how the weighted sum of a neuron's input signal is translated to an output signal.

The major distinguishing feature of ANNs is that they accumulate experience during the learning process and adapt to new situations using the knowledge learned during the learning process. Because the learning process in ANNs is facilitated by the repeated adjustment of numerical weights, weighted connections are regarded as the most fundamental form of long-term memory in these networks.

Based on the training scenario neuron organization and neuron connections, several topologies of neural networks (NNs) have formed. Multilayer perceptron (MLP) and radial basis function (RBF) networks have proven to be the greatest useful among the numerous forms of NNs in diverse applications. The activation functions of the hidden layer are the fundamental difference between these two forms of NN. The linear, logistic sigmoid, and bipolar sigmoid (hyperbolic tangent) activation functions are employed in RBF networks, whereas the linear, logistic sigmoid, and bipolar sigmoid (hyperbolic tangent) activation functions are used in MLP networks (Kankal, Akpınar, Kömürcü, & Özşahin, 2011).

In general, there is a trade-off between the RBF network's stronger resilience and MLP's larger accuracy improvements. The RBF network is substantially more resilient to adversarial noise because to its non-linear character. MLP, on the other hand, is an acronym for deep learning in NNs, which uses several hidden layers to improve accuracy.

MLP is a universal estimator because, being a feedforward NN it has the generalizing capacity to approximate nearly any function with a high degree of accuracy. The MLP architecture is made up of three layers, as seen in Figure 2.6:

i. Input layer, where the data are introduced into the NN
ii. Hidden layer(s) where the data are processed
iii. Output layer where the results of given inputs are obtained.

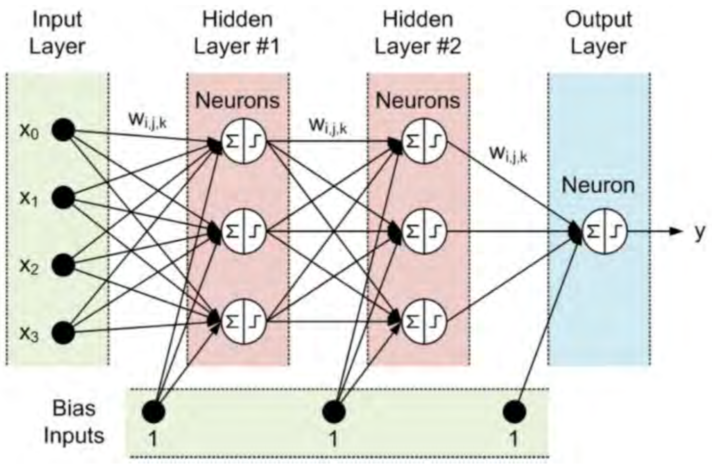


Figure 2.6: MLP Architecture

Each of these levels contains a number of processing units, each of which is fully interconnected using weighted connections to units in the next tier. There are a number of nodes in each stratum. Every input is multiplied by the connectivity weight of each node. The sum of the product is sent through an activation function to obtain the output of each neuron, while the bias input is coupled to each neuron to stabilize the origin of the activation function for improved learning.

MLP networks are typically used to execute supervised learning tasks that require an iterative training procedure to alter the network's connection weights. To achieve a particular level of estimating accuracy, multiple passes are usually required. The standard error back propagation procedure is used to adjust the correction weight, which minimizes the overall error with the gradient descent approach. (Raza & Khosravi, 2015).

2.4.5.2 Support vector regression

Techniques based on kernels for example, Kernel principal component analysis (PCA). Machine learning algorithms such as Kernel Fisher discriminant analysis (KFD), Bayes point machines, Gaussian processes and SVMs (support vector machines) have advanced significantly. Kernel-based strategies map data across higher-dimensional feature spaces in the expectation that the data will be segregated or have better organization in the greater spaces. SVMs are the best-known member of Kernel-based approaches, which can either categorize the input data or capture complicated relationships in the input data as an extension to the nonlinear model of the generalized portrait algorithm. SVR refers to SVMs that deal with function approximation and forecasting, and support vector classification refers to SVMs that deal with classification problems (SVC). SVC is transformed to SVR with only just few minor changes. As a powerful function estimation technique centered on statistical learning theory, SVM can be advanced to form SVR (Bian, Han, Du, Jaubert, & Li, 2016).

2.4.5.3 Deep learning module

RNN has recently shown considerable promise in a wide range of human pursuits, including speech recognition, language modelling, and translation (Siegelmann, 1995). RNNs are good at processing sequence tasks because they repeat the same process for each member of the input sequence at each time step, keeping a state in their hidden units that implicitly contains information about the sequence's history (LeCun, Bengio, & Hinton, 2015).

For short sequences, the basic RNN technique performs well. When the series is long, however, important concerns such as the disappearing gradient, inflating gradient, and long-term dependency problem occur(Bengio, Simard, & Frasconi, 1994). The LSTM structure was presented as an upgraded version of RNN to tackle these concerns. In many applications, RNN paired with LSTM structure captures long-term dependencies in a more generic and effective way than basic RNN, resulting in superior overall performance (Greff, Srivastava, Koutník, Steunebrink, & Schmidhuber, 2016). The updating of the current state in basic RNN and LSTM is based on the prior state. For some applications, the output at the present time-step may be influenced by both the past and subsequent states. As a result, bidirectional LSTM (BiLSTM) was presented as a solution to the above-mentioned problems. Each input sequence is presented forwards and backwards to two distinct recurrent networks in a bidirectional LSTM, and the output is generated depending on the hidden state of both RNNs.

2.4.6 Hybrid models

The majority of models and power price forecasting strategies in the literature are considered hybrid strategies, which include two or more methods or techniques from the aforementioned list. For example, Neural Networks and Box Jenkins models can be merged to form AleaModel.

2.5 Progress in using AI to forecast electricity price

This section summarizes current developments in artificial intelligence-based power price predictions. In addition, a summary of all methodologies based on country, year, and accuracy is gathered to ensure the consistency and existence of this research from the beginning to the present.

Fundamental models, statistical models, and artificial intelligence-based approaches are the three types of forecasting methodologies that can be categorized according to the forecasting framework. Among the various techniques, AI-based electricity price forecasting approaches have gained a lot of traction in recent years because they provide a significant advantage of ensuring a certain level of prediction accuracy compared to the high variability of independent and dependent variables in the statistical model (Hernandez et al., 2014). It is possible to find concise reviews of recent methodologies and techniques for price forecasting based on AI technology. An in-depth examination of the majority of published methodologies, including stochastic models, artificial intelligence models, and regression models (Panapakidis & Dagoumas, 2016).

ARIMA and GARCH are two common time series modelling techniques. They are extremely capable and can serve as a model for others. Additionally, they can be combined with other models to form hybrid models. The use of extensive mathematical formulas is the major reason why these methods have grown so popular. In time series models, past values are employed with the premise that the quantity progression follows a specified path. Furthermore, as a prediction step, the patterns extension is employed in this model to predefine a future time period. In AR, the time series models ARMA and ARIMA are compared. AR (1) with jumps, AR (1) with log jumps, and many other sub models have been proposed. In(Diongue, Guegan, & Vignal, 2009) A comparison of simulations between a SARIMA-GARCH and a k-factor GIGARCH process was conducted. The test research used data from the EEX market for a month and only lagged price values were used in the models.

Some of the most common ANN-based models in the recent literature are included but not limited to specifically, Multi-Layered Perceptron's (MLPs), Feed Forward Neural Networks (FFNNs), Radia Basis Function Networks (RBFNs), Support Vector Machines (SVMs), Fuzzy Neural Networks (FNNs), Probabilistic Neural Networks (PNNs), Recurrent Neural Networks (RNNs) and Self-Organizing Maps (SOMs). MLPs training speed, simplicity and effectiveness have made them to be the most common network(Abedinia, Amjady, Shafie-Khah, & Catalão, 2015).

Shallow learning models have shown better performance compared to statistical models in terms of error minimization and some other factors. In the field of load forecasting, Support Vector Machine (SVM) (Shuai, Song, & Wang, 2017; Zhao, Dong, Li, & Wong, 2007) has been applied to predict ranges of nonlinear quantities and perform feature selection. Support vector regression (SVR) (Dhillon, Rahman, Ahmad, & Hossain, 2016), artificial neural network (ANN) (Elfahham, 2019; Hamilton, Abeygunawardana, Jovanović, & Ledwich, 2018), and regression tree are the main shallow machine learning models that were commonly used in forecasting system. Besides, the work in (M. Du, Zhao, Liu, & Zhu, 2021) proposed a hybrid of SVR and gray wolf optimization to forecast life cost of power transformer. A hybrid model based on SVR and ANN is proposed in (Lv et al., 2020) by adopting new signal decomposition and correlation analysis technique to predict electricity price for next 24-hours. Furthermore, in (Pourdaryaei, Mokhlis, Illias, Kaboli, & Ahmad, 2019) a hybrid approach of ANFIS and Backtracking Search algorithm (BSA) was proposed for electricity price forecasting and feature selection. Besides, a multi-objective binary-valued backtracking search algorithm (MOBBSA) and ANFIS approach has been employed which is the exemplary method of shallow learning. Nevertheless, over-fitting and gradient disappearance have been the common challenges in shallow machine learning models. It can be seen that previous techniques seemed to be less feasible for day-ahead EPF due to limited compatibility with big data and perplexing nonlinear problems (Alazab et al., 2020).

Recently, deep learning has been widely applied in the fields of artificial intelligence and big data due to its ability to learn effective feature representations from a large amount of input data and excavate the profound features of the data. The model in (Lv et al., 2020) is mainly based on distributed representation, bidirectional gated recurrent unit (Bi-GRU), and attention mechanism where the Bi-GRU layer processes the past and the future information simultaneously to fully extract temporal and nonlinear features from input data for the improvement of forecasting accuracy. Meanwhile, the work in (Zhang et al., 2020) combined the deep belief network (DBN), LSTM RNN, and convolutional neural network (CNN) to extract complicated nonlinear features. Meanwhile, the work in (Haq & Ni, 2019) applied signal decomposition and correlation analysis technique in the DBN model. In (Bedi & Toshniwal, 2019), to forecast electricity load the authors developed multi-input and multi-output LSTM models whereas the time is confined. Deep learning algorithm with a recurrent feedback network framework is called Recurrent neural network RNN has the ability to perform more overarching and entire modelling of time series compared to other conventional AI algorithms, considering the terrestrial correlation of time series. Through the training process of RNN, the gradient explosion and gradient disappearance problems can be solved with LSTM (J. Du, Vong, & Chen, 2020; M. Wang, Wang, Lu, Lin, & Wang, 2019). To predict the day-ahead electricity price, LSTM has been adopted as electricity price prediction, for the Australian market in the Victoria region and Singapore market (Manner, Fard, Pourkhanali, & Tafakori, 2019). Moreover, single gated recurrent units (GRU) network structure has been investigated for prediction purpose. The simple neuron structure of GRU had shown to contribute to lesser processing time compared to LSTM network (Ugurlu, Oksuz, & Tas, 2018). LSTM had shown to exhibit better performance compared to SVM, ANN and RNN in terms of forecasting accuracy (C. Fan, Sun, Zhao, Song, & Wang, 2019; Ugurlu, Tas, Kaya, & Oksuz, 2018).

Alternatively, deep learning algorithms have increasingly become popular in the disciplines of artificial intelligence and big data due to its ability to generate efficient classification approximations from a huge volume of input data and extract the data's underlying properties. The model in (Lv et al., 2020) focused on distributed depiction, bidirectional gated recurrent unit (BiGRU) and learning algorithm with the BiGRU layer processing past and prospect information concurrently to fully extract chronological and nonstationary features from input data with the goal of improving forecasting performance. Meanwhile, to extract difficult nonlinear characteristics, (Zhang et al., 2020) incorporated the deep belief network (DBN), LSTM RNN, and convolutional neural network (CNN). The DBN model was used in (Haq & Ni, 2019) to use signal processing and correlation analysis techniques. In addition, (Bedi & Toshniwal, 2019) created a multi-input and multi-output LSTM model for forecasting electricity demand. When evaluating the aerial correlation of dataset, it seems to be that a deep learning algorithm with a recurrent feedback framework called Recurrent neural network RNN has the capacity to accomplish more overarching and entire designing of time series than other traditional AI algorithms. The gradient inflation and gradient vanishing issues could be handled using LSTM through the RNN training procedure (J. Du et al., 2020; M. Wang et al., 2019). As a result, LSTM has been used to anticipate day-ahead power prices for the Victoria region of Australia and the Singapore market (Manner et al., 2019). Furthermore, the network topology of single gated recurrent units (GRU) has been explored for prediction purposes. When compared to an LSTM network, the GRU's simple neuron topology has been proven to lead to a faster processing time (Ugurlu, Oksuz, et al., 2018). In a nutshell, LSTM has been demonstrated to perform better in terms of forecasting accuracy than SVM, ANN, and RNN (C. Fan et al., 2019; Ugurlu, Tas, et al., 2018). As a result, the analysis of time-series data for deep learning model in EPF has been an active subject of research for decades.

2.6 Summary

This chapter presents an analysis into price forecasting approaches that have been used in prior work in the unregulated setting. Because of the fast changes in the composition of power markets, market participants must estimate prospective prices with a high level of precision in order to increase profit.

Considering the most relevant input variables in electricity price is a difficult process due to the dependence of energy demand and price on many aspects. A robust model with high accuracy has been presented to solve the issues inherent in power price forecasting. Statistical methods, artificial intelligence-based methods, and hybrid methods for electricity price prediction are discussed in length in this chapter which is showed below in Table 2.1.

Table 2.1: Review on most recent machine learning methods

Compared Model	Electricity market	RMSE	MAPE (%)	Limitations/Challenges
SVM(Yan et al., 2016) LSSVM(Yan et al., 2016)	Mid-term PJM electricity market	No RMSE	11.7491 10.9722	Accuracy in spike price forecasting considerably low by using the proposed machine learning methods. Optimization of forecasting accuracy in the **spike price area is the main challenge of the study**.
Pc4 (Maciejowska & Weron, 2015) Auto Regressive (AR) (Maciejowska & Weron, 2015)	Short-and mid-term electricity market APX. UK	0.970 4.270	6.24 6.24	In Summer, the electricity price does not react to the significant decrease in demand. It is challenging to relate the forecasting **performance** of **demand** combined with **natural gas** when applying **statistical** approach.
ARIMA(Hu et al., 2020) DBN(Hu et al., 2020)	Mid-Long Term Electricity Consumption Wuhan, China	0.068 0.008	5.140 3.278	Data **analysis is limited** since short-term prediction is challenging.
ANN PSO (Hybrid)(Heydari et al., 2018)	Mid-term load power North American electricity market	No RMSE	1.9	ANN PSO method is not feasible to handle **large data set of nonlinear data**.
CNN (Zhang et al., 2020)	Day-ahead PJM electricity market	1.76	0.082	Limited discussion on **time series data analysis** and statistical reliability.
EEMD-LSTM_SMBO (Zhou et al., 2019)	Day-ahead PJM electricity market	0.89	2.47	**Uncertain accuracy** due to limited variables considered in the prediction model.

Although various ways to forecasting energy market price have been used in the literature, they all have certain limitation. For example, one of the primary issues of data-driven techniques is that there are so many control parameters, many of which are quite sensitive, making initialization of their values extremely difficult. Although several machine learning approaches mentioned above have been used to anticipate electricity price forecasting, new methodology is still necessary to produce more accurate forecasts. Furthermore, long-term forecasting has been explored in the majority of the aforementioned studies. However, because the seasonal pattern of energy consumption is seasonal, short and midterm forecasting will be more beneficial in the competitive electricity market for legitimate strategic planning. Time series analysis is critical for making this forecasting dependable and precise, and the deep learning method has considerably improved forecasting accuracy.

Autocorrelation is important in forecasting to evaluate the correlation between the current value of a variable and any previous values that it would have direct exposure in time series for performing regression analysis. One of the assumptions of regression analysis is that the data is devoid of autocorrelation. This can be troublesome since attempting to perform a regression analysis on data with autocorrelation will result in false results. So, this is one of the gaps discovered from the above literature review.

Secondly, feature selection, over-fitting and gradient disappearance have been found the common challenges in these models. It can also be seen that ML techniques seemed to be less feasible for day ahead EPF due to limited compatibility with big data and perplexing nonlinear problems. Alternatively, deep learning algorithms such as bidirectional gated recurrent unit (BiGRU), deep belief network (DBN), LSTM, RNN, and convolutional neural network (CNN) have increasingly become popular in the field of EPF due to their ability to generate efficient classification approximations from a huge volume of input data and extract the data's underlying properties (Lv et al., 2020; Zhang et al., 2020). However, while processing EPF data with the mentioned deep learning models, there have found unusual price spike and gradient vanishing problem found from the above literature review.

Finally, comprehensive statistical analysis has not been considered in those methods to obtain highly accurate forecasting outcomes. Hence, it is important to validate the performance and set benchmark of the proposed deep leaning models for accurate EPF which is another scope of addressing the limitation found in the above literature.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

The technique of the forecasting algorithms used for short and midterm electricity price forecasting in Australia's five most important economic zones is presented in this chapter. The time series analysis performed with respect to ADF test to check stationarity of the time series and optimization of the time series using LSTM, which is recurrent in nature, and LISHT takes up a large portion of this chapter. The autocorrelation of the residual of the training dataset is noted in time series analysis to make the timeseries suited for deep learning simulation. Australia’s five most important economic states electricity data considered for the proposed methodologies. One of the main reasons to choose Australian electricity market as Australian Energy Market Operator (AEMO) Provide most classified electricity data which includes renewable and non-renewable energy source data to determine electricity price in every 5 minutes. Besides, Australia has five different states which are vast and varieties of weather condition, economic status, energy resources consist of various pattern of the electricity data. Such as New South Wales: temperatures can get quite high in the northwest and very low throughout the southern mountainous region but overall, the state has a pretty humid climate. The eastern portion of NSW is mostly warm and has significant rainfall, ranging from humid subtropical on the northern shore. However, more than half of the state is desert or semi-arid. So, Electricity consumption will be high at daytime and bit lower in night in this region. Queensland has pleasant winters season and hot summers. In summer and winter average temperature changes from 9°C to 29°C, while those in Cairns typically range from 17°C to 31°C. From December to February is summer season in South Australia and the usual temperature there is 28°C, though it frequently exceeds 30°C. Since a large portion of the northern region of the state is made up of desert, there are often quite warm conditions there. These places frequently become cold during night. So, during daytime the demand is very higher of this state. Tasmania has a moderate marine atmosphere, which means weather don't change all that much over the year yet, it can get both warm and cold. Tasmania experiences its warmest weather from 17° to 23°C in December, January, February, and March. Winter lasts from May through August, with average lows of 3°-11°C. Victoria experiences Varity of seasons than other parts of Australia, here includes hot summers, pleasant springs and autumns and brisk winters. From June to August there's winter, summer continues from December to February, and spring season is from September to November. The coolest and wettest months are June July and October, respectively.

3.2 Processing the features of time series data

Based on the different weather conditions and demand of the electricity consumers of various states of Australia time series electricity price data collected for the proposed methodologies.

3.2.1 Autocorrelation of the model's forecasting reliability

Conventional time series data may contain missing values, outliers and high dimensional data. These factors contribute to unstable forecasting performance. Therefore, pre-processing is required to solve the abovementioned problems. This work emphasized on linear trend-based equation for features processing. The linear trend approach is able to perform effectively with trend and depict it without any assumptions. Besides, the residual seasonality, peak-off peak hour and renewable energy trend can distinguish any time series dataset.

Let $h_1, h_2, \dots\dots, h_n$ be the time-series data. The following is the definition of a nonlinear regression model of order m is denoted by

$$h_t = f(g_t, \theta) + \epsilon_t \tag{3.1}$$

where $g_t = (h_{t-1}, h_{t-2} \ldots \ldots h_{t-m}) \in \mathbb{R}^m$ made of m values of h_t, θ is the parametric vector and ϵ_t is the residual. After the model has been built, machine learning or deep learning approaches can be used to find the function $f(*)$. Root mean square error (RMSE) and mean absolute error (MAE) are the most often used indicators for regression performance evaluation of a forecasting model. Nevertheless, both regression performance evaluators only indicate the accuracy of the observed and estimated values. Since they are unable to analyse the fitness of time series data in the proposed forecasting model, the residuals are employed to assess this dedicatedly. In other words, the forecasting model's residuals of regression analysis for normal distribution and autocorrelation are estimated by function $\widehat{\epsilon_t}$, where $\hat{h}_t$ is the predicted value.

$$\widehat{\epsilon_t} = h_t - \hat{h}_t \tag{3.2}$$

The presumption of no autocorrelation in the residuals might make the forecasting vulnerable as there may not be exploration on all available data in the training process. In other words, the reliance of the residuals indicates that the model did not well fit the time-series data and that there is important data remaining that must be investigated.

The auto correlation function (ACF) plot and the Ljung–Box Q test for residual autocorrelation are two important techniques for determining the presence of autocorrelation in the residuals (Brockwell, Brockwell, Davis, & Davis, 2016). More analytically, by calculating the linear correlation of every residual in various lags, $\hat{\epsilon}_{t-1}$, $\hat{\epsilon}_{t-2}$, . . . the ACF can be obtained in which the temporal autocorrelation is depicted by ACF, and the Ljung‑Box Q test is a "portmanteau" test. The null hypothesis H_0 that ''a sequence of residuals does not exhibit autocorrelation for a specified number of lags L", is proved technically with respect to other hypothesis H_1 that ''some autocorrelation coefficient is nonzero.'' Eq. (3.3) defines the Ljung‑Box Q test statistic in more detail,

$$Q = s(s+2)\sum_{k=1}^{M}\frac{\rho_k^2}{s-k'} \tag{3.3}$$

where Eq. (3.4) indicates at lag-k, autocorrelation coefficients ρ_k are,

$$\rho_k = \frac{\sum_{i=1}^{s-k}(h_i - \bar{h})(h_{i+k} - \bar{h})}{\sum_{i=1}^{s}(h_i - \bar{h})^2} \tag{3.4}$$

with $\bar{h} = \frac{1}{s}\sum_{i=1}^{s} h_i$ under H_0 the statistic Q asymptotically follows a $g_{(M)}^2$ distribution. The model shows autocorrelation and reject the zero hypothesis H_0 if,

$$Q > g_{(1-\alpha,M)}^2 \tag{3.5}$$

where the critical value of the Chi-square distribution is defined for significance level α, or critical level $p = 1 - \alpha$, known as p value.

3.2.2 Stationarity and no stationarity

Autocorrelation, long memory, fractal and multi-fractal properties are the features of time-series that appear so frequently that they are referred to as stylized facts. The main disadvantage of working with values of price time series is that they follow a random walk process from the standpoint of stochastic processes. The coefficient of autocorrelation is ρ_k, with $k > 1$ are statistically remarkable for many lags L and the first-order autocorrelation coefficient ρ_1 is equal to one (Stavroyiannis, 2019). This kind of time series are called unit root time-series or integrated of order one which are expressed by I (1). Modelling the levels of these series under such conditions is unproductive since the residuals of the models show redundancy, which putting the entire framework of statistical validity in jeopardy. In order to examine these series effectively, they must be stationary which is essential for the advent of a new forecasting model.

Assume that $F_h(h_{t_1+\tau}, \ldots\ldots, h_{t_n+\tau})$ is the total distribution algorithm of the intrinsic joint distribution of $\{h_t\}$ at times $t_1 + \tau, \ldots\ldots, t_1 + n$ then the stochastic process $\{h_t\}$ is strictly stationary if

$$F_h\left(h_{t_1+\tau}, \ldots\ldots, h_{t_n+\tau}\right) = F_h\left(h_{t_1}, \ldots\ldots, h_{t_n}\right) \tag{3.6}$$

for all $\tau, t_1 \ldots. t_n \in \mathbb{R}$ and $n \in N$. Nevertheless, the stationarity of time series is reduced resulting to weak covariance stationarity (Brockwell et al., 2016). A stochastic process becomes covariance stationary when the mean is constant, the second moment is finite, and the covariance function relies on the difference between t_1 and t_2. Hence, the auto-covariance needs to be denoted with one variable, i.e.,

$$cov_{hh}(t_1, t_2) = cov_{hh}(t_1 - t_2, 0) \tag{3.7}$$

where cov_{hh} is the auto-covariance of the y_t series to summarize stationarity based on statistical features of the stochastic process. It has been a general hypothesis that many procedures such as statistical assessment, modelling and prediction become simpler when adopted the stationary processes. The partial autocorrelation function provides a resolution once the problem has been detected, where the lag-k coefficient $\phi_{k,k}$ is displayed by the indicated formula in Eq. (3.8),

$$\begin{cases} \phi_{k,k} = \dfrac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}} \\ \phi_{kj} = \phi_{k-1,j} - \phi_{k,k}\phi_{k-1,k-j}, \end{cases} \tag{3.8}$$

for $k > 1$ and $\phi_{1,1} = \rho_1$. Clearly, if there is unit root throughout the series, that is $\rho_1 = 1$, the first-order partial autocorrelation coefficient $\phi_{1,1}$ will become one. Commonly, the initial coefficient is statistically significant while the rest are insignificant(Shaman, 2010).

Then, the first series should be characterized by the first differences of the series, defined by following equation,

$$\triangle_t = h_t - h_{t-1} \tag{3.9}$$

Therefore, the first difference of the time series in stationarity obtained can be represented with integrated of order zero which is $I(0)$. However, crossover of different non-stationarities could present while computing the time-series data there such as unit-roots, structural pause, level up-downs, seasonal trend or a shifted variance. When the series is non-stationary (I (1)), the typical transformation is to take the first differences and transform it to stationary series (I (0)), whereas if the series contains structural breaks or a changing variance due to crises, a nonlinear BoxCox transformation will be the best solution (Osborne, 2010). As normality is an essential criterion for various statistical procedures, a BoxCox transformation provides a mechanism to turn non-normal data into a normal pattern. The following equation (3.10) defines one-parameter Box-Cox transformation as,

$$h_t = \begin{cases} \dfrac{h_t^{\lambda} - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \ln h_t, & \text{if } \lambda = 0. \end{cases} ; \tag{3.10}$$

where nonzero Box-Cox transformations are used for $\lambda = -3, -2, -0.5, 0, 0.5, 1$ and 2. The rule $\lambda = 0$ is followed by majority of the time series; therefore, the returns which are the first logarithmic differences are used to attain stationarity in these series,

$$r_t = \ln h_t - \ln h_{t-1} \approx \frac{h_t - h_{t-1}}{h_{t-1}} \tag{3.11}$$

the last expression being the percentage change or returns (Brockwell et al., 2016).

3.2.3 Augmented Dickey Fuller (ADF) test

The proposed pre-processing module for greatly improving the accuracy and durability of a deep learning algorithm for time series prediction is discussed in this part, based on well-known statistical concept and estimation for stationarity and non-stationarity qualities. Generally, the components of the dataset are not-stationary when a machine learning or deep learning model is applied to estimate the time-series. This implies that they may have unit roots and some order of integration. It is worth noting that the Augmented Dickey–Fuller (ADF) test has the ability to identify a unit root in a time series data (Brockwell et al., 2016; Pal & Prakash, 2017). The model is subjected to the testing method as following Eq. (3.12),

$$\triangle_{h_t} = \alpha + \beta t + \gamma h_{t-1} + \sum_{i=1}^{k-1} \delta_i \triangle_{h_{t-i}} + \epsilon_t \qquad (3.12)$$

Where α represents a constant, β is the coefficient of trend and $\gamma = (\rho_1 - 1)$ where ρ_1 denotes the first-order autocorrelation coefficient. It is notable that k is the lag order of the autoregressive determined so that the residuals ϵ_t have no serial correlation. There has a stochastic random walk process, if $\alpha = 0$ and $\beta = 0$, while if $\alpha \neq 0$ and $\beta = 0$, here the stochastic process is with drift. The unit root test is employed to evaluate statistical importance under the null hypothesis. H_0:$\{\gamma = 0 \text{ that is } \rho = 1\}$ versus the nonzero hypothesis H_1: $\{\gamma < 0 \; that \; is \; \rho < 1\}$.

Recursively taking the first differences in (9) or returns in (11) until the sequence is made stationary depending on the nature of the series. The autocorrelation in the model's residuals will be reduced when using a series of transformation based on first difference and returns. This means that the forecasting method will be considerably better at explaining the data because it captures all conceivable nonlinearities, assuring the model's accuracy and efficacy.

The flow-chart for the framework is shown below in Figure 3.1. Firstly, the time-series data is imported. The ADF test is then used to determine if the sequence levels are non-stationary, or if they have a unit root in time series. If the sequence is stochastic, the dataset will be continuously converted using first differences or returns till the resultant series becomes stationary in Steps 4–7. The newly modified time-series data is then utilised to train the forecasting model in Step 8.

Step 1: Input time-series data.

Step 2: Assess unit root test (ADF).

Step 3: If (U*nit root exists in time-series*) **then**

Step 4: **repeat** */* Time-series is not I (0)*/*

Step 5: Convert the time series based on differences (9) or returns (11). */* non-stationary/**

Step 6: Assess unit root test (ADF).

Step 7: Until (*Stationarity exists in time-series.*)

Step 8: By using converted time-series, train the prediction model.

Step 9: **else.**

Step 10: By using real time-series, train the prediction models. */* Time-series is I (0)*/*

Step 11: On the training sample, estimate the residuals. */* stationary/**

Step 12: **If** (*Autocorrelation exists in residuals.*) then

Step 13: Convert the time series based on differences (9) or returns (11).

Step 14: Using the converted time-series, retrain the forecasting model.

Step 15: **end if**

Step 16: end if

If the data is stationary, on the other hand, the levels of time-series are employed to train the forecasting model in Step 10. The errors of the estimation method on the learning algorithm are employed for further analysis and testing. It is noticeable that a training is performed with a series which has a unit root. When the predicted values becoming near to the real values for any realistic model, then, presence of strong autocorrelation factors mark the model as unproductive (Livieris, Stavroyiannis, Pintelas, & Pintelas, 2020). Therefore, ACF plots and/or the Ljung–Box Q test is used to investigate autocorrelation within residuals of the dataset in step 11. Eventually, if the residuals have autocorrelation, the recommended transformation is performed to the training phase and the algorithm is retrained utilizing the newly transformed dataset according to steps 13–14. It is noticeable that if the series levels are stationary and the residuals on the training dataset indicate no autocorrelation, there is no need to reform the series because it will result in catastrophic phenomena of over-differencing. To put it in another way, over-differencing makes the entire mechanism "non-invertible," and thus lacked an endless autoregressive expression. In the form of a flowchart, Figure 3.1 depicts an insight of the intended structure.

Finally, if the classifier is trained with a transformed series with no autocorrelation in residuals, the inverse transformation will be used in the model's forecasts to obtain the forecast for the levels of the exact time-series.

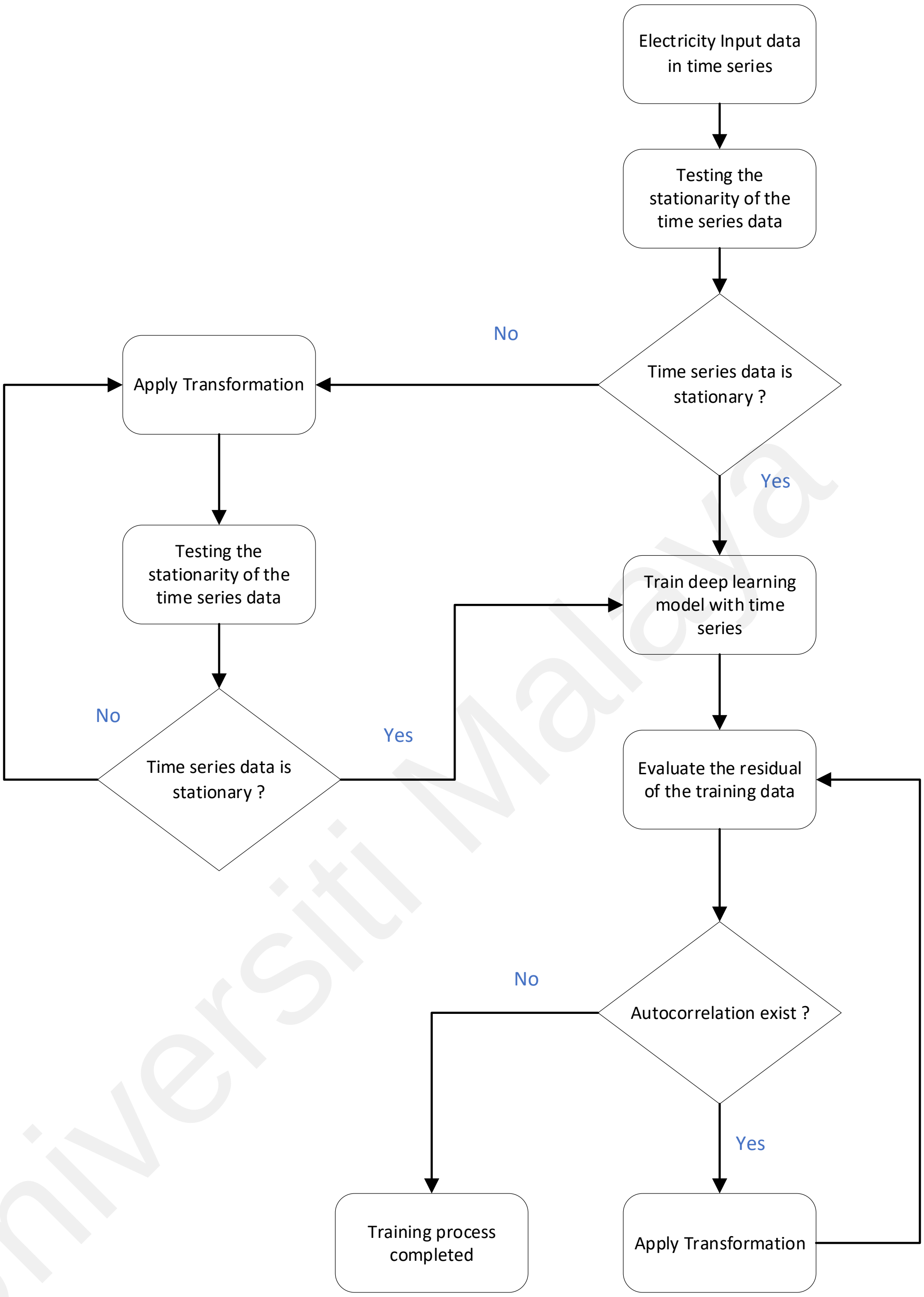


Figure 3.1: Proposed framework of time series analysis

3.3 The proposed LSTM+LISHT

In order to process the long sequence of time series data, LSTM recurrent neural network (RNN) is proposed with the aim to overcome the problem of vanishing gradient and gradient explosion that can occur in conventional RNN. The input gates and output gates are replaced by memory/forget gates in the hidden layer of LSTM RNN which include memory space and information flows process for long historical time series shown in Figure 3.2.. In this work, a fully-connected layer follows the sequential input layer and LSTM layers in the forecasting module. The results obtained from the two nodes are combined and used as feed for the following fully-connected layer after crossing the LSTM layer. The forecasting output is then calculated. The nodes in the forecast module's fully connected layer comprise the features retrieved from input training data. As a result, the number of neurons in each layer determines the number of features to retrieve (Zhang et al., 2020).

Simple gradient descent methods are often used to find global minimum/saddle points, where the defined configurations reach training loss zero or near to zero, even when the data and labels are randomized before training. However, this behaviour is desirable, but always not universal. The training ability of deep neural network (DNN) is directly and indirectly influenced by the factors like network architecture, the choice of optimizer, variable initialization, and most importantly the type of non-linearity function to be used in the architecture. All these factors are considered in this methodology for predicting spikes.

Conventional activation functions such as ReLU and Swish are less feasible for large negative input values and also may suffer from the dying gradient problem due to zero-hard rectification. Therefore, it is essential to adopt a better activation function to overcome those limitations. In this work, a non-parametric function, called Linearly Scaled Hyperbolic Tangent (LISHT) for neural networks (NNs) is employed in this model as referred (13). The LISHT activation function is utilized to scale the non-linear hyperbolic tangent (Tanh) function through a linear function and tackle the dying gradient problem.

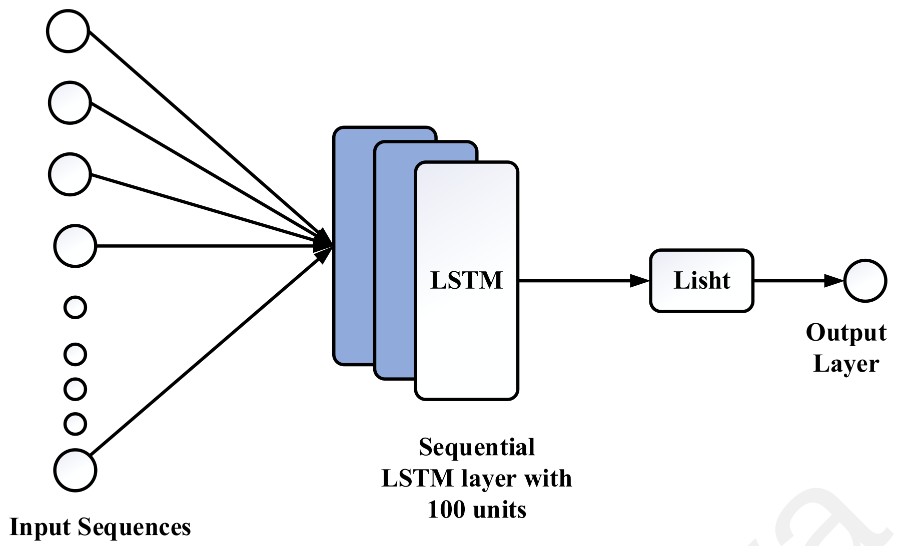


Figure 3.2: LSTM+Lisht Prediction model architecture

3.3.1 Recurrent Neural Network (RNN)

The feedforward neural network is a generalisation of the RNN, which incorporates an internal memory. Because RNN is recurrent in nature, it accomplishes the equivalent function for each input data. However, the output of the given inputs is dependent on the previous calculations. The output is replicated and transmitted back into the recurrent network when it is created. RNN makes a judgement based on the present input and the output acquired from the prior input (Figure 3.3). Unlike feedforward neural networks, RNNs may process sequences of inputs using their internal state (memory). This RNN is appropriate if there is unsegmented, linked handwriting recognition or voice recognition. In an RNN network, all of the inputs are connected to one another, which distinguishes it from other neural networks.

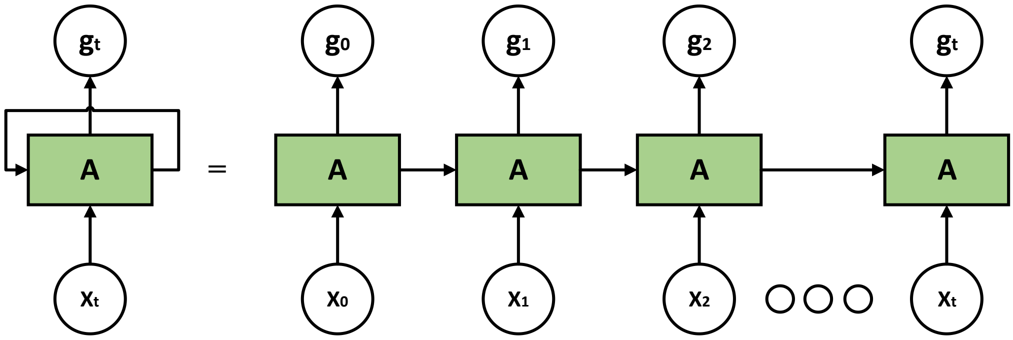


Figure 3.3: An unrolled recurrent neural network

From the series of the input data, first comes x_0 which becomes g_0 at the output. After that together with x_1 it creates input for next stage. Which means the input for further proceedings are g_0 and x_1. Likewise, x_2 together with g_1 prepare for the next unit input and so on. Basically, during the time of training RNN keeps remembering the context for further step (Figure 3.3).

The formula for the current state is expressed in Eq (3.16):

$$g_t = f(g_{t-1}, x_t) \tag{3.16}$$

Applying activation function shown below in Eq (3.17):

$$g_t = \tanh\,(w_{gg}g_{t-1} + w_{xg}g_t) \tag{3.17}$$

Where w stands for weight, whereas h stands for the solitary hidden vector, w_{gg} is the weight from the past hidden unit, w_{gx} is the weight from the present input sequence, $tanh$ is the activation function, to the range [-1.1].

The output in Eq (3.18) given below:

$$y_t = w_{gy}g_t \tag{3.18}$$

3.3.2 Long-short-term memory (LSTM) neural network

The LSTM improves RNN to avoid the difficulties of diminishing gradient and gradient inflation, LSTM RNN is presented to handle a long sequence of time series data (Hochreiter & Schmidhuber, 1997). Memory/forget gates substitute inlet and outlet gates in the LSTM RNN's hidden layer, which includes internal memory and data flow processes for extensive historical time series. Figures (3.4 – 3.5) depicts the structure of an LSTM RNN for data flows.

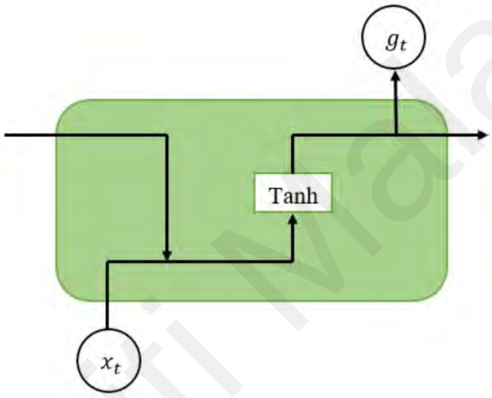


Figure 3.4: Recurrent neural unit (RNN architecture)

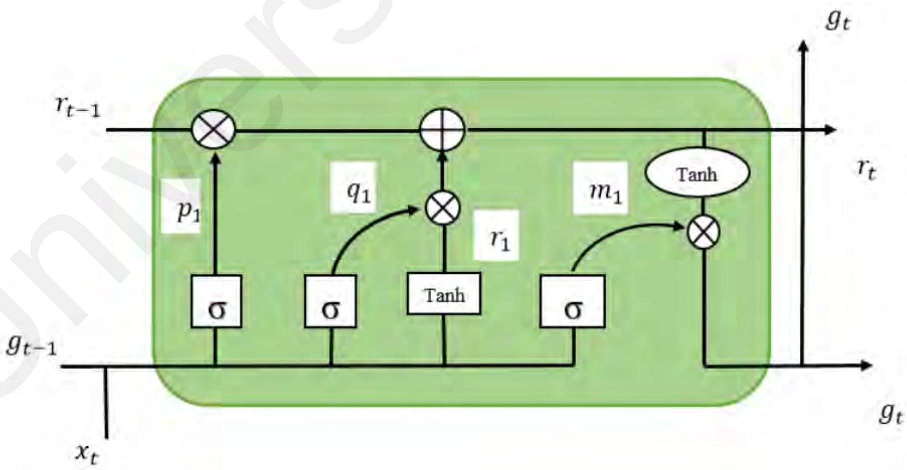


Figure 3.5. LSTM architecture

In Figure 3.5, the hyperbolic tangent activations and sigmoid function are *tanh* and σ respectively. For t moment, x_t and r_t are respectively input and cell states, likewise for the moment of $t - 1$ the cell states and output are respectively r_{t-1} and g_{t-1}. Here x_t & g_{t-1} are used as input of the forget gate p_t to get the output values which is a sigmoid function that keep in mind the essential information r_{t-1}. To determine cell states r_t it uses x_t & g_{t-1} as input which is the output value of the input by a combination of a candidate cell state r_{t_0}. Based on r_t with a sigmoid function and *tanh* function to adjust the output value of g_t of the output gate m_t is utilized. The mathematical description is as follows in Eqs (3.19 – 3.24):

$$p_t = \sigma\left(W_p . \ [g_{t-1}, x_t] + b_p \right) \tag{3.19}$$

$$q_t = \sigma\left(W_q \ . \ [g_{t-1}, x_t] + b_q\right) \tag{3.20}$$

$$\bar{r}_t = tanh(W_r \ . \ [g_{t-1}, x_t] + b_r) \tag{3.21}$$

$$r_t = p_t \cdot r_{t-1} + q_t \cdot r_t \tag{3.22}$$

$$m_t = \sigma(W_m \cdot [g_{t-1}, x_t] + b_m) \tag{3.23}$$

$$g_t = m_t \cdot tanh(r_t) \tag{3.24}$$

The residual error between forecasted and actual values can be reduced by error indemnification module. Validation set errors should be interconnected with the inputs to acknowledge the error adjusting trends, but the exception is for the features which are extracted from the prediction module.

3.3.3 The proposed spike detection method

Simple gradient descent methods are often used to find global minimum/saddle points, where the defined configurations reach training loss zero or near to zero, even when the data and labels are randomized before training. However, this behaviour is desirable, but always not universal. The training ability of DNN is directly and indirectly influenced by the factors like network architecture, the choice of optimizer, variable initialization, and most importantly the type of non-linearity function to be used in the architecture. All these factors are considered in this methodology for predicting spikes.

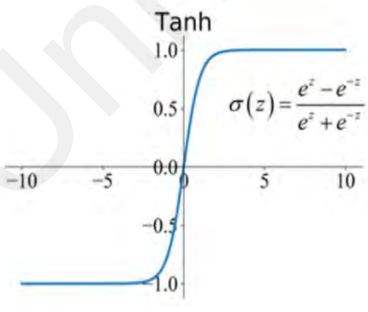
Some known activation functions, such as ReLU and Swish, are incompatible with high negative input data and may experience the dying gradient phenomenon caused by zero-hard inversion. Therefore, it is important to find an effective activation function to overcome those limitations. In this work, for LSTM network a non-parametric function, called Linearly Scaled Hyperbolic Tangent (LISHT) is employed in this model as referred. The LISHT activation function is utilized to scale the non-linear Hyperbolic Tangent (Tanh) function through a linear function and tackle the dying gradient problem.

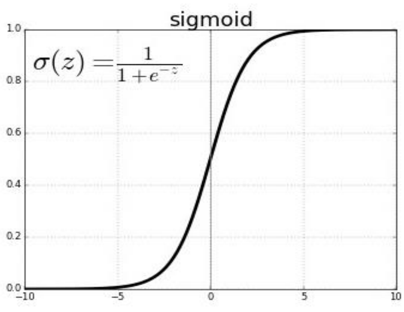
From Eq. (3.13) Let an input vector be a ∈ R^d, and each hidden layer is capable to transform its input vector by applying a nonlinear mapping from the q^{th} layer to the $(q+1)$ th layer as following:

$$a = \tau^0$$

$$\sum_{l=1}^{N^q} w_{kl}^q \tau_l^q + o_k^q = c_k^{q+1} \tag{3.13}$$

$$\phi\left(c_k^{q+1}\right) = \tau_k^{q+1}$$





(a) **(b)**

Figure 3.6: Differences of different activation layers

LiSHT is a non-parametric linearly scaled hyperbolic tangent activation layer (Elfwing, Uchibe, & Doya, 2018) that have unrestricted upper limits property on the right hand side of activation curve. LiSHT has the advantage of positive activation that does not identically propagate for all inputs, which solves gradient problem at back propagation and contributes to faster training of deep neural network. The LISHT activation function is calculated by multiplying the $Tanh$ function to its input x and defined as the Eqs. (3.14-3.15). where $g(x)$ is a hyperbolic tangent function.

$$\phi(x) = x \cdot g(x) \tag{3.14}$$

$$g(x) = Tanh(x) = \frac{exp^{x} - exp^{-x}}{exp^{x} - exp^{-x}} \tag{3.15}$$

3.4 The proposed improved BTE-GRU model

In the field of AI machine learning machine learning have significant role. Machine learning techniques investigate data, gain overview through it and make decision on the new information. A multi-layered design of computation is used in deep learning which is unlike machine learning. Several problems are unable to solve through machine learning where deep learning can solve very easily. In this research machine learning algorithm used with deep learning to get highest accuracy of electricity price forecasting.

3.4.1 Bagged trees ensemble

Bagged Trees Ensemble (BTE) algorithm is adopted to generate several bootstrap samples and trains classifiers on the new learning sets. Then, BTE algorithm computes the mean predictions for a sequential output or performs a plurality for a class outcome. Assuming the training set is defined as $A\{(p_m, q_m), m \epsilon 1,2, \ldots\ldots, M\}$with q_m represents either a class label or numerical response. For an p input q, could be predicted using Ø (p, A), where Ø (p, A) is a single learning set predictor. Assuming there is a series of learning sets $\{A_R, R = 1, \ldots\ldots, M\}$each having M number of distinct samples chosen from A. The purpose will be to utilize $\{A_R\}$to obtain more accurate predictor than the single learning set predictor Ø (p, A). Replacing Ø (p, A) with the average of Ø (p, A_R) over R for a numerical value of q, i.e., $Ø_D(p)$ = E_A Ø (p, A) is an apparent method of performing the task. The subscript D in $Ø_D$ signifies aggregation, and E_A represents the anticipation over A in that equation. Most of the time we have one learning set, however, bootstrap samples can be created from A (A^C) which might be used to replicate a similar procedure leading to $Ø_D$ with replacement, such that $Ø_D(p)$ = $average$ Ø (p, A^C) (Hong Wang, Xu, & Zhou, 2015).

The classifiers are also defined as regression trees (decision trees). In this work, the proposed bagged tress implemented 5 folds cross-validation. Then, the number of RT (chosen N = 30) and the minimal leaf size (selected G_{min} = 8) are applied. Each regression tree was built using a bootstrap sample selected uniformly from the input data. Further, the bagging method averaged the learners' outputs to obtain a single forecast. This technique is called bagging and Table 3.1 presents the bagging algorithm applied in this work.

Table 3.1: Bagged tree ensemble algorithm

Input	Training dataset = $\{(p_k, q_k), k \in\ 1,2, \ldots\ldots, m\}$ A base learning algorithm using regression trees (RT) and the number of learning cycles, j.	
Process	for $j = 1, \ldots.., \mathrm{J}$ $\mathrm{TD_j} = \mathrm{Bootstrap\ (TD)}$	Create bootstrapped samples from training dataset with replacement.
Output	$RT_j(p, q) = GTD_j$	Bootstrap sample for training dataset (TD), $\mathrm{TD_j}$; Train regression trees (RT_j).

	$q* = \frac{1}{J}\sum_{j=1}^{J} RT_j(p*, q)$	The output of the trained base learners are averaged.

3.4.2 Gated recurrent unit (GRU)

As RNN is recurrent in nature, it works much the same way for all inputs, while the output of the input data is dependent on the previous calculations. After generating the output data, it is replicated and revert back into the recurrent network unit. RNN count the present input and the output acquired from the last input while it makes logical decisions. RNNs can utilize the internal state (memory) to evaluate input variables, which is different from feedforward neural network (Nguyen, Duong, & Le, 2020; Tokgöz & Ünal, 2018). In recurrent neural network, all of the inputs are connected with one another, which distinguishes it from the other neural networks.

In general, the RNN has an issue with inflating and erasing gradients (Pan et al., 2020). The most familiar and used Recurrent Neural Network (RNN) elements are GRU and LSTM. RNNs have a reverse connectivity which has significant detrimental impact on model performance, which can't see in CNNs, GRU deals with these difficulties. GRU is a more robust RNN framework, designed for long-range dynamic feature dependencies. Besides, a GRU architecture requires less training time, with typically competitive results to an LSTM. The input and forget gates are fused into a single update gate in GRU's core structure(Manotumruksa, Macdonald, & Ounis, 2020; Shakiba & Zhou, 2020). The GRU architecture contains two gates layers: the reset (Y) and an update (Z) gate, whereas LSTM architecture includes three gates (Borovkova & Tsiamas, 2019; Zhou et al., 2019).

In this work, the input and forget gate in GRU are merged to update gate and hidden state reset gate as result it takes less time to process the data. The equations of the GRU cell adopted in this work are shown in Eqs. (3.25-3.29). A multi-layer GRU is adopted due to faster training process and smaller number of parameters required.

$$p_t = \sigma(W_p . [g_{t-1}, x_t]) \tag{3.25}$$

$$q_t = \sigma(W_q . [g_{t-1}, x_t]) \tag{3.26}$$

$$\bar{r}_t = \emptyset(W_r . [p_t \times g_{t-1}, x_t]) \tag{3.27}$$

$$r_t = (I - q_t) \times g_{t-1} + q_t \cdot \bar{r}_t \tag{3.28}$$

$$y_t = \sigma (W_o \cdot r_t) \tag{3.29}$$

where x_t , g_{t-1}, g_t , p_t , q_t , $\bar{r}_t$ and y_t are the input vector, the state memory variable at previous moment, the state memory variable at current moment, the state of reset gate, the state of update gate, the state of the current candidate set and the output vector at current moment respectively. On the other hand, W_p , W_q, W_r , W_o are the weight matrices for the corresponding inputs of the network activation functions while I represent the identity matrix. Then, backpropagation (BP) algorithm is employed to train and adjust the system parameters of the GRU RNN, such as the weights and biases.

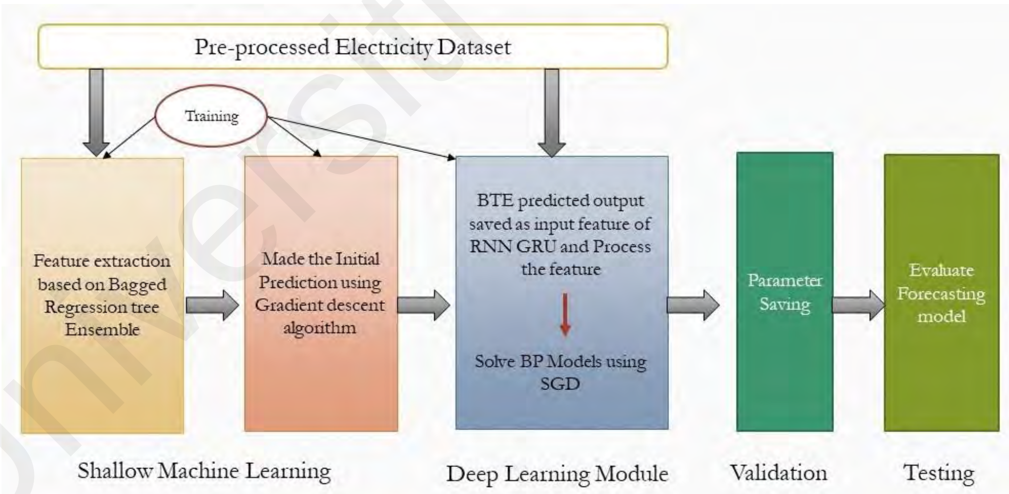


Figure 3.7: Flow diagram of the proposed Improved BTE-GRU model

Therefore, this method proposes an optimized GRU consisting of RNN with BTE forecasting model for electricity price prediction which is displayed in Figure 3.7. The BTE is applied to predict nonlinear data which is further optimized by using GRU RNN network.

3.5 Summary

In this research, new prediction methods for day-ahead and midterm price forecasting in competitive power markets is proposed. The proposed methods are divided into two parts: one is time series analysis and preparation of an effective time series for better results from deep learning simulation and the other is the development of an optimised LSTM module for price prediction with big data time series that includes LISHT for spike prediction.

Firstly, the time series has taken from five regions of Australia's national energy market with five minutes of interval, the time series are tested for unit root through ADF test which is one of the most familiar methods of checking the data is stationarity or not. If there is unit root, then data will convert into first difference and return series which makes the time series reliable for better performance in deep learning. Besides is the time series do not show unit root then the time series will process through an autocorrelation check in residuals of the time series. Therefore, if there exists autocorrelation in residuals of time series it will convert the time series through box cox transformation to remove the outliers which could make the time series vulnerable and prediction accuracy become ineffective. Finally, the transformed time series will process through a deep learning module where RNN used as the unit cell and more enhanced by using LSTM cell along with a hyperbolic tangent layer. This proposed methodology has applied on NSW, QLD, SA, TAS and VIC regions different time series data. The simulation results show its efficacy and reliability which is comparatively better than other performed methods for short term electricity price forecasting.

Secondly, the BTE method has been explored for long range of prediction. The preprocessed data was trained with shallow machine learning module which called bagged decision tree and combined or optimized with gated recurrent unit (GRU). The predicted value from BTE utilized as the input variable for GRU which is the advanced form of RNN. Unlike LSTM, GRU performs better for large sequence of time series data. During training and testing while considered monthly prediction or more than week prediction the BTE+GRU model is very handy. Besides the time for simulation is less than the typical midterm forecasting through other deep learning methods. Five most important economic zone of Australia also used as case study and the input features were same as the short-term forecasting method.

Finally, it can be concluded that for short term electricity forecasting LSTM along with LISHT method worked remarkably where prediction accuracy reflects the efficacy, on the other hand for mid-term electricity forecasting BTE method which is optimized by GRU performed significantly noteworthy. Both of the method improved the forecasting accuracy, which can be seen through the result of the experiment.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction

The findings of the suggested time series analysis and forecasting method for short and mid-term competitive market prices are presented in this chapter. This time series data was extracted from five locations of Australia's national energy market, each with a five-minute interval. The time series were verified for unit root using the ADF test, which is one of the most used methods for determining whether or not data is stationary. If the data has a unit root it means the data is not stationary, to solve this and make time series data stationary the algorithm will convert it into a first difference and return series, which will make the time series more dependable for deep learning. In addition, if the time series does not have a unit root, it will be subjected to an autocorrelation check in the residuals. As a result, if autocorrelation occurs in time series residuals, the time series will be converted using the box cox transformation to remove outliers, which could make the time series vulnerable and prediction accuracy useless. Finally, the modified time series will be processed through a deep learning module with RNN as the unit cell and an LSTM cell coupled with a hyperbolic tangent layer to improve performance. This suggested methodology has been tested on various time series data from the New South Wales, Queensland, South Australia, Tasmania, and Victoria. The simulation results demonstrate its efficacy and reliability, which is superior to that of other methods.

4.2 Experimental results of pre-processed time series data

In this research, the time series dataset includes electricity price, demand, and renewable energy supply of Australia’s most important five economic zones. The electricity market data covers the duration from 1 September 2020 to 31 May 2021, which is obtained from AEMO ("Australian Energy Market Operator," 2022). The data were divided into training and test set consisting of hourly electricity price data as tabulated in Table 4.1:

Table 4.1: Seasonal Training Dataset

Seasons	Training set (Short-term)	Training set (Mid-term)
	24 hours forecasting	720 hours forecasting
Sep-Oct-Nov (Spring)	Oct (696 hours)	Sep-Oct (1440 hours)
Dec-Jan-Feb (Summer)	Jan (696 hours)	Dec-Jan (1440 hours)
Mar-Apr-May (Autumn)	Apr (696 hours)	Mar-Apr (1440 hours)

There have been no missing data in any of the time-series and outlier prices were not eliminated in order to preserve the characteristics of every series, even though these prices are the consequence of rare events. Table 4.2 presents the descriptive analysis for every training dataset and testing dataset, such as the measurements of minimal, max, average, sample variance (std. dev.), median, skewness, and kurtosis for illustrating the distribution's nature.

Table 4.2: Descriptive analysis of time series data

Region	Data	Minimum	Maximum	Mean	SD	Median	Skewness	Kurtosis
NSW	Training	19.62	54.89	37.98	7.31	38.21	0.306	0.071
	Testing	33.80	58.23	42.86	7.16	52.08	0.900	-0.451
QLD	Training	-39.38	152.0	39.32	18.09	37.85	1.145	5.726
	Testing	25.94	49.40	37.24	6.70	38.03	0.162	-0.735
SA	Training	-114.91	139.50	32.80	35.95	39.62	-0.987	2.344
	Testing	35.60	66.18	45.99	6.91	44.17	1.292	2.130
TAS	Training	-25.26	154.55	53.31	20.54	50.32	0.806	3.02
	Testing	39.30	68.37	52.13	6.55	49.57	0.735	0.525
VIC	Training	-120.54	305.07	45.11	27.87	44.44	1.738	16.784
	Testing	36.09	61.56	45.02	6.45	44.25	1.209	1.386

Using the ADF unit root test, the proposed framework employed to the National Electricity Market (NEM) price time-series in Australia to determine whether the training data are stationary or not. The outputs of the ADF unit root test for the training data of Australia's five states series under investigation are shown in Table 4.3. Considering the t-statistics (t-stat) and the corresponding p values the null hypothesis H_0: ''the levels possess a unit root and are non-stationary'' is accepted for time series.

Table 4.3: ADF unit root test for the training data

Series	NSW	QLD	SA	TAS	VIC
***t* static**	-37.16	-71.36	-28.82	-39.42	-24.58
***p* value**	0.001*	0.001*	0.001*	0.001*	0.001*

In the sequel, the ADF test was run on a time-series to see if the unit root existed, as per the provided framework. The outcomes of the ADF unit root test for the training data of all time-series datasets are shown in Table 4.3. The (*) indicates statistical impact at the 5% critical threshold. Clearly, it's worth of noting that all p values are almost zero, the null hypothesis H_0 is rejected. This test shows the overall time series data behaviour and trend. Here maximum value of P, 0.999 and minimum 0.001. By using MATLAB econometric modeler tool, it is found that the P value obtained 0.001 and it rejects the null hypothesis.

As a result, the time series are "appropriate" for training a deep learning model with minimal autocorrelation in the errors, and a significant boost in forecasting accuracy is anticipated when comparing to same model trained with the non-transformed series. In order to evaluate the performance of the proposed model, the regression ability is assessed using mean absolute error (MAE) and root mean square error (RMSE). Besides that, another four key performance indicators are also employed: Accuracy (Acc), F1-score

(F1), Sensitivity (Sen), Specificity (Spe), Positive Predicted Values (PPV) and Negative Predictive Values (NPV)(Raj, Babu, VL, Varalatchoumy, & Kathirvel, 2022; Vantuch, Mišák, & Stuchlý, 2016) which are indicated by the following Eqs. 4.1-4.6.

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{FP}'} \tag{4.1}$$

$$\mathrm{F}_1 = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}'} \tag{4.2}$$

$$\mathrm{Spe} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}'} \tag{4.3}$$

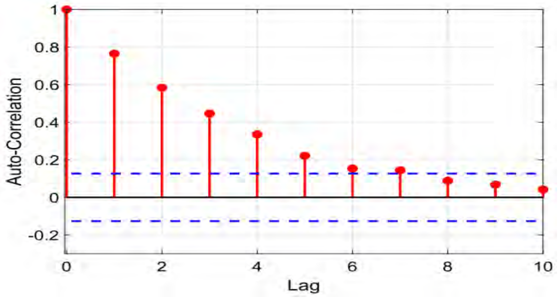
$$\mathrm{Spe} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}'} \tag{4.4}$$

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}'} \tag{4.5}$$

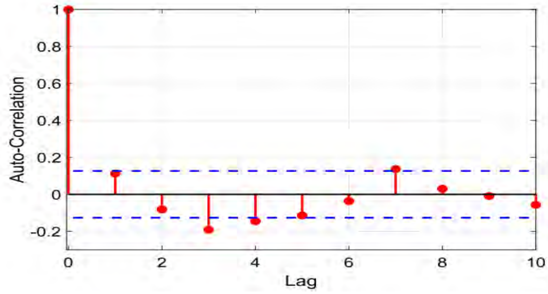
$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}} \tag{4.6}$$

In this case, TP represents the frequency of prices that were successfully identified as raised, the number of prices that were successfully detected as having a decreasing value is denoted by TN, FP is the amount of prices that were incorrectly detected as being increased, whereas FN denotes the quantity of prices that were incorrectly detected as being dropped. Furthermore, the area under curve (AUC) statistic, considered one of the most important classification metrics which has been incorporated in the assessment and is shown using the receiver operating characteristic (ROC) curve. The ROC curve is made by comparing the true positive rate (Sensitivity) against the false positive rate (Specificity) at different cut-off values.

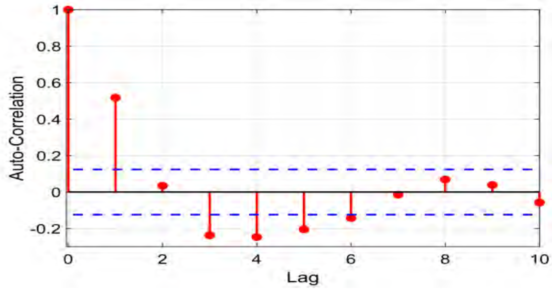
In the following Figures, the forecasting performance of prediction techniques is investigated by using the Auto-Correlation Function (ACF) plot, where the Ljung–Box Q test is performed to detect for autocorrelation in the residuals. This checks that all predictive model has correctly fit with time series and the data sets are evenly dispersed, steadily over time independent, and well-fit. The Ljung– Box Q test is a ‘‘portmanteau’’ test which analyse the null hypothesis H_0 that ‘‘a series of residuals exhibits no autocorrelation for a fixed number of lags L,’’ which is the opposite of another hypothesis H_1 that ‘‘some autocorrelation coefficient is non-zero’’. Figures 4.1(a) to 4.1(j) displayed the ACF graphs of LSTM+LISHT model for the electricity price data for five different states of Australia. The ACF graphs of the prediction model are trained with typical time-series showing the high spikes were observed in several lags displayed in Figures. 4.1(a), 4.1(c), 4.1 (e), 4.1 (g), 4.1(i). which shows that such model's estimation may be inaccurate. The ACF plot spikes at lag 1 then slowly decays to lag 10. From lag 1 to 4 spikes are too high and cut off at the significant band 0.2, Which shows that the significant autocorrelation presents in the residual of trained data. On the contrary, from the Figures (4.2, 4.4, 4.6, 4.8, 4.10) it is shown that the spikes from lag 2 immediately go down under or between the significant band. Therefore, the autocorrelation in the residual does not exist in the trained data and is statistically sound for the evaluation of time series. In summary, all ACF plots of the LSTM+LiSHT generated using the converted time series Figures (4.2, 4.4, 4.6, 4.8, 4.10), show that the residuals do not have autocorrelation. This can be further verified by the results obtained from the Ljung‐ Box Q test (Table 4.4) where the transformed time series data using the BoxCox transformation is free from autocorrelation.



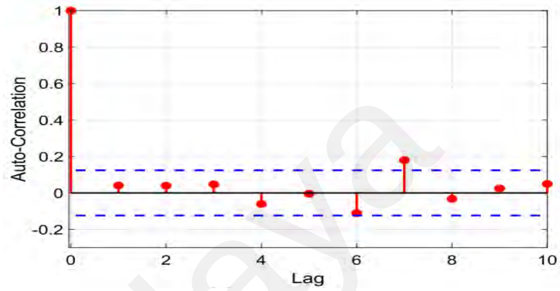
(a) Residual before transform NSW



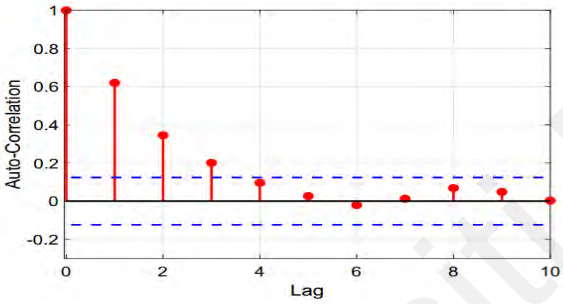
(b) Residuals after transform (box-cox) NSW



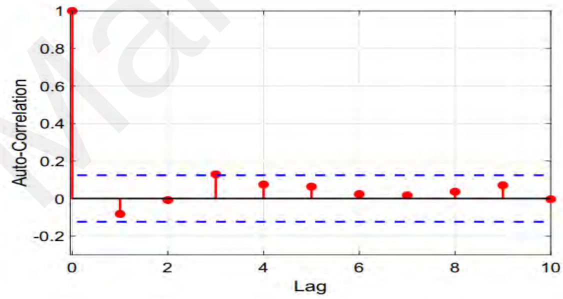
(c) Residual before transform QLD



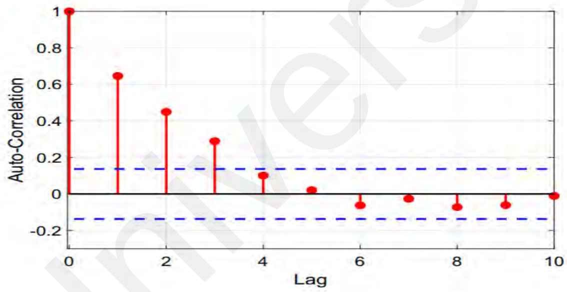
(d) Residuals after transform (box-cox) QLD



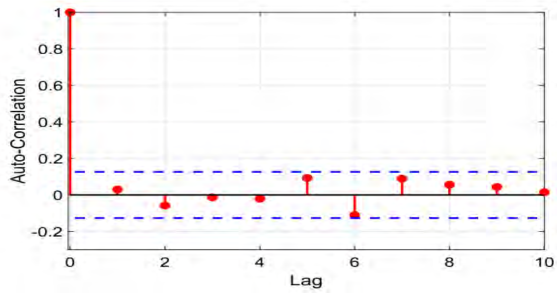
(e) Residual before transform SA



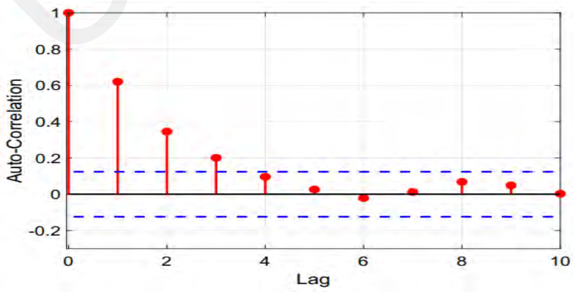
(f) Residuals after transform (box-cox) SA



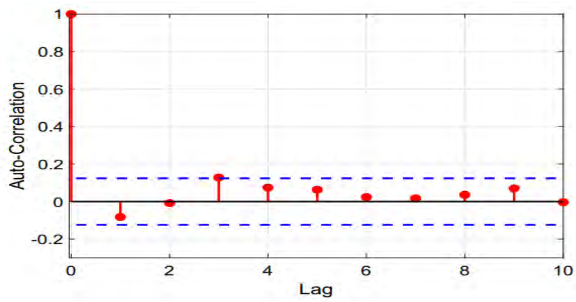
(g) Residual before transform TAS



(h) Residuals after transform (box-cox) TAS



(i) Residual before transform VIC



(j) Residuals after transform (box-cox) VIC

Figure 4.1:Autocorrelation of residuals before and after transform

Table 4.4: represent the result of the Ljung–Box Q test using *L* = 10

Condition	Forecasting	Autocorrelation existence
Before transformed	1 day	Yes
	1 month	Yes
After transformed	1 day	No
	1 month	No

In this work, it has been established theoretically and experimentally that the time-series data are ‘‘appropriate’’ for developing a deep learning model, which is one of the contributions of this study. In another way it can be said that this work has developed a new framework that can discover effective time series data for training a deep learning model. This will lead to a stable and reliable forecasting model. On the other hand, if the series fails to meet the desired criteria, it is deemed "unsuitable," and any attempts to develop a solid prediction model would most likely be useless. Therefore, this work is a beginning point for development of any prediction methodology for various time series forecasting. If the starting dataset are unstable or non-stationary, the work done for developing the forecasting model could be meaningless. Furthermore, it can be justified that this work has developed an innovative and comprehensive framework that allows any unstable time-series to be transformed to a stable condition by conducting a boxcox transformation method. It can be seen that the proposed transformation has successfully eliminated the "unsuitable" data, avoiding the costly and time-consuming "trial and error" method. Besides that, it is noticeable that one of the most interesting properties of our suggested framework is that this method can be simply modified to encompass a broader scientific domain of time-series forecasting operations without requiring any further adjustments or limits. More specifically, the recommended method uses statistic and economic tests to conduct an optimal pre-processing phase for utilising the internal structure of the timeseries. Finally, it is seen that while deep learning models are well accepted for time series, the proposed framework significantly enhance the performance. However, more study is being done to see which from these approaches may be implemented more effectively a priori based on the properties for every time-series in order to get better forecasting performance. For accomplishing the prerequisite diagnosis and appropriate time transformation methodology, a complex pre-processing framework that refers to the inherent time-series particular traits such as stationarity, heteroskedasticity, seasonal cycles, and shifting variance can be used.

4.3 Experimental results for proposed method 1

The efficacy of the BiLSTM and LSTM+LISHT prediction model for the energy price dataset is compared in Tables 4.6 - 4.11. A confusion matrix is commonly known as an error matrix, a table used in machine learning to do classification which enables the visualisation of classification quality (Raj et al., 2022). Each column represents the instances in the forecasted price while each row represents the actual price. The TP, FP, TN, and FN data is being used to concisely display the confusion matrix of binary classification shown below in table 4.5.

Table 4.5: Confusion matrix of binary classification

Predicted Price / Actual Price	1	0
1	TP	FN
0	FP	TN

From the confusion matrix value ACC, AUC, F1, Spe, Sen, PPV and NPV calculated from the Eqns 4.1 – 4.6. In spring referring to Table 4.6, the BiLSTM model's average accuracy (ACC) in NSW is 0.958 for short-term one-day predicting and 0.989 for midterm forecasting, whereas in QLD the accuracy is 1 and 0.763 for short and midterm forecasting, respectively. Likewise in SA, TAS and VIC the short-term accuracy is 0.958, 0.916, 0.958 respectively and for midterm 0.976, 0.832, 0.896 respectively. The sensitivity and specificity also show the high sensitivity for the accuracy. Additionally, Tables 4.9 - 4.11 showed the average accuracy results (ACC) for proposed LSTM+LISHT. In NSW referring to Table 4.9, the short-term accuracy is 1 and the midterm is 0.984 in the spring. For short and mid forecasting in QLD, the accuracy is 1 and 0.870, respectively. Along with the above results, the short term and midterm accuracy for SA, TAS, and Victoria are 0.958, 0.958,1 and 0.923, 0.970, 0.997 respectively. Furthermore, throughout the summer and autumn, both techniques (BiLSTM and LSTM+LISHT) exhibit higher accuracy in short term forecasting than midterm forecasting. Finally, it can be sum up with the results of binary classification that some cases yet its very less amount that the short-term algorithm performs good with midterm data, but it is not adequate for using mid-term forecasting as standard. There need to find another algorithm to predict midterm forecasting.

Table 4.6: Average performance comparison of short and mid-term BiLSTM for the five different dataset in spring

Time Series	Horizon	ACC	AUC	F1	Sen	Spe	PPV	NPV
NSW	1 day	0.958	0.890	0.923	0.857	1	1	0.944
	1 month	0.989	0.923	0.909	1	0.988	0.833	1
QLD	1 day	1	0.956	1	1	1	1	1
	1 month	0.763	0.543	0.794	1	0.563	0.659	1
SA	1 day	0.958	0.876	0.923	1	0.944	0.857	1
	1 month	0.976	0.964	0.985	1	0.893	0.970	1
TAS	1 day	0.916	0.934	0.857	0.75	1	1	0.888
	1 month	0.832	0.675	0.846	1	0.690	0.733	1
VIC	1 day	0.958	0.785	0.933	1	0.941	0.875	1
	1 month	0.896	0.923	0.612	1	0.887	0.441	1

Table 4.7: Average performance comparison of short and mid-term BiLSTM for the five different dataset in summer

Time Series	Horizon	ACC	AUC	F1	SEN	SPE	PPV	NPV
NSW	1 day	0.791	0.567	0.871	1	0.2857	0.772	1
	1 month	0.763	0.679	0.841	1	0.368	0.725	1
QLD	1 day	0.916	0.875	0.875	1	0.882	0.777	1
	1 month	0.809	0.879	0.888	1	0.204	0.799	1
SA	1 day	0.916	0.873	0.8	0.8	0.947	0.8	0.947
	1 month	0.503	0.439	0.348	1	0.427	0.210	1
TAS	1 day	0.9583	0.894	0.933	1	0.941	0.875	1
	1 month	0.734	0.895	0.519	1	0.690	0.351	1
VIC	1 day	0.916	0.872	0.928	1	0.818	0.866	1
	1 month	0.965	0.829	0.688	1	0.964	0.525	1

Table 4.8: Average performance comparison of short and mid-term BiLSTM for the five different dataset in autumn

Time Series	Horizon	ACC	AUC	F1	Sen	Spe	PPV	NPV
NSW	1 day	0.708	0.594	0.787	1	0.363	0.65	1
	1 month	0.994	0.813	0.963	0.952	0.998	0.975	0.996
QLD	1 day	0.583	0.673	0.5	1	0.473	0.333	1
	1 month	0.990	0.703	0.915	0.9	0.996	0.931	0.994
SA	1 day	0.958	0.862	0.857	0.75	1	1	0.952
	1 month	0.876	0.753	0.774	0.975	0.848	0.642	0.991
TAS	1 day	0.875	0.798	0.903	1	0.7	0.823	1
	1 month	0.969	0.848	0.779	1	0.967	0.638	1
VIC	1 day	0.958	0.735	0.8	0.666	1	1	0.954
	1 month	0.94	0.897	0.943	1	0.879	0.893	1

Table 4.9: Average performance comparison of short and mid-term LSTM+LISHT for the five different dataset in spring

Time Series	Horizon	ACC	AUC	F1	Sen	Spe	PPV	NPV
NSW	1 day	1	0.945	1	1	1	1	1
	1 month	0.984	0.802	0.852	1	0.983	0.742	1
QLD	1 day	1	0.957	1	1	1	1	1
	1 month	0.870	0.768	0.678	0.76	0.894	0.612	0.944
SA	1 day	0.958	0.896	0.909	1	0.947	0.833	1
	1 month	0.923	0.853	0.927	1	0.851	0.864	1
TAS	1 day	0.958	0.965	0.933	0.875	1	1	0.941
	1 month	0.970	0.910	0.977	0.997	0.921	0.958	0.995
VIC	1 day	1	0.962	1	1	1	1	1
	1 month	0.997	0.955	0.9583	1	0.997	0.92	1

Table 4.10: Average performance comparison of short and mid-term LSTM+LISHT for the five different dataset in summer

Time Series	Horizon	ACC	AUC	F1	Sen	Spe	PPV	NPV
NSW	1 day	0.958	0.902	0.956	0.916	1	1	0.923
	1 month	0.997	0.895	0.952	0.909	1	1	0.997
QLD	1 day	1	0.964	1	1	1	1	1
	1 month	0.969	0.847	0.954	0.917	0.997	0.994	0.957
SA	1 day	1	0.940	1	1	1	1	1
	1 month	0.973	0.877	0.551	1	0.973	0.380	1
TAS	1 day	0.916	0.969	0.666	0.666	0.952	0.666	0.952
	1 month	0.972	0.861	0.957	0.987	0.966	0.929	0.994
VIC	1 day	1	0.937	1	1	1	1	1
	1 month	0.986	0.887	0.862	1	0.985	0.758	1

Table 4.11: Average performance comparison of short and mid-term LSTM+LISHT for the five different dataset in autumn

Time Series	Horizon	ACC	AUC	F1	Sen	Spe	PPV	NPV
NSW	1 day	1	0.873	1	1	1	1	1
	1 month	0.995	0.905	0.952	0.952	0.997	0.952	0.997
QLD	1 day	1	0.968	1	1	1	1	1
	1 month	0.998	0.784	0.888	1	0.998	0.8	1
SA	1 day	1	0.935	1	1	1	1	1
	1 month	0.990	0.879	0.968	0.974	0.993	0.962	0.995
TAS	1 day	0.958	0.790	0.8	1	0.954	0.666	1
	1 month	0.968	0.857	0.830	1	0.965	0.709	1
VIC	1 day	1	0.906	1	1	1	1	1
	1 month	0.989	0.869	0.934	0.914	0.996	0.955	0.992

The classification performance of both prediction models was also improved utilizing our proposed methodology. More specifically, both BiLSTM and LSTM+LISHT models were biased in case they were trained with the traditional time-series. In contrast, the trade-off between sensitivity and specificity as well as between positive and negative predictive values of both models was considerably increased in case the models were trained with the first differenced. It is worth noticing that both BiLSTM and LSTM+LISHT models exhibited the highest classification performance in case they were trained with the transformed time series.

Figures 4.2 - 4.3 shows the training and testing data of BiLSTM model for short term and midterm respectively, where for short term training performance evaluated from 0 to

696 hours training and tested for 24 hours. At the same time for short term forecasting 1440 hours used for training and 720 hours for testing. Actual and trained curve showed in the left portion of the graph and right portion displaying the Actual and tested curves. Similarly, in the Figures 4.4 - 4.5 displaying the training and testing data curve for proposed LSTM+LISHT where short term training performance evaluated from 0 to 696 hours training and tested for 24 hours. At the same time for short term forecasting 1440 hours used for training and 720 hours for testing as well.

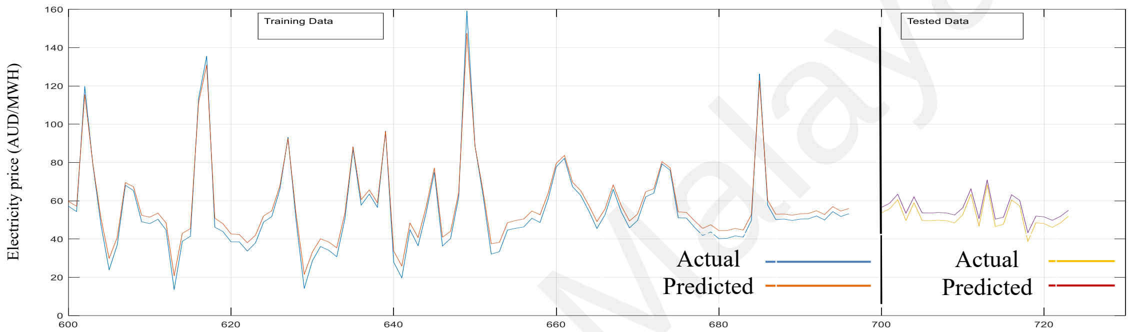


Figure 4.2: Transformed (box-cox) time series training and tested performance for BiLSTM 24 hours

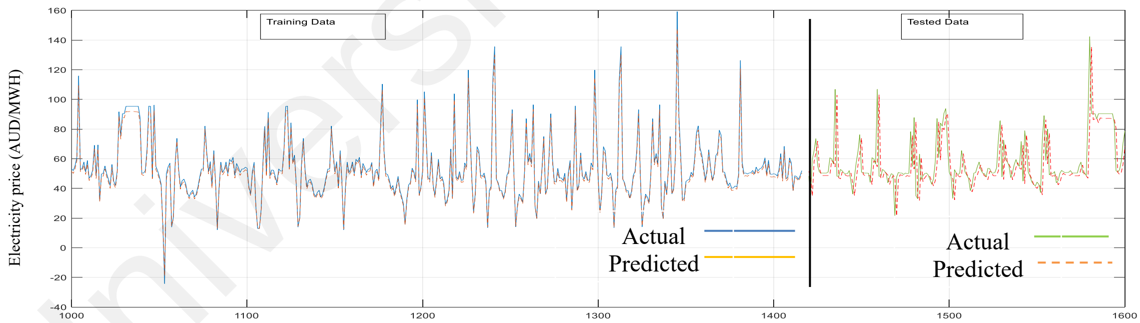


Figure 4.3: Transformed (box-cox) time series training and tested performance for BiLSTM 720 hours

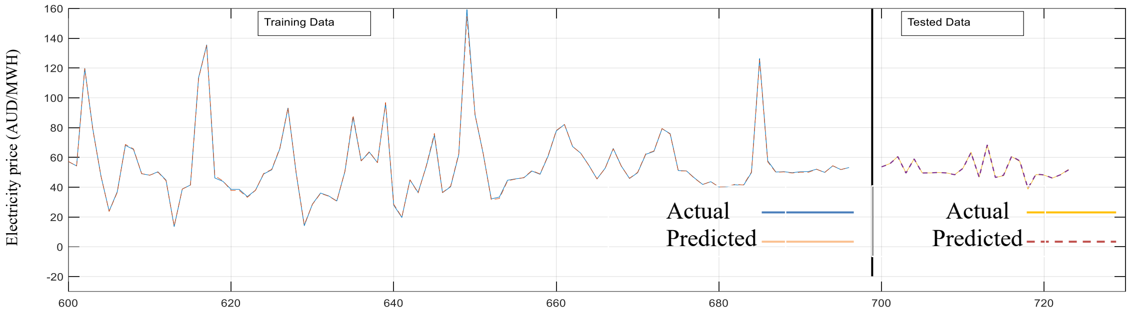


Figure 4.4: Transformed (box-cox) time series Training and Tested performance for LSTM+LISHT 24 hours

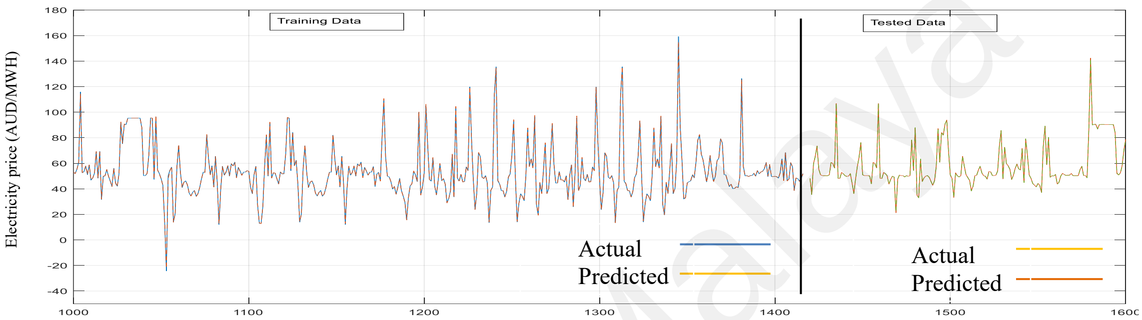


Figure 4.5: Transformed (box-cox) time series Training and Tested performance for LSTM+LISHT 720 hours

Moreover, the LSTM+LISHT model trained with the box-cox transformed time-series reported the best classification performance for all values of window size m and the best regression performance for $m = 24, 720$. The LSTM+LISHT model trained with the time series reported the best performance relative to classification and regression accuracy, respectively. Root mean square error (RMSE) is to evaluate the forecasting precision and ability of the point prediction results and mean average error (MAE) conveys the absolute average forecasting deviation of trains and targets. The formula for RMSE and MAE given below in the Eqs (4.7 – 4.8) where x_t is actual value and y_t predicted value.

$$RMSE = \sqrt{\frac{1}{N}\sum_{t\epsilon N}(x_t - y_t)^2} \tag{4.7}$$

$$MAE = \frac{1}{N}\sum_{t\epsilon N}\frac{|x_t - y_t|}{x_t} \quad (4.8)$$

The proposed deep learning forecasting model (LSTM RNN and LISHT) is compared with four other deep learning models. The simulation results of one day (24 hours) for five different states over various seasons are presented in the following table 4.12 – 4.14. Firstly, for Spring 24 hours testing data presented that the minimum RMSE value in constructed model is 0.9939 and the maximum RMSE value is 2.7418. Besides for Summer and Autumn the minimum RMSE consecutively 0.4431, 0.1700 and maximum RMSE 4.3231, 3.1852 respectively.

Table 4.12: Benchmarking performance of LSTM+LISHT in Spring comparing with other methods of forecasting

Regions	Error	BILSTM	LSTM+GRU	GRU	LSTM	LSTM+LISHT
NSW	RMSE	2.7418	2.1190	1.5455	1.3222	0.9939
	MAE	2.0013	1.6080	1.0110	0.8655	0.6663
QLD	RMSE	1.9158	0.4233	0.3492	0.3430	0.3409
	MAE	1.6612	0.3495	0.2735	0.2784	0.2796
SA	RMSE	1.2140	0.4159	0.4188	0.4136	0.3956
	MAE	0.9332	0.3463	0.2706	0.3713	0.3260
TAS	RMSE	2.2763	0.2995	0.3116	0.3797	0.2613
	MAE	2.1398	0.2237	0.2348	0.3128	0.2082
VIC	RMSE	1.6994	0.5295	0.4455	0.4607	0.3181
	MAE	1.5652	0.4299	0.3344	0.3849	0.2608

Table 4.13: Benchmarking performance of LSTM+LISHT in Summer comparing with other methods of forecasting

Regions	Error	BILSTM	LSTM+GRU	GRU	LSTM	LSTM+LISHT
NSW	RMSE	4.3231	0.8878	0.8254	0.6725	0.4431
	MAE	4.1249	0.7626	0.6944	0.4768	0.3294
QLD	RMSE	5.0952	0.8136	0.7929	0.7014	0.6678
	MAE	4.7455	0.5504	0.4904	0.4690	0.5200
SA	RMSE	2.6170	0.6744	0.3688	0.3288	0.2799
	MAE	2.3100	0.5309	0.2812	0.2329	0.2037
TAS	RMSE	0.2640	0.1572	0.1663	0.1913	0.1264
	MAE	0.1977	0.1345	0.1389	0.1482	0.1005
VIC	RMSE	1.8957	0.7112	0.6130	0.3681	0.2658
	MAE	1.4428	0.5306	0.4997	0.2872	0.2049

Table 4.14: Benchmarking performance of LSTM+LISHT in Autumn comparing with other methods of forecasting

Regions	Error	BILSTM	LSTM+GRU	GRU	LSTM	LSTM+LISHT
NSW	RMSE	3.1852	0.4122	0.2081	0.2430	0.1700
	MAE	2.8768	0.3114	0.1596	0.1762	0.1328
QLD	RMSE	13.1399	1.0516	0.9660	0.7012	0.6369
	MAE	12.8242	0.8189	0.7071	0.5704	0.5518
SA	RMSE	8.7743	0.7941	0.5878	0.8797	0.4911
	MAE	8.2195	0.5888	0.4640	0.7332	0.3966
TAS	RMSE	1.8318	0.5200	0.3491	0.2837	0.2269
	MAE	1.7760	0.3706	0.2443	0.2240	0.1914
VIC	RMSE	1.3571	0.3930	0.2851	0.1837	0.1721
	MAE	1.1868	0.3004	0.2196	0.1382	0.1449

As BiLSTM and LSTM+GRU are deterministic models, the RMSE value of BiLSTM is 3.4167 average and the RMSE value of LSTM+GRU is 1.13 average from all the three seasons and all five states of Australia. The calculation of average RMSE for each of the tested model showed below.

$$Average\ RMSE = \frac{AvgRMSE(Spring) + AvgRMSE(Summer) + AvgRMSE(Autumn)}{Number\ of\ seasons\ observed\ (3)}$$

The average RMSE value of GRU is 0.8596 and the average RMSE value of LSTM is 0.7459. The average RMSE value of constructed LSTM+LISHT model is 0.5356, obtained from all the three seasons.

The proposed model compared with LSTM, GRU, LSTM+GRU and BILSTM, where the average RMSE of the individual state calculated from the formula below,

$$Average\ RMSE = \frac{RMSE(Spring) + RMSE(Summer) + RMSE(Autumn)}{Number\ of\ seasons\ observed\ (3)}$$

The average RMSE improved respectively by 24.83%, 35.69%, 53.1%, 63.75% in NSW, 3.36%, 14.52%, 16.69%, 86.7% improved respectively in QLD, 4.35%, 5.54%, 4.88%, 67.41% improved respectively in SA, 31.18%, 16.14%, 12.75%, 88.52%

improved respectively in TAS and 30.95%, 28.6%, 39.92%, 81.28% improved in VIC respectively. From Table 4.12 - 4.14 shows that the proposed model has the minimum RMSE on the contrast of other methods of deep learning itself. Some other methods Minimum RMSE displayed in the following Table 4.15, where proposed method improved 60% from CNN-LSTM, 96.16% from Gated-FCN, 98.55% from BiGRU, 98.96% from ARIMA and 99.11% from SMA model.

Table 4.15: Comparison Chart for Minimum RMSE

Error	RMSE	Improvements
SMA(Naz et al., 2021)	13.42	99.11%
ARIMA(Naz et al., 2021)	11.45	98.98%
BiGRU(Naz et al., 2021)	8.23	98.55%
Gated-FCN(Naz et al., 2021)	3.12	96.16%
CNN-LSTM(Kim & Cho, 2019)	0.30	60%
The proposed LSTM+ LiSHT	0.12	

4.4 Experimental results for proposed method 2

In this research, electricity demand and price data were obtained from Australian Energy Market Operator (AEMO) from August 2020 to May 2021 to develop the proposed mid-term EPF framework. Test dataset includes the hourly data from January 2021 to May 2021. In the meantime, the training dataset is arranged according to the train and test dataset. The training dataset includes 5 prior months to the forecasting weeks. This work proposed two types of forecasting: 1-week forecasting and 2 weeks forecasting. Table 4.16 briefly shows the sample arrangement of training dataset to forecast electricity price for the month of January and February. The arrangement is modified accordingly to forecast the months of March, April and May 2021.

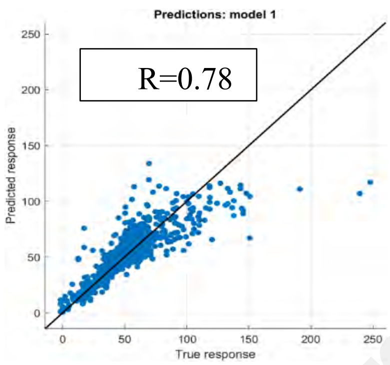
Table 4.16: Examples of training data arrangement for the EPF

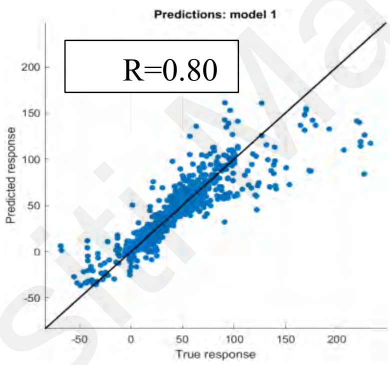
Training dataset	**Testing dataset**	
5 months data	1 week forecasting	2 weeks forecasting
Week 1, Aug 2020 – Week 4, Dec 2020	Week 1, Jan 2021	Week 1 – 2, Jan 2021
Week 3, Aug 2020 – Week 2, Jan 2021	Week 3, Jan 2021	Week 3- 4, Jan 2021

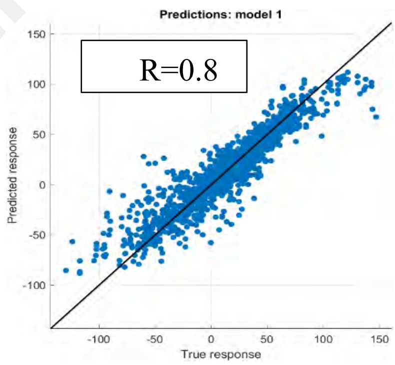
It is argued that the data points used for developing the forecasting model should be strongly correlated with each other. Hence, the correlation coefficient R of the actual and predicted output of the model is computed to assess the feasibility of implementing BTE model. The purpose of analysing regression model is to extract significant relationships between the forecast variable of interest and the predictor variables. A perfect forecasting modelling will produce a correlation coefficient R value of 1. Figures 4.6 (a) - 4.6 (e) showed a regression value, R of 0.78, 0.80, 0.88, 0.76, and 0.87 for Australia's five economic states (NSW, QLD, SA, TAS, VIC) respectively when applying BTE model. As can be seen, the regression value obtained when implementing conventional BTE model is in the range between 0.76 to 0.87 which is inadequate in forecasting complex time series data. Hence, in order to improve the accuracy of the forecasting model the data was further transferred to RNN which incorporates GRU for further optimization. The following Table 4.16 displaying the R values for BTE and BTE+GEU respectively.

Table 4.17: Comparison of regression correlation coefficient between two different models

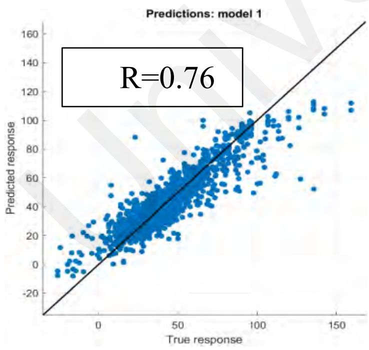
States of Australia	R (BTE model)	R (Proposed BTE+GRU model)
NSW	0.78	0.99
QLD	0.80	0.99
SA	0.88	0.98
TAS	0.76	0.99
VIC	0.87	0.99

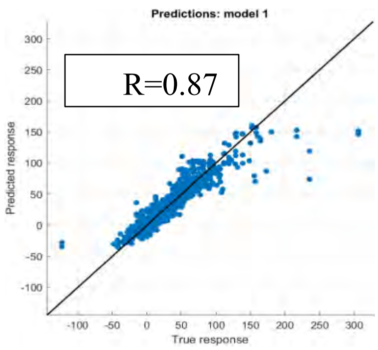






(a) NSW BTE testing model **(b) QLD BTE testing model** **(c) SA BTE testing model**





(d) TAS BTE testing model **(e) VIC BTE testing model**

Figure 4.6: Correlation coefficient R for BTE algorithm

The two weeks forecasting results are compared in Figure 4.7. After applying the proposed BTE and GRU model, the correlation coefficient of R for NSW, QLD, SA, TAS, VIC has improved significantly to 0.9961, 0.9995, 0.9800, 0.9996, 0.9996 respectively (Table 4.17). This shows that the proposed forecasting model achieved a correlation coefficient, R approximately 1 which means that the proposed forecasting model manage to correlate the data points better as compared to BTE model in Figures 4.6 (a - e). Thus, high value of R contributed to better performance in (MAPE) and (RMSE). In this work, the accuracy of the proposed point forecasting model is evaluated by computing the MAPE and RMSE. While MAPE conveys the absolute average forecasting deviation of trains and targets. The MAPE formula given below in the Eq (4.7), where R_t is the actual time series data and P_t is the forecasted data:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{R_t - P_t}{R_t}\right| \tag{4.7}$$

As can be seen from Figure 4.7, the proposed BTE+GRU produced the smallest value of RMSE and MAPE values compared to other methods such as BiLSTM, LSTM+GRU, and LSTM for all the five states NSW, QLD, SA, TAS, VIC. It can be concluded that the most effective method in forecasting the electricity price in this work is the proposed BTE+GRU model where BTE and GRU are incorporated in the RNN architecture. Meanwhile, Table 4.18 tabulated the average performance evaluation of the proposed BTE+GRU method for 1 week and 2 weeks forecasting. The results show that the RMSE and MAPE values are about the same for both types of forecasting interval which means that the forecasting model is feasible to solve 1 week and 2 weeks forecasting problem. Eventually, accurate information on the electricity price forecasting will contribute to effective management in the deregulated electricity market, complex renewable energy and emission policy objectives.

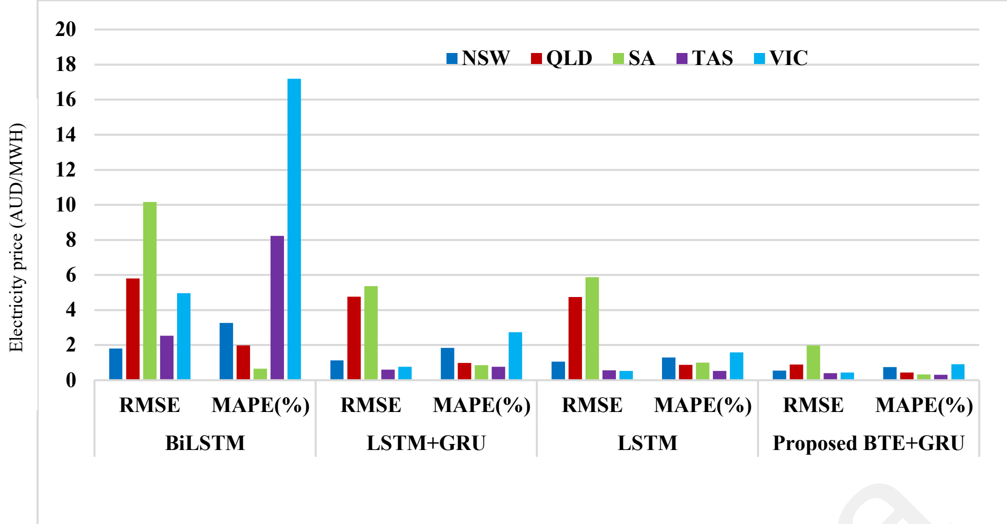


Figure 4.7: The RMSE and MAPE for EFP using several deep learning methods

Table 4.18: Average performance evaluation of the proposed BTE+GRU method

Region	1 week forecasting		2 weeks forecasting	
	RMSE	MAPE	RMSE	MAPE
NSW	0.294784	0.671307	0.552075	0.745342
QLD	0.513686	0.428665	0.895746	0.441092
SA	2.263616	0.36288	1.992935	0.326579
TAS	0.404107	0.300019	0.409577	0.318366
VIC	0.308608	0.895352	0.430335	0.905478

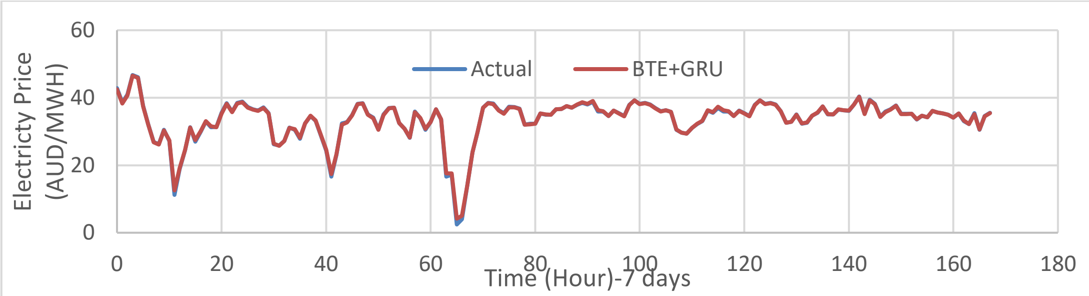
As tabulated in Table 4.19, the proposed model is benchmarked with several methods to measure the effectiveness of the proposed BTE and GRU model. As can be seen, the proposed BTE and GRU model produced the lowest mean RMSE and mean MAPE values as compared to other methods which are 0.36 and 0.55 respectively. The work in (Karabiber & Xydis, 2019) adopted machine learning approach with Trend and Seasonal Components (TBATS) which adopted trigonometric technique supports forecasting of daily seasonality by applying maximum likelihood estimation. However, TBATS method does not permit the adoption of external regressors. The computation of TBATS+ANN, ANN+ARIMA, TBATS+ARIMA, TBATS+ANN+ARIMA methods were reported in (Karabiber & Xydis, 2019) by using Denmark electricity market. It can be seen that the average RMSE for the four methods applied in (Karabiber & Xydis, 2019) is significantly high compared to methods applied in this work that adopted deep learning methods such as LSTM, LSTM+GRU, BiLSTM and the proposed model. This justifies the importance of adopting deep learning method in developing an accurate forecasting model.

Table 4.19: Performance evaluation of the proposed method and other methods

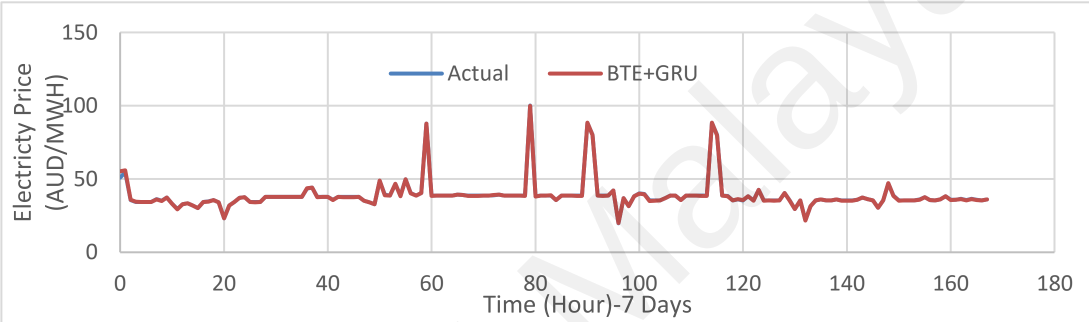
Model	RMSE				MAPE			
	Mean	Min	Max	Std Dev	Mean	Min	Max	Std Dev
The proposed BTE+GRU	0.36	0.12	0.96	0.27	0.55	0.17	1.684	0.39
LSTM	2.55	0.18	18.21	5.09	1.06	1.00	2.81	0.96
LSTM+GRU	2.52	0.33	16.32	5.24	1.43	0.48	5.05	1.28
BiLSTM	5.05	0.84	16.53	4.37	8.27	8.64	59.34	14.57
TBATS+ANN(Karabiber & Xydis, 2019)	40.21	8.88	174.2	25.71	33.46	7.53	136.3	21.5
ANN+ARIMA(Karabiber & Xydis, 2019)	38.05	8.07	168.8	24.06	31.92	5.90	165.6	21.14
TBATS+ARIMA (Karabiber & Xydis, 2019)	37.5	9.94	176.4	26.41	31.11	8.10	162.7	21.86
TBATS+ANN+ARIMA (Karabiber & Xydis, 2019)	36.44	8.04	164.2	24.34	30.06	6.63	148.2	20.34

Figures 4.8 (a) and 4.8 (b) show the examples of 1-week forecasting results while Figures 4.8 (c) and 4.8 (d). show the examples of 2-week forecasting results when using the proposed model for two different states in Australia. The electricity price fluctuations for different states of Australia differ due to different demand, supply and energy resources. The results justified that the proposed forecasting model can generate comparatively accurate forecasting results and the deviation between the curve of the proposed BTE+GRU model and the actual load curve is considered the lowest compared

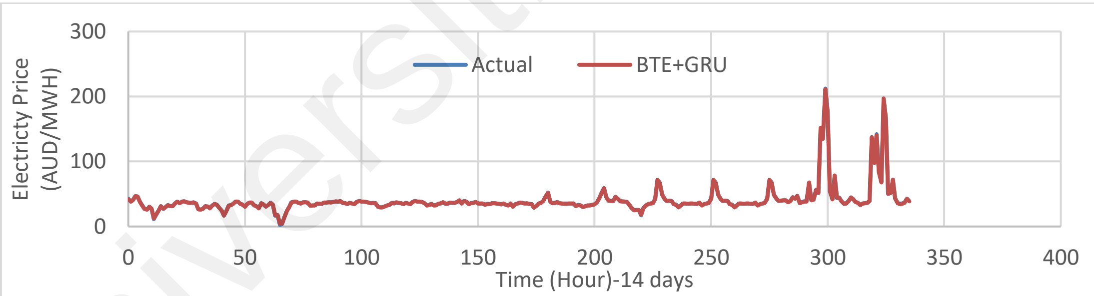
to other methods as tabulated in Table 4.18.



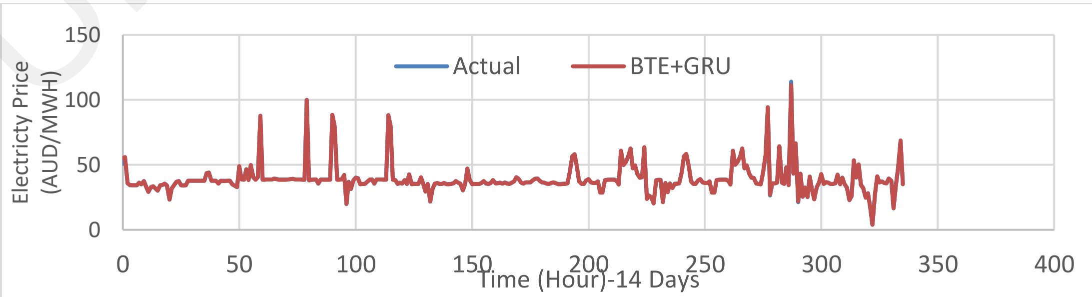
(a) One week forecasting model for NSW



(b) One week forecasting model for TAS



(c) Two weeks forecasting model for NSW



(d) Two weeks forecasting model for TAS

.

Figure 4.8: One and two week forecasting of NSW and TAS

Figures 4.9 - 4.11 presents the 2 weeks forecasting results from the month of January to May 2021 in an hourly basis for three economical states in Australia such as New South Wales, Tasmania and Victoria. At most points, the conventional BTE model was not able to forecast the spikes which justifies the inadequacy of implementing conventional BTE method in EPF. Despite the complex nonlinearity in the trend of electricity price, the proposed model which incorporated BTE and GRU managed to forecast the spikes more accurately and seems to fit the actual data to a satisfactory degree. Hence, this justified the contribution of the proposed model in solving mid-term EPF problem.

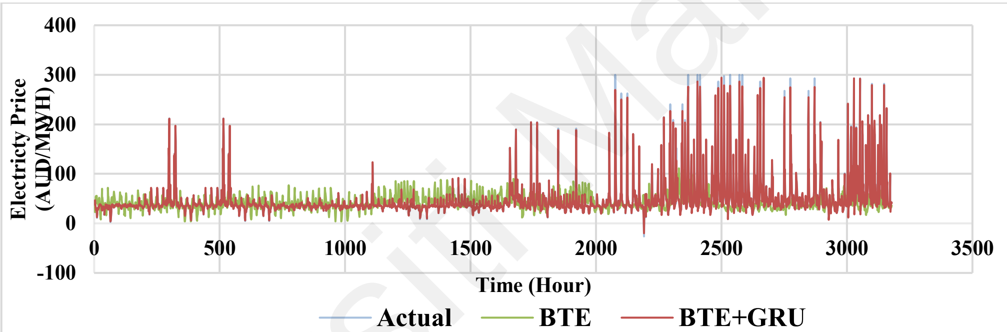


Figure 4.9: Forecasting model comparison for NSW

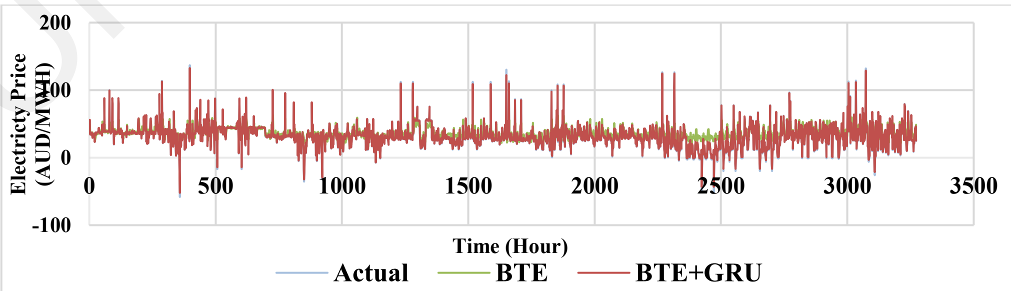


Figure 4.10: Forecasting model comparison for TAS

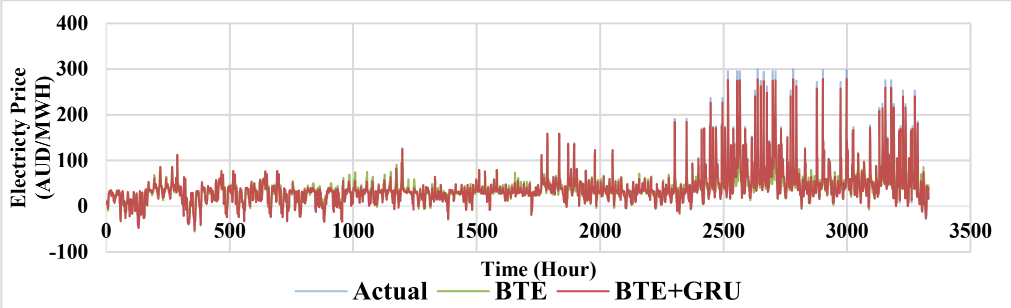


Figure 4.11: Forecasting model comparison for VIC

Moreover, the MAPE of the proposed forecasting model is benchmarked with previous works that adopted different electricity market. It can be summarized that the proposed forecasting model for Australian electricity market is feasible with MAPE value in the range of 0.17% to 1.68% as compared to previous works with MAPE values of approximately 11.74% in (Yan et al., 2016), 6.24% in (Maciejowska & Weron, 2015), 3.27% to 5.14% in (Hu et al., 2020), 1.9% in (Heydari et al., 2018), 2.4% in (Zhou et al., 2019). The MAPE value was computed by averaging the MAPE values for the five states of Australia that are focused in this research. This suggests that the proposed forecasting model is feasible for multi-regional mid-term electricity price forecasting.

4.5 Summary

In this work, it has been established theoretically and experimentally that the time-series data are ‘‘appropriate’’ for developing a deep learning model, which is the main contribution of this study. In another way it can be said that this research developed a new framework which can find out an effective time series for training a deep learning model, which can perform for a stable and reliable forecasting model. The term "appropriate" denotes that the time-series data has satisfied the stated scientific requirements and is adequate for training a forecasting model. If, on the other hand, the series fails to meet the desired criteria, it is deemed "unsuitable," and any attempts to develop a solid prediction model would most likely be useless. Therefore, this research is a beginning point for construction of any prediction methodology for various time series forecasting.

If the starting dataset are unstable or non-stationary the work done for developing a forecasting model could waste. It's a positive signal for machine learning researcher to devote intellectual effort in constructing a forecasting framework. Furthermore, this research developed an innovative and comprehensive framework that allows any unstable time-series to be transformed to stable by conducting a transformation focused on boxcox method. However, this approach is familiar, but it is not proved where the best utilization of the technique is. Many methodologies relied on a "trial and error" logic, which is ineffective and inefficient, particular in circumstances where expensive and time-consuming real-world initiatives are aimed at developing precise and trustworthy prediction model. In this research it's shown that these equations successfully eliminated these "unsuitable" data, avoiding the costly and time-consuming "trial and error" method.

It's noticeable that one of the most interesting properties of our suggested framework is this may simply be modified to encompass a broader scientific domain of time-series forecasting operations without requiring any further adjustments or limits. More specifically, the recommended method uses statistic and economic tests to conduct an optimal pre-processing phase for utilising the internal structure of the timeseries. Finally, it is seen that while deep learning models are well accepted for time series, the proposed framework significantly enhance the performance. However, more study is being done to see which from these approaches may be implemented more effectively a priori based on the properties for every time-series in order to get improved forecasting performance. For accomplishing that a prerequisite diagnosis and the appropriate time transformation methodology, a complex pre-processing framework refers to the inherent time-series particular traits such as stationarity, heteroskedasticity, seasonal cycles, and shifting variance could be used.

CHAPTER 5: CONCLUSIONS AND FUTURE WORKS

5.1 Conclusions

Time-series prediction and analytics is universally perceived as the most difficult data mining challenges. Most time-series prediction techniques in the research intend to use machine learning and deep learning methods in order to improve efficacy over established or existing methods. Eventually, a time series data analysis has done successfully in this work and improved deep learning methods have been applied to develop an effective and trustworthy deep learning forecasting model for short term and midterm electricity price forecasting.

The developed forecasting model consists of pre-processed and post trained data analysis in terms of statistical reliability of time series. An augmented dickey fuller test has performed to check the presence of stationarity and nonstationary data of the time series electricity data before training. Additionally, while training is completed with deep learning module the residual calculated and check the autocorrelation of the residuals. So, it can be said that the autocorrelation of the residuals has been evaluated successfully to make the data prepare for deep learning approach. In our study the time series data found stationary and good to fit with the deep learning model, but we have found some autocorrelation in residuals which is fixed by transforming data through box-cox transformation technique.

Significant experiment has done with time series electricity data from Australia's five most vital economic zones in proposed machine learning and deep learning techniques. The models were tested on their capacity to anticipate time-series pricing and the accuracy of their predictions. Short-term Forecasting: improved the mean RMSE by 60% - 99.11%. Mid-term forecasting improved the mean RMSE by 86.15% - 97.64%In terms of binary classification and regression analysis, where the minimum RMSE is 0.12 among the three seasons of five different economic states of Australia, the suggested deep learning model with LSTM+LISHT performed remarkably well. For midterm forecasting a BTE method has been utilized which is optimized by gated recurrent unit (GRU) for one and two-week prediction. The prediction accuracy of midterm forecasting is very significant such as average RMSE 0.36 and average MAPE 0.55 which is mentioned in the table 4.18. Overall, it can be claimed that the research has found an improved forecasting performance in the area of deep learning as well as machine learning. There have been some statistical and machine learning technique which has performed better and set a benchmark along with those our forecasting method showed an immense significance and ensures the forecasting accuracy of the deep learning model's which set a new benchmark for the electricity price forecasting technique. This research indicates that the suggested technique significantly increased the accuracy and dependability of a deep learning model's forecasting performance for short and midterm electricity forecasting. It can be claimed that the proposed method can employ for any deep learning framework, further optimized and reconfigured deep learning methods can have performed more better.

5.2 Future work

This will be used in future studies to ensure that the proposed framework is compatible with any regression algorithm. Other field of future research will be to compare our empirical strategy for new profit and return-based performance measurements. The proposed methodologies will be used in other time series forecasting and long-term forecasting need to explore. Finally, it's a fascinating concept to use our proposed framework to anticipate anomaly detection in order to "catch" exceptions or other unusual data that might indicate predicting fragility.

REFERENCES

Abedinia, O., Amjady, N., Shafie-Khah, M., & Catalão, J. P. (2015). Electricity price forecast using combinatorial neural network trained by a new stochastic search method. *Energy Conversion and Management, 105*, 642-654.

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon, 4*(11), e00938.

Alazab, M., Khan, S., Krishnan, S. S. R., Pham, Q.-V., Reddy, M. P. K., & Gadekallu, T. R. (2020). A multidirectional LSTM model for predicting the stability of a smart grid. *IEEE Access, 8*, 85454-85463.

Almeshaiei, E., & Soltan, H. (2011). A methodology for electric power load forecasting. *Alexandria Engineering Journal, 50*(2), 137-144.

Amjady, N. (2001). Short-term hourly load forecasting using time-series modeling with peak load estimation capability. *IEEE Transactions on Power Systems, 16*(3), 498-505.

Andruszkiewicz, J., Lorenc, J., & Weychan, A. (2019). Demand price elasticity of residential electricity consumers with zonal tariff settlement based on their load profiles. *energies, 12*(22), 4317.

Australian Energy Market Operator. (2022). from https://aemo.com.au/energy-systems/electricity/national-electricity-market-nem/data-nem/data-dashboard-nem

Bedi, J., & Toshniwal, D. (2019). Deep learning framework to forecast electricity demand. *Applied Energy, 238*, 1312-1326.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks, 5*(2), 157-166.

Bian, X.-Q., Han, B., Du, Z.-M., Jaubert, J.-N., & Li, M.-J. (2016). Integrating support vector regression with genetic algorithm for CO2-oil minimum miscibility pressure (MMP) in pure and impure CO2 streams. *Fuel, 182*, 550-557.

Borovkova, S., & Tsiamas, I. (2019). An ensemble of LSTM neural networks for high-frequency stock market classification. *Journal of Forecasting, 38*(6), 600-619.

Brockwell, P. J., Brockwell, P. J., Davis, R. A., & Davis, R. A. (2016). *Introduction to time series and forecasting*: Springer.

Bunn, D. W. (2000). Forecasting loads and prices in competitive power markets. *Proceedings of the IEEE, 88*(2), 163-169.

Chatfield, C. (2000). *Time-series forecasting*: CRC press.

Chenyu, D., Ruihua, Y., & Yuandong, L. (2018). *ARL Study of Second Order Autocorrelation Residual Control Chart and its Application in MAP.* Paper presented at the 2018 37th Chinese Control Conference (CCC).

Ciarreta, A., Muniain, P., & Zarraga, A. (2017). Modeling and forecasting realized volatility in German–Austrian continuous intraday electricity prices. *Journal of Forecasting, 36*(6), 680-690.

Dhillon, J., Rahman, S. A., Ahmad, S. U., & Hossain, M. J. (2016). *Peak electricity load forecasting using online support vector regression.* Paper presented at the 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE).

Diongue, A. K., Guegan, D., & Vignal, B. (2009). Forecasting electricity spot market prices with a k-factor GIGARCH process. *Applied Energy, 86*(4), 505-510.

Du, J., Vong, C.-M., & Chen, C. P. (2020). Novel efficient RNN and LSTM-like architectures: Recurrent and gated broad learning systems and their applications for text classification. *IEEE transactions on cybernetics*.

Du, M., Zhao, Y., Liu, C., & Zhu, Z. (2021). Lifecycle cost forecast of 110 kV power transformers based on support vector regression and gray wolf optimization. *Alexandria Engineering Journal, 60*(6), 5393-5399.

Elfahham, Y. (2019). Estimation and prediction of construction cost index using neural networks, time series, and regression. *Alexandria Engineering Journal, 58*(2), 499-506.

Elfwing, S., Uchibe, E., & Doya, K. (2018). Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks, 107*, 3-11.

Eydeland, A., & Wolyniec, K. (2002). *Energy and power risk management: New developments in modeling, pricing, and hedging* (Vol. 97): John Wiley & Sons.

Fan, C., Sun, Y., Zhao, Y., Song, M., & Wang, J. (2019). Deep learning-based feature engineering methods for improved building energy prediction. *Applied Energy, 240*, 35-45.

Fan, G.-F., Wei, X., Li, Y.-T., & Hong, W.-C. (2020). Forecasting electricity consumption using a novel hybrid model. *Sustainable Cities and Society, 61*, 102320.

Gocheva-Ilieva, S. G., Voynikova, D. S., Stoimenova, M. P., Ivanov, A. V., & Iliev, I. P. (2019). Regression trees modeling of time series for air pollution analysis and forecasting. *Neural Computing and Applications, 31*(12), 9023-9039.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems, 28*(10), 2222-2232.

Hamilton, G., Abeygunawardana, A., Jovanović, D. P., & Ledwich, G. F. (2018). *Hybrid model for very short-term electricity price forecasting.* Paper presented at the 2018 IEEE Power & Energy Society General Meeting (PESGM).

Haq, M. R., & Ni, Z. (2019). A new hybrid model for short-term electricity load forecasting. *IEEE Access, 7*, 125413-125423.

Hernandez, L., Baladron, C., Aguiar, J. M., Carro, B., Sanchez-Esguevillas, A. J., Lloret, J., & Massana, J. (2014). A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings. *IEEE Communications Surveys & Tutorials, 16*(3), 1460-1495.

Heydari, A., Keynia, F., Garcia, D. A., & De Santoli, L. (2018). *Mid-term load power forecasting considering environment emission using a hybrid intelligent approach.* Paper presented at the 2018 5th International Symposium on Environment-Friendly Energies and Applications (EFEA).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735-1780.

Hong, Y.-Y., Taylar, J. V., & Fajardo, A. C. (2020). Locational Marginal Price Forecasting Using Deep Learning Network Optimized by Mapping-Based Genetic Algorithm. *IEEE Access, 8*, 91975-91988.

Hu, C., Yuan, H., Zhang, J. J., Yan, J., Gao, T., Zhao, H., . . . Zhou, P. (2020). *Mid-Long Term Electricity Consumption Forecasting Analysis Based on Cyber-Physical-Social System Architecture.* Paper presented at the 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE).

Huang, X., Hong, S. H., & Li, Y. (2017). Hour-ahead price based energy management scheme for industrial facilities. *IEEE Transactions on Industrial Informatics, 13*(6), 2886-2898.

Kabaila, P., Farchione, D., Alhelli, S., & Bragg, N. (2021). The effect of a Durbin–Watson pretest on confidence intervals in regression. *Statistica Neerlandica, 75*(1), 4-23.

Kankal, M., Akpınar, A., Kömürcü, M. İ., & Özşahin, T. Ş. (2011). Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables. *Applied Energy, 88*(5), 1927-1939.

Karabiber, O. A., & Xydis, G. (2019). Electricity price forecasting in the Danish day-ahead market using the TBATS, ANN and ARIMA methods. *energies, 12*(5), 928.

Keles, D., Scelle, J., Paraschiv, F., & Fichtner, W. (2016). Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. *Applied Energy, 162*, 218-230.

Kim, T.-Y., & Cho, S.-B. (2019). Predicting residential energy consumption using CNN-LSTM neural networks. *Energy, 182*, 72-81.

Kraft, E., Keles, D., & Fichtner, W. (2020). Modeling of frequency containment reserve prices with econometrics and artificial intelligence. *Journal of Forecasting, 39*(8), 1179-1197.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature, 521*(7553), 436-444.

Li, X., Wu, P., & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. *Information Processing & Management, 57*(5), 102212.

Lillard, L. A., & Acton, J. P. (1981). Seasonal electricity demand and pricing analysis with a variable response model. *The Bell Journal of Economics*, 71-92.

Liu, S., Zhang, C., & Ma, J. (2017). *CNN-LSTM neural network model for quantitative strategy analysis in stock markets.* Paper presented at the international conference on neural information processing.

Livieris, I. E., Pintelas, E., & Pintelas, P. (2020). A CNN–LSTM model for gold price time-series forecasting. *Neural Computing and Applications, 32*(23), 17351-17360.

Livieris, I. E., Stavroyiannis, S., Pintelas, E., & Pintelas, P. (2020). A novel validation framework to enhance deep learning models in time-series forecasting. *Neural Computing and Applications, 32*(23), 17149-17167.

Lv, P., Liu, S., Yu, W., Zheng, S., & Lv, J. (2020). EGA-STLF: A hybrid short-term load forecasting model. *IEEE Access, 8*, 31742-31752.

Maciejowska, K., & Weron, R. (2015). Short-and mid-term forecasting of baseload electricity prices in the UK: The impact of intra-day price relationships and market fundamentals. *IEEE transactions on power systems, 31*(2), 994-1005.

Manner, H., Fard, F. A., Pourkhanali, A., & Tafakori, L. (2019). Forecasting the joint distribution of Australian electricity prices using dynamic vine copulae. *Energy Economics, 78*, 143-164.

Manotumruksa, J., Macdonald, C., & Ounis, I. (2020). A Contextual Recurrent Collaborative Filtering framework for modelling sequences of venue checkins. *Information Processing & Management, 57*(6), 102092.

Muglia, C., Santabarbara, L., & Grassi, S. (2019). Is Bitcoin a Relevant Predictor of Standard & Poor's 500? *Journal of Risk and Financial Management, 12*(2), 93.

Mulder, M., & Scholtens, B. (2013). The impact of renewable energy on electricity prices in the Netherlands. *Renewable energy, 57*, 94-100.

Naz, A., Javaid, N., Asif, M., Javed, M. U., Ahmed, A., Gulfam, S. M., . . . Choi, J.-G. (2021). Electricity Consumption Forecasting Using Gated-FCN With Ensemble Strategy. *IEEE Access, 9*, 131365-131381.

Nguyen, V.-B., Duong, M.-T., & Le, M.-H. (2020). *Electricity Demand Forecasting for Smart Grid Based on Deep Learning Approach.* Paper presented at the 2020 5th International Conference on Green Technology and Sustainable Development (GTSD).

Osborne, J. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research, and Evaluation, 15*(1), 12.

Ozer, I., Efe, S. B., & Ozbay, H. (2021). A combined deep learning application for short term load forecasting. *Alexandria Engineering Journal, 60*(4), 3807-3818.

Pal, A., & Prakash, P. (2017). *Practical time series analysis: master time series data processing, visualization, and modeling using python*: Packt Publishing Ltd.

Palomino, K., Reyes, F., Núñez, J., Valencia, G., & Herrera Acosta, R. (2020). Wind speed prediction based on univariate ARIMA and OLS on the Colombian Caribbean Coast.

Pan, M., Zhou, H., Cao, J., Liu, Y., Hao, J., Li, S., & Chen, C.-H. (2020). Water level prediction model based on GRU and CNN. *IEEE Access, 8*, 60090-60100.

Panapakidis, I. P., & Dagoumas, A. S. (2016). Day-ahead electricity price forecasting via the application of artificial neural network based models. *Applied Energy, 172*, 132-151.

Pappas, S. S., Ekonomou, L., Karamousantas, D. C., Chatzarakis, G., Katsikas, S., & Liatsis, P. (2008). Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models. *Energy, 33*(9), 1353-1360.

Pourdaryaei, A., Mokhlis, H., Illias, H. A., Kaboli, S. H. A., & Ahmad, S. (2019). Short-term electricity price forecasting via hybrid backtracking search algorithm and ANFIS approach. *IEEE Access, 7*, 77674-77691.

Pourdaryaei, A., Mokhlis, H., Illias, H. A., Kaboli, S. H. A., Ahmad, S., & Ang, S. P. (2019). Hybrid ANN and artificial cooperative search algorithm to forecast short-term electricity price in de-regulated electricity market. *IEEE Access, 7*, 125369-125386.

Raj, R. J. S., Babu, S. A., VL, H. J., Varalatchoumy, M., & Kathirvel, C. (2022). *Implementing Multiclass Classification to find the Optimal Machine Learning Model for Forecasting Malicious URLs.* Paper presented at the 2022 6th International Conference on Computing Methodologies and Communication (ICCMC).

Raza, M. Q., & Khosravi, A. (2015). A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews, 50*, 1352-1372.

Rodriguez, C. P., & Anders, G. J. (2004). Energy price forecasting in the Ontario competitive power system market. *IEEE transactions on power systems, 19*(1), 366-374.

Shakiba, F. M., & Zhou, M. (2020). Novel Analog Implementation of a Hyperbolic Tangent Neuron in Artificial Neural Networks. *IEEE Transactions on Industrial Electronics*.

Shaman, P. (2010). Generalized Levinson–Durbin sequences, binomial coefficients and autoregressive estimation. *Journal of multivariate analysis, 101*(5), 1263-1273.

Shuai, Y., Song, T., & Wang, J. (2017). Integrated parallel forecasting model based on modified fuzzy time series and SVM. *Journal of Systems Engineering and Electronics, 28*(4), 766-775.

Siegelmann, H. T. (1995). Recurrent neural networks. *Computer Science Today*, 29-45.

Singh, N., Mohanty, S. R., & Shukla, R. D. (2017). Short term electricity price forecast based on environmentally adapted generalized neuron. *Energy, 125*, 127-139.

Stavroyiannis, S. (2019). Can Bitcoin diversify significantly a portfolio? *International Journal of Economics and Business Research, 18*(4), 399-411.

Tokgöz, A., & Ünal, G. (2018). *A RNN based time series approach for forecasting turkish electricity load.* Paper presented at the 2018 26th Signal Processing and Communications Applications Conference (SIU).

Ugurlu, U., Oksuz, I., & Tas, O. (2018). Electricity price forecasting using recurrent neural networks. *energies, 11*(5), 1255.

Ugurlu, U., Tas, O., Kaya, A., & Oksuz, I. (2018). The financial effect of the electricity price forecasts' inaccuracy on a hydro-based generation company. *energies, 11*(8), 2093.

Urtnasan, E., Park, J.-U., & Lee, K.-J. (2020). Automatic detection of sleep-disordered breathing events using recurrent neural networks from an electrocardiogram signal. *Neural Computing and Applications, 32*(9), 4733-4742.

Vantuch, T., Mišák, S., & Stuchlý, J. (2016). *Power quality prediction designed as binary classification in AC coupling Off-Grid system.* Paper presented at the 2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC).

Wang, H., Liu, Y., Zhou, B., Li, C., Cao, G., Voropai, N., & Barakhtenko, E. (2020). Taxonomy research of artificial intelligence for deterministic solar power forecasting. *Energy Conversion and Management, 214*, 112909.

Wang, H., Xu, Q., & Zhou, L. (2015). Large unbalanced credit scoring using lasso-logistic regression ensemble. *PloS one, 10*(2), e0117844.

Wang, M., Wang, Z., Lu, J., Lin, J., & Wang, Z. (2019). E-lstm: An efficient hardware architecture for long short-term memory. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 9*(2), 280-291.

Weron, R. (2007). *Modeling and forecasting electricity loads and prices: A statistical approach* (Vol. 403): John Wiley & Sons.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International journal of forecasting, 30*(4), 1030-1081.

Westland, J. C., Mou, J., & Yin, D. (2019). Demand cycles and market segmentation in bicycle sharing. *Information Processing & Management, 56*(4), 1592-1604.

Wu, F., Cattani, C., Song, W., & Zio, E. (2020). Fractional ARIMA with an improved cuckoo search optimization for the efficient Short-term power load forecasting. *Alexandria Engineering Journal, 59*(5), 3111-3118.

Yan, X., & Chowdhury, N. A. (2013). Mid-term electricity market clearing price forecasting: A hybrid LSSVM and ARMAX approach. *International Journal of Electrical Power & Energy Systems, 53*, 20-26.

Yan, X., Song, Y., & Chowdhury, N. A. (2016). *Performance evaluation of single SVM and LSSVM based forecasting models using price zones analysis.* Paper presented at the 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC).

Yin, J., Xiao, P., Li, J., Liu, Y., Yan, C., & Zhang, Y. (2020). Parameters analysis of sample entropy, permutation entropy and permutation ratio entropy for RR interval time series. *Information Processing & Management, 57*(5), 102283.

Yuan, C., Liu, S., & Fang, Z. (2016). Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM (1, 1) model. *Energy, 100*, 384-390.

Zhang, R., Li, G., & Ma, Z. (2020). A Deep Learning Based Hybrid Framework for Day-Ahead Electricity Price Forecasting. *IEEE Access, 8*, 143423-143436.

Zhao, J. H., Dong, Z. Y., Li, X., & Wong, K. P. (2007). A framework for electricity price spike analysis with advanced data mining methods. *IEEE Transactions on Power Systems, 22*(1), 376-385.

Zhou, S., Zhou, L., Mao, M., Tai, H.-M., & Wan, Y. (2019). An optimized heterogeneous structure LSTM network for electricity price forecasting. *IEEE Access, 7*, 108161-108173.

Zimmermann, H.-G., Neuneier, R., & Grothmann, R. (2001). Multi-agent modeling of multiple FX-markets by neural networks. *IEEE transactions on neural networks, 12*(4), 735-743.