Efficient Use of Deep Learning and Machine Learning for Load Forecasting in South African Power Distribution Networks

Sibonelo C. Motepe
201512867

Thesis submitted in fulfilment of the requirements for the degree *Doctorate of Electrical and Electronic Engineering* at the Faculty of Engineering at the

University of Johannesburg, Auckland Park Campus

Supervisor: Dr Ali N. Hasan

Co-supervisors: Prof. Bhekisipho Twala

Prof. Riaan Stopforth

June 2019

To my son, Kanalelo Letlotlo Motepe, you were born to be Great. Never let anyone tell you that you are not capable of doing something.

Abstract

Title: Efficient Use of Deep Learning and Machine Learning for Load Forecasting in South African Power Distribution Networks

Author: Sibonelo Cobry Motepe

Promoter: Dr Ali Hasan

School: Electrical Engineering Science

Faculty: Engineering and Built Environment

Degree: Doctorate of Engineering (Electrical and Electronic Engineering)

Keywords: Load forecasting, deep learning techniques, machine learning techniques, South African power systems, power distribution networks, maintenance planning

Load forecasting, which is the act of anticipating future loads, has been shown to be important in power system network planning, operations and maintenance. Artificial Intelligence (AI) techniques have been shown to be good tools for load forecasting. Load forecasting can assist power distribution utilities maximise their revenue through optimising maintenance planning. With the dawn of the smart grid, first world countries have moved past the customer's point of supply and use smart meters to forecast customer loads. These recent studies also utilise recent state of the art AI techniques such as deep learning techniques. Weather parameters are such as temperature, humidity and rainfall are usually used as parameters in these studies. South African load forecasting studies are outdated and recent studies are limited. Most of these studies are from 2010, and dating backwards to 1999. Hence they do not use recent state of the art AI techniques. The studies do not focus at distribution level load forecasting for optimal maintenance planning. The impact of adjusting power consumption data when there are spikes and dips in the data was not investigated in all these South African studies. These studies did not investigate the impact of weather parameters on different South African loads and hence load forecasting performance.

Data is one of the key components in the development of AI models. The integrity of the data is key to developing accurate models. Hence, knowing the integrity of the data one is using to train and test their AI models becomes important. The integrity of data can be compromised at different points before an end user accesses the data. The data can also have noise or patterns that may seem out of the norm. Determining South African power consumption data integrity at end-user level using AI techniques has not been investigated. The impact of performing data clean-up on South African load forecast has not been studied.

In this thesis, three case studies (Substation A, Substation B and Substation C) are presented to overcome the shortfalls stated previously in this section and, contribute to body of knowledge in the field of load forecasting and AI. The first contribution is the introduction of a novel load forecasting system utilising state of the art deep learning and machine learning techniques for South African power distribution networks. The proposed system included a module that determines data integrity using Mamdani-Type fuzzy logic. The module evaluates data integrity by looking at three different fault types: data lost faults, spikes and out of bounds fault, and multiple entries of the same variable. The flags are raised for field or database repairs for low data integrity detections. The data with high integrity were used to forecast load using the load forecasting module. The machine learning/deep learning load forecasting module has machine learning/deep learning models deployed. These models are trained before deployment, with the best performing model per application being deployed. The best model is chosen from the four techniques used, namely, adaptive neuro-fuzzy inference system (ANFIS), optimally pruned extreme learning machine (OP-ELM), deep belief networks (DBN) and long short-term memory recurrent neural networks (LSTM-RNN). The best performing load forecasting model was determined using three performance measures, the symmetric mean absolute percentage error (sMAPE), mean absolute error (MAE) and root mean square error (RMSE). The load forecast are then used to inform the distribution planned power outages as part of maintenance planning or electrification.

The second contribution is the introduction of deep learning techniques in South African load forecasting. In both case studies, it was found that deep learning techniques outperform machine learning techniques. DBN achieved the lowest load forecasting error in the first case study (Substation A) with an sMAPE of 0.0785323 (3.93 %), MAE of 0.0306600 (3.07 %) and

RMSE of 0.0429001 (4.29 %). LSTM-RNN achieved the lowest load forecasting error in the second case study with an sMAPE of 0.065859 (3.29 %), MAE of 0.04598 (4.6 %) and RMSE of 0.055058 (5.51 %). In the third case study, an LSTM-RNN model achieved the lowest load forecasting error with an sMAPE of 0.2307 (11.54 %), MAE of 0.0896 (8.96 %) and RMSE of 0.14065 (14.07 %)

The third contribution is two-fold. The first part is investigating the effects of 'cleaning' loading data for dips and spikes. The second part is investigating how the temperature affects the performance of machine learning and deep learning models in forecasting South African distribution networks loads. It was found that both machine learning and deep learning models, in the two case studies, generally achieved their best performance without 'cleaning' the loading data. The impact of temperature as an input variable in the development of load forecasting models was also investigated. In the first case (Substation A) and third case (Substation C) it was found that the machine learning techniques achieved their lowest load forecasting errors without temperature in their model's development. The deep learning techniques achieved their best performance with the inclusion of temperature in their model's development. In the second case study all the models achieved the best performance without the inclusion of temperature, with the exception of OP-ELM.

Acknowledgements

Thanks firstly go to God, who has made all of this possible. His mercy is enduring. Another big thank you to my late mother, Puseletso Letia, whose upbringing, hard work and sacrifices have enabled me to be at this point today. *Mme wa ngwana* o *tshwara thipa ka bohaleng.*

I would like to thank Prof. Bhekisipho Twala for his guidance and supervision, and for helping me get to the finish line.

I would also like to thank Dr Ali Hasan and Prof. Riaan Stopforth for allowing me to complete this study under their supervision. I appreciate their guidance, patience, push, the late e-mails and not to forget the jokes.

I would like to thank my wife, Nthabiseng Motepe, for her patience and support throughout my studies. Also great thanks to my son, Kananelo Motepe, for being patient with me when I stole some of his time.

I would also like to thank all my relatives and friends who have contributed positively towards my life while I was undertaking this journey. With a special thanks to Thapelo Tsotetsi, Raphepheng Motepe, Tholoana Hlungwani, Dipontsho Motepe, Vusi Hlungwani, Mr and Mrs Mohlakoana, Thabang Nkadimeng, Mahlatsi Malatji, Ben Mabizela and Ndivhuwo Mulimeli.

I would also like to give special thanks to my late friend Teboho Maloka for the discussions we had as undergraduate students which encouraged me to pursue postgraduate studies.

Publications Resulting from this Research

Eight publications have been developed as an output of this research. These publications are listed below.

A. Conference Proceedings

1. S. Motepe, B. Twala and R. Stopforth, "Determining South African Distribution Power System Big Data Integrity Using Fuzzy Logic: Power Measurements Data Application," Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech), Bloemfontein, South Africa, 2017, pp. 139-143
2. S. Motepe, Ali N. Hasan and R. Stopforth, "South African Distribution Networks Load Forecasting Using ANFIS," IEEE Power Electronics Drivers and Energy Systems (PEDES) Conference, Chennai, India, 2018, pp. 1-6
3. S. Motepe, Ali N. Hasan and R. Stopforth, "Power Distribution Networks Load Forecasting Using Deep Belief Networks: The South African Case," IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT) Proceedings, Amman, Jordan, 2019, pp. 507-512
4. S. Motepe, Ali N. Hasan, B. Twala and R. Stopforth, "South African Power Distribution Network Load Forecasting Using Hybrid AI Techniques: ANFIS and OP-ELM," accepted and to be presented at the joint Aegean Conference on Electrical Machines and Power Electronics, and Optimization of Electrical & Electronic Equipment Conference (ACEMP-OPTIM), Istanbul, Turkey, 2019
5. S. Motepe, Ali N. Hasan, B. Twala and R. Stopforth, "Using Deep Learning Techniques for South African Power Distribution Networks Load Forecasting," accepted and to be presented at the joint Aegean Conference on Electrical Machines and Power Electronics, and Optimization of Electrical & Electronic Equipment Conference (ACEMP-OPTIM), Istanbul, Turkey, 2019

B. Journal Publications

6. S. Motepe, B. Twala, Q-G Wang and R. Stopforth, "Determining Distribution Power System Loading Measurements Accuracy Using Fuzzy Logic," Procedia Manufacturing, vol. 7, 2017, pp 435-439

7. S. Motepe, Ali N. Hasan and R. Stopforth, "Improving Load Forecasting Process for a Power Distribution Network Using Hybrid AI and Deep Learning Algorithms," IEEE Access, vol. 7, 2019, pp. 82584 – 82598
8. S. Motepe, Ali N. Hasan, B. Twala and R. Stopforth, "Effective Load Forecasting For Large Power Consuming Industrial Customers Using Long Short-Term Memory Recurrent Neural Networks," Journal of Intelligent & Fuzzy Systems, vol. 37, no. 6, pp. 8219-8235, 2019, DOI:10.3233/JIFS-190658

Nomenclature

ANFIS – Adaptive Neuro-Fuzzy Inference System

ANN – Artificial Neural Network

ARIMA – Autoregressive Integrated Moving Average

ARMA – Autoregressive Moving Average

ARMAX – Autoregressive Moving Average with Exogenous Variable

BP-NN – Back Propagation Neural Network

DBN – Deep Belief Network

Dx – Distribution

ELM – Extreme Learning Machine

Gx – Generation

IPP – Independent Power Producer

LSTM – Long Short-Term Memory

LSTM-RNN – Long Short-Term Memory Recurrent Neural Network

MAE – Mean Absolute Error

MRSR – Multi-Response Sparse Regression

MTS – Main Transmission Substation

OP-ELM – Optimally Pruned Extreme Learning Machine

RBM – Restricted Boltzmann Machine

RMSE – Root Mean Square Error

RNN – Recurrent Neural Network

SAIDI – System Average Interruption Duration Index

SAIFI – System Average Interruption Frequency Index

SVM – Support Vector Machine

SVR – Support Vector Regression

sMAPE – Symmetric Mean Absolute Percentage Error

Tx – Transmission

List of Figures

List of Tables

Table of Contents

Chapter 1 – Introduction

“If you spend too much time thinking about a thing, you’ll never get it done. Make at least one definite move daily toward your goal.”

Bruce Lee

1.1. Introduction

South Africa is a developing country located at the southern tip of Africa. In the year, 1994 South Africa became a democratic country. Upon becoming democratic, the country set out to make electricity available to all its citizens. This ambition has led to the electrification of over 12 000 schools and over 5.2 million homes [1]. The government's plan is to achieve universal access by 2025/2026. South Africa has a vertically integrated electricity sector, with Eskom as the main player across the value chain. Eskom supplies 95 % of South Africa's power [2]. A large portion of this power is produced using coal plants. Private entities produce the remaining percentage. The company is the only power transmitter in the country, with the role of power distribution shared between Eskom and municipalities. There have been recent talks of unbundling Eskom, which may potentially introduce more players across the value chain.

1.2. Electrical Power Distribution Networks

In South Africa, distribution networks are power systems networks at voltages up to 132 kV. These networks are constructed to deliver power to end users [3]. The power is produced in generation plants and is then transformed to a higher voltage and low current for transmission, typically over long distances. This transformation is to reduce the required infrastructure and the technical losses. The power then gets to a transmission substation, also termed main transmission substation (MTS), where it is transformed to a lower voltage. It is then transported to distribution substations, which distribute the power to customers, including power redistributors. Figure 1 gives a high-level example of a power system network from generation to distribution, up to the distribution substation higher voltage bus bar.

There is a drive in SOUTH AFRICA to diversify its energy mix and become greener. At the end of October 2016, 2.8 GW of renewable independent power producers' (IPPs) plants had already been connected to the South African grid at transmission and distribution levels [4]. The South African government plans to have 17 800 MW supplied by renewable sources by 2030 [4]. This drive has led to the grid code requiring N-1 firmness. Due to this and partially also due to aging, Eskom has paid special attention to distribution and transmission maintenance [5]. Globally, utility distribution networks are also seeing a growth in renewable generation sources penetration. The importance of a reliable distribution system is thus paramount.

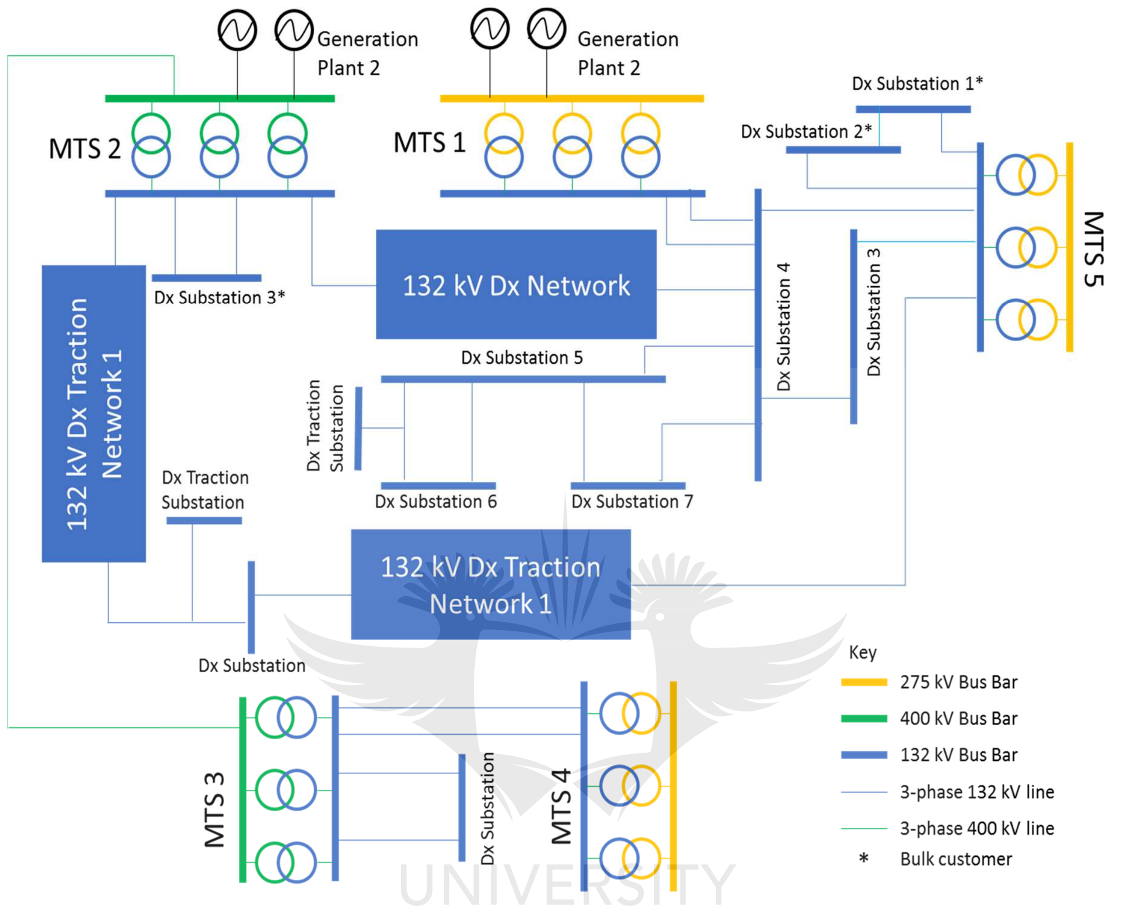

Figure 1: Illustration of a Power System Network from Generation to Distribution

One of the key performance measures of a distribution network is its reliability. The reliability is measured by the System Average Interruption Frequency Index (SAIFI) and the System Average Interruption Duration Index (SAIDI) [6]. SAIFI simply put is a measure of how frequently customers' power supply is disrupted over a set period of time, usually a year, and SAIDI measures how long customers' power is disrupted. These two measures are given by (1) and (2) [6]:

$$SAIDI = \frac{\sum r_i n_i}{n_k} \tag{1}$$

$$SAIFI = \frac{\sum n_i}{n_k} \tag{2}$$

where r_i is each incident's restoration time, n_i is the number of incidents and n_k is the number of interrupted customers. Network power outages, whether planned or unplanned, contribute to an increased SAIDI. Lack of maintenance can lead to an increase in the number of failures. Power utilities, therefore, need to manage these key performance areas (KPIs), to not only keep their customers satisfied with their service, but also to maximise revenue. This is achieved by having most of the utility's customers supplied for the maximum possible time, with a minimal number of interruptions.

Distribution networks can be constructed into a number of different configurations. Some of the common configurations are the single-end radially fed feeder, a doubly fed feeder with a normally open point and lastly a ring network configuration. The single-end radially fed feeder has a single feeder from the main substation. This main feeder can have branches or sub-feeders which have customers connected to it. However, in this configuration power flow is in one direction from the supply substation to the end user with no other alternative supply options. A fault in this type of network means that customers beyond the fault will be without power. Should the fault cause the main breaker at the substation to open, the whole feeder can be without power. The doubly fed feeders can be regarded as two single-end radially fed feeders supplied from two different substations. The feeders, however, have a point that connects them. This point is called a normally open point and does not allow power to flow between the networks under normal operations. Closing this point connects the two networks and, therefore, allows power to flow from one of the sources through its feeder to supply the second feeder's network. The ring network configuration comprises sub-networks that are interconnected and are supplied from the same main substation from two separate feeders or more. The network has interconnecting feeders that connect the different networks. These interconnecting feeders usually have breakers that allow the networks to be disconnected when required. A network with this configuration typically operates with the breakers closed, allows power to flow across the two sub-networks. These distribution networks can be made up either of underground cables or overhead lines. The underground cables are usually utilised in urban areas and the overhead lines in rural areas.

1.3. Maintenance Overview

A distribution network consists of multiple components/equipment such as transformers, cables, meters, bus bars, circuit breakers, poles and other support structures. These components need to be maintained and/or periodically replaced to ensure that the power system network is operational and reliable. Maintenance can be classified as preventative or corrective maintenance. Preventative maintenance is maintenance conducted before a component fails, to prevent its failure and impact to the rest of the network. This type of maintenance can also be seen as planned maintenance, as it is planned for and conducted at a predetermined time. Corrective maintenance is maintenance that is conducted once a breakdown has occurred. This type of maintenance is conducted to repair or replace a failed component. This maintenance is also conducted to correct the impact a failed component has had on the power system network. Various topics in power grid maintenance have been studied by various researchers [7], [8]. *Xie et al.* proposed a power grid maintenance scheduling intelligence arrangement based on power flow forecasting [7]. Their system had a load forecasting module that was based on historical data to forecast the power bus load. Distribution maintenance time scheduling optimisation has been studied using genetic algorithms. The authors stated power flow constraint as one of the optimisation problem constraints [9]. A maintenance optimisation problem can be established where the target is to reduce SAIDI and SAIFI. In the studies mentioned in this section, one of the key aims was to reduce maintenance-related costs. In [10] the authors predict load peaks to determine load flows in restricted transmission networks for maintenance scheduling. Load forecasting can tell how much power will be flowing in different parts of the power system in a future period. This knowledge can assist in developing a plan to have minimal customer’s power supply interrupted when parts of the power network need to be switched off, to be maintained.

1.4. Electrical Load Forecasting Overview

Load forecasting is the act of anticipating the future load and has been an interesting topic for multiple researchers for over six decades [11], [12]. Load forecasting is divided into short-term load forecasting, that looks at forecasts over an hourly to weekly period, medium-term load forecasting looks at two weeks to three months ahead load forecasting and long-term load forecasting looks at yearly forecast periods [1], [13]. Load forecasting has many useful applications, such as network planning and capacity planning. Load forecasting is also useful in maintenance and operations. In distribution utilities, maintenance should be scheduled to have a minimum number of customers disrupted at one time. Having access to accurate load forecasts can give utilities a view of when distribution substations or feeders will be lightly loaded and when to schedule maintenance outages. The load can at times be transferred onto another feeder if there exists back-feeding capabilities with another feeder in the network. A similar approach can be followed where another transformer in a substation may need to carry the load from a transformer that has to undergo maintenance. Figure 2 gives an illustration of a distribution substation and feeders with different customer loads, to explain the above-mentioned points. The normally open point can be used to connect the reticulation network supplied by feeder 1 and feeder 2. This is achieved by closing the normally open switch. The breaker on the feeder that needs to be worked on can be opened at the supply voltage's lower voltage side. A similar approach can be followed to isolate one of the transformers. Here breakers on both sides of the transformer that is to be maintained are opened, or isolators for the bus bar section to be maintained are opened. In both cases, the forecasted load of both feeders or transformers will help determine how much additional load the feeder or transformer that will remain in service needs to handle and if this equipment will be able to handle this additional load. Power systems simulation software can be used to determine the system behaviour under this reconfiguration.

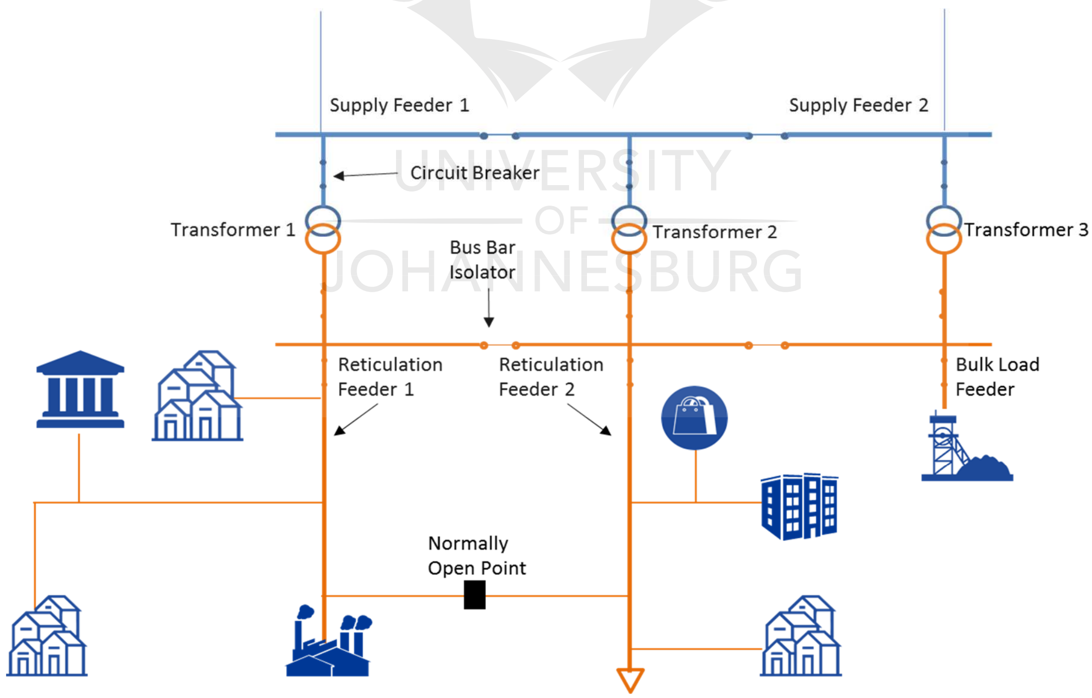

Figure 2: Distribution network and different distribution loads

1.5. Recent Load Forecasting Applications and Studies

Recent load forecasting studies in developed countries have moved past the customer supply point and use recent state art artificial intelligence (AI) techniques [14], [15]. Appliance usage patterns were incorporated for load forecasting using a fuzzy logic approach [15]. The smart grid was one of the common drivers of movement past the customer supply point. Deep learning techniques are commonly deployed in most of these recent studies. In [14] long short-term memory (LSTM) was used to forecast residential loads using Canadian households and their 19 appliances data. Long short-term memory recurrent neural networks (LSTM-RNN) were also used to forecast Australian residential loads using publicly available smart meter data [16]. Enhanced deep networks (cycle-based LSTM and time-dependency convolutional neural networks) have been used for medium-term load forecasting [13]. Deep learning techniques were used in this study due to the inability of swallow artificial neural networks (ANN) to accurately conduct the complicated and complex medium-term load forecasting. Deep belief network (DBN) and Copula-DBN were used in [17] to forecast hourly loads in the USA (Texas and Arkansas). In [18] LSTM, combined with characteristic load decomposition application in the data pre-processing, was found to outperform conventional load forecasting approaches. Other recent studies applied classical and ensemble techniques in forecasting loads. Seasonal ARIMA was used to conduct short-term load forecasting for optimal operations planning of electric distribution systems [19]. Autoregressive integrated moving average (ARIMA) was used as it requires fewer variables than other common load forecasting techniques. A holographic ensemble technique was used to forecast Guangzhou's (China) and New England's (USA) total and daily peak load [20].

A number of load forecasting studies have been undertaken in South Africa [21], [22], [23], [24], [25], [26]. However, the study of the application of AI in South African load forecasting is still in its infancy and limited. Some of the studies are outdated. *Ijumba and Hunsley's* study is from the late 1990s and does not consider the latest state of the art AI techniques [22]. The study focused on residential loads and had separate models to forecast weekday and weekend loads. The study of the application of deep learning techniques on South African load forecasting is almost non-existent. *Marwala* studied the South African total consumption load forecasting using AI. His work also looked at main drivers, related to the industrial and mining indices, as drivers for total load growth [27]. Load forecasting in South African distribution networks has not been pursued at distribution substation and substation feeder level. The load forecasting performance of different state of the art machine learning and deep learning for different consumer types has not been studied. *Yuill et al.* studied South African load forecasting for optimal generation scheduling [23], [24]. The researchers focus on forecasting loads 30 minutes ahead over a day. The authors include temperature and humidity in their study. In one of the recent South African load forecasting studies, ANN was used to forecast the net energy consumption [28]. In another South African load forecasting study, *Inglesi* used a vector error correction model to forecast the aggregate South African electricity demand [29].

1.6. Machine Learning and Deep Learning

AI had its inception in the 1950s, emerging from computing, psychology, mathematics, engineering and cybernetics. AI's main objective is to develop a system that achieves human-like competence and intelligence in completing complex tasks [30]. The term, AI was introduced in 1956 by *McCarthy*. The early stages of AI focused on developing theorem proving and game playing programs. Modern AI is focused on techniques for human-like reasoning, planning, learning, language and pattern recognition [31]. Alan Turing proposed the Turing test in 1950. The Turing test is a test of intelligence. A computer passed the test if a human interacting with it was not able to distinguish if a response was from a computer or a human being. AI researchers have however not devoted time to passing the Turing test, due to the belief that it is less important to duplicate an example as opposed to understanding the underlying principles [32]. To deduce that a program thinks like a human, some understanding of how a human thinks is required. This understanding can be established through three approaches: introspection, psychological experiments and brain imaging. Cognitive science links the AI computer models and psychological experiments to develop precise and testable human mind theories [32]. Machine learning is a subset of AI that is based on learning processes, which allow the evolution of machines without the change of algorithms [33].

With the advancements in technology, computational power has increased. This has led to an uptake in the application of deep learning techniques. Deep learning techniques are special AI techniques that have multiple layers, that enable them to learn more features from the data. These techniques have been popular in language processing and computer vision [34].

Their use has also spread to other areas and sub-areas, such as load forecasting, cancer detection, voice assistants such as Apple's Siri [34], [35], [36]. Facebook's DeepFace, which uses deep learning for human facial recognition, has achieved 97.35 % accuracy in identifying human faces versus a human's accuracy of 97.5 % [37]. Self-driving cars are another technology where AI is playing a pivotal role. Google's Waymo has successfully tested their self-driving cars on different USA public roads, in different cities [38]. Tesla sells its cars with the hardware required for autonomous driving. Their cars have been tested on public roads [39]. Deep learning enables these cars to identify objects, allowing the cars to: navigate without driving into objects, follow road signs, drive in lanes, etc. A number of companies have developed AI platforms to enable the development of AI models and to provide processing power. These platforms include Alibaba cloud, Amazon web services respectively by Alibaba and Amazon.

These companies have also integrated AI in their business processes to improve how they serve their clients. Alibaba deploys deep learning for different tasks such as: to recommend products to its customers, to attend to client queries using chatbots and to deliver orders to their customers using drones [40]. General Electric Healthcare deploys machine learning to improve patient outcomes. One of the areas this company has applied deep learning in is the improvement of x-ray technologies [41].

1.7. Data Integrity

At the heart of AI are the data. Machine learning and deep learning techniques learn from data. These techniques are therefore as good as the data given to them to learn from. This is also true for human beings. Given that a child grows up being taught without visuals that a cat is an animal, with fur, two eyes, and two ears and does not grow to be bigger than 30 CM in height. If this child is shown a picture of a puppy, what would he or she classify the animal in the picture as? If the puppy in the picture is smaller than 30 cm, with fur and with both eyes and ears are in place, the child might call the puppy a cat. This decision will be based on the information given to the child about a cat. In this case, even though the data about the cat was correct, it was limited. In a similar manner shortage of training data or data lacking integrity can lead to AI models giving unexpected or incorrect results. Data integrity can be defined as information reliability, usability, relevance and quality. Lack of data integrity can be due to multiple sources such as non-functional measuring instrument or computer, misunderstanding, technology limitations and faults during data transmission. These sources can lead to data that are [42]:

- Incomplete
- Noisy
- Inconsistent

1.8. Problem Statement

South African distribution power utilities are experiencing a financial strain. Non-technical losses are on the rise, aging equipment requires maintenance or upgrades, and new customers need to be connected to the grid. This work requires scheduled power outages to enable safe working conditions. These power outages typically lead to a loss of revenue due to customers being without supply. Further financial losses can be experienced from other sources, such as rescheduling work that need to be conducted by contractors due to lack of planning. Not carrying out these outages can lead to a loss of revenue from frequent equipment failures. These failures can lead to expensive emergency work being required, damages to the environment, injuries to humans, etc.

As challenges increase for distribution utilities, limited strategies are developed and deployed to plan utility upgrades and maintenance. There was, therefore, a need to:

- Introduce and utilise state of the art techniques to optimise distribution operations
- Achieve accurate distribution load forecasting to drive optimal maintenance planning

1.9. Research Objectives and Contributions

This research's novelty is the introduction of a unique South African distribution networks load forecasting system that utilises state of the art machine learning and deep learning techniques. The research further contributes to the body of knowledge by introducing the application of load forecasting in South Africa using deep learning techniques. The comparison of load forecasting performance of state of the art AI techniques for different South African distribution customer types, and the impact of data clean-up and weather parameters is another novel contribution of this research. The research also introduces AI in determining distribution loading data integrity. The research's four contributions are discussed next:

1. Introduction of a unique South African distribution networks load forecasting system that utilises state of the art machine learning and deep learning techniques

Why are the current methods insufficient?

Current load forecasting research in South Africa is focused on total demand and very short-term for demand and supply balancing in power dispatching. There is currently no research work being conducted at a distribution substation level with a forecast window long enough to influence maintenance planning. The current research work is outdated and does not investigate recent state of the art techniques.

How does the current system work?

The current systems in S.A. do not use AI techniques for load forecasting to improve maintenance planning. Utility engineers sometimes use naïve methods to forecast loads to inform network configuration and maintenance scheduling. Naïve methods are globally outdated in load forecasting and usually consider limited data to inform a decision.

What needs to be done?

A system that takes advantage of the technological development, such as powerful and recent state of the art machine and deep learning techniques, needs to be introduced in distribution load forecast.

How does this solve the problem?

Recent state of the art AI techniques, such as deep learning, can utilise larger quantities of data to efficiently forecast distribution networks loads. These load forecast can then be utilised to improve utility maintenance planning.

2. Introduction of the application of deep learning techniques in South African load forecasting

Why are the current methods insufficient?

Deep learning techniques have not yet been applied in South African load forecasting and South African distribution networks load forecasting.

How does the current system work?

Other machine learning techniques have been studied/applied in total demand load forecasting and day ahead load forecasting for optimal power dispatch.

What needs to be done?

Deep learning techniques need to be studied and applied in South Africa load forecasting to harness their ability to learn more features.

How does this solve the problem?

Deep learning techniques can lead to high accuracy load forecasts.

3. A novel comparative study of sophisticated AI techniques' performance on different South African distribution customers. An investigation of the impact of data clean-up and the inclusion of geographical temperature on the performance of these techniques per customer type is also studied

Why are the current methods insufficient?

A comparative study of AI application on different South African distribution customer types has not yet been studied. The impact of temperature and data clean-up on the performance of recent AI techniques for different South African distribution load types has also not been investigated.

How does the current system work?

Temperature has been included in a South African load forecasting study and is usually considered to have an impact on the load forecast accuracy. The impact of excluding temperature against including it in the development of AI techniques models and on different South African loads has not yet been studied. The impact of cleaning dips and spikes from different South African distribution customer types power consumption/loading data on AI techniques' performance has also not been studied.

What needs to be done?

Data preparation and sourcing weather data can be daunting tasks. Access to weather data is sometimes not possible or the data need to be purchased if it is to be used commercially. Hence, the impact of these two parameters on the load forecasting of different distribution customer types needs to be well understood before committing time and money to them.

How does this solve the problem?

This understanding can lead to utilities achieving high accuracy load forecasting without the need to invest (time and money) in these two areas or being able to justify the investment.

4. Introduction of a novel AI based process to determining distribution loading data integrity

Why are the current methods insufficient?

AI has not been applied to determine South African power systems distribution network data integrity.

How does the current system work?

Business Intelligence and other visual tools are implemented to determine loading data integrity. This can be a time-consuming process depending on the amount of data being analysed.

What needs to be done?

With data sizes growing utilities can utilise AI techniques in their process to determine their loading data integrity.

How does this solve the problem?

This can save utilities time and money by freeing up their staff's time, having a view of areas in the distribution network that have data integrity issues and have the issues addressed to enable data utilisation to help take improved business decisions.

1.10. Thesis Layout

Chapter 2 presents an overview of AI. Examples of application of machine learning and deep learning techniques are given in this chapter. The techniques used in this study, ANFIS, OP-ELM, LSTM-RNN and restricted Boltzmann machine's DBN, as well as their advantages and disadvantages, are also presented. The chapter also discusses the concept of learning and intelligence. Classification, forecasting and prediction are also briefly defined.

Chapter 3 presents the proposed load forecasting system. The experimental setup is presented together with the three performance measures that are used to measure the AI model's performance in this study. These performance measures are the following error measurements; symmetric mean absolute percentage error (sMAPE), mean absolute error (MAE) and the root mean square error (RMSE). The statistical significance test used to measure if a significant difference exists in the performance of the different techniques is also presented. This chapter further presents an AI approach to determine loading data integrity.

Chapter 4 gives the first experimental case study on distribution substation A, which is a power redistributor customer. The substation has two 40 MVA, 88/11 kV transformers. The distribution network overview and data used for experiments are presented. The results for this case study are presented and discussed.

Chapter 5 presents the second distribution substation case study (substation B). The substation supplies a single industrial customer through multiple supply points and at different voltage levels (132/22 kV) to substation A. The results of the experiments for this substation are presented and discussed.

Chapter 6 presents the third case study (Substation C). The substation is a power redistributor supplied power at 132 kV. The load forecasting performance of the four techniques' models in load forecasting are presented. A comparison between the findings of the three case studies is also presented in this chapter.

Chapter 7 gives concluding remarks through a summary of the findings. Recommendations for future work on machine learning and deep learning techniques application in distribution power systems and distribution load forecasting are suggested.

1.11. Chapter Summary

This chapter introduced distribution power systems networks. A typical power system setup was also outlined. A load forecasting literature review was presented, together with a South Africa specific load forecasting literature review. Machine learning and deep learning concepts were introduced. The problem statement was presented together with the research objective. The research objective presented and discussed the research’s novel contribution. The layout of this thesis was also outlined in this chapter.

Chapter 2 – Artificial Intelligence

"The measure of intelligence is the ability to change."

Albert Einstein

2.1. Introduction

Chapter 1 introduced distribution power system networks, load forecasting and AI aspects. The chapter also presented the research problem statement and contributions. A load forecasting literature review was presented, which included a South African specific load forecasting literature review.

This chapter introduces machine learning and deep learning, as well as the specific machine learning and deep learning techniques used in this research. A machine learning and deep learning literature review is also presented in this chapter. This literature review looks at the application of machine learning and deep learning in different fields. The concepts of machines that are capable of learning and are intelligent towards human levels are presented.

2.2. Knowledge, Intelligence and Learning

Human beings use knowledge to help them make everyday decisions. Knowledge is key in building AI systems. Knowledge can be defined as the state of knowing or a human being accumulating a body of facts and principles. Knowledge has familiarity with procedures, rules, ideas, abstraction, customs, facts, etc. Knowledge has three key basic concepts, namely, a data set, a form of information and belief or hypothesis. Knowledge can be one or all of these concepts in a slightly different form. Data is a raw form of observation, and knowledge is organised data and procedures which have some useful purpose. Information is data plus its meaning. Information can be classified as knowledge if it can create more information and become a part of some action. Knowledge is a true justified belief as opposed to just a coherent expression (belief) or a belief that may or may not be true but is supported by some fact (hypothesis). The relation between knowledge and intelligence lies in that, to be intelligent one needs access to and possession of knowledge [31].

The Oxford Dictionary defines intelligence as “the ability to acquire and apply knowledge and skills”. Intelligence’s exact definition is not known. There has however been a number of different definitions of intelligence [31], [43].

“Intelligence = ability to accomplish complex goals”

“Intelligence is the ability to learn, to deal with different situations, to acquire, understand and apply knowledge and to analyse and reason.”

As there are different goals that different species can try to obtain, quantifying intelligence becomes a futile exercise. For example, who is more intelligent between a chess Grandmaster and a chief executive officer (CEO) who turns around a multibillion-Rand company from near bankruptcy? Due to different goals, it is not easy to deduce who is more intelligent than the other. This challenge is because the Grandmaster may or may not be able to turn around a company, the CEO may or may not be able to defeat the chess Grandmaster in chess. In addition the CEO may not be able to turn around a company which is in the same state but in a different sector. If however, there exists a third person who can accomplish both tasks and is better at one of the tasks, say company turnarounds in different industries, it may be safe to say that she or he is "more intelligent" than the original CEO. Since there are spectrums of intelligence, the argument of whether an entity is intelligent or not in borderline cases is at times not worthwhile. An example is who, between a toddler in a pre-school soccer team or a professional football player like Lionel Messi who plays for one of the biggest soccer teams in the world, has the ability to play soccer. Before discussing the 'Artificial' part of 'Artificial Intelligence', it is important to also state three earliest and most accepted definitions of intelligence which are [31]:

"Intelligence is a state grasping the truth, involving reason, concerned with action about what is good or bad for human beings…."

"The ability to learn or understand from experience, the ability to acquire and retain knowledge and the ability to respond quickly and successfully to a new situation, use of the faculty of reason in solving problems, directing the conduct effectively. "

"The test of the first rate intelligence is the ability to hold two opposite ideas in the mind at the same time and still retain the ability to function."

'Artificial', refers to something that is not natural or is not real. Human intelligence can, therefore, be regarded as 'real intelligence' as humans naturally develop their intelligence. AI is created using mathematics, engineering, computers, data, etc. and is not regarded as real intelligence [31].

As stated previously, to be regarded as intelligent one must possess knowledge. Learning is key to acquiring knowledge. Thus, for a machine to be intelligent it has to have the ability to learn. This is where the concept of machine learning comes in. Learning has two key features

– skill refinement and knowledge refinement. Skill refinement is the improvement of a skill by repetitive execution of the same task. Knowledge acquisition is skill improvement by being able to gain knowledge or remembering past experience. An entity can learn in five different ways, by [31]:

i. Memorisation
ii. Taking advice
iii. Induction
iv. Deduction
v. Analogy

Memorisation, also known as rote learning, is repeatedly storing a task's data for use in performing the task in the future [44]. In this case, whenever a task is performed, the memory is used to determine the best option from the possible options based on what has been observed to work well and what has been observed in the past. **Learning by taking advice** involves an external person giving advice, guidance or instructions to the person seen as learning. An example of this is a baby learning from his or her parents. This type of learning is common throughout a person's life. Likewise, when an engineer writes a piece of code to run on a computer, the code can be seen as a set of instructions that tell the computer what to do [44]. **Induction** means generalising from specific instances [31]. This case can be likened to the observations of a number of BMW (Bavarian Motor Works) vehicles and then being able to tell when presented with a picture of a car if it is a BMW or not, and if it is a BMW which model it is. **Deductive learning** is based on improving performance by exploiting the previous problem-solving experiences [44]. Here an entity uses what they found to work or not to work in solving previous problems to solve a current problem. **Analogy** is a strong inference tool which human reasoning and speech are full of. In inference, there is usually an underlying meaning based on the mapping of concepts which seem to be dissimilar [44]. An example is when one says, “He kept an eagle's eye on the opponent". Which means he was closely watching the opponent and not that that he took an eagle’s eye and put it on the opponent.

2.3. Application of Deep Learning

The rise of computational power and access to big data has enabled progress in deep learning techniques. These techniques have been popular in language processing and computer vision, and have for example, led to computers being able to interpret handwritten text with high accuracy [36], [45], [46]. The use of deep learning techniques has spread to other topics such as load forecasting, solar energy forecasting, and electricity price forecasting [47], [48], [49]. The most popular deep learning techniques are DBN, convolutional neural networks (CNN) and recurrent neural network (RNN). LSTM-RNN's performance has been shown to supersede or match the state of the art techniques [50]. LSTM has also been shown to outperform CNN in energy consumption forecasting and natural language processing [50], [51]. DBN differs from CNN and RNN in that they combine unsupervised learning and supervised learning in their learning process. DBN has been applied in energy load forecasting, very short-term wind power prediction, wind speed forecasting and photovoltaic power forecasting [52], [53], [54], [55]. CNN was found to have comparable performance to RNN and DBN, but outperformed SVR, in energy load forecasting [52].

LSTM has been used to forecast wafer lots' short-term cycle time for production planning and control in semiconductor wafer manufacturing [56]. In [57] the authors use LSTM to automatically generate conversations. The LSTM models are deployed as chat bots. In another application, LSTM was used to recognise single Chinese character font [58].

DBNs have been applied in various areas including medicine, text recognition, computer vision, and many other applications. *Jemimma et al.* utilised DBN to segment and classify brain tumours [59]. DBN has also been used to classify pedestrians, bikes, motorcycle and vehicle classification using a small training data set [60]. In [61] DBN was used to diagnose wind turbine gearbox faults. DBN has also been used to detect intruders on a computer network [62].

Deep learning techniques have been applied in domestic violence identification [63]. Victims in critical need were identified by analysing Facebook posts and identifying important words that might indicate critical posts. Deep convolutional neural networks were used to classify music [64]. The aim of this study was to investigate the performance of the deep CNN when the data contains noise. CNN has also been used to detect breast cancer using crowd-sourced images [65]. CNN has also been applied to determine human personality traits from text [46].

2.4. Application of Machine Learning

Machine learning has been widely used across multiple fields such as engineering, power systems, medicine, economics, insurance, gaming, social media, law, emergency response, search and rescue, online shopping, voice assistance, robotics and military.

Machine learning is the process of machines learning from data to be able to give an answer similar to that which a human could give. There are various machine learning techniques such as ANFIS, OP-ELM, ANN, autoregressive moving average (ARMA), ARIMA, autoregressive moving average with exogenous variable (ARMAX) and support vector machine (SVM). The detailed manner in which these techniques learn is different, but follows a similar principle based on the learning approach. The general learning approach is to present a set of data inputs and targets to the machine learning model. The model then tries to predict the output and learns by adjusting its parameters based on the error between the predicted and target value. These machine learning techniques sometimes suffer from a number of challenges during training or application. Examples of challenges are overfitting, not finding an optimal solution due to getting stuck on local minima, unexplainability, and many others. Researchers have come up with workarounds to these challenges. These workarounds include combining two or more techniques to leverage their strengths. An example is the combination of fuzzy logic and ANN to form an ANFIS, which enables the manner in which the model gets to its output to be explainable.

ANFIS and other machine learning techniques, including ARIMA, ARMAX, ANN and evolutionary algorithms have been applied in load forecasting [1], [21], [25], [66], [67], [68], [69], [70], [71]. ANFIS was found to be more superior to these techniques in a number of these studies [21], [70], [71]. ANFIS has also been applied in rainfall prediction, robot stabilisation control and image tracking, DC motor speed control, navigation systems, power electronics converter stages open circuit fault diagnosis [72], [73], [74], [75], [76]. The performance of ARIMA, ANFIS and an ensemble of the two techniques was compared for rainfall prediction. The two individual methods showed better performance than their ensemble, each in one of the two respective towns that were under study [72].

Computational complexity has been observed to increase with an increase in the data used. Due to high computational power not always being available, there was an inclination to not use non-linear models as broadly. This decline was despite the models' overall good performance [77]. *Huang et al*. introduced extreme learning machines (ELM), which is an algorithm that reduces neural networks' computational time and model structure selection of neural networks [78]. ELM has been applied to solve various problems such as predicting stock volatility, forecasting wind generation, analysing power utility non-technical losses, forecasting electricity prices and load forecasting [27], [79], [80], [81], [82]. In [27] OP-ELM was found to outperform ANFIS, ARMA, support vector regression (SVR), ANN and ELM in South African electricity demand forecasting. *Wang et al.* found that multi-kernel ELM outperformed basic-ELM, single kernel ELM, SVM and back propagation neural network (BP-NN) in most cases, in predicting the Hong Kong Exchange stock volatility [79].

Another application of machine learning in stock exchange is the study conducted by *Khoza et al.* on stock price prediction for the Johannesburg Stock Exchange stocks [83]. Here rough set theory was used to extract a set of reducts and trading rules to predict stock prices. An ensemble of rough set theory and multi-layer perceptron has been used to predict the direction South Africa's GDP will turn [84]. Ensemble techniques have also been applied in photovoltaic (PV) systems maximum power point tracking (MPPT). *Farayola et al.* investigated and compared the performance of ANNs combined with rational quadratic gaussian process regression (RQGPR), ANN combined with linear SVM regression and ANN in MPPT [85]. The ensemble of ANN and RQGPR was found to achieve the best results. Machine leaning techniques: random forest, boosting and SVM, were applied to detect fake identities on twitter accounts [86]. This research shows that features used to detect fake accounts created by bots are not similar to those required to detect fake accounts created by humans. The machine learning techniques used here achieved the best accuracy of 49.75 %, which is lower than the accuracy one would achieve when taking a guess. SVM was also applied to identify endangered tree species in Dukuduku forest in South Africa [87].

2.5. Techniques Applied in This Research

Four techniques are applied in this research for load forecasting, two machine learning techniques (ANFIS and OP-ELM) and two deep learning techniques (RBM-DBN and LSTM-

RNN). Fuzzy logic is an additional technique used in this research to determine the integrity of the data, which the data used in this research was a subset.

2.5.1. Neuro-Fuzzy Systems

Neuro-Fuzzy Systems also referred to as Adaptive Neuro-Fuzzy Inference Systems were introduced as a combination of ANN and Fuzzy Logic to combine their strengths and overcome their weakness. ANN cannot represent knowledge and is not easy to explain and Fuzzy Logic is not able to learn from data. The technique was introduced in the 1990s by Jang [88], [89], [90].

2.5.1.1. Fuzzy Logic

Fuzzy logic is an expert system that needs to be taught by an expert or from expert knowledge and cannot learn from data [91]. Fuzzy logic was developed in 1930 by Jan Łukasiewicz, a Polish philosopher and logician. Mamdani and Tagaki-Sugeno (TS) are the two popular fuzzy logic inference systems. The fuzzy logic inference system is applied in four key steps:

Step 1: Input variable fuzzification

Step 2: Evaluation of rules

Step 3: Aggregating rule outputs

Step 4: Defuzzification

The key difference between these two inference systems is the last two steps. The TS output function is a linear function or a constant, whereas the Mamdani output is a value based on the rule's inputs and their antecedent. The Mamdani rule can be summarised as (3) [92]:

$$R_n\text{: if } z \text{ is } A_n \text{ AND } x \text{ is } B_n \text{ then } y_n = c_n \tag{3}$$

The TS output is defined by (4). R_n is the rule, A_n is the antecedent of input z, B_n is the antecedent of input x, c_n is the output, b_n is the bias, a_i the consequent parameter vector and $n=1,2,\dots.i$.

$$R_n\text{: if } z \text{ is } A_n \text{ then } y_n = a_n{}^T z + b_n \tag{4}$$

In data-driven applications the Takagi-Sugeno, which is also used in ANFIS, is more popular. The fuzzy logic aggregated output, y, is given by (5), where α_i is the *ith* rule's degree of fulfilment. The TS model can be regarded as a piecewise smooth linear approximation of the non-linear function. This is due to the TS parameters being local linear models of non-linear systems [1].

$$y = \frac{\sum_{i=1}^{n} \alpha_i(z) y_i}{\sum_{i=1}^{n} \alpha_i(z)} \tag{5}$$

2.5.1.2. Artificial Neural Networks

ANNs are models designed to work like a human brain, with neurons being connected to each other using synaptic weights. Neurons can have a single output and multiple inputs. An activation function and sum of bias and weighted values are used to map the output to the input as given by (6). This mapping can be seen in the structure of a neural perceptron given in Figure 3. The aim of the model training is to minimise the error between the target value and the model output by adjusting synaptic weights through an iterative process. The iteration stops when the error is acceptable or a set number of iterations is reached.

$$y = \sum_{i=1}^{n} w_i \, z_i + b_i \tag{6}$$

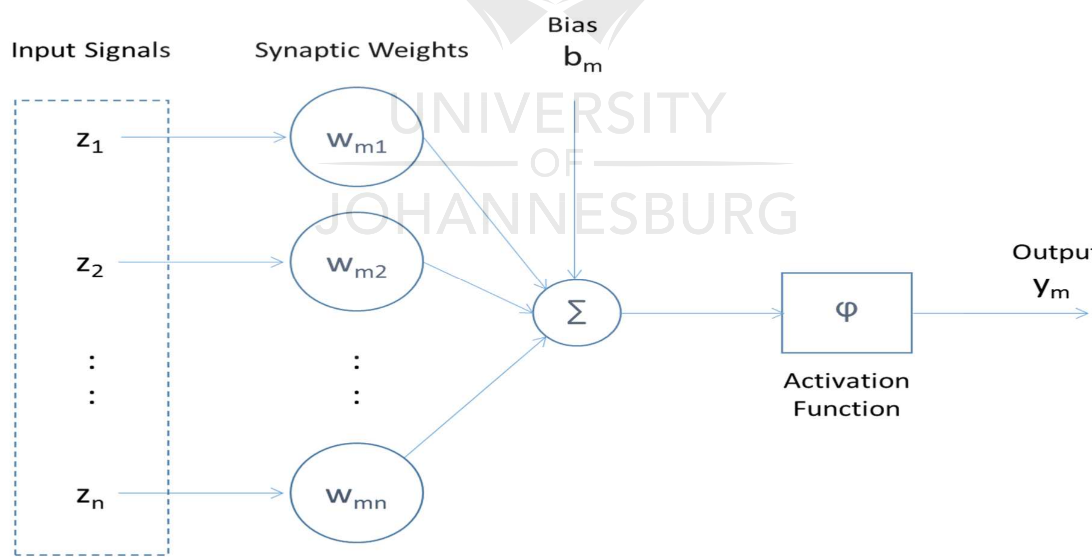

Figure 3: A structure of neural perceptron

The error is given by (7) and the synaptic weights are updated using (8) [1].

$$E(w) = \frac{1}{2} \sum_{j}^{N} \left\| y(z_j, w) - t_j \right\|^{2} \tag{7}$$

$$w_m^{s+1} = w_m^s - \lambda \nabla E^s(w_m^s) \tag{8}$$

In (6) and (7), *z* is the input, *y* the output, *b* the bias , *w* the synaptic weights, *t* the target value and *E* the total error. In (8) *s* is the iteration step, *m* is the weight index, λ the learning rate and the gradient $\nabla E(w)$ is given by (9).

$$\nabla E(w) = \left[\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, \dots \dots, \frac{\partial E}{\partial w_m}\right] \tag{9}$$

2.5.1.3. Adaptive Neuro-Fuzzy Inference System

ANFIS is seen to be adaptive due to its ability to learn. Gradient descent can be used to train this technique's models instead of expert knowledge. The first layer in the ANFIS structure, shown in Figure 4, is made up of adaptive nodes. These nodes compute the membership degree of the input in the antecedent Gaussian fuzzy sets. Equation (10) gives the Gaussian membership function commonly used. Here *g* is the centre of the Gaussian function and δ is the variance of the Gaussian membership function [1].

$$\mu A_{ij}(z_j, g_{ij}, \delta_{ij}) = \exp\left(-\frac{(z_i - g_{ij})^2}{2\delta_{ij}^{\ 2}}\right) \tag{10}$$

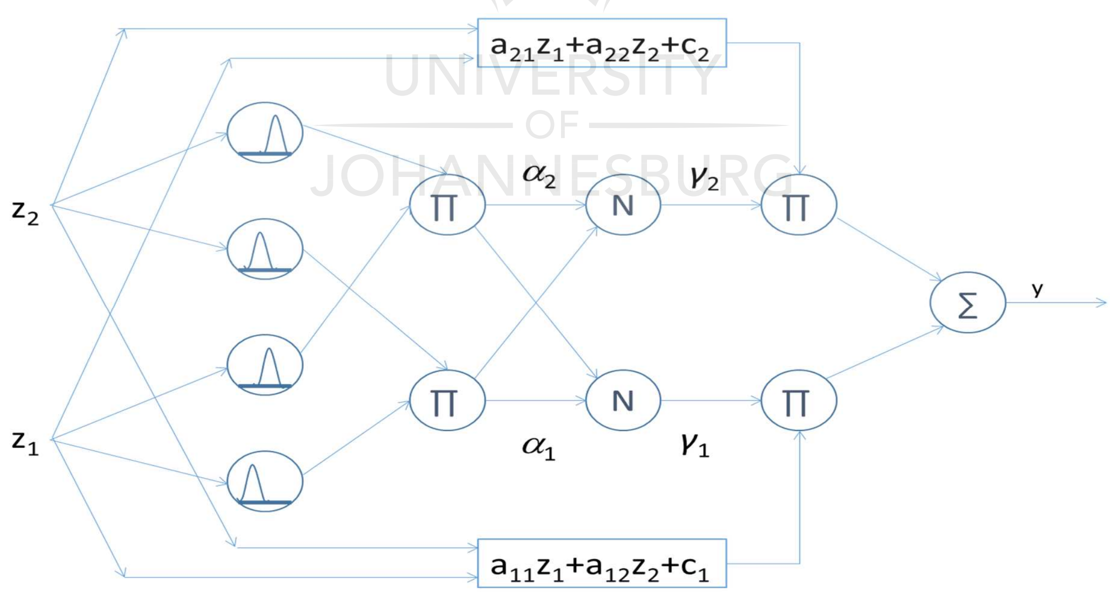

Figure 4: ANFIS structure

In the second layer the fuzzy *AND* operator is applied, and then the mean operator is attained with the normalisation (*N*) and summation ($\sum$). Equation (11) gives the TS output, in relation to the input.

$$y = \sum_{i=1}^{N} \gamma_i(z)\,(a_i^T z + c_i) \tag{11}$$

Here *z* is the input, a_i the consequent parameter, c_i is the bias and γ_i is given by (12).

$$\gamma_i(z) = \frac{\prod_{j=1}^{p} \exp(-(z_i - g_j^i)^2(z)/2\delta_{ij}^2)}{\sum_{i=1}^{N}\prod_{j=1}^{p} \exp(-(z_i - g_j^i)^2(z)/2\delta_{ij}^2)} \tag{12}$$

The TS follows a hybrid training process with a least square estimator and gradient descent method. This method starts with finding an optimal number of rules and then partitions the input space to be equally divided, with widths and slopes of functions having enough overlaps. There is a forward and backward pass in the training. In the forward pass, the neuron outputs calculated from input data are used to determine the consequent parameters. In the backward pass, backpropagation is applied and the error signals are used to update the antecedent parameters.

2.5.2. Optimally Pruned Extreme Learning Machines

After the introduction of ELM by *Huang et al.* in the mid- 2000s. *Miche et al.* proposed optimal pruning to improve the technique. This improvement was to address the drawback in approximating correlated and irrelevant variables included in the training set [77]. The optimal pruning is achieved by marginalising the irrelevant neurons of the ELM built network. To illustrate how the model learns, suppose there is a training set z_i, where *i=1,…..n*, with a target vector t_i the ELM's goal is to decrease the training error function *E* to be as low as possible. The ELM can then be represented by (13).

$$\sum_{j=1}^{k} f(w_j, b_j, x_i)\beta_j = t_i \tag{13}$$

Here w_j is the input weight vector that connects the *jth* hidden neuron and the input, β_j is the output weight that connects the *jth* hidden neuron and the output and the *jth* hidden node's bias is represented by *b*j. If the ELM model can estimate the data sample with an error of zero, that is $\sum_{j=1}^{n}\|y_i - t_i\| = 0$, a w_j, b_j and β_j exist such that $\sum_{j=1}^{k} f(w_j, b_j, x_i)\beta_j = y_i$, *i=1,….n*. Equation (13) can thus be re-written in a compact form as (14):

$$H\beta = T \tag{14}$$

$$H = \begin{bmatrix} f(w_1, b_1, x_1) & \cdots & f(w_k, b_k, x_1) \\ \vdots & \cdots & \vdots \\ f(w_1, b_1, x_n) & \ldots & f(w_k, b_k, x_n) \end{bmatrix}_{n \times k} \tag{15}$$

H is the hidden layer output matrix, which can be written as (15). It has been demonstrated that the input weight and hidden bias do not require tuning [82]. Hence, *H* can be left unchanged after assigning random values to it at the beginning of the training. If *H* is a square matrix, hidden nodes can be randomly assigned and the output weights can be calculated through the inversion of *H.* The ELM can thus estimate the data sample with an error of zero. *H* is, however, in most cases not invertible due to not being a square matrix. In this case, a w_j, b_j and β_j such that $H\beta = T$ may not exist. The ELM training process then corresponds to solving a least square problem. Here, the hidden layer output matrices' generalised inversion can be utilised to determine the ELM weight between the output and hidden layer, unlike in conventional Neural Networks. This operation is called the Moore-Penrose. The weights are then given by (16):

$$\beta = H^{*}T = (HH^{T})^{-1}HT^{T} \tag{16}$$

Where H^* is matrix *H's* Moore-Penrose generalised inverse [93]. The OP-ELM training algorithm can be summarised in five key steps:

Step 1: Assignment of weights (w_j) and bias (b_j,) at random, *j=1… k*

Step 2: Determine *H,* the hidden layer's output matrix

Step 3: Determine β*,* the output weight using (13)

Step 4: Use multi-response sparse regression (MRSR) to rank the neurons

Step 5: Use the Leave-One-Out validation method to select an optimal number of neurons

2.5.3. Deep Belief Networks

A restricted Boltzmann machine (RBM) is a generative stochastic neural network model that learns the probability distribution over the inputs [35]. An RBM has a Boolean layer of hidden neurons and a binary-valued layer of neurons [94]. A basic structure of the RBM is given in Figure 5 [95]. Here the RBM consists of a visible layer, *n*, with *j* units and a hidden layer, *h*, with *i* units.

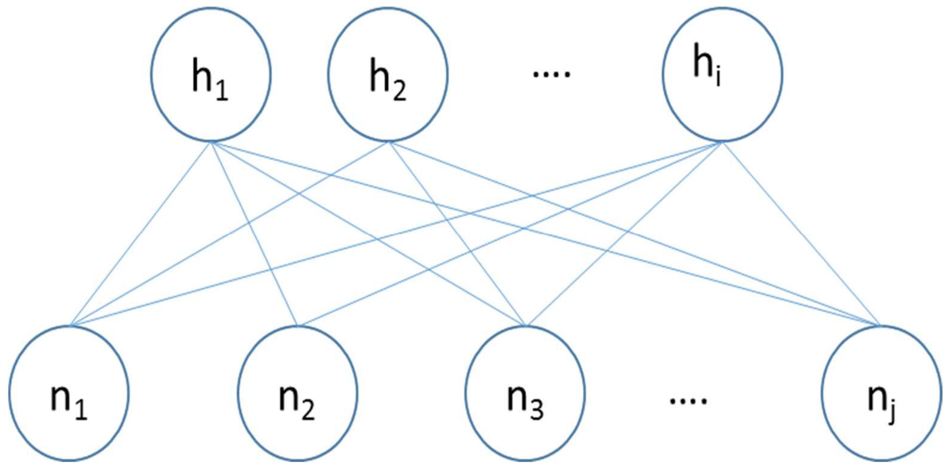

Figure 5: An RBM basic structure

DBNs were introduced by Geoffrey Hinton [96]. DBN uses RBM as its building block. RBMs are stacked to form DBN layers, with the hidden input of a previous layer (*n-1*) being the visible input of the layer of the next layer (n). Figure 6 shows an illustration of an *n*-layered DBN structure [95]. The DBN has no connection of neurons in the same layer. However, there are symmetrical and bidirectional connections between the layers [94]. The joint distribution over the hidden and visible units is defined by (17).

$$P(n,h) = \frac{e^{-E(n,h)}}{\sum_n \sum_h e^{-E(n,h)}} \quad (17)$$

Where *E(n,h)* is the energy function and is expressed as (18).

$$E(n,h) = -\sum_{j=1}^{k_n} \alpha_j n_j - \sum_{i=1}^{k_h} \beta_i h_i - \sum_{j=1}^{k_n} \sum_{i=1}^{k_h} h_i W_{ij} n_j \quad (18)$$

In (17) and (18) W_{ij} is the weight matrix of the links between the visible unit, n_j, and the hidden unit, h_i, α_j and β_i are the respective biases for the two layers. The hidden and visible unit conditional probabilities, given that they are conditionally independent are given in (19) and (20). If the hidden and visible unit's values are limited between 0 and 1, the conditional probabilities are respectively given by (21) and (22). Where *Sigmoid()* presents the logistic sigmoid function [95].

$$p(n|h) = \prod_j p(n_j|h) \quad (19)$$

$$p(h|n) = \prod_i p(h_i|n) \quad (20)$$

$$p(n_j = 1 | h) = sigmoid(\alpha_j + \sum_{i=1}^{k_h} W_{ij} h_i) \tag{21}$$

$$p(h_j = 1 | n) = sigmoid(\beta_i + \sum_{j=1}^{k_n} W_{ij} n_j) \tag{22}$$

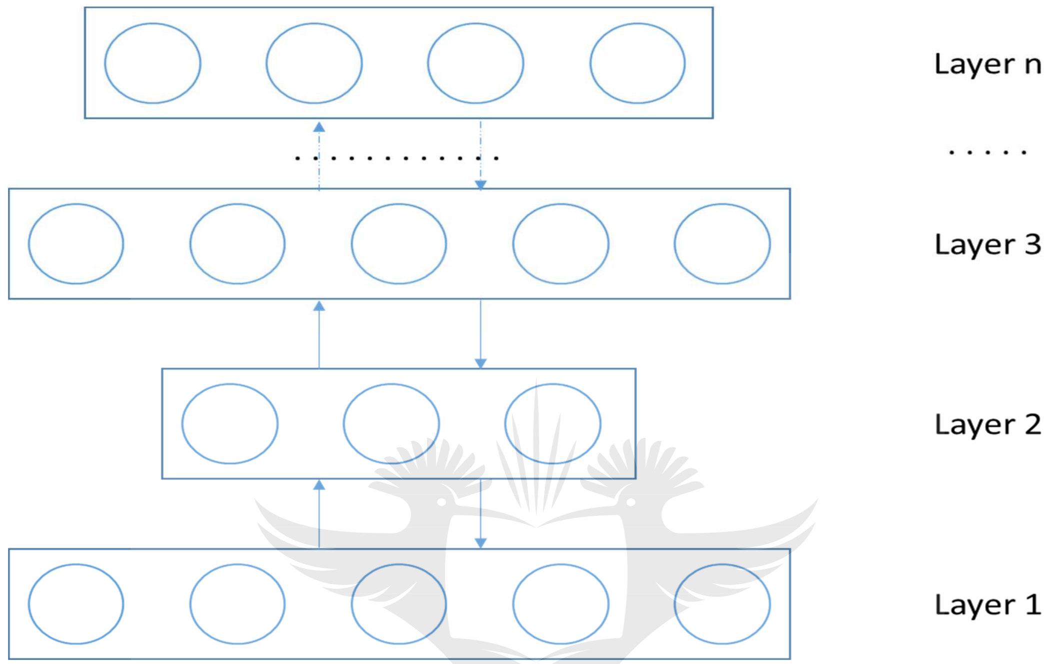

Figure 6: An illustration of a DBN with n layers

The training of a DBN can be summarised in two steps:

Step 1. (Pre-Training): Unsupervised learning where the DBN is trained by contrastive divergence to reduce the set of features. Here the input data is used to determine the visible and hidden state. The model parameters' (weights and biases) initial values are also determined in this step. This step is also known as greedy learning.

Step 2. (Fine Tuning): The model uses supervised training to train an appended layer to the pre-trained network in Step 1. The supervised training is achieved using backpropagation.

Optimisation techniques, such as scaled conjugate gradient algorithm, Levenberg-Marquardt (LM) algorithm and one step secant algorithm, can be used in Step 2 to optimise the training [97].

2.5.4. Long Short-Term Memory Recurrent Neural Network

Recurrent neural networks (RNN) are artificial intelligent techniques that use temporal information to estimate their output. The unit achieves this by having an edge, called a recurrent edge, which is connected to an adjacent time step. A basic structure of an RNN is given in Figure 7. By stacking several recurrent neural network units, a deep recurrent neural network is attained. RNNs' foundation was introduced in the early 1980s by John Hopfield [98]. They became more popular around the late 1980s and early 1990s [99], [100].

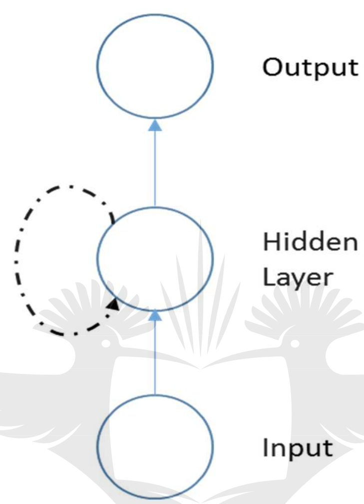

Figure 7: An RNN Basic structure

LSTM-RNNs were introduced to overcome the long-term dependency challenges experienced in other RNN architectures. This challenge can be described as the fading of previously learned patterns with time. The LSTM has a memory cell that enables it to keep its memory over time. Non-linear gating units manage the memory cell's information flow. The LSTM updates and erases the internal state vectors through the interaction of the input z_t and previous step's output, with the intermediate state, h_{t-1}, and cell state, s_{t-1}, where t is the time step. The structure of an LSTM block is given in Figure 8 [101]. The non-linear gates (input gate (i_t) and output gate (o_t)) and the forget gate (f_t) are respectively given by (23) to (25). Equation (26) gives the input node (g_t).

$$i_t = \sigma(W_{iz}z_t + W_{ih}h_{t-1} + b_i) \tag{23}$$

$$o_t = \sigma(W_{oz}z_t + W_{oh}h_{t-1} + b_o) \tag{24}$$

$$f_t = \sigma(W_{fz}z_t + W_{fh}h_{t-1} + b_f) \tag{25}$$

$$g_t = \emptyset(W_{gz}z_t + W_{gh}h_{t-1} + b_g) \tag{26}$$

Where σ, is the sigmoid function, W_{iz}, W_{ih}, W_{oz}, W_{oh}, W_{fz}, W_{fh}, W_{gz}, and W_{gh} are the network's activation functions' corresponding inputs weight matrices, and $\emptyset$ is the tanh function. The state and cell state at time step t are shown in (27) and (28), with $\odot$ representing element-wise multiplication.

$$s_t = g_t \odot i_t + s_{t-1} \odot f_t \tag{27}$$

$$h_t = \emptyset(s_t) \odot o_t \tag{28}$$

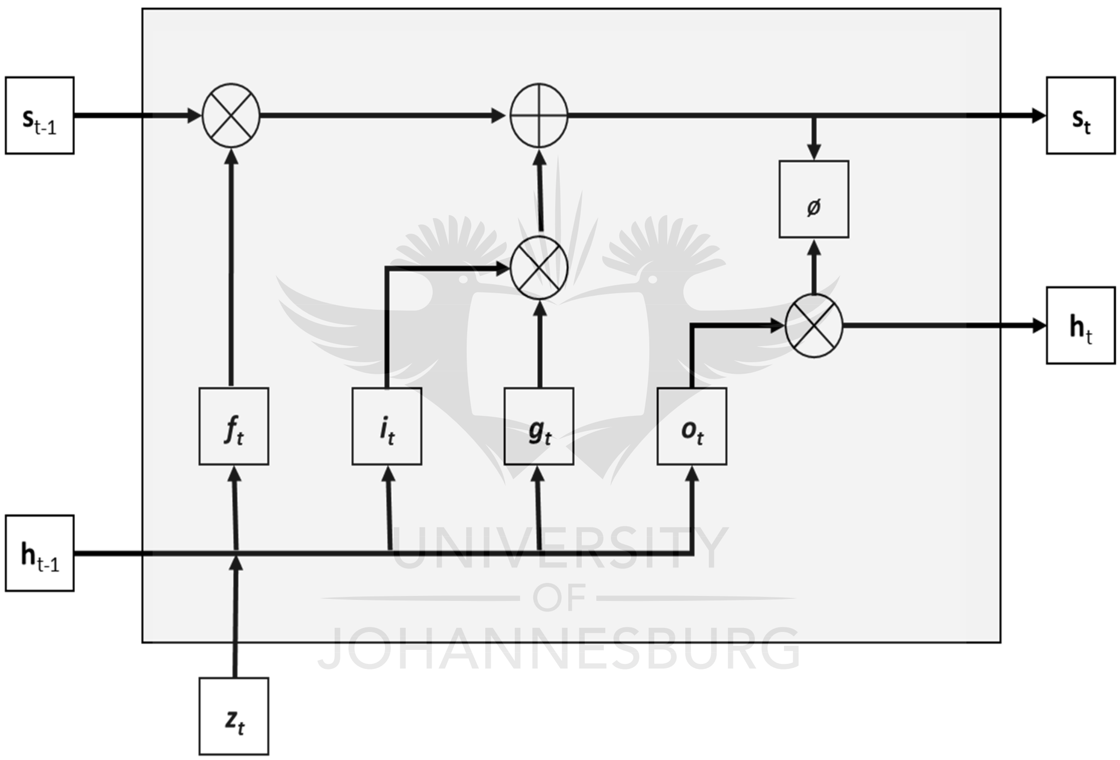

Figure 8: A LSTM unit

2.6. Forecasting, Prediction, Classification and Regression

Prediction is defined as determining a value or range of values that a certain sample is likely to have [42]. Forecasting is a prediction of future values over time. If it does not involve time it cannot be termed forecasting [27]. Classification and regression are regarded as prediction problems; with the difference, being classification is used to predict an individually distinct value and regression predicting a continuous value. An example of classification is a program being able to classify a German Shepherd as a dog after observing many images of different dog types. The problem studied in this research is a regression problem, where the load is being predicted over a period of time. Because the load prediction involves time, it is termed load forecasting.

2.7. Chapter Summary

This chapter presented an overview of AI. The concepts of knowledge, intelligence and learning were also presented. A literature review of machine learning and deep learning techniques, with some of the applications of these techniques across different fields, was presented. The machine learning (ANFIS and OP-ELM) and deep learning (LSTM-RNN and DBN) techniques used in this research were also presented. Some of their structures were presented, along with their basic form's advantages and shortfalls, and how these shortfalls have been overcome. Forecasting, prediction, classification and regression were also defined in this chapter.

Chapter 3 – Proposed Load Forecasting System

"The best thing about the future is that it comes one day at a time."

Abraham Lincoln

3.1. Introduction

The previous chapter introduced the AI concepts, i.e. knowledge, intelligence and learning. A machine learning and deep learning literature review was also presented in chapter 2. The literature review presented the application of machine learning and deep learning techniques in various fields including science and engineering. Machine and deep learning techniques used in this study were introduced, together with their advantages and disadvantages.

This chapter presents the proposed load forecasting system. The system implements a data integrity analysis module, which uses fuzzy logic to determine loading data integrity before training load forecasting models. The data classified to have low integrity raise an alarm for further investigations and system repairs. This is useful for a utility preparing for the fourth industrial revolution, to have data with high integrity, which they can use to draw insights from and to optimise their processes. The experimental approach is also presented in this chapter. The three performance measures (sMAPE, MAE and RMSE) used to measure the performance of the four AI techniques' load forecasting models are also presented. The statistical significance test is also presented. This test measures the significance in the difference the load forecasting results obtained.

3.2. Determining Distribution Network Data Integrity

A power system network consists of multiple measuring devices that measure different parameters such as voltage, current, active power and reactive power. The measurement of these parameters allows for the monitoring of a power system network. These measurements are usually also stored for utilisation in other applications such as network optimisation, and network planning. This data can be seen as big data. Big data can be defined as data that move fast, do not fit conventional database architecture or are too big [102]. Big data have also been defined as "*high-volume, high-velocity, high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making*" [103]. For this big data to be applied they have to be usable. The data can however have errors which not only affects operations, but also makes it complicated to use the data. As an example, metering and recording errors have led to non-technical losses [104]. Researchers have conducted work in power systems measurements' accuracy. However, the focus of their studies was on the measuring devices and the algorithms used in the measuring technique [105], [106], [107], [108]. It has however been shown that data integrity challenges can happen beyond the measuring point [109]. Figure 9 illustrates how data attacks at different parts of the network can lead to network challenges such as non-optimal power flow [92]. Data attacks can be defined as accidental faults, such as hardware or software glitches, or malicious third party interaction with the data [92] . The end user may not be aware of such attacks and may need to evaluate the integrity of the data before utilising it. In [110] data line faults, out of bounds fault, data lost faults and spike faults were presented as four key error scenarios. Various methods have been applied to protect data integrity. These methods also included looking at data before getting to the end-user [111], [112], [113].

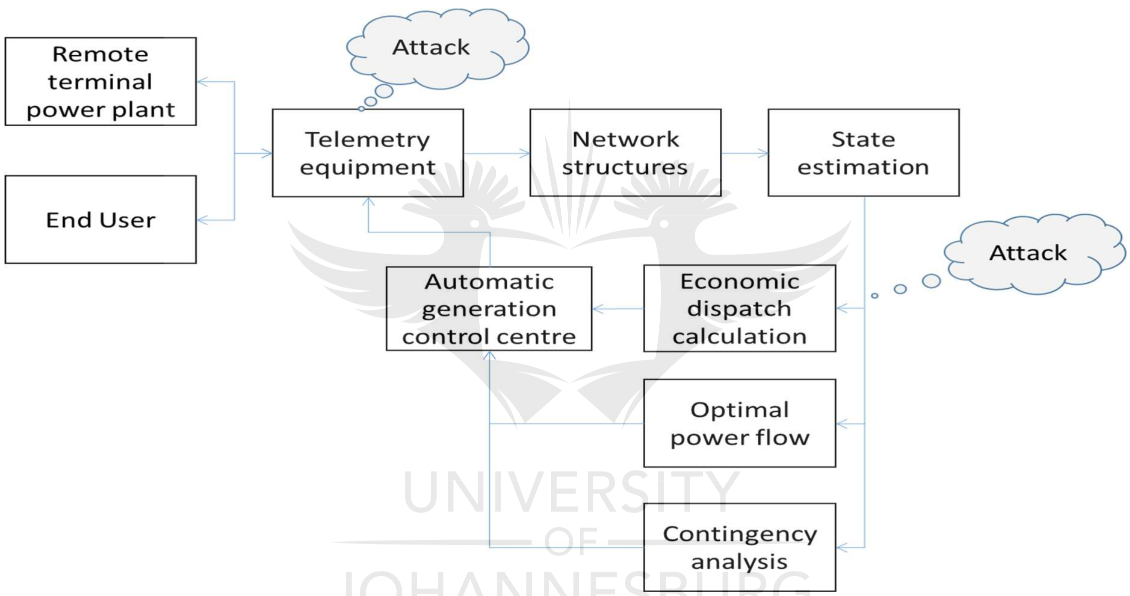

Figure 9: Optimal power flow data integrity attacks

Fuzzy logic was used to determine the data integrity of the loading/power consumption data for the power system network in which the substations used for case studies in this research are located. The data integrity was determined with two aims: the first was to determine which distribution networks had the highest data integrity for potential case studies in this research and the second was to enable the utility to be aware of areas with data integrity issues. This second aim was to enable the utility to investigate these issues further and address them. The use of an AI technique, fuzzy logic, was to enable the process of determining the data integrity to be less time-consuming as opposed to manually analysing the data. The Mamdani-type fuzzy model was used for the experiments. The model used three variables – number of zeroes (data lost faults), very large values (spikes and out of bounds fault) and a number of different entries of the same variable in the same file, to classify the loading data to either have high integrity, low integrity or undetermined. The data used were from a customer network centre (CNC) that had 36 substations. CNCs are demarcated areas in which certain parts of power system networks fall under. The transformer or feeder loading data were contained 201 files. The loading data were stored as 30 minutes average power consumption (active and reactive power) values. An annual loading file therefore had 17250 line entries. Each file had eight variables, which included the date, time, recorder id, etc. These files were between 2 MB and 5 MB in size, with a CNC's loading data being therefore between 400 MB and 1 GB for a year. Looking at a province or country level, and at other measurements such as voltage or current over longer periods, one can see how power distribution networks' data can be regarded as big. The model used had to be able to determine the data integrity of different load types. Figure 10 shows a week's load profile of both an industrial and a residential customer [92]. It can be seen from Figure 10 that industrial customers typically have a large power demand and which has a less periodic profile, as opposed to residential load types. The model rules were created and the fuzzy logic model developed.

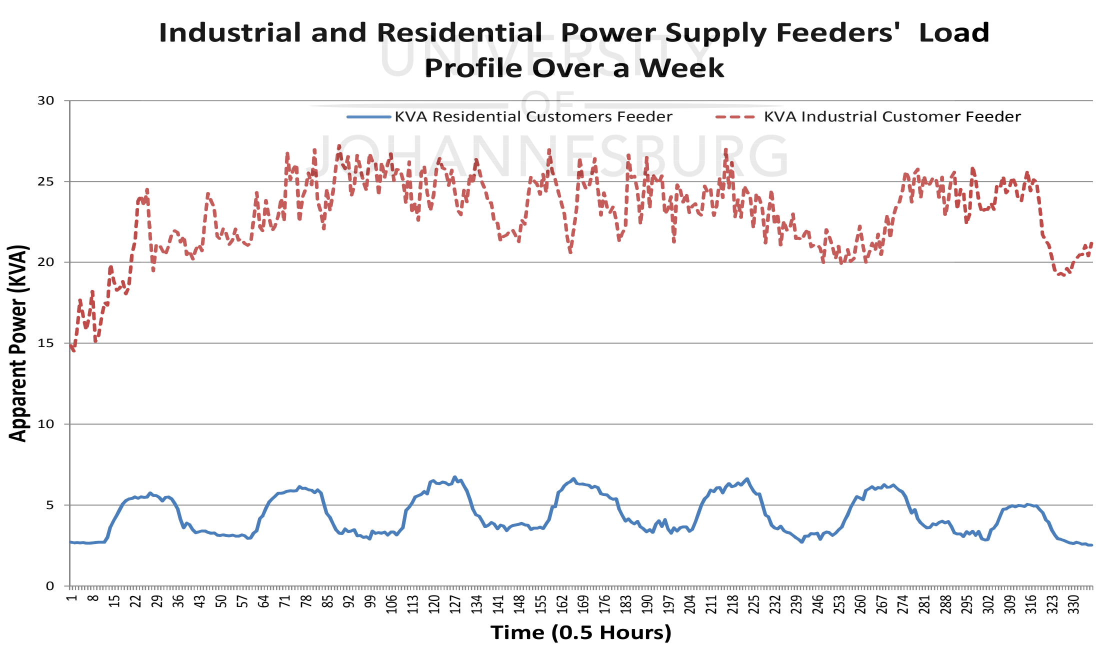

Figure 10: One week's load profile of an industrial as well as a residential customer

Figure 11 shows the process followed to determine data integrity in the experiment. This process can be summarised in six key steps as follows:

Step 1: Download data from the server to store locally

Step 2: Determine the number of files, read and store file names

Repeat Step 3 to Step 6 until all files are analysed

Step 3: Load data files

Step 4: Normalise data

Step 5: Evaluate data integrity using the Fuzzy Logic model

Step 6: Store results in excel file

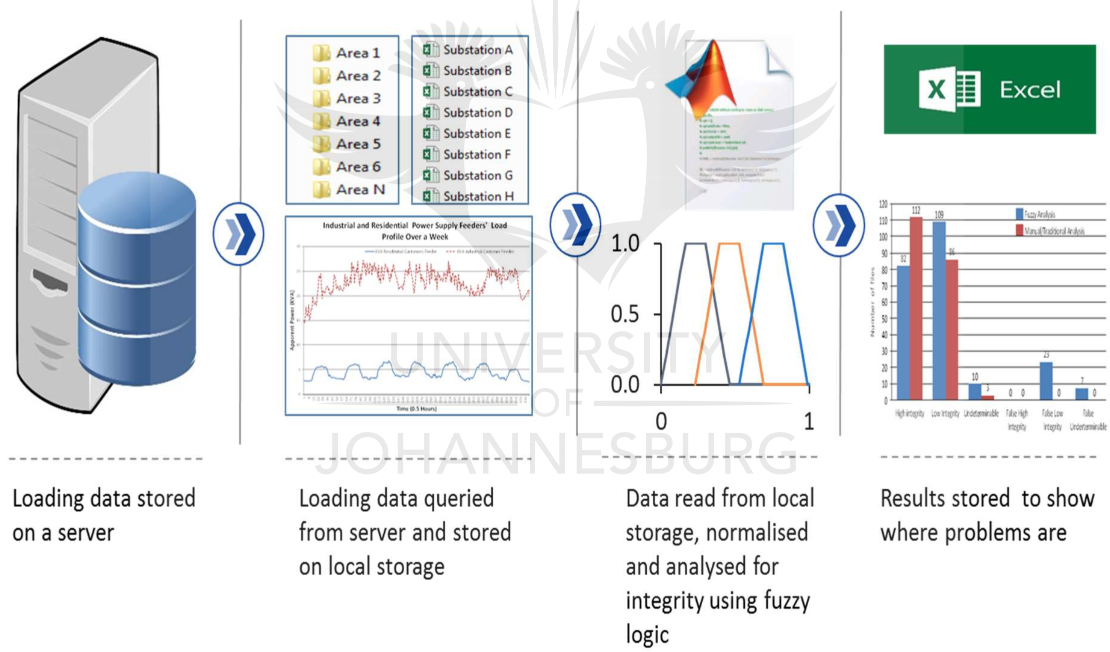

Figure 11: High-level process followed to determine distribution loading data integrity

The CNC data were analysed and the results captured. The results were compared to those obtained from analysing the data manually. The manual analysis involved visual analysis (such as analysing data plots), load balancing and cross checking of files. This manual analysis did not include looking the power operations activity logs. The results are shown in Figure 12 [92].

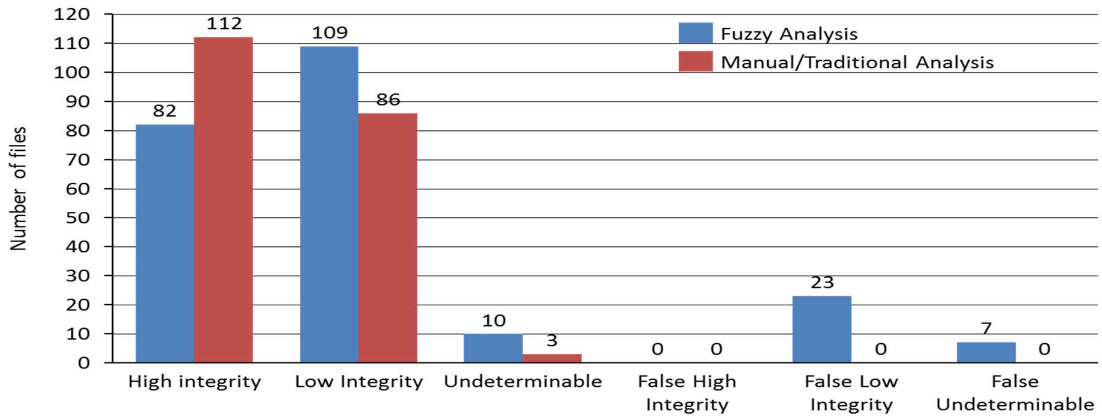

Figure 12: Distribution network loading data integrity analysis

The model falsely classified 11 % of the files as having low integrity and 3 % of the files were falsely classified as undeterminable. The main reason for these false classifications were found to be n-1 or n-2 firmness of certain substations. This setup was more prevalent in industrial customer's dedicated substations. Here a customer typically has multiple transformers that supply different customer's sub-distribution loads. Certain transformers are able to shift their loads to another transformer(s) or there is a backup transformer for all the installed transformers.

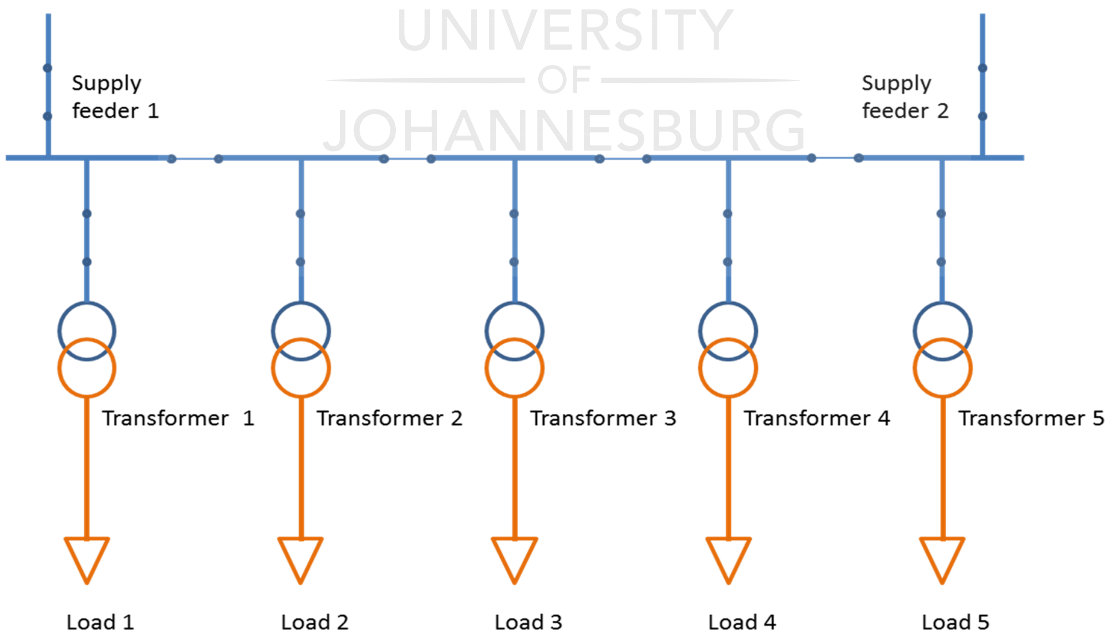

Figure 13: Example of industrial customer substation setup

Figure 13 shows such an example, where transformers 1 and 2 can supply the same loads on the customer side, with transformers 3 and 4 also supplying the same load [114]. Transformer 5 acts as a backup that supports both transformer sets. This leads to the backup transformer not being loaded at times, and hence a lost data (zeroes) integrity issue gets picked up in the AI model's analysis. This is despite this issue being a result of the power system's normal operation and not being a real data integrity issue. To have the ability to detect that such setups with n-1 or n-2 firmness exist, file crosscheck mechanism can be added. Here a corresponding feeder and transformer were picked up using their names, and their data were analysed to confirm interrelation and similarity in the data analysis results. In cases where the feeder and transformer data correspond, the data were re-classified as being accurate. From the example used here, this check would show that the loading on transformer 5 and that on load 5 correspond, and hence their data despite being classified as having low integrity, it do not. This approach took approximately 1.3 seconds to perform an analysis on a single file using MATLAB. The manual analysis in comparison took 10-15 minutes. The analysis was conducted on an Intel i5 (2.5 GHz) PC with 4 GB RAM. This approach can therefore be deployed to determine loading data accuracy, and thus save utility engineers' and technicians' time.

3.3. The Proposed Load Forecasting System

The proposed AI load forecasting system is given in Figure 14. The loading data are collected from field equipment via power meters. The power measurements are logged from the field meters to a database via fibre or microwave communication channels. The data integrity is then determined using the approach presented in Section 3.2. The data classified to have low data integrity or classified as undetermined are analysed further to determine the faults' root causes leading to low data integrity. These root causes are then addressed. The AI system uses machine learning and deep learning techniques, deployed in the AI load forecasting module, to conduct the load forecasts. The models are trained and conduct these load forecasts for substations that have feeders and transformers with loading data that have high integrity. These models should be trained and tested offline before being deployed in this system's AI load forecasting module. The load forecasts are used to run power systems simulations using simulation packages such as DigSilent Power Factory and PSS®E. The outcome of the simulations for the different networks inform the maintenance schedule, from a network loading perspective, when is the best time to schedule maintenance power outages. Other parameters such as resources, spares availability and customer extraordinary needs are then used together with the load forecasts to optimise the outage schedule.

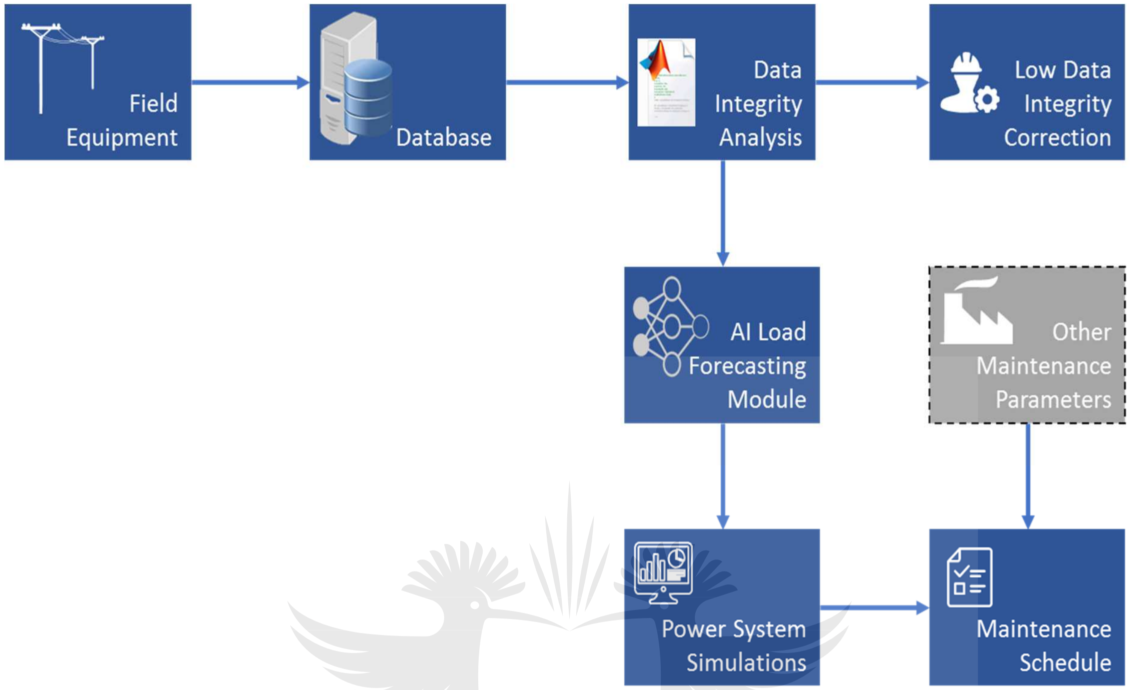

Figure 14: Proposed AI distribution network load forecasting system

3.4. Experimental Approach

Following the evaluation of the data's integrity, three networks with high data integrity were selected for the experiments. Two of these networks are supplied by the same transmission substation, but are supplied at different voltage levels. The third network is supplied by a different transmission substation. These networks will be described in detail in Chapters 4 to 6, respectively. The three different substations' power consumption data, also termed loading data, were collected for a period of 2012 to 2016 (August 2012 to May 2016, for case study 1, January 2012 to September 2015, in case study 2 and January 2012 to December 2015 for case study 3). The experiments' training and testing periods were the same in both case studies in this research. The apparent (total) power was used in this research. The data were stored on a database and could be downloaded to local storage for use. The loading was stored as 30 minutes average power consumption values. The data were normalised to be between *0* and *1* using (29).

$$z_{norm} = \frac{z - z_{min}}{z_{max} - z_{min}} \tag{29}$$

Where z in the input value being normalised, z_{min} is the minimum input value, z_{max} is the maximum input value and z_{norm} is the normalised value. Temperature data were requested from the South African Weather Services. These data were also normalised using (29). The input variables for training were separated into two groups for the each of the experiments as shown in Table 1.

Table 1: Experiment input variables

Input variables group	**Inputs**
Group A	Loading t-2 years Corresponding time of day Peak or non-peak period indicator Loading t-1 year Loading t-2 weeks
Group B	Loading data t-2 years Temperature t-2 years Load corresponding time of day Peak or non-peak period indicator Loading t-1 year Temperature t-1 year Loading t-2 weeks Temperature t-2 weeks

Temperature was used as an input variable in only one of the two input variables groups. This was to determine its impact on load forecast accuracy, per technique used. The other variables used in developing the models were the two-week loading data; two years before the target forecast period, a year before the target forecast period, and two weeks before the target forecast period. The corresponding time of day was also included as an input variable. The values are mapped to be between *0* and *1.* Here *0* corresponds to the value stored at *00:30*, the day's first 30 minutes average loading, and 1 corresponds to the value at *00:00*, which is the day's last 30 minutes average loading. An indicator for peak and non-peak periods was also used as an input variable, with a *0* indicating non-peak period and a *1* indicating the peak period. For this study the winter peak periods were taken as *06:00 to 09:00* for the morning peak and *18:00 to 21:00* for the evening peak. The training and testing were conducted for a winter period. Winter was chosen due to its favourability for maintenance by the utility mainly due to reduced rainfall and thunderstorms. These conditions provide benefits such as ease of performing work in non-rainy conditions, ease of access to mountainous areas and gravel road and reduced lightning strike risks. The datasets used for training the models were different to those used in testing the models. The different techniques' models were trained differently based on their properties and their training procedure. The following subsections summarise how each of the models were trained in respect of the three cases studied.

3.4.1. ANFIS

A trial and error approach was used to experiment with ANFIS tuning parameters. These parameters determine the number of membership functions, overlaps, number of rules and the rules. Four main parameters were established and are used to report the results in this document. This was due to the similarity in their results, in relation to other parameters experimented with. These four parameters were used for training all ANFIS models in this research.

3.4.2. OP-ELM

For OP-ELM the key parameter that was tuned was the model's dimensions. This tuning was conducted through adjusting the model's number of hidden nodes. The model's performance was then tested. Due to optimal pruning the ultimate number of hidden nodes is determined from the leave one out method, and the exact number of hidden nodes cannot always be obtained across different corresponding sub-experiments. However, for consistency and comparison of such cases, the closest number of hidden nodes obtainable were used in the corresponding experiments.

3.4.3. DBN

A single layer DBN was used in the experiments, with the number of hidden units being varied. The results will show why more layers were not explored, based on the performance with an increase in the number of units and versus the other techniques explored.

3.4.4. LSTM

LSTM models were trained with the number of hidden stacked LSTM units being varied per sub-experiments. The different corresponding sub-experiments in both case studies had the same number of hidden units.

3.5. Performance Measures

To measure how well an entity performs, different performance measures can be used. The performance measure used depends on the specific activity whose performance is of interest. In Section 1.3 it was mentioned that the performance of a distribution network can be measured in terms of the network's availability and the number of customer interruptions, using SAIDI and SAIFI respectively. Generation units' performance can be measured by their efficiency, which can be a measure of the electrical energy output in relation to the input thermal energy [115]. In AI application for classification, the aim is usually to correctly classify what is being presented to an AI model. The AI model may need to correctly classify a signal, a picture, an electrical fault, etc. Here, what one would typically measure is how well the AI model classifies the intended object. *Hasan* used four performance measurements to determine how well AI models predicted dam water level and energy consumption by classification [2]. The author used classification and misclassification accuracy as primary performance measures. He used the RMSE and mean square error as secondary performance measures. *Taigman et al.* used the mean recognition accuracy as a performance measure to determine how well their AI model, DeepFace, was able to classify human faces [37]. In forecasting, one would want to measure how accurately his/her model can predict the variable of interest over time. A common way to achieve this is to measure the error [1]. The common error measurements used in load forecasting studies are RMSE [116], [117], [118], MAE [116], [117], mean absolute percentage error (MAPE) [117], [119], [120], sMAPE [21], [121], and mean percentage error (MPE) [120]. Even though it is not common, some authors have also used the normalised versions of these error measurements. *Dong et al.* used normalised root square mean error and normalised mean absolute error to measure CNN performance in load forecasting [117]. The sMAPE is preferred over the MAPE due to challenges that affect the traditional MAPE. These are challenges such as: unequal errors being obtained when the forecasted value is greater than or is less than the target value by the same absolute value and the percentage error can become very large when the target value is very small [121], [122].

Three error measurements were used to determine the performance of the load forecasting models in this research. These measurements were the sMAPE, MAE and RMSE. These error measurements are, respectively, given by (30) to (32) [95].

$$sMAPE = \frac{2}{N}\sum_{k=1}^{N}\frac{|Y_k - T_k|}{|Y_k| + |T_k|} \tag{30}$$

$$MAE = \frac{\sum_{k=1}^{N}|Y_k - T_k|}{N} \tag{31}$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N}(Y_k - T_k)^2}{N}} \tag{32}$$

Where Y_k is the k^{th} forecasted value, T_k is the k^{th} target value and N is the total number of forecasts. The sMAPE in (30) calculates an error between 0 % and 200 %, after multiplication by the 100 % factor. To get a percentage value between 0 % and 100 %, the "2" is removed before multiplying with the 100 % factor.

3.6. Statistical Significance Test

When using performance measures stated in section 3.6 to determine the performance of AI techniques' models, different levels of errors are obtained. The residual differences may reflect particular sample differences as opposed to that of the population data it is sampled from [27]. The importance of a statistical significance assessment of the difference in the error is thus evident. One of the ways to conduct this test is through a t-test. This test checks if two samples are from the same population using the variance and the mean of the two samples. A significance value, also called a *p*-value, is calculated from the test. An acceptable value of p in academic studies is $p = 0.05$. If the p-value is less than 0.05 there is a significant difference in the samples being compared and thus the model with the lowest error has the best performance.

3.7. Chapter Summary

This chapter presented the proposed AI distribution load forecasting system. The load forecasts from the load forecasting module are inputs into maintenance planning/scheduling. The system also has a data integrity analysis module that uses fuzzy logic to determine the integrity of loading data, before the data go into the load forecasting model. The data classified by this model to have low integrity raise alarms for further investigations and system repairs. The experimental approach followed in this research was also presented. Three performance measures used to measure the performance of load forecasting models in this research were also presented. These measures are the sMAPE, MAE and RMSE. The statistical significance test was also presented. This test measures the significance in the load forecasting performance difference of the different techniques' models.

Chapter 4 - 1st Case Study: Distribution Substation A - Power Redistributor Load

"The African renaissance is about creating a new Africa, about ending poverty and oppression, regaining dignity."

Thabo Mbeki

4.1. Introduction

The proposed AI distribution network load forecasting system was presented in the previous chapter. Different components of this system were also presented. These components included the module that determines the loading data's integrity. The experimental approach, together with the performance measures used in this study were also outlined.

This section presents a case study of a power redistributor customer. The customer's substation, therefore, has different types of customers connected to it. This section gives an overview of the distribution network that this substation is a part of. An overview of the substation under study is also presented. Four techniques were trained and tested using the substation's historical power consumption/loading data. These historical data are also described in this section. The load forecasting test results for the different machine learning and deep learning techniques are also presented and discussed.

4.2. Case Study A Distribution Network Overview

The distribution substation under study is located in the North-Eastern part of South Africa. The substation was commissioned in 2012, and had loading data available from August 2012 until May 2016. This relatively recent commissioning date further illustrates the country's electrification drive. The substation is connected to the grid via a 275 kV main transmission substation's 88 kV lower voltage network side. This connection is at a T-off section on a line towards another 88 kV distribution network. The network is given in Figure 15 with the lines representing three phase systems [101]. The network is interconnected with other networks and generation sources via main transmission substations connections and other distribution networks. These interconnections are not fully shown in Figure 15. The substation under study is an 88/11 kV, 80 MVA substation with two 40 MVA transformers. The customer has their own distribution network that reticulates power to their different customers. There are power measuring devices on different parts of the substation. The key points of measurements are: the incoming feeder, the transformer primary sides and point of supply on the lower voltage side. The measuring point utilised in this study was the one on the main supply feeder. This was to forecast the substation's total load. The substation under study is also referred to as Substation A in the remainder of this document.

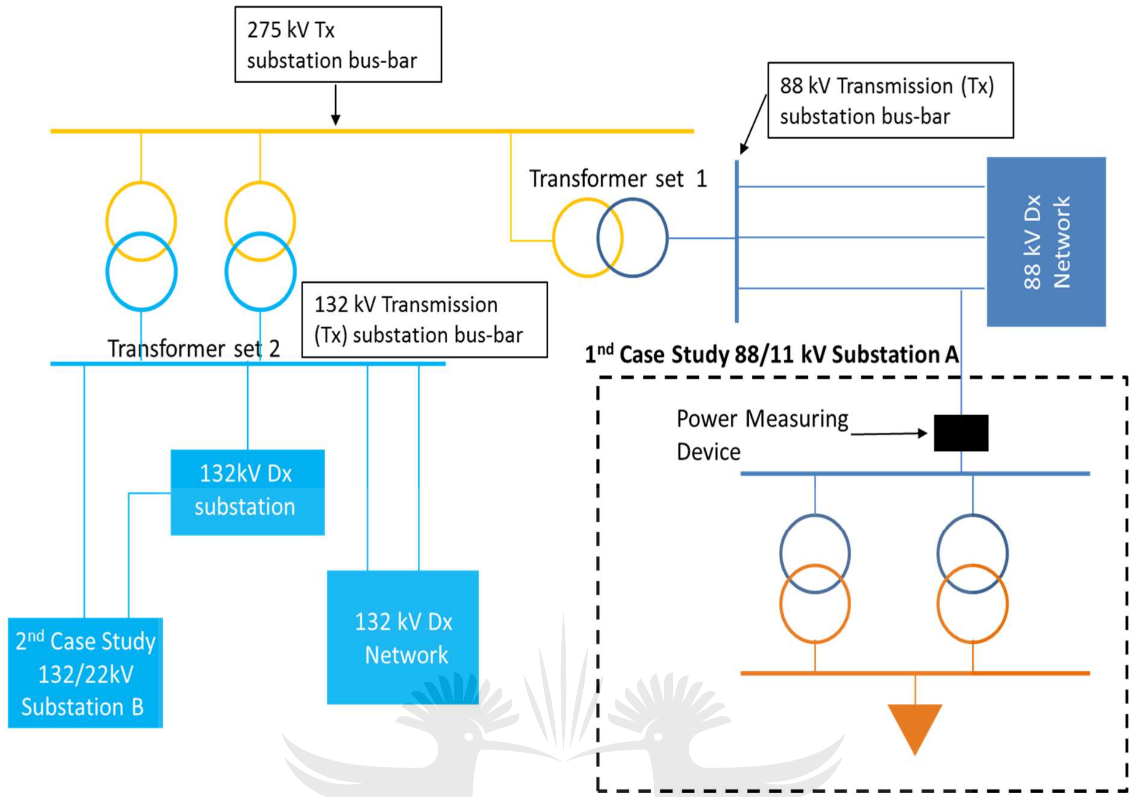

Figure 15: The distribution network where the redistributor substation under study is located

4.3. Data Description

As mentioned in Section 4.2., the data were collected for the period from the year 2012 to 2016. The loading data were observed to have a period of low loading just after commissioning in August 2012. This was observed to have only occurred in this period. The data in this period were disregarded from this study and all data processing excluded it. The data were also observed to have dips in them, which were where the load went to zero. The actual reasons for these dips were not known with certainty, but could have potentially been from trips or planned breaker operations. It was unknown if the data should be used as is or have these dips and spikes cleaned out. The impact of cleaning these dips out on the forecasting performance was also unknown. Cleaning out these dips can also be cumbersome and time-consuming, and counter-progressive when working toward making the load forecasting system have limited human intervention. Part of the experiments was then to evaluate the impact of the data clean-up versus non-cleaned up/raw data. The data were, hence, cleaned up to remove these dips. These zero values were replaced with the previous week's corresponding time's values. The raw loading data and cleaned data are shown in Figure 16 and Figure 17.

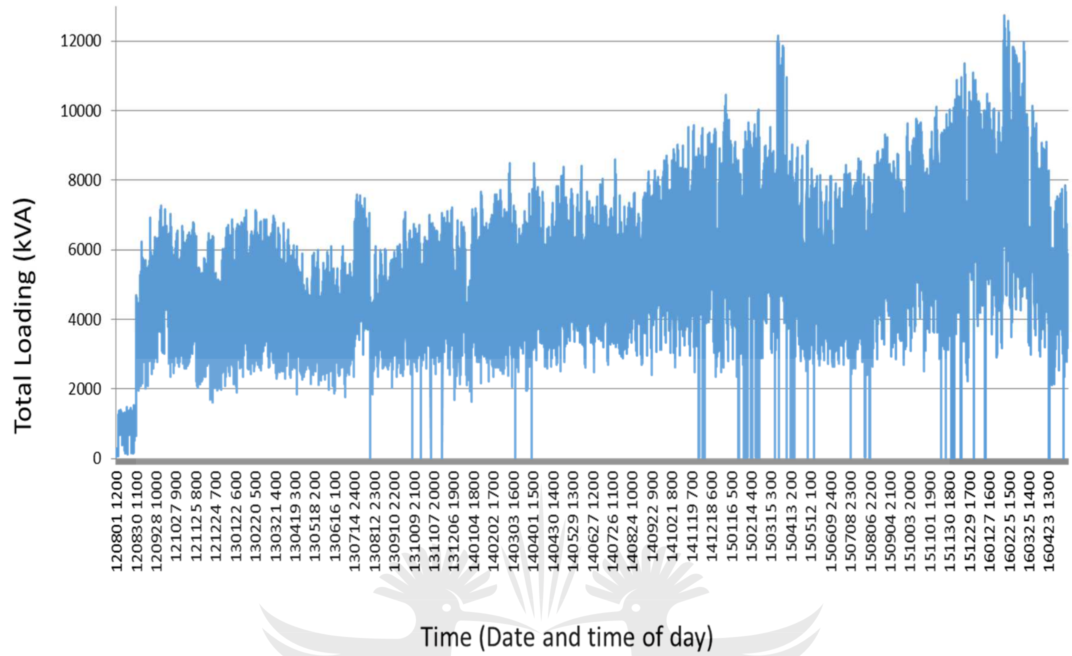

Figure 16: Plot of Substation A's incoming feeder raw loading data

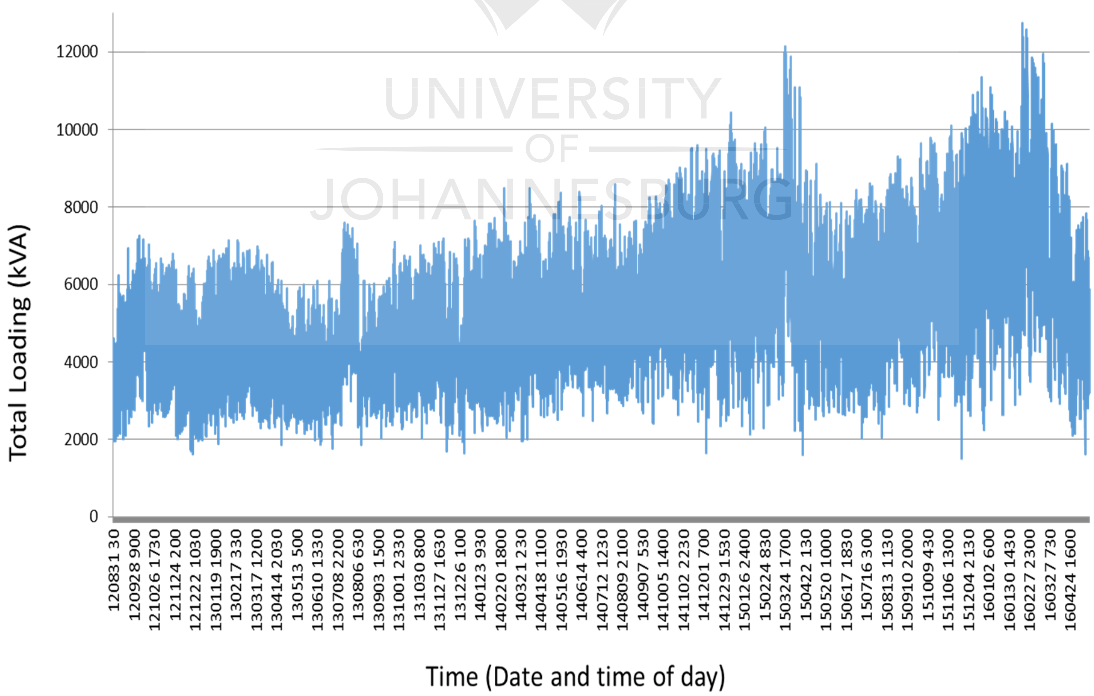

Figure 17: Plot of Substation A's incoming feeder cleaned-up loading data

The loading data had a periodic pattern that stood out when looking at the load profile over shorter periods. Figure 18 and Figure 19 show the daily and two-week load profile, respectively.

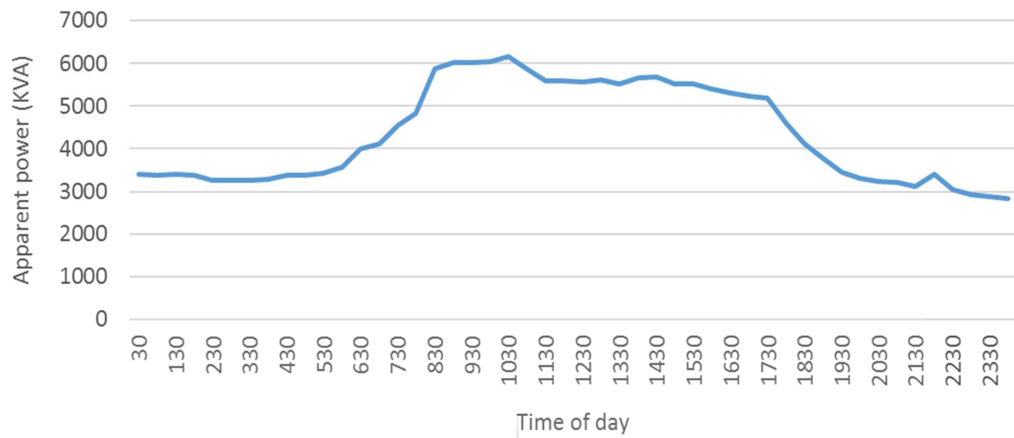

Figure 18: Substation A's day load profile for 15th June 2015

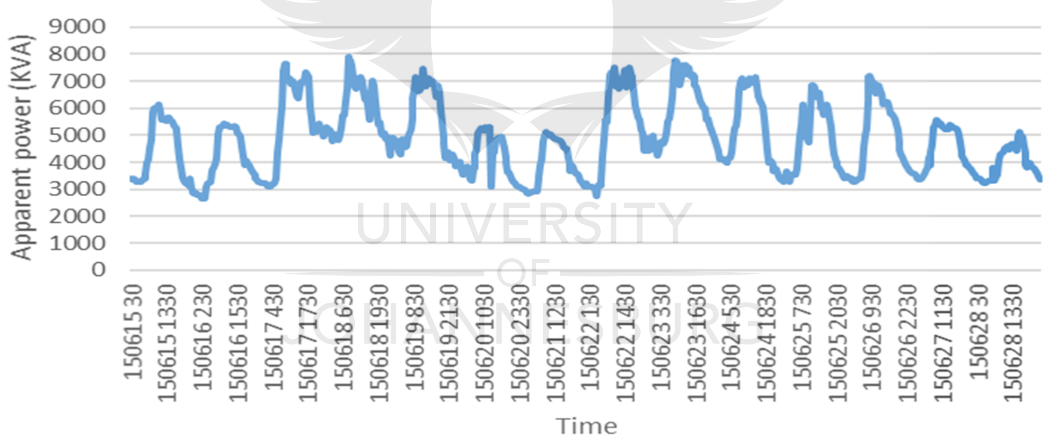

Figure 19: Substation A's two-week load profile for the 15th to 28th June 2015

The weather-related data were also requested from the South African Weather Services. The requested data included temperature, rainfall, humidity and wind speed. As explained previously, temperature is the most commonly used weather parameter in load forecasting studies. This research used it as the first weather parameter to investigate and establish if weather parameters have an impact on South African distribution load forecasting. Figure 20 and Figure 21, respectively show a daily and a two-week temperature profile corresponding to the load profiles in Figure 18 and Figure 19.

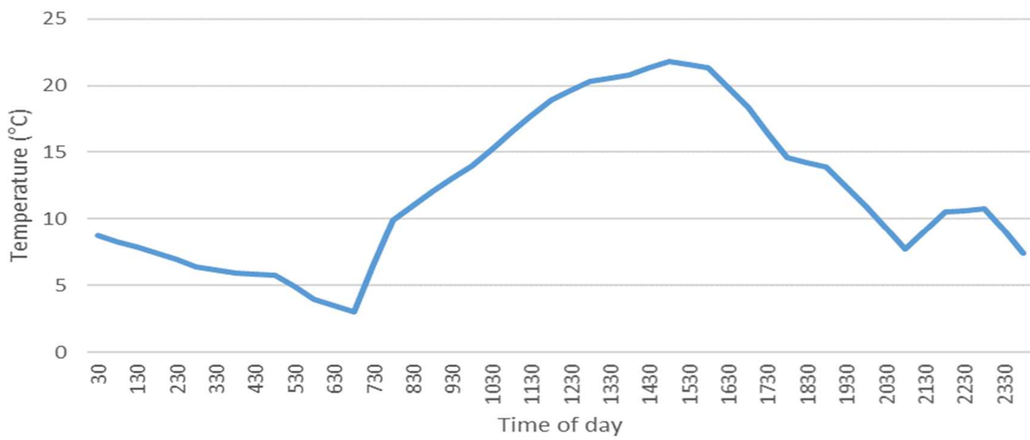

Figure 20: Day temperature profile for study location for 15th June 2015

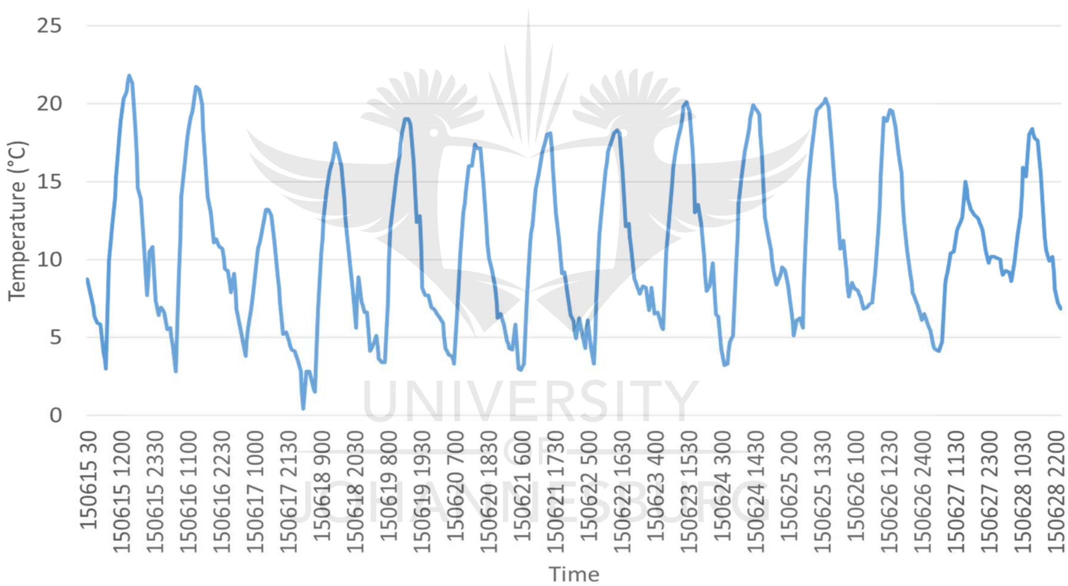

Figure 21: Two-week temperature profile for study location for the 15th to 28th June 2015

4.4. Experiment Results and Results Discussion

The different experiments were conducted following the approach described in Section 3.4. The load forecasting test results are presented in the following subsections.

4.4.1. ANFIS Results

The results showed that lower errors were obtained with the models trained with raw loading data. The first input variables group (Group A), which did not include temperature, was observed to lead to models with lower errors as compared to input variable Group B. This observation was in both cases with the cleaned and raw data. The lowest obtained test errors with raw and cleaned data are highlighted in Table 2 and Table 3, respectively. The lowest obtained errors as shown in bold in Table 2, were an sMAPE of 0.138483 (6.92 %), MAE of 0.052392 (5.23 %) and RMSE of 0.071799 (7.18 %). The lowest obtained errors by ANFIS models developed with cleaned loading data are bolded in Table 3.

Table 2: ANFIS models' test errors with raw loading data

Input variables	Model tuning parameters	Performance		
		sMAPE	*MAE*	*RMSE*
Group A	1	0.1495145	0.0567013	0.0804968
	2	0.17506	0.066914	0.092543
	3	**0.138483**	**0.052392**	**0.071799**
	4	0.1495145	0.0567013	0.0804968
Group B	1	0.1836643	0.0696844	0.0920817
	2	0.177639	0.068296	0.092137
	3	0.162643	0.05886	0.07956
	4	0.1968307	0.0728959	0.0964678

Table 3: ANFIS models test errors with cleaned-up loading data

Input variables	Model tuning parameters	Performance		
		sMAPE	*MAE*	*RMSE*
Group A	1	0.226631	0.064804	0.092812
	2	0.279851	0.078866	0.109888
	3	**0.207322**	**0.059294**	**0.081476**
	4	0.226631	0.064804	0.092812
Group B	1	0.280986	0.079387	0.104689
	2	0.27447	0.078297	0.104904
	3	0.243374	0.066815	0.090639
	4	0.295661	0.079239	0.105947

The statistical significance test was conducted between four sets of results as shown in Table 4, using the t-test. The aim was to determine if there was a significant difference or not between the different results obtained from the tests. The statistical significance test showed that there was no significant difference between the respective results, which gave lowest forecast errors, respectively, with and without temperature as input variable. This was an observed finding with both the raw and cleaned up data results, respectively. There was, however, a significant difference between the forecasted load results with the lowest error and those with the highest error, respectively for the models developed using raw data. The t-test also showed that there was a significant difference between the forecast results of lowest obtained forecast errors with the raw and the cleaned data. Therefore, in this case of a distribution power redistributor, ANFIS load forecasting models can be trained without temperature as one of the input variables. Since the lowest error was observed without temperature as an input variable and there was no significant difference to the lowest error results obtained with temperature as an input variable, the use of temperature should be avoided for this case. The loading data also do not need to be cleaned up to obtain a high load forecasting accuracy and should therefore not be cleaned.

Table 4: ANFIS models' load forecast t-test results for Substation A

Compared model's results	**P-value**
Lowest errors: Cleaned vs Raw data	0.0000000000
Cleaned data lowest errors: Input Group B vs Input Group A	0.3808930563
Raw data lowest errors: Input Group B vs Input Group A	0.1570281141
Raw Data: Lowest vs Highest error	0.0000006557

The plot of the two-week ahead test load forecast for the model that obtained the lowest error is shown in Figure 20 versus the target load for the test experiment. The load forecasts can be seen to follow the target load closely. However, some portions were overestimated and relatively fewer parts were underestimated.

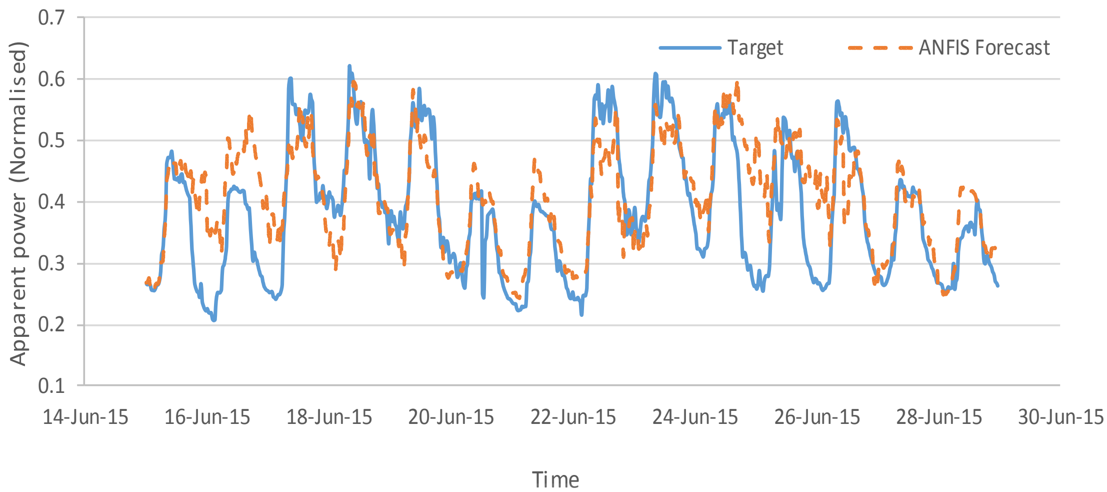

Figure 22: ANFIS lowest test error model's two-week ahead load forecast vs target load

4.4.2. OP-ELM Results

The lowest errors were obtained with raw data used for model development. The lowest error was obtained without temperature as an input variable. Lower errors were observed with a lower number of hidden units. The results are given in Table 5 and Table 6, with the lowest errors with the raw and cleaned bolded. The lowest attained errors were an sMAPE of 0.1315575 (6.58 %), MAE of 0.0491749 (4.92 %) and RMSE of 0.0657673 (6.58 %).

Table 5: OP-ELM models test errors with raw loading data

Input variables	**Hidden nodes**	**Performance**		
		sMAPE	***MAE***	***RMSE***
Group A	**10**	**0.1315575**	**0.0491749**	**0.0657673**
	55	0.1406914	0.0528443	0.0721915
	80	0.1542054	0.0568670	0.0779820
	100	0.1645267	0.0592804	0.0812588
Group B	8	0.1364757	0.0511859	0.0682780
	58	0.1596773	0.0587954	0.0754893
	103	0.1484154	0.0552860	0.0712436
	158	0.1793395	0.0681535	0.0885394

Table 6: OP-ELM models' test errors with cleaned-up loading data

Input variables	Hidden nodes	Performance		
		sMAPE	***MAE***	***RMSE***
Group A	**10**	**0.2014222**	**0.0562507**	**0.0752450**
	50	0.2171528	0.0601624	0.0818262
	100	0.2294828	0.0617002	0.0826178
	110	0.2319736	0.0641421	0.0835023
Group B	8	0.2004182	0.0564162	0.0759865
	58	0.2264838	0.0629379	0.0841188
	108	0.2645317	0.0714689	0.0917274
	158	0.2996661	0.0777273	0.0992270

The statistical significance test was conducted and the results are captured in Table 7. The results showed that a significant difference exists between all the compared load forecast results, respectively. There was therefore a significant difference between the lowest error results obtained with and without temperature as an input variable. This finding was observed for results with models developed using the raw and the cleaned data, respectively. The lowest and the highest load forecast errors results with raw data were also found to be significantly different. Therefore, for this case of a distribution power redistributor, OP-ELM load forecasting models should be developed without temperature as one of the variables. The loading data do not need to be cleaned up to obtain a high load forecasting accuracy. This was due to the best performance being obtained by a model developed without temperature as an input variable and with raw loading data.

Table 7: OP-ELM models' load forecast t-test results for Substation A

Compared models' results	P-value
Lowest errors: Cleaned vs Raw data	0.00
Cleaned data lowest errors: Input Group B vs Input Group A	0.0001007222
Raw data lowest errors: Input Group B vs Input Group A	0.0003847982
Raw Data: Lowest vs Highest error	0.0000798700

The two-week test load forecast was plotted against the target load for the lowest attained error, which was with the raw data. The forecasted load was seen to follow the target load, with a number of over estimates more prevalent than under estimates as shown in Figure 23.

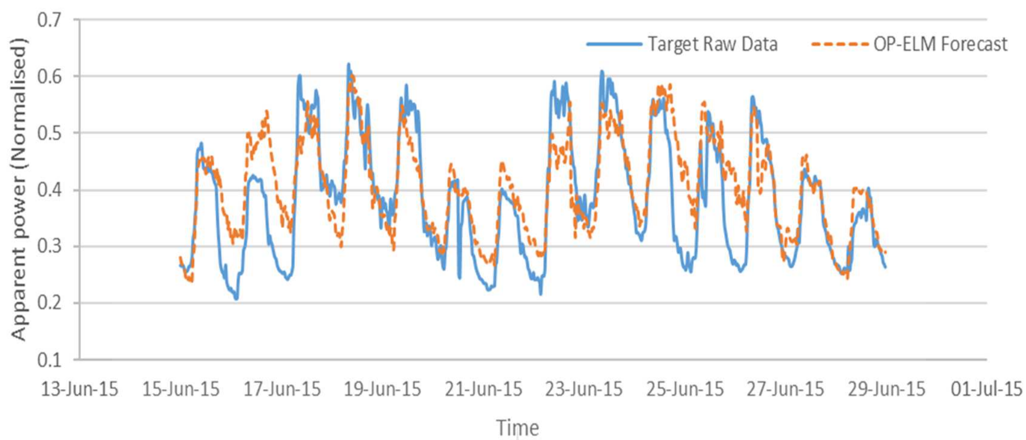

Figure 23: OP-ELM lowest test error model's two-week ahead load forecast vs target load

4.4.3. DBN Results

The load forecasting test errors for the DBN models are presented in Table 8 for experiments with raw data and Table 9 for experiments with cleaned up data. For experiments without temperature as a variable, i.e. variable Group A, it was observed that the errors increased with an increase in the number of hidden units. After a certain number of hidden units the error stopped increasing and became constant. The error with temperature as an input variable did almost the opposite to errors with variable Group A. Here the errors decrease with an increase in the number of hidden units and then start increasing after reaching the lowest error. The lowest obtained errors, bolded in the Tables 8, were an sMAPE of 0.0785323 (3.93 %), MAE of 0.0306600 (3.07 %) and RMSE of 0.0429001 (4.29 %). These lowest errors were obtained with temperature used as an input variable in the experiment and with raw data. The inclusion of temperature was, therefore, seen to lead to lower errors. The lowest errors with non-cleaned data is bolded in Table 9.

The statistical significance test was conducted and showed that a significant difference exists between all the compared load forecast results, respectively. As can be seen in Table 10 the p-value was less than 0.05 for all the cases. There was, therefore, a significant difference between the lowest error results obtained with and without temperature as an input variable.

This finding was observed for results with raw and cleaned data, respectively. The lowest and the highest forecast error results with raw data were also found to be significantly different. Therefore, for this case of forecasting a distribution power redistributor's load, DBN models should be trained with temperature as one of the variables. The loading data does not need to be cleaned up to obtain a high load forecasting accuracy. This was due to the best performance being obtained with temperature and raw loading data. Figure 24 shows the two-week ahead load forecasting results, which attained the lowest error, plotted against the target load.

Table 8: DBN models' test errors with raw loading data

Input variables	**Number hidden units**	**Performance**		
		sMAPE	***MAE***	***RMSE***
Group A	4	0.1248296	0.0466474	0.0600733
	8	0.1280208	0.0478700	0.0625895
	9	0.1261236	0.0470927	0.0608853
	10	0.1279107	0.0478198	0.0625180
	11	2.0000000	0.3837948	0.3979794
	12	2.0000000	0.3837937	0.3979785
	13	2.0000000	0.3837937	0.3979785
	14	2.0000000	0.3837937	0.3979785
	15	2.0000000	0.3837937	0.3979785
	16	2.0000000	0.3837937	0.3979785
	32	2.0000000	0.3837937	0.3979785
Group B	4	0.1624387	0.0629639	0.0763579
	8	0.0815791	0.0317031	0.0431172
	9	0.0800122	0.0311290	0.0429411
	10	0.0795810	0.0309429	0.0427489
	11	0.0790919	0.0307875	0.0426642
	12	0.0790417	0.0307235	0.0425964
	13	0.0788268	0.0307212	0.0426948
	14	0.0789635	0.0307617	0.0426318
	15	**0.0785323**	**0.0306600**	**0.0429001**
	16	0.0787966	0.0306738	0.0425675
	32	2.0000000	0.4043417	0.4168377

Table 9: DBN models' test errors with cleaned up loading data

Input variables	Number hidden units	Performance		
		sMAPE	*MAE*	*RMSE*
Group A	**4**	**0.19381**	**0.053706**	**0.069287**
	8	0.193659	0.053843	0.07006
	9	0.193488	0.053816	0.070244
	10	0.195002	0.054301	0.071079
	11	0.195413	0.054335	0.071006
	12	2.0000000	0.300239	0.323178
	13	2.0000000	0.300239	0.323178
	14	2.0000000	0.300239	0.323178
	15	2.0000000	0.300239	0.323178
	16	2.0000000	0.300239	0.323178
	32	2.0000000	0.300239	0.323178
Group B	4	0.28181	0.079471	0.091373
	8	0.196174	0.054889	0.073198
	9	0.194538	0.054675	0.073365
	10	0.195685	0.054717	0.072961
	11	0.195084	0.054702	0.072975
	12	0.195787	0.054657	0.07245
	13	0.195351	0.054596	0.072592
	14	0.195662	0.054694	0.072842
	15	0.193782	0.054397	0.072643
	16	0.195299	0.054625	0.072978
	32	2.0000000	0.300239	0.323178

Table 10: DBN models' load forecast t-test results for Substation A

Compared results	P-value
Lowest errors: Cleaned vs Raw data	0.0000000000
Cleaned data lowest errors: Input Group B vs Input Group A	1.05432×10^{-68}
Raw data lowest errors: Input Group B vs Input Group A	7.38277×10^{-05}
Raw Data: Lowest vs Highest error	0.0000000000

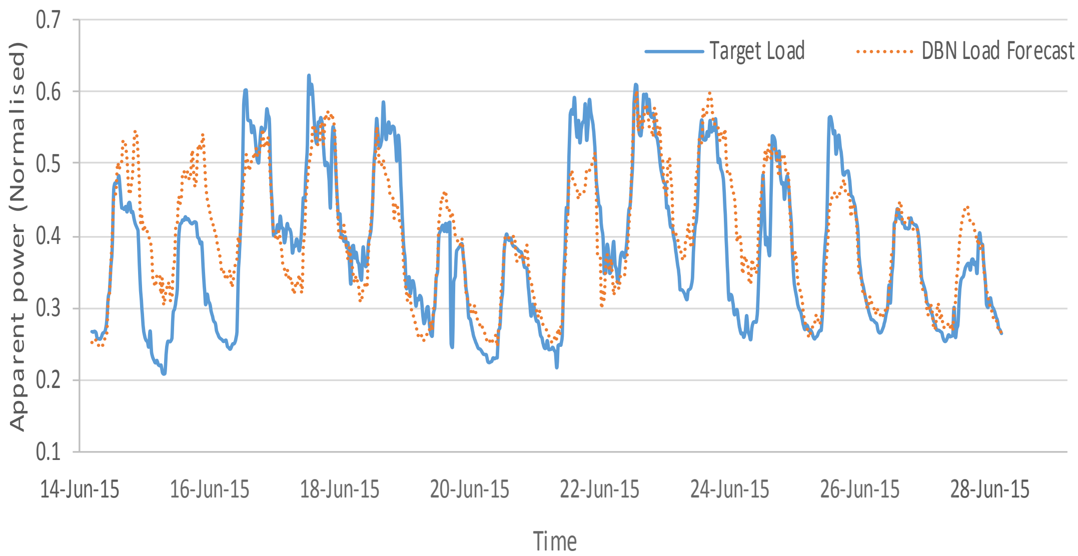

Figure 24: DBN lowest test error model's two-week ahead load forecast vs target load

4.4.4. LSTM-RNN Results

The results for raw and cleaned data are presented in Table 11 and Table 12, respectively. The lowest error was obtained with raw data and input variable Group B. This performance was achieved with the hidden number of units in the lower region of the number of units in the experiments. An sMAPE of 0.1268958 (6.34 %), MAE of 0.0477758 (4.78 %) and RMSE of 0.0632759 (6.33 %), bolded in Table 11, were the lowest obtained errors. The error did not consistently increase with an increase in the number of hidden units, however, lower errors were generally observed at the lower end of the number of hidden units. The lowest obtained error with the cleaned up data was also with input variable Group B as shown in Table 12. Hence, the inclusion of temperature led to lower forecast errors.

Table 11: LSTM-RNN models' test errors with raw loading data

Input variables	Number hidden units	Performance		
		sMAPE	*MAE*	*RMSE*
Group A	60	0.1278800	0.0482684	0.0641278
	67	0.1333182	0.0508165	0.0682295
	336	0.1306934	0.0497548	0.0671990
	470	0.1425678	0.0529337	0.0668141
	538	0.1394558	0.0536316	0.0722546
	672	0.1375375	0.0520812	0.0704330
Group B	60	0.1311405	0.0495421	0.0674328
	67	**0.1268958**	**0.0477758**	**0.0632759**
	336	0.1338203	0.050850	0.0649749
	470	0.1343266	0.0510782	0.0629332
	538	0.1957023	0.0696217	0.0818892
	672	0.1486360	0.0566232	0.0755845

Table 12: LSTM-RNN models' test errors with cleaned-up loading data

Input variables	Number Hidden Units	Performance		
		sMAPE	*MAE*	*RMSE*
Group A	60	0.2076085	0.0605033	0.0817285
	67	0.2144895	0.0626986	0.0843537
	336	0.2011207	0.0563641	0.0733492
	470	0.2540580	0.0748961	0.0936343
	538	0.2235020	0.0638470	0.0837461
	672	0.2127313	0.0595342	0.0764751
Group B	**60**	**0.1909078**	**0.0533111**	**0.0695006**
	67	0.1916753	0.0546699	0.0739185
	336	0.2071473	0.0585931	0.0787505
	470	0.2076085	0.0605033	0.0817285
	538	0.2144895	0.0626986	0.0843537
	672	0.2011207	0.0563641	0.0733492

The statistical significance test results from the conducted t-test are captured in Table 13. The p-value was less than 0.05 for all the cases, indicating a significant difference between all the compared load forecast results, respectively. There was therefore a significant difference between the lowest error results obtained with and without temperature as an input variable. This finding was observed for comparison between the results with raw and cleaned data, respectively. The lowest and highest forecast error results with raw data were also found to be significantly different. Therefore, for this case of forecasting a distribution power redistributor's load, LSTM-RNN models should be developed with temperature as one of the variables. The loading data does not need to be cleaned up to obtain a high load forecasting accuracy. This was due to the best performance being obtained by a model developed with temperature as an input variable and using raw loading data. A significant difference test also showed that these results were better than the results without temperature as an input variable.

Table 13: LSTM-RNN models' load forecast t-test results for Substation A

Compared results	**P-value**
Lowest errors: Cleaned vs Raw data	0.0000000000
Cleaned data lowest errors: Input Group B vs Input Group A	1.05432×10^{-68}
Raw data lowest errors: Input Group B vs Input Group A	7.38277×10^{-05}
Raw Data: Lowest vs Highest error	0.0000000000

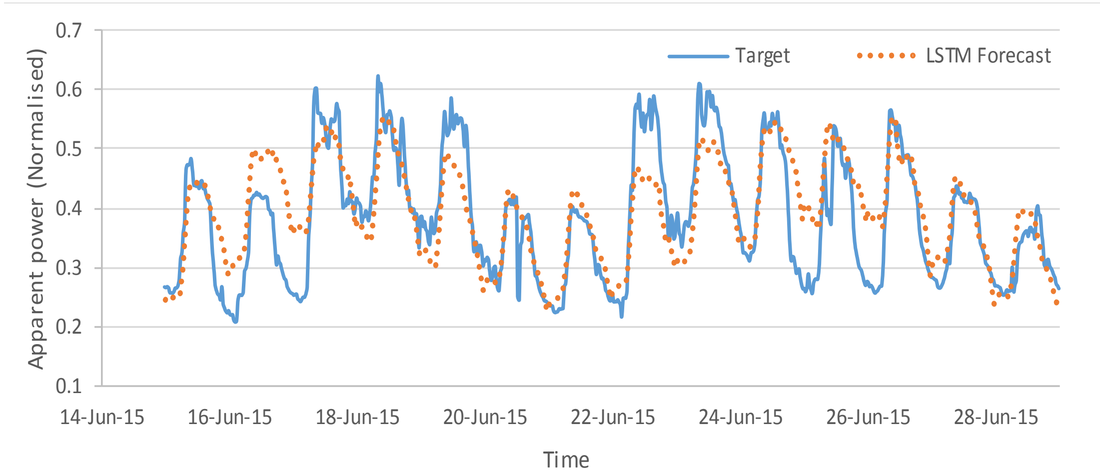

Figure 25: Substation A LSTM-RNN lowest test error model's two-week ahead load forecast vs target load

4.4.5. Results Discussion

The model's lowest obtained error results for each machine learning and deep learning technique are summarised in Table 14. The load forecasting results, corresponding to these lowest errors per model, were compared using the t-test. All these lowest errors were obtained by models developed using raw loading data. The models for all four techniques, therefore, do not need to be trained with cleaned data to achieve the lowest errors. The t-test results are captured in Table 15. The p-value for all comparisons was found to be less than 0.05, which meant all the results were significantly different from each other. The best performance was, therefore, obtained using DBN with temperature as an input variable. The two deep learning techniques were found to outperform the machine learning techniques, with ANFIS achieving the worst performance. The machine learning techniques performed better without the inclusion of temperature as opposed to the deep learning techniques. This could be due to deep learning techniques' ability to generally perform better with more data by extracting more features from the data. To forecast a distribution network load for a redistributor or a similar load profile machine learning and deep learning techniques can therefore both be used. In cases where weather data is not available and accuracies around 95 % are acceptable, machine learning techniques may be deployed. These can be cases such as forecasting load of feeders for an outage, where feeders can typically be overloaded by 10 % to 25 % for a number of hours depending on their condition. Deep learning techniques' can still be used in the absence of weather data. It was observed that the errors obtained with the deep learning models without the use of temperature were still lower than those obtained with machine learning techniques. To achieve better performance, therefore, deep learning techniques can be deployed with temperature as one of the input variables.

Table 14: Summary of first case study's lowest errors per model

Technique	Input variables	Performance		
		sMAPE	***MAE***	***RMSE***
ANFIS	Group A	0.138483	0.052392	0.071799
OP-ELM	Group A	0.1315575	0.0491749	0.0657673
DBN	**Group B**	**0.0785323**	**0.0306600**	**0.0429001**
LSTM- RNN	Group B	0.1268958	0.0477758	0.0632759

Table 15: Substation A different techniques' lowest error models' load forecast t-test results

Compared models' lowest errors	P-value
ANFIS vs OP-ELM	0.0054268064
ANFIS vs DBN	5.9475×10^{-06}
ANFIS vs LSTM	1.40419×10^{-20}
OP-ELM vs DBN	0.0005862928
OP-ELM vs LSTM	4.05716×10^{-21}
DBN vs LSTM	0.0031420008

4.5. Chapter Summary

This chapter presented the first case study in this research. This case study researched the performance of deep learning and machine learning techniques on forecasting the load of a distribution customer that redistributes power. The customer has an 88/11 kV, 80 MVA substation. The impact of cleaning up loading data and the inclusion of temperature in the model development was also investigated. It was found that uncleaned loading data led to machine learning and deep learning models with the best load forecasting performance. machine learning techniques' models' performance was not the best without the inclusion of temperature in the development of the models. The deep learning techniques' models' were found to achieve their best load forecasting performance when temperature is utilised in their development.

Chapter 5 - 2nd Case Study: Distribution Substation B - Industrial Large Power End User Load

"If you're walking down the right path and you're willing to keep walking, eventually you'll make progress."

Barack Obama

5.1. Introduction

The previous chapter presented the first case study on an 88/11 kV, 80 MVA redistributor, distribution substation network. Deep learning techniques were found to achieve lower errors than the machine learning techniques. Raw/uncleaned loading data led to machine learning and deep learning models which achieved the lowest errors in all experiments. The inclusion of temperature in the development of machine learning models was found not to improve their load forecasting performance. Whereas the inclusion of temperature in the development of deep learning to models led to lower forecasting errors.

This section presents the second case study. The substation in this study is a dedicated industrial large power user substation. The network this substation is located in is also described in this section, as well as the substation setup. The data used to train and test the models is discussed. The load forecasting results are also presented and discussed. A discussion and comparison of the models in the two case studies is also presented.

5.2. Case Study B Distribution Network Overview

The distribution substation under study, which will be referred to as Substation B, is also located in the North-Eastern part of South Africa. The substation is connected to the same main transmission substation as the substation A from the first case study. The substation is however connected through the MTS 275/132 kV transformer set's bus bar. The MTS also supplies an 88 kV distribution network. Substation B is in a ring network with one other substation. The two substations are each supplied from the MTS through a single feeder, and have an interconnecting feeder which connects their 132 kV bus bars. The interconnecting feeder can be opened or closed from either of the two substations. This configuration, shown in Figure 26, is also a cost-effective network firmness mechanism. Here, if one of the feeders is off due to failure or planned maintenance, the remaining feeder can potentially supply the two substations with power. An example is if feeder 1 is off, substation B can be supplied from the MTS through feeder 2 and then the interconnecting feeder from the other substation's bus bar. Substation B has five 132/22 kV, 40 MVA transformers, each with its own point of supply. From an analysis of the transformer's data it was observed that the transformers, possibly through an internal configuration, may have a grouping in terms of the loads they supply power to. Transformers 1 and 2 had the same load profile and, transformers 3 and 4 also had the same load profile to each other, which led to the deduction that these groups of transformer supply the same load. Transformer 5 was observed to be a backup transformer that supports both groups of transformers. Similar to the first case study there are power measuring devices on the incoming feeders and the primary side of each transformer. Power measuring devices are also installed at each point of supply on the lower voltage side.

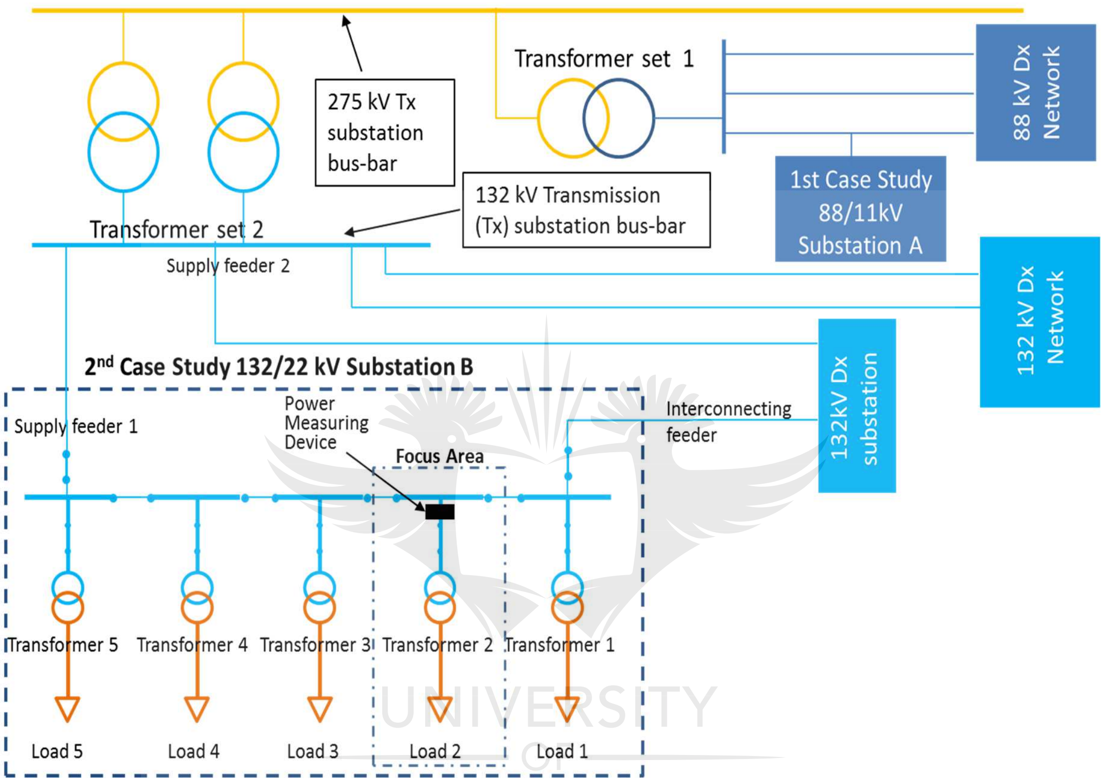

Figure 26: The distribution network in which the bulk large power user substation under study is located

5.3. Data Description

The loading data were collected for the period between January 2012 and September 2015. Hence the winter periods for the year 2012 to the year 2015 were captured. The data were observed to have irregular dips to zero load. The possible causes of these zero loading values are similar to those described in section 4.3. for case study 1. The data were cleaned up as in the previous case and, following the same rationale the experiments were conducted with raw and cleaned-up data. Figure 27 shows the plot of the collected raw loading data and Figure 28 shows the plot of the cleaned-up loading data. The daily and two-week load profiles are respectively shown in Figure 29 and Figure 30. The same temperature data used in case study 1 were used in this case study as the stations were closest to the same weather station in relation to other weather stations. This also allowed a good way to observe the impact of the temperature on the different distribution load types.

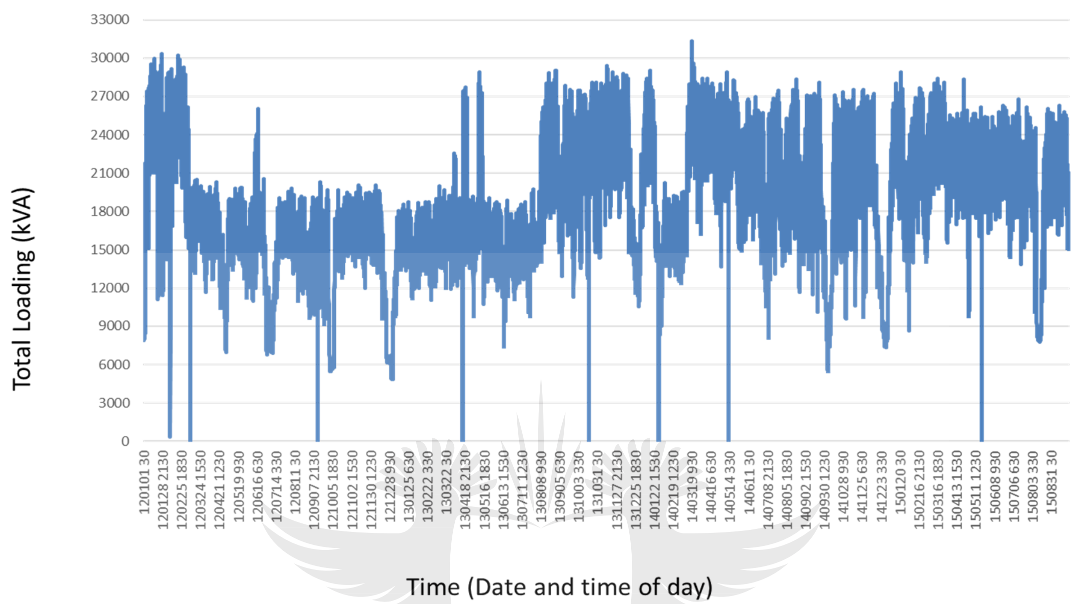

Figure 27: Plot of transformer 2 raw loading data

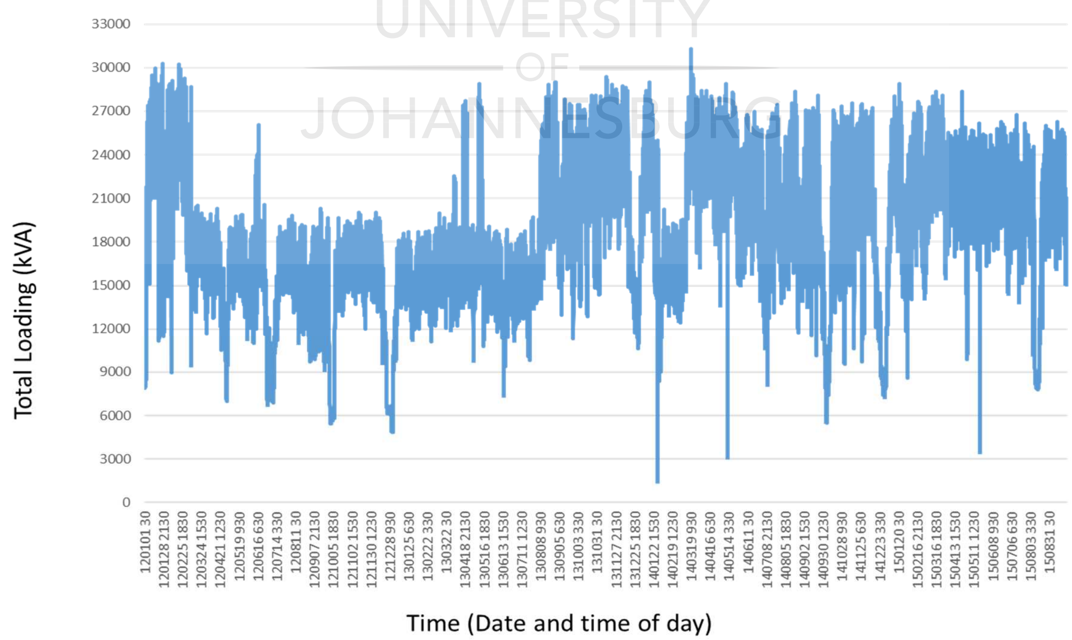

Figure 28: Plot of transformer 2 cleaned-up loading data

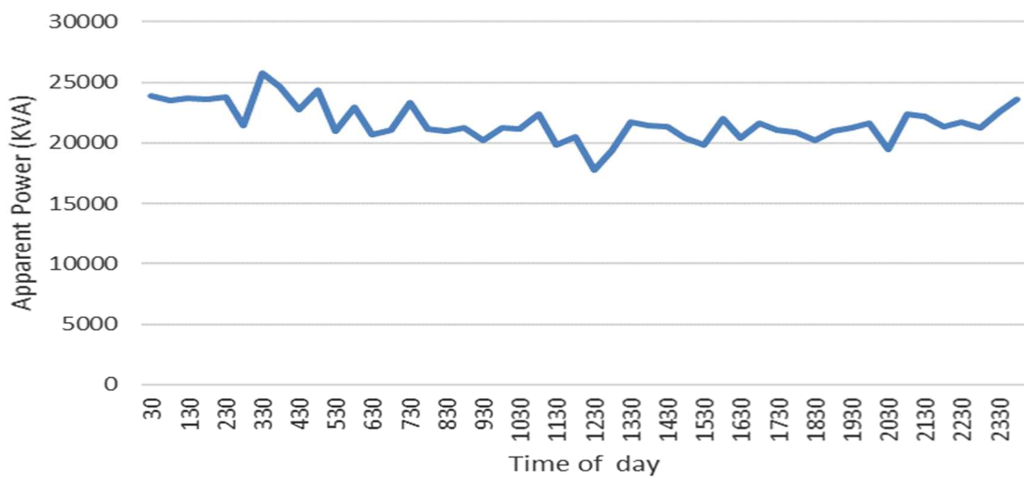

Figure 29: Transformer 2's day load profile for 15th June 2015

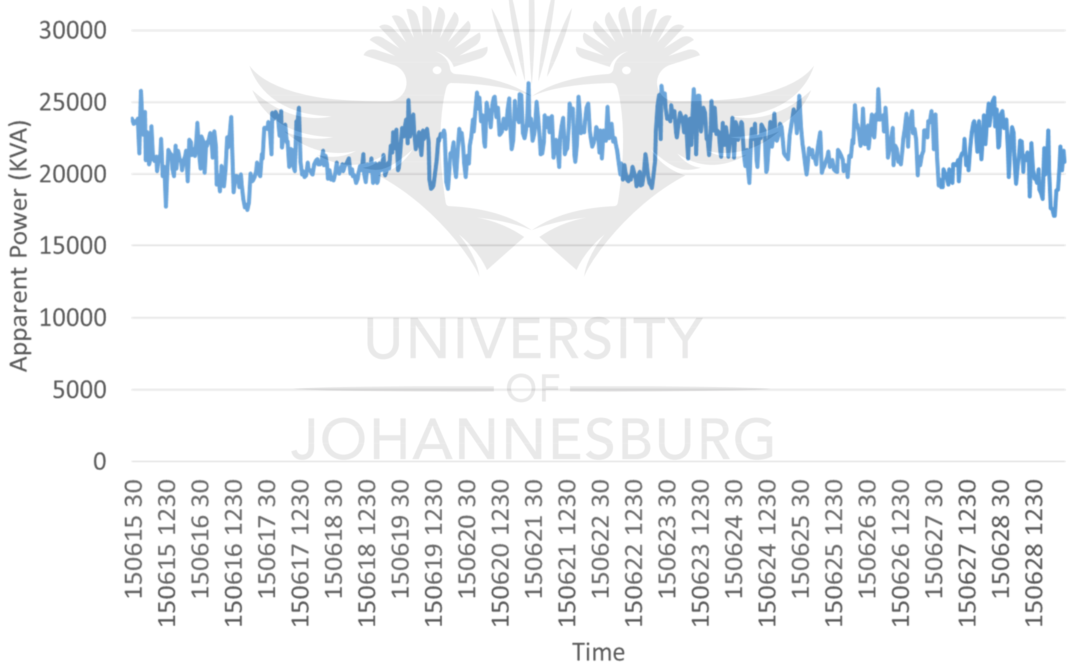

Figure 30: Transformer 2's two-week load profile for the 15th to 28th June 2015

5.4. Experiment Results and Discussion

The different experiments were conducted following the approach described in Section 3.4. The test results are presented in the following subsections.

5.4.1. ANFIS Results

The ANFIS model's test results are captured in Table 16 and Table 17, respectively, for experiments with non-cleaned and cleaned data. From the results it can be seen that the lowest error was obtained with the non-cleaned data and input variable Group A. This model obtained the following load forecasting test errors; an sMAPE of 0.078875 (3.94 %), MAE of 0.0548 (5.48 %) and RMSE of 0.067278 (6.73 %). This model's results are shown in Figure 31, plotted with the target load. The models trained with Group A input variables achieved lower load forecasting test errors than those trained with Group B input variables. The inclusion of temperature, therefore, did not lead to better performance. Cleaning up the data also did not lead to a better performance.

Table 16: Substation B ANFIS models' performance with non-cleaned loading data

Input variables	**Model tuning parameters**	**Performance**		
		sMAPE	***MAE***	***RMSE***
Group A	1	0.086554	0.05962	0.075169
	2	0.233798	0.156265	0.221828
	3	**0.078875**	**0.0548**	**0.067278**
	4	0.088155	0.06063	0.07585
Group B	1	0.121348	0.081983	0.104547
	2	0.09584	0.065327	0.080201
	3	0.08605	0.059333	0.073461
	4	0.126163	0.086611	0.114288

The t-test revealed that there was a significant difference in all compared results except for one comparison. This was the comparison between the results with the lowest and highest load forecasting errors, respectively, with raw loading data. There was a significant difference between the results with the lowest errors with input variable Groups A and B, respectively, for both cleaned and non-cleaned loading data. This observation means that the results obtained without temperature as an input variable were better than those obtained with temperature as an input variable. The t-test results are given in Table 18.

Table 17: Substation B ANFIS models' performance with cleaned loading data

Input variables	Model tuning parameters	Performance		
		sMAPE	***MAE***	***RMSE***
Group A	1	0.09473	0.064364	0.080462
	2	0.236399	0.157662	0.211377
	3	**0.088012**	**0.060583**	**0.073518**
	4	0.0921	0.062804	0.078134
Group B	1	0.12329	0.083074	0.105467
	2	0.105026	0.070778	0.087629
	3	0.089798	0.061756	0.076607
	4	0.128889	0.086846	0.108592

Table 18: ANFIS models' load forecast t-test results for Substation B

Compared models' results	P-value
Lowest errors: Cleaned vs Raw data	3.49697×10^{-60}
Cleaned data lowest errors: Input Group B vs Input Group A	0.008872329
Raw data lowest errors: Input Group B vs Input Group A	6.87824×10^{-18}
Raw Data: Lowest vs Highest errors	0.068706595

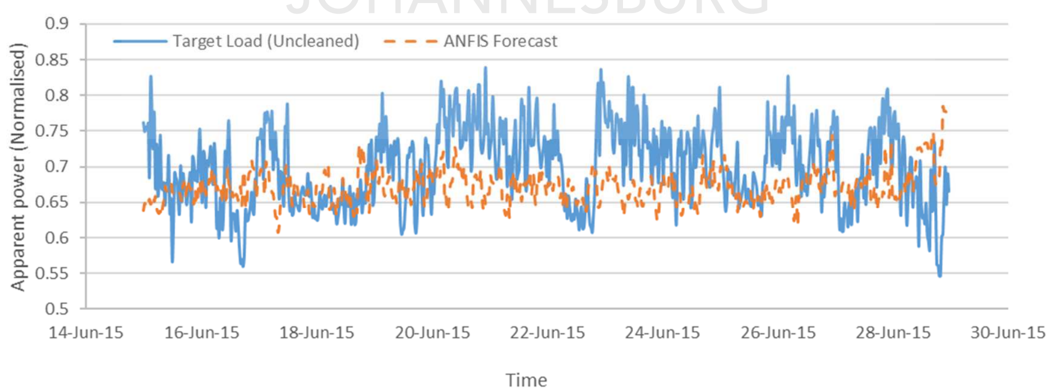

Figure 31: ANFIS lowest test error model's two-week ahead Substation B load forecast vs target load

5.4.2. OP-ELM Results

Table 19 and 20 present the load forecasting test results' errors, with the lowest errors shown in bold. This model achieved an sMAPE of 0.076639 (3.83 %), MAE of 0.053231 (5.32 %) and RMSE of 0.065293 (6.53 %). The two-week ahead test load forecasting results are shown, plotted against the target load, in Figure 32. The errors increased with an increase in the number of hidden units. This was however not the case with cleaned data and Group B input variables. Here the errors increased with an increase in the number of hidden units and then dropped again after 108 hidden units at 158 hidden units.

Table 19: OP-ELM models' performance with non-cleaned loading data

Input variables	Number of hidden nodes	Performance		
		sMAPE	*MAE*	*RMSE*
Group A	10	0.077614	0.053866	0.065600
	55	0.091838	0.061893	0.097532
	80	0.100575	0.067402	0.096671
	105	0.114489	0.073351	0.116556
Group B	**8**	**0.076639**	**0.053231**	**0.065293**
	58	0.088769	0.061137	0.076092
	103	0.100233	0.06852	0.089016
	158	0.107616	0.072069	0.093743

Table 20: OP-ELM models' performance with cleaned-up loading data

Input variables	Number of hidden nodes	Performance		
		sMAPE	*MAE*	*RMSE*
Group A	**10**	**0.0827701**	**0.056119**	**0.068664**
	50	0.092224	0.061758	0.075695
	70	0.101389	0.067256	0.089437
	80	0.120470	0.077187	0.138783
Group B	8	0.090344	0.060570	0.077942
	58	0.112712	0.076688	0.148648
	108	0.123649	0.078600	0.115015
	158	0.111106	0.073698	0.097400

Table 21 gives the t-test results for the OP-ELM model's results. All the compared results had a significant difference to each other. For non-cleaned data, the lowest error results from the model where temperature was used as a variable outperformed that where temperature was not used as a variable. With cleaned up data it was the opposite case, where the inclusion of temperature did not improve the forecasting accuracy.

Table 21: OP-ELM models' load forecast t-test results for Substation B

Compared model's results	P-value
Lowest errors: Cleaned vs Raw data	2.8778×10^{-107}
Cleaned data lowest errors: Input Group B vs Input Group A	7.51278×10^{-08}
Raw data lowest errors: Input Group B vs Input Group A	0.016049816
Raw Data: Lowest vs Highest errors	0.002344021

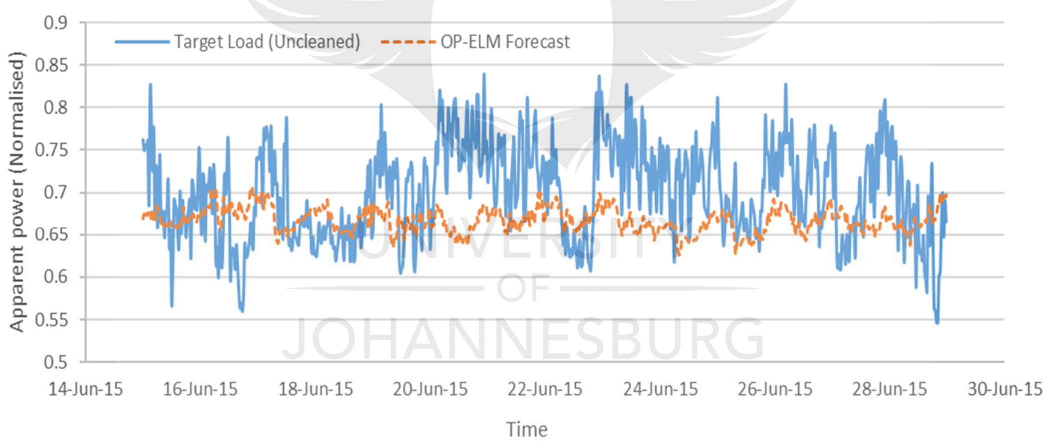

Figure 32: OP-ELM lowest test error model's two-week ahead Substation B load forecast vs target load

5.4.3. DBN Results

The load forecasting test results for the DBN models are captured in Table 22 and Table 23. The lowest error was achieved by a model trained with raw data without temperature as part of the input variables. The sMAPE achieved here was 0.074553 (3.73 %), with a MAE of 0.051886 (5.19 %) and RMSE of 0.06325 (6.33 %). Lower errors were observed with a lower number of hidden units. The error stagnated after increasing past a different number of hidden units for each sub-experiment. With non-cleaned data the error stagnated from nine and 13 hidden units, respectively, with variable Group A and B. With cleaned data this behaviour was observed, respectively, from nine and ten hidden units for input variables Group A and Group B. A t-test conducted showed that the results with the lowest errors in each respective experiment, with cleaned and raw data, had the best performance. The t-test results are presented in Table 24. The experiments also showed that the exclusion of temperature led to a better performance with DBN in forecasting transformer 2's load.

Table 22: DBN models' performance with non-cleaned loading data

Input variables	**Number hidden units**	**Performance**		
		sMAPE	***MAE***	***RMSE***
Group A	4	0.084358	0.058398	0.071928
	8	**0.074553**	**0.051886**	**0.06325**
	9	0.355518	0.300051	0.305125
	10	0.355518	0.300051	0.305125
	11	0.355518	0.300051	0.305125
	12	0.355518	0.300051	0.305125
	13	0.355518	0.300051	0.305125
	14	0.355518	0.300051	0.305125
	15	0.355518	0.300051	0.305125
	16	0.355518	0.300051	0.305125
	32	0.355518	0.300051	0.305125
Group B	4	0.098332	0.067465	0.082154
	8	0.076923	0.053481	0.065309
	9	0.077413	0.053808	0.065873
	10	0.076501	0.053202	0.065006
	11	0.07527	0.052338	0.064215
	12	0.077017	0.053483	0.06562
	13	0.355518	0.300051	0.305125
	14	0.355518	0.300051	0.305125
	15	0.355518	0.300051	0.305125
	16	0.355393	0.299922	0.305005
	32	0.355518	0.300051	0.305125

Table 23: DBN models' performance with cleaned loading data

Input variables	Number hidden units	Performance		
		sMAPE	*MAE*	*RMSE*
Group A	4	0.096254	0.064928	0.079892
	8	**0.081501**	**0.055427**	**0.067506**
	9	0.376716	0.314994	0.320321
	10	0.376716	0.314994	0.320321
	11	0.376715	0.314993	0.320319
	12	0.376716	0.314994	0.320321
	13	0.376716	0.314994	0.320321
	14	0.376716	0.314994	0.320321
	15	0.376716	0.314994	0.320321
	16	0.376716	0.314994	0.320321
	32	0.376716	0.314994	0.320321
Group B	4	0.082468	0.05608	0.068781
	8	0.083146	0.056521	0.069206
	9	0.081904	0.055714	0.068124
	10	**0.080669**	**0.054858**	**0.067393**
	11	0.376715	0.314993	0.320319
	12	0.376716	0.314994	0.320321
	13	0.083201	0.056469	0.069171
	14	0.082861	0.05626	0.068955
	15	0.376716	0.314994	0.32032
	16	0.37661	0.314884	0.320219
	32	0.376716	0.314994	0.320321

Table 24: DBN models' load forecast t-test results for Substation B

Compared models' results	P-value
Lowest errors: Cleaned vs Raw data	0.0000000000
Cleaned data lowest errors: Input Group B vs Input Group A	0.025323222
Raw data lowest errors: Input Group B vs Input Group A	7.6528×10^{-07}
Raw Data: Lowest vs Highest errors	0.0000000000

Figure 33 shows a plot of the load forecast results for the best performing model against the target load. The forecasted load was observed to track the target load. However the forecast load did not vary much and was more flat in comparison to the target load.

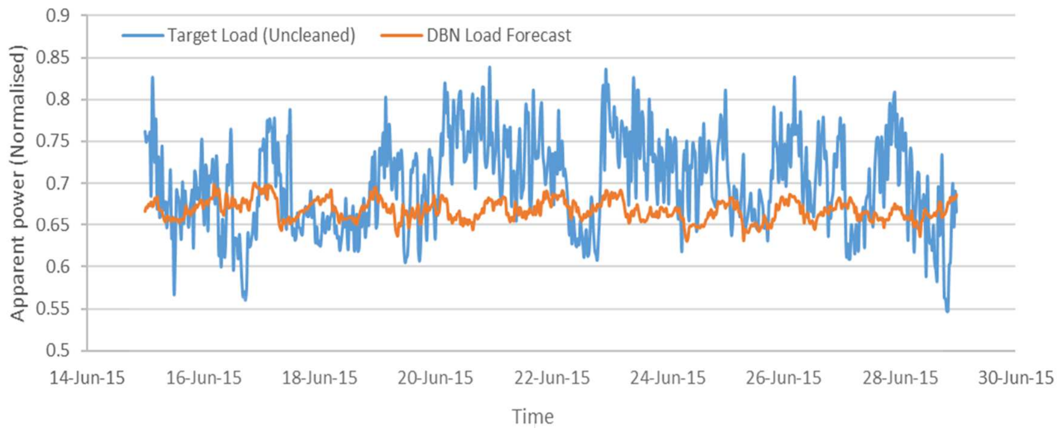

Figure 33: DBN lowest test error model's two-week ahead Substation B load forecast vs target load

5.4.4. LSTM-RNN Results

The LSTM models with the highest number of hidden layers in the experiments gave the lowest test errors, with both raw and cleaned data. These errors are bolded in Table 25 and Table 26. The lowest errors were obtained with a model developed with uncleaned data. In both cases, with cleaned and uncleaned loading data, temperature was not an input variable in the model development. The lowest obtained values for the different performance measures are an sMAPE of 0.065859 (3.29 %), MAE of 0.04598 (4.6 %) and RMSE of 0.055058 (5.51 %). The load forecasting results which gave these performance measures are plotted against the target load in Figure 34. The statistical significance, t-test, results are given in Table 27. The statistical significance test showed these results were the best results for LSTM-RNN. For this type of distribution customer, the best performance when using LSTM-RNN can therefore be obtained with uncleaned data and without temperature data.

Table 25: LSTM-RNN models' performance with non-cleaned loading data

Input variables	Number hidden units	Performance		
		sMAPE	***MAE***	***RMSE***
Group A	60	0.077626	0.053888	0.06557
	67	0.075899	0.052738	0.064054
	336	0.067377	0.047037	0.057051
	470	0.067405	0.047036	0.0567
	538	0.072826	0.050744	0.062387
	672	**0.065859**	**0.04598**	**0.055058**
Group B	60	0.068902	0.04807	0.059121
	67	0.069627	0.048554	0.059077
	336	0.083265	0.057628	0.070237
	470	0.067387	0.047039	0.057631
	538	0.071501	0.049845	0.061517
	672	0.066452	0.046385	0.056338

Table 26: LSTM-RNN models' performance with cleaned loading data

Input variables	Hidden layers	Performance		
		sMAPE	***MAE***	***RMSE***
Group A	60	0.076185	0.051969	0.063520
	67	0.077460	0.052804	0.064516
	336	0.072048	0.049197	0.059428
	470	0.073190	0.049974	0.060671
	538	0.076781	0.052359	0.064038
	672	**0.070594**	**0.048218**	**0.057900**
Group B	60	0.074273	0.050665	0.062063
	67	0.074493	0.050818	0.062484
	336	0.071403	0.048768	0.059263
	470	0.084471	0.057327	0.070240
	538	0.076295	0.052030	0.064442
	672	0.071717	0.048970	0.060011

Table 27: LSTM-RNN models' load forecast t-test results for Substation B

Compared models' results	P-value
Lowest errors: Cleaned vs Raw data	0.0000000000
Cleaned data lowest errors: Input Group B vs Input Group A	2.76×10^{-175}
Raw data lowest errors: Input Group B vs Input Group A	9.7376×10^{-125}
Raw Data: Lowest vs Highest errors	0.0000000000

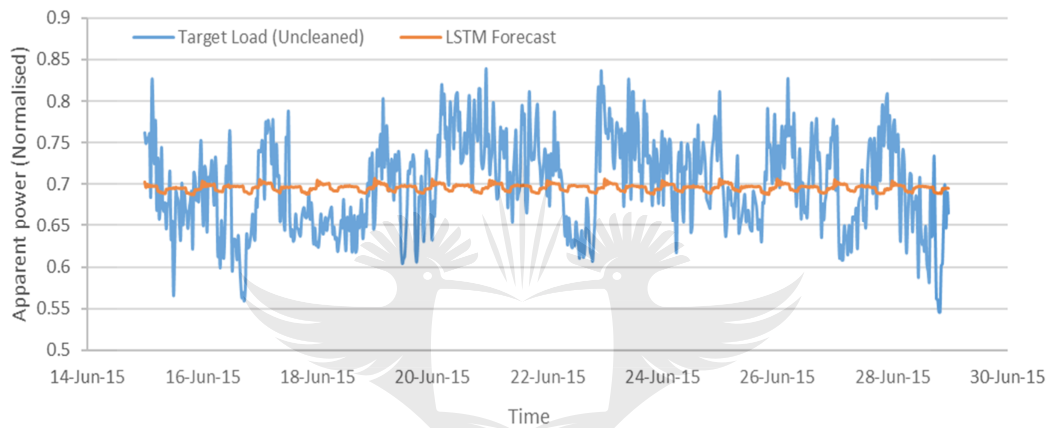

Figure 34: LSTM-RNN lowest test error model's two-week ahead Substation B load forecast vs target load

5.4.5. Results Discussion

Deep learning techniques were found to have better load forecasting performance than the machine learning techniques. This was confirmed via a statistical significance test on each technique's models' results with the lowest error. These lowest errors and the t-test results are presented in Table 28 and Table 29, respectively. It was, however, found that there was no significant difference between the ANFIS and DBN load forecasting results. All these techniques' best performing models were obtained with raw data and Group A variables. This observation was with the exception of OP-ELM, which attained its best performing model with Group B variables. The exclusion of temperature, therefore, did not influence the performance of the models negatively. To forecast load for this type of distribution load customer profile, the LSTM-RNN model should be developed with Group A variable to achieve the best performance.

Table 28: Summary of second case study's lowest errors per model

Technique	Input variables	Performance		
		sMAPE	***MAE***	***RMSE***
ANFIS	Group A	0.078875	0.0548	0.067278
OP-ELM	Group B	0.076639	0.053231	0.065293
DBN	Group A	0.074553	0.051886	0.06325
LSTM-RNN	**Group A**	**0.065859**	**0.04598**	**0.055058**

Table 29: Substation B different techniques' lowest error models' load forecast t-test results

Compared model's lowest errors	P-value
ANFIS vs OP-ELM	1.4661×10^{-06}
ANFIS vs DBN	0.07521253
ANFIS vs LSTM	4.193×10^{-109}
OP-ELM vs DBN	4.6263×10^{-08}
OP-ELM vs LSTM	1.06×10^{-246}
DBN vs LSTM	0

5.5. Chapter Summary

This chapter presented the second case study in this research. This case study was conducted on one of the five, 132/22 kV, 40 MVA, transformers in large industrial power user's dedicated distribution substation. The performance of machine learning techniques (ANFIS and OP-ELM) and deep learning techniques (DBN and LSTM-RNN) were investigated in forecasting this distribution substation's load. The impact of temperature and data clean-up on the load forecasting performance was also investigated. It was found that deep learning techniques outperform machine learning techniques. All the techniques' models, apart from OP-ELM, performed better without the inclusion of temperature in their development as opposed to when the temperature was included. All the techniques' models achieved their best load forecasting performance when the raw uncleaned used in their development.

Chapter 6 - 3rd Case Study Distribution Substation 3 - Power Redistributor Load

"Sometimes when you innovate, you make mistakes. It is best to admit them quickly, and get on with improving your other innovations."

Steve Jobs

6.1. Introduction

The previous chapter presented the second case study on a 132/22 kV distribution large power consuming customer's substation. Here the loading data used were from one of the customer's five 40 MVA transformers. Deep learning techniques' models were found to achieve lower errors than the machine learning techniques. All the techniques' models, apart from OP-ELM, achieved their lowest errors without the inclusion of temperature in their development. All the four techniques' models achieved their lowest errors when developed with uncleaned data. An LSTM-RNN's model achieved the lowest load forecasting error in case study 2.

This chapter presents the third case study. The load used in this case study is that of a power redistributor which is located in a different distribution network to that in the first two case studies. This distribution network and loading data are described in this chapter. The load forecasting performance of machine learning techniques, ANFIS and OP-ELM, and deep learning techniques, DBN and LSTM-RNN, is investigated. This investigation's results are presented and discussed. A comparison of the three case studies' findings is also presented.

6.2. Case Study C Distribution Network Overview

The substation in this case study is also located in the North-Eastern part of South Africa. The substation is located in a neighbouring town, which is approximately 30 kilometres away from the town the Substation A and Substation B are locate in. The substation in this case study is referred to as Substation C in the rest of this document. Substation C is connected to the grid through a 400/132 kV transmission substation, which also supplies other distribution substations and networks at 132 kV. The distribution network overview is shown Figure 35. The customer is supplied via a switching substation and has its own distribution substation with transformers to step down the voltage for distribution. A switching substation is a substation that does not have transformers. The switching substation is also connected to the transmission substation through a 132 kV substation it is in a ring with. There are two feeders from the switching substation to the point of supply. The customer has different types of consumers it supplies with the power it redistributes. The power, which is supplied by Substation C, is measured at point of supply.

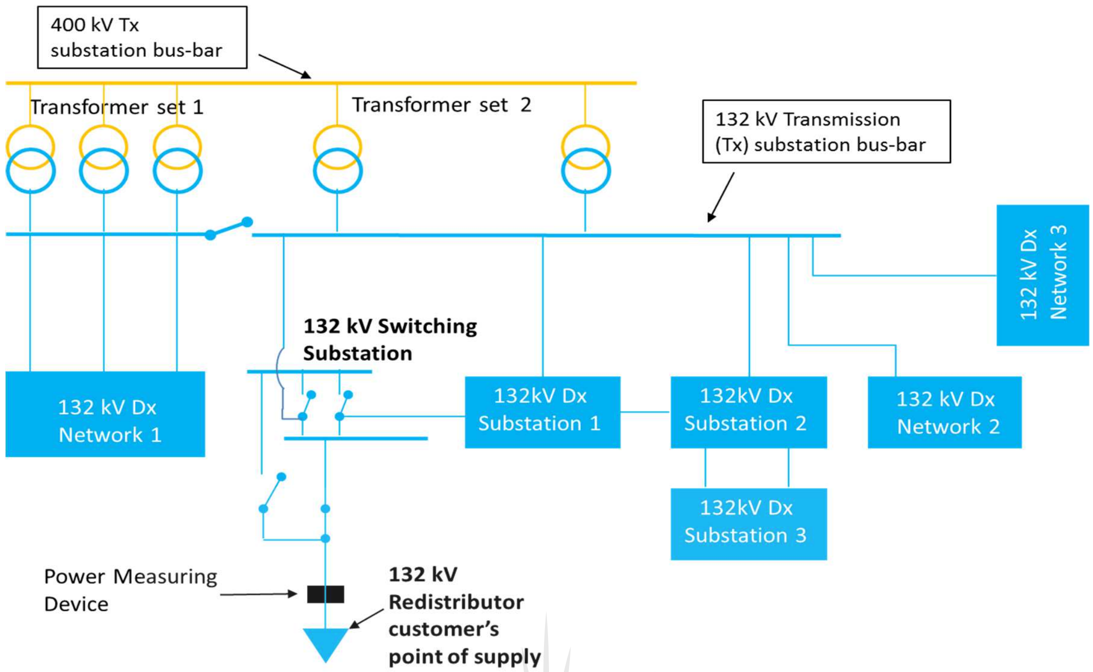

Figure 35: The distribution network the power redistributor's switching substation under study is located

6.3. Data Description

The loading data were collected for a period of four years between January 2012 and December 2015. These data were also stored in 30 minutes averages. The loading data in this case study also had dips and were cleaned-up to remove these dips. The plots of the raw/uncleaned data and the cleaned data are shown in Figure 36 and Figure 37, respectively. From these plots, it can be seen that there was a significant increase in the power consumption, by above 100 %, from January 2015.

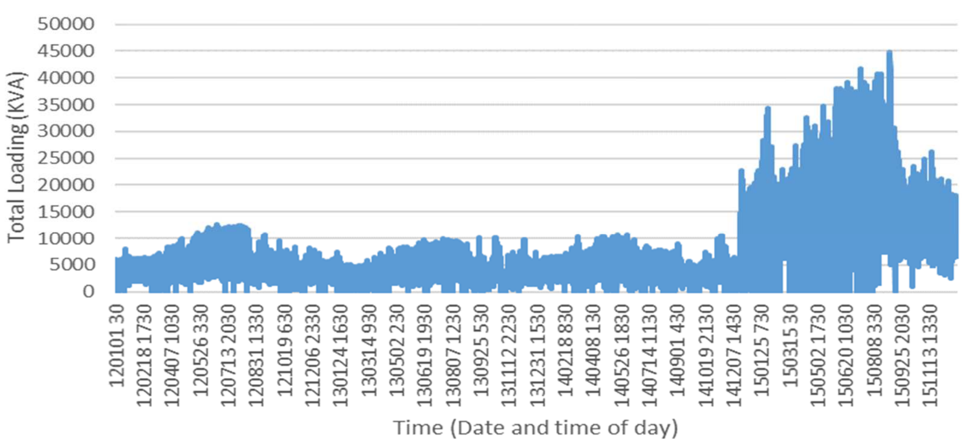

Figure 36: Plot of Substation C's raw loading data

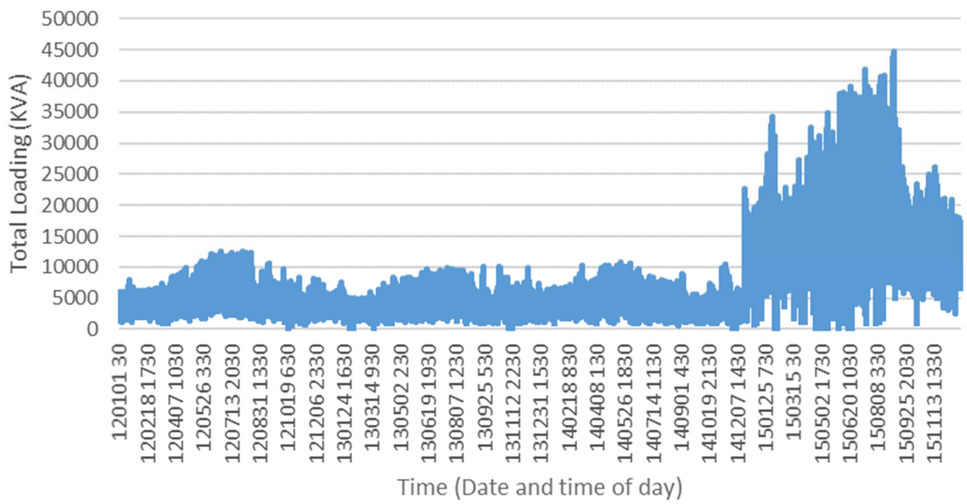

Figure 37: Plot of Substation C's cleaned-up loading data

The one day and the two-week load profiles are, respectively, presented in Figure 38 and Figure 39. Both these graphs are plotted with uncleaned loading data. It can be seen from these figures that the power consumption typically fluctuated from a loading of 5 000 and 7 000 KVA during the early non-peak hours of the day, to over 35 000 KVA during the evening peak. The temperature data used in this case study is the same data used in the first two case studies. The only difference is that this substation in this case study is located in the same town as the weather station used to measure the temperature.

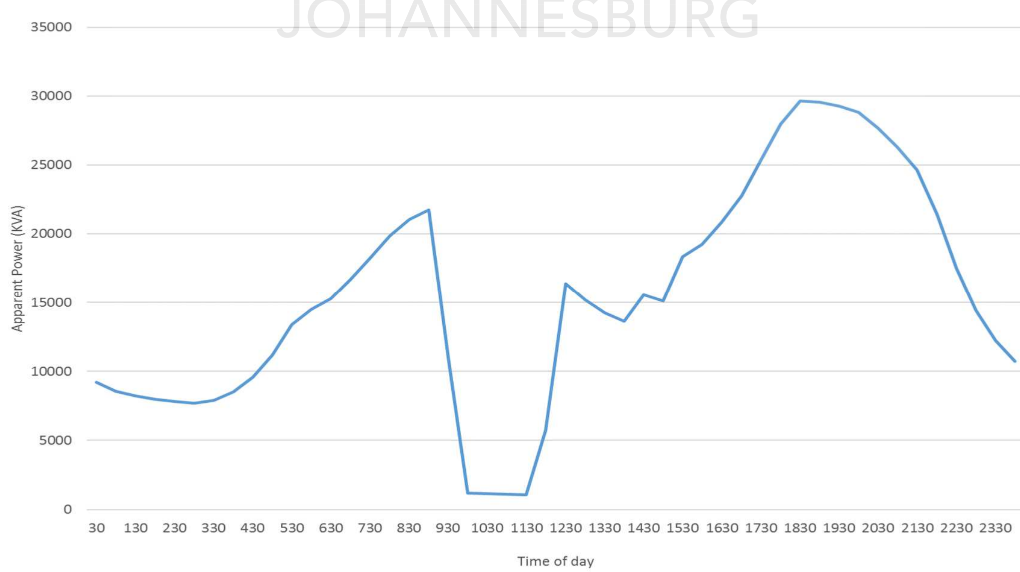

Figure 38: Substation C's day load profile for 15th June 2015

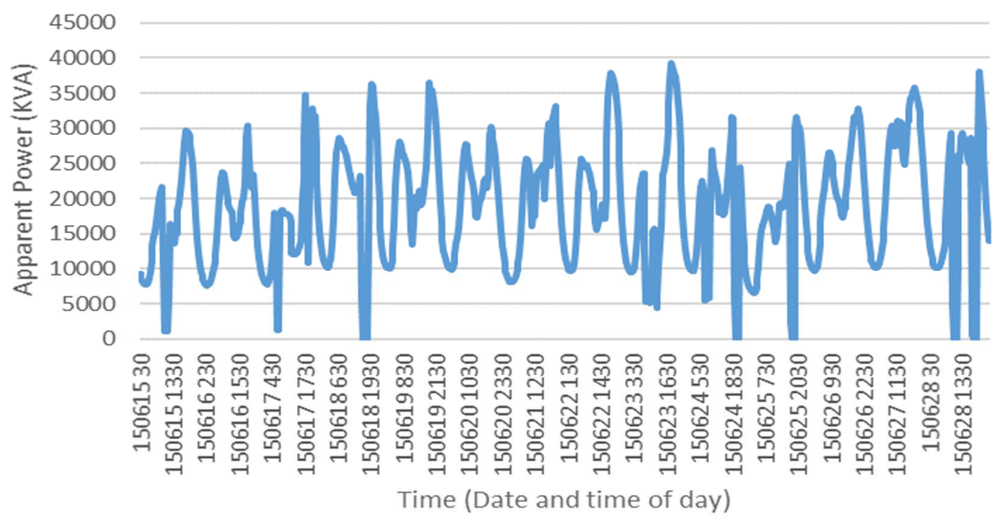

Figure 39: Substation C's two-week load profile for the 15th to 28th June 2015

6.4. Experiment Results and Discussion

The different techniques' models were developed as per the approach described in Section 3.4. The load forecasting performance of these models was determined. These performance results are presented and discussed in this Section.

6.4.1. ANFIS Results

The load forecasting performance test results are captured in Table 30 and Table 31 for models developed with non-cleaned and cleaned loading data, respectively. The models developed with cleaned loading data attained the lowest load forecasting errors. The inclusion of temperature did not lead to an improved accuracy with both non-cleaned and cleaned loading data. The lowest attained errors with both the models developed with non-cleaned and cleaned loading data are bolded in Tables 30 and 31. The lowest attained load forecasting errors were an sMAPE of 13.05 %, MAE of 10.09 % and RMSE of 14.99 %, as shown in Table 31. The statistical significance test was conducted and the results are presented in Table 32. The t-test showed that there is a significant difference between all the compared results, except for the results with and without temperature for models with cleaned data. The load forecasting performance of the models developed with cleaned loading data were thus not significantly different.

Table 30: Substation C ANFIS models' performance with non-cleaned loading data

Input variables	Model tuning parameters	Performance		
		sMAPE	***MAE***	***RMSE***
Group A	1	0.3280877	0.1244868	0.1763246
	2	**0.3548188**	**0.1181713**	**0.1698449**
	3	0.3485234	0.1591555	0.3033930
	4	0.356657	0.128948	0.182569
Group B	1	0.509364	0.184724	0.272112
	2	0.919458	0.765156	1.386215
	3	0.4223565	0.1656288	0.2677004
	4	0.470634	0.179244	0.26091

Table 31: Substation C ANFIS models' performance with cleaned loading data

Input variables	Model tuning parameters	Performance		
		sMAPE	***MAE***	***RMSE***
Group A	1	0.261473	0.103331	0.150772
	2	0.259367	0.101062	0.153101
	3	**0.260919**	**0.100845**	**0.149861**
	4	0.261473	0.103331	0.150772
Group B	1	0.299051	0.114858	0.166327
	2	0.875142	0.79492	2.006465
	3	0.299748	0.11146	0.160826
	4	0.299051	0.114858	0.166327

Table 32: ANFIS models' load forecast t-test results for Substation C

Compared model's results	**P-value**
Lowest errors: Cleaned vs Raw data	1.94743×10^{-10}
Cleaned data lowest errors: Input Group B vs Input Group A	0.124759926
Raw data lowest errors: Input Group B vs Input Group A	2.05816×10^{-9}
Raw Data: Lowest vs Highest error	5.29115×10^{-18}

The two-week load forecast is shown plotted along the target load in Figure 40. The load forecast was seen to follow the target load closely, except for when the loading suddenly reduced.

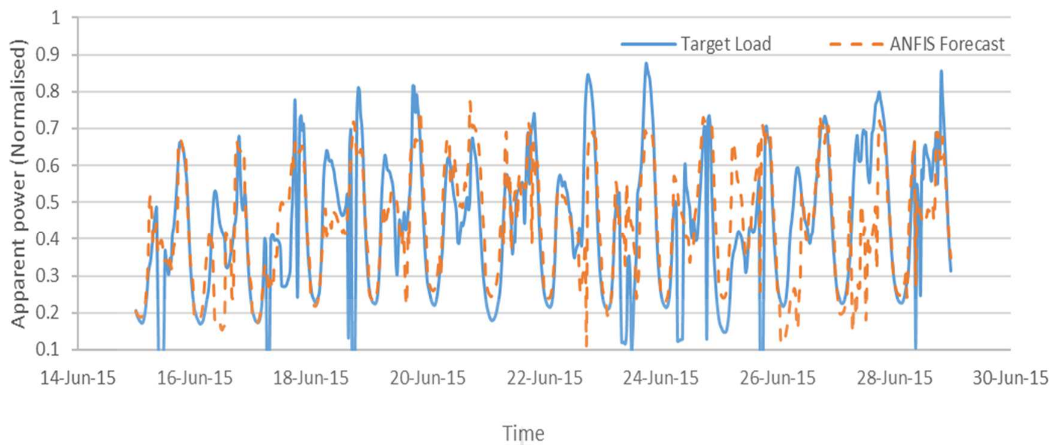

Figure 40: ANFIS lowest test error model's two-week ahead Substation B load forecast vs target load

6.4.2. OP-ELM Results

The OP-ELM models were tested for a two-week ahead load forecast and the performance for was recorded. This performance is presented in Table 33 and Table 34, respectively. It was found that the OP-ELM models achieved their lowest errors when developed with cleaned loading data as opposed to the uncleaned data. The inclusion of temperature in the models' development was found to lead to increased errors. The lowest achieved errors were therefore with a model developed with cleaned loading data and without temperature. This model had sMAPE of 0.226413 (11.32 %), MAE of 0.09043 (9.04 %) and RMSE of 0.143772 (14.38 %). These errors were achieved by a model with ten hidden nodes and are bolded in Table 34. The lowest attained errors was with a model developed with cleaned data and ten hidden nodes. These lowest attained errors are bolded in Table 33. Models with 100 and 110 hidden nodes were not attainable when using cleaned data with input variables Group A.

Table 33: Substation C's OP-ELM models' load forecasting performance with non-cleaned loading data

Input variables	Number of hidden nodes	Performance		
		sMAPE	*MAE*	*RMSE*
Group A	**10**	**0.279146**	**0.105113**	**0.16436**
	55	0.375057	0.156518	0.232913
	80	0.515501	0.229931	0.38793
	100	0.483081	0.221968	0.396976
Group B	8	0.285174	0.106607	0.161916
	58	0.410711	0.144622	0.203808
	103	0.578441	0.208359	0.275639
	158	0.595474	0.212417	0.312686

Table 34: Substation C's OP-ELM models' load forecasting performance with cleaned up loading data

Input variables	Number of hidden nodes	Performance		
		sMAPE	*MAE*	*RMSE*
Group A	**10**	**0.226413**	**0.09043**	**0.143772**
	50	0.347886	0.22436	0.482956
	100	-	-	-
	110	-	-	-
Group B	8	0.235541	0.092093	0.141977
	58	0.281593	0.108865	0.169301
	108	0.376353	0.149048	0.232213
	158	0.626704	0.213157	0.29973

The t-test showed that the results that attained the lowest errors with and without temperature had no significant difference. This was observed with both the experiments with cleaned and non-cleaned data. Since the errors were still lower without the use of temperature, OP-ELM models for this type of load profile should be developed without the use of temperature. The two-week load forecast for the OP-ELM model that achieved the lowest error is plotted against the target load in Figure 41.

Table 35: OP-ELM models' load forecast t-test results for Substation C

Compared model's results	P-value
Lowest errors: Cleaned vs Raw data	2.31535×10^{-32}
Cleaned data lowest errors: Input Group B vs Input Group A	0.759253005
Raw data lowest errors: Input Group B vs Input Group A	0.11326398
Raw Data: Lowest vs Highest error	0.007879207

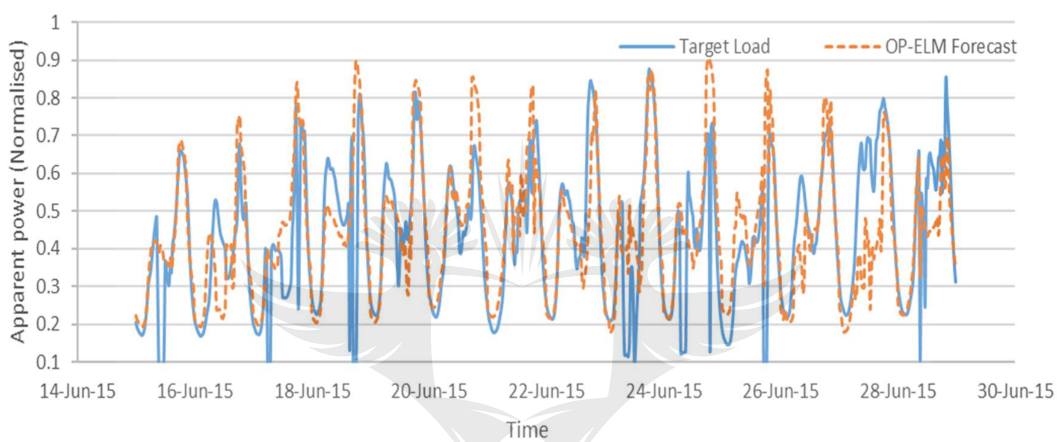

Figure 41: OP-ELM lowest test error model's two-week ahead Substation C load forecast vs target load

6.4.3. DBN Results

Tables 36 and 37, respectively, show that DBN models achieved their lowest load forecasting errors when developed without temperature in both experiments with non-cleaned and cleaned loading data. A model that attained the lowest error was developed with cleaned loading data, using input variables Group B and had nine hidden units. This models' load forecasting performance results were an sMAPE of 0.24054 (12.03 %), MAE of 0.092952 (9.3 %) and RMSE of 0.145891 (14.6 %) as shown in bold in Table 37.

Table 36: DBN models' load forecasting performance with non-cleaned loading data for substation C

Input variables	Number hidden units	Performance		
		sMAPE	***MAE***	***RMSE***
Group A	4	0.398581	0.148483	0.180647
	8	2	0.418499	0.461072
	9	2	0.418499	0.461072
	10	-	0.418499	0.461072
	11	2	0.418499	0.461072
	12	-	0.418499	0.461072
	13	2	0.418499	0.461072
	14	-	0.418499	0.461072
	15	-	0.418499	0.461072
	16	-	0.418499	0.461072
	32	-	0.418499	0.461072
Group B	4	0.344318	0.128328	0.164776
	8	0.31624	0.118985	0.160006
	9	0.297192	0.11072	0.159004
	10	2	0.418499	0.461072
	11	2	0.418499	0.461072
	12	2	0.418499	0.461072
	13	**0.296264**	**0.110234**	**0.158867**
	14	2	0.418499	0.461072
	15	2	0.418499	0.461072
	16	2	0.418499	0.461072
	32	-	0.418499	0.461072

Table 37: DBN models' load forecasting performance with cleaned loading data for substation C

Input variables	**Number hidden units**	**Performance**		
		sMAPE	***MAE***	***RMSE***
Group A	4	0.280238	0.108589	0.1449
	8	1.999999	0.430297	0.469002
	9	0.45363	0.182622	0.227979
	10	1.601264	0.401537	0.437211
	11	2	0.430297	0.469002
	12	2	0.430297	0.469002
	13	-	0.430297	0.469002
	14	-	0.430297	0.469002
	15	-	0.430297	0.469002
	16	-	0.430297	0.469002
	32	-	0.430297	0.469002
Group B	4	0.284593	0.111641	0.146439
	8	0.279181	0.107977	0.1433
	9	**0.24054**	**0.092952**	**0.145891**
	10	0.241173	0.092986	0.146564
	11	1.619861	0.386188	0.424417
	12	0.764644	0.224283	0.259192
	13	0.376284	0.132758	0.186961
	14	2	0.430297	0.469002
	15	2	0.4302970	0.469002
	16	2	0.430297	0.469002
	32	-	0.430297	0.469002

The t-test was conducted and it showed that there was a significant difference in the results that gave the lowest error for experiments with cleaned data. The model developed with temperature as an input variable and cleaned loading data was the best performing DBN load forecasting model. The t-test results for the DBN models are shown in Table 38. The forecasted load that gave the lowest DBN load forecasting error is plotted against the target load in Figure 42.

Table 38: DBN models' load forecast t-test results for Substation C

Compared model's results	P-value
Lowest errors: Cleaned vs Raw data	6.894×10^{-55}
Cleaned data lowest errors: Input Group B vs Input Group A	8.21832×10^{-11}
Raw data lowest errors: Input Group B vs Input Group A	0.057609591
Raw Data: Lowest vs Highest error	0.057609591

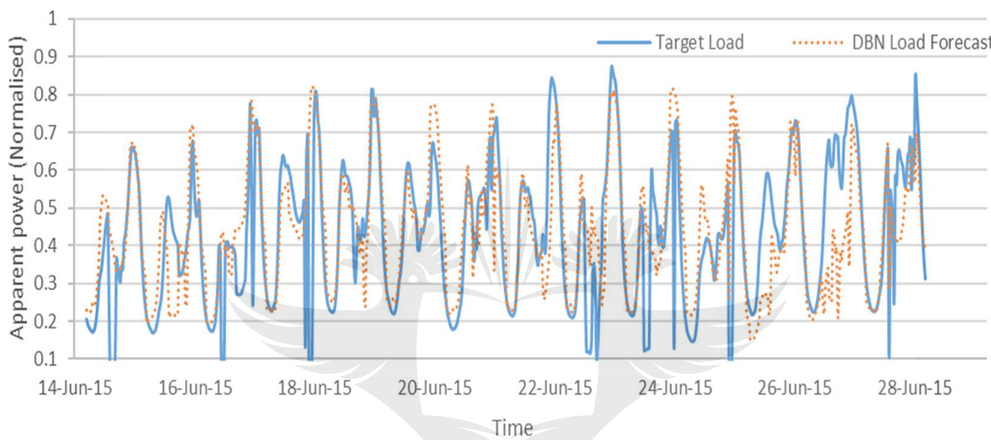

Figure 42: DBN lowest test error model's two-week ahead Substation C load forecast vs target load

6.4.4. LSTM Results

LSTM models were trained and tested with cleaned and uncleaned data as detailed out in Section 3.4. These models' load forecasting performance is presented in Table 39 and Table 40, respectively. From these results it was observed that the lowest load forecasting error was attained by a model developed with cleaned loading data and input variable Group 2. This model achieved an sMAPE of 0.2307 (11.54 %), MAE of 0.0896 (8.96 %) and RMSE of 0.14065 (14.07 %).

Table 39: LSTM-RNN models' performance with non-cleaned loading data

Input variables	Number hidden units	Performance		
		sMAPE	***MAE***	***RMSE***
Group A	60	0.286864	0.106295	0.162173
	67	0.347221	0.121265	0.16158
	336	0.298951	0.109542	0.15679
	470	0.301678	0.112097	0.154771
	538	0.288156	0.105127	0.152514
	672	0.293449	0.106803	0.151341
Group B	60	0.383348	0.141005	0.192572
	67	0.39196	0.146102	0.179556
	336	0.344087	0.122243	0.170315
	470	**0.263784**	**0.098621**	**0.153207**
	538	0.291305	0.10337	0.152825
	672	0.32681	0.121577	0.159715

Table 40: LSTM-RNN models' performance with cleaned loading data

Input variables	Hidden layers	Performance		
		sMAPE	***MAE***	***RMSE***
Group A	60	0.258684	0.102794	0.156705
	67	0.278574	0.107823	0.152683
	336	0.356105	0.139668	0.167961
	470	0.265542	0.099076	0.149288
	538	0.340204	0.132768	0.171631
	672	0.24577	0.101353	0.152643
Group B	60	0.251145	0.100358	0.155083
	67	0.272416	0.102874	0.153517
	336	0.242085	0.094544	0.143096
	470	0.299866	0.114675	0.146952
	538	0.412537	0.171554	0.208968
	672	**0.230693**	**0.089595**	**0.14065**

All load forecasting results compared using the t-test were found to be significantly different to each other. This finding was with the exception of the results obtained by models developed with and without temperature data, for experiments with cleaned loading data. The t-test results are presented in Table 41. The two-week load forecasting results that gave the lowest LSTM load forecasting error are plotted against the target load in Figure 43.

Table 41: LSTM-RNN models' load forecast t-test results for Substation B

Compared model's results	P-value
Lowest errors: Cleaned vs Raw data	$9.86346E \times 10^{-05}$
Cleaned data lowest errors: Input Group B vs Input Group A	0.346937152
Raw data lowest errors: Input Group B vs Input Group A	8.06653×10^{-19}
Raw Data: Lowest vs Highest error	3.11834×10^{-12}

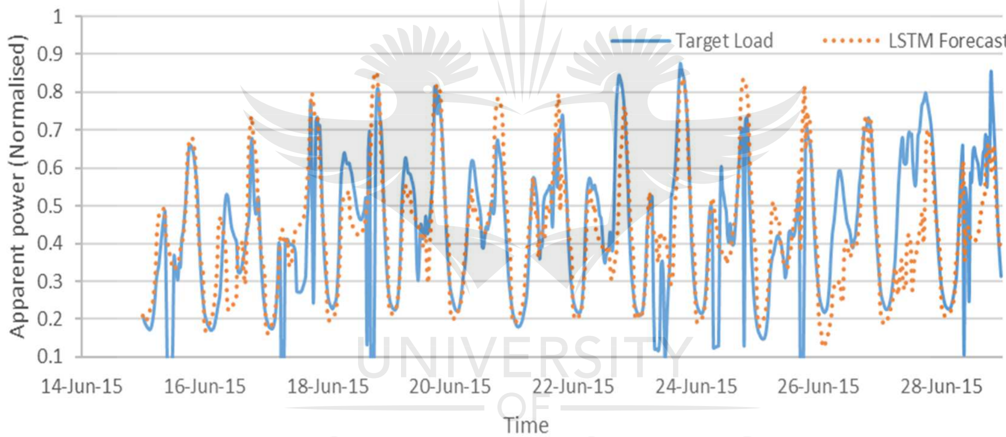

Figure 43: LSTM-RNN lowest test error model's two-week ahead Substation C load forecast vs target load

6.4.5. Results Discussion

A deep learning technique, LSTM-RNN, was found to achieve the highest load forecasting accuracy. All the techniques' models achieved their lowest errors when developed with cleaned-up loading data. The deep learning techniques all achieved their lowest errors with models developed with temperature as an input variable. The opposite behaviour was observed for machine learning techniques, where the lowest errors were attained without the use of temperature. The lowest obtained errors with each of the technique are presented in Table 42. The lowest attained errors ANFIS, were the higher than the other three techniques lowest errors. DBN had the second highest errors, followed by OP-ELM.

Table 42: Summary of third case study's lowest errors per model

Technique	Input variables	Performance		
		sMAPE	*MAE*	*RMSE*
ANFIS	Group A	0.260919	0.100845	0.149861
OP-ELM	Group A	0.226413	0.09043	0.143772
DBN	Group B	0.24054	0.092952	0.145891
LSTM-RNN	**Group B**	**0.230693**	**0.089595**	**0.14065**

The results for the statistical significance tests between the different techniques' models' results that led to their lowest attained errors are presented in Table 43. It was observed that the LSTM model's results were significantly different to those of the other three techniques. There was no significant difference between the DBN results and the ANFIS results. There was a significant difference between the rest of the techniques' results. Therefore the model whose results achieved the lowest error between the compared models had superior performance. Hence, for this case study, based on the t-test: LSTM can be regarded to be superior than all the three techniques, OP-ELM to be more superior than DBN and ANFIS and, ANFIS and DBN comparable.

Table 43: Substation C different techniques' lowest error models' load forecast t-test results

Compared model's lowest errors	P-value
ANFIS vs OP-ELM	0.01315299
ANFIS vs DBN	0.42192947
ANFIS vs LSTM	0.00345749
OP-ELM vs DBN	0.00548995
OP-ELM vs LSTM	6.5567E-21
DBN vs LSTM	3.7124E-18

6.5. Case Studies' Results Comparison

In Substation A, in the first case study a DBN model obtained the best performance. This model was trained with Group B input variables, which included temperature as an input variable. This model achieved an sMAPE of 3.93 %, MAE of 3.07 % and RMSE of 4.29 %. With Substation B, in the second case study, an LSTM model trained with Group B input variables achieved the lowest error with an sMAPE of 3.29 %, MAE of 4.6 % and RMSE of 5.51 %. In the third case study, an LSTM model achieved the lowest load forecasting error with an sMAPE of 0.2307 (11.54 %), MAE of 0.0896 (8.96 %) and RMSE of 0.14065 (14.07 %). Deep learning techniques thus achieved the best load forecasting performance in all three case studies. This could be further observed from deep learning techniques achieving the 2nd best load forecasting performance in the first and the second case study, respectively. The machine learning techniques' models were found to be less accurate than those of deep learning techniques in all case studies. This observation was with the exception of OP-ELM and DBN in the third case study. In this case study OP-ELM achieved a higher accuracy than DBN. ANFIS' best load forecasting model in each case study achieved the highest error in relation to the other techniques' best load forecasting models. However, the ANFIS results were found not to be significantly different to the DBN results in the 2nd and 3rd case study. This finding was despite ANFIS achieving a higher load forecasting error in both cases. DBN results were significantly different and better than those of OP-ELM in the 2nd case study. Following these results we can deduce that deep learning techniques can be regarded as the more efficient techniques for load forecasting in South African distribution networks. In both the 1st and the 2nd cases, the best performing models were attained with non-cleaned loading data. Hence the applications of the load forecasting system can be implemented by using uncleaned loading data from the database as is, as long as the data has an acceptable integrity. This finding was opposite with Substation C, where all the techniques' lowest errors were obtained by models developed with cleaned data. From the three cases it can be generalised that both machine learning and deep learning techniques can achieve high load forecasting accuracies without cleaning up loading data. However, other factors, such as significant load growth, combined with the dips in loading may lead to a need to clean up the loading to achieve higher accuracies as in case study 3.

In the 1st and 3rd case study, it was found that the deep learning techniques achieved their highest accuracy when developed with input variable Group B, which included temperature. Here the machine learning techniques achieved lower errors without the inclusion of temperature. In the second case study, only the OP-ELM models achieved higher accuracies with the inclusion of temperature in the model development. Generalising across the three cases it can be stated that machine learning techniques models generally achieve low load forecasting errors without the inclusion of temperature in their development. It can be further stated that deep learning techniques' models achieve lower errors with the inclusion of temperature. This observation can be regarded to be more applicable to power redistributor load types than it is to large industrial customer load types. Thus, for a redistributor customer type or a customer with a load profile similar to that in case study 1 and 3, temperature should be used in to develop deep learning load forecasting models as it can lead to better performance. For a customer type or a customer with a load profile similar to that in case study 2, temperature should not be used when using deep learning techniques to forecast the customer's load.

6.6. Chapter Summary

This chapter presented the third case study in this research. The case study was conducted using a power redistributor that is supplied power at 132 kV through a switching substation. The load forecasting performance of machine learning techniques, ANFIS and OP-ELM, and deep learning techniques, DBN and LSTM, was investigated on this substation load. The impact of using temperature data and cleaning up loading data for dips was also investigated. It was found that all the techniques achieved the lowest load forecasting error with models developed with cleaned loading data. The deep learning techniques achieved their lowest error when temperature data was used to develop their models. An LSTM load forecasting model achieved the lowest load forecasting error.

The chapter further compared the three case studies' findings. It was found that in the first two cases using raw data, i.e. not cleaning up the data for spikes and dips, led to the techniques' models' best load forecasting results. The opposite was true in the third case study. Temperature was seen to affect performance based on the machine learning/deep learning technique used and the type of load that was being forecasted. It was however, observed that the inclusion of temperature generally led to deep learning techniques' models with low errors. The use of raw/non-cleaned loading data generally led to machine learning/deep learning load forecasting models with low forecasting errors.

Chapter 7 – Conclusions and Recommendations

"No country can really develop unless its citizens are educated"

Nelson Mandela

7.1. Introduction

This study was conducted to introduce an artificial-intelligence based system that can be implemented for load forecasting South African distribution power system networks' loads to improve maintenance planning. The study's rationale, objectives and contributions were presented in Chapter 1. The load forecasting literature review was also presented in this chapter. Chapter 1 furthermore introduced machine learning and deep learning concepts. Chapter 2 introduced AI concepts and the application of machine learning and deep learning techniques. The machine learning and deep learning techniques used in this research were presented in this chapter. Chapter 3 presented the proposed load forecasting system and the experimental approach. Chapters 4, 5 and 6 presented the two case studies in this research. This includes the experiments conducted, the results obtained and a discussion of these results.

This chapter concludes the document. Hence, the findings of this study and each chapter's conclusions are presented in this chapter. This discussion is followed by an assessment of the thesis' novel contributions. Future work recommendations are made and then the paper is concluded with final remarks.

7.2. Conclusions

It was mentioned that South Africa has an extensive electrification program which aims to achieve universal access by 2025/2026. However, South African distribution power utilities are faced with multiple challenges over and above the need to connect customers to the grid. These are challenges such as a financial strain, rising non-technical losses and aging equipment that requires maintenance or upgrades. Maintenance work and connection of customers to the grid require scheduled power outages to enable personnel to work safely on the power system equipment. These outages can mean a further loss of revenue if not optimally planned for. Not carrying out these outages and hence, not conducting maintenance, can lead to loss of revenue from frequent equipment failures, emergency work, etc. Development and deployment of limited strategies by distribution utilities in planning for utility equipment upgrades and maintenance can lead to utilities in worse financial situations. The literature review showed that there are limited studies on the application of state of the art AI techniques in South African load forecasting, with studies focusing on distribution networks almost non-existent. It also became evident that recent state of the art deep learning techniques have not been applied in South African load forecasting. The problem statement, hence, led to the following key objectives:

- Introduce and utilise state of the art techniques to optimise distribution operations
- Achieve accurate distribution load forecasting to drive optimal maintenance planning

How the novel contributions following these objectives were achieved will be discussed in Section 7.3.

An AI literature review was presented. This included the introduction of related concepts such as learning and intelligence. The literature review also presented machine learning and deep learning applications in various areas such as robotics, medicine, natural language processing, power systems, power electronics and computer vision. The machine learning and deep learning techniques utilised in this study, together with their advantages and disadvantages, were presented. These techniques are ANFIS, OP-ELM, LSTM-RNN and restricted Boltzmann machine's DBN.

An AI distribution network load forecasting system was proposed. The system had a number of modules, including the module dealing with the measurement and collection of the loading data, the module that determines the loading data's integrity and the hybrid AI/deep learning load forecasting module. The experimental setup was presented together with the three performance measures that were used to measure the AI model's performance. These performance measures are the following error measurements: sMAPE, MAE and RMSE. The statistical significance test to measure the significance in the performance of the different techniques was also presented. An AI data integrity analysis module was presented as a part of the proposed load forecasting system.

Three case studies were conducted on three different South African distribution substations in this research. The substation used in the first case study (Substation A) is a distribution customer who redistributes power. While the second case study was conducted using an industrial large power user's substation (Substation B). The third case study (Substation C) was also a power redistributor customer. In all three case studies, the machine learning and deep learning techniques' load forecasting models were compared to each other across two main experiments, one with "cleaned" loading data and the other with uncleaned/raw data.

These two experiments had two sub-experiments, where one sub-experiment included temperature in the development of models, and the other one did not include temperature. Statistical significance tests were conducted to evaluate the difference in the performance of each techniques' models.

In the first two case studies, with Substation A and Substation B, it was found that the best performance was obtained with uncleaned loading data. It was also found that the best load forecasting model in both cases was a deep learning model. The best performing model in the first case study, a DBN, achieved an sMAPE of 3.93 %, MAE 3.07 % and RMSE of 4.29 %. An LSTM-RNN model achieved the best performance in the second case study. In the third case study, with Substation C, the machine learning techniques' models with the lowest load forecasting error were those developed without the use of temperature. With the deep learning techniques these models with the lowest error where those developed with temperature as an input variable. Both machine learning and deep learning techniques achieved their lowest errors with models developed with cleaned-up loading data. LSTM-RNN achieved the lowest load forecasting errors in the third case study with an sMAPE of 0.2307 (11.54 %), MAE of 0.0896 (8.96 %) and RMSE of 0.14065 (14.07 %). Deep learning techniques overall gave the higher load forecasting accuracies over the machine learning techniques. The cleaning up of loading data was also generally seen to not lead to improved load forecasting performance. The observation of the opposite behaviour in the third case study could be as a result of the drastic change in consumption that was observed, or a combination of this and the amount of zero load consumption data points this substation had.

7.3. Assessment of Thesis' Novel Contributions

This research work's novel contributions were presented in Section 1.8 to solve the challenges presented in Chapter 1. This section will present how each of these novel contributions was achieved.

1. Introduction of a unique South African distribution networks load forecasting system that utilises state of the art machine learning and deep learning techniques

The unique system for forecasting South African distribution network loads using state of the art machine learning and deep learning techniques was introduced in Section 3.3. In Chapters

4 to 6 load forecasting models were developed and tested using real South African distribution substations' data. machine learning and deep learning techniques were found to be effective in forecasting South African distribution loads and can therefore be used to improve Dx maintenance planning.

2. Introduction of the application of deep learning techniques in South African load forecasting

Two deep learning techniques, DBN and LSTM-RNN were introduced in Section 3.4.3. and Section 3.4.4, respectively. The load forecasting performance of these two deep learning techniques on South African distribution networks was investigated via three different distribution substation case studies in Chapter 4, Chapter 5 and Chapter 6. The deep learning techniques were found to outperform machine learning techniques.

3. A novel comparative study of sophisticated AI techniques' performance on different South African distribution customers. The impact of data clean-up and the inclusion of geographical temperature on the performance of these techniques per customer type is also studied

Section 4.4.5, Section 5.4.5., Section 6.4.5 and Section 6.5 compared the performance of machine learning and deep learning techniques for the three case studies in this research. These case studies were conducted on three different South African distribution customers. This comparison included an analysis of the impact of using loading data that has been cleaned up to remove dips versus data that was not modified before the development of the models on the model's performance. It was found that the machine learning and deep learning techniques' models generally achieved their highest accuracy when developed using the non-cleaned loading data. The impact of including the geographical temperature as a variable in the machine learning and deep learning models' development was investigated and the findings presented in Section 4.4.5, Section 5.4.5., Section 6.4.5 and Section 6.5 . A summary of these findings was also presented in Section 7.2 of this chapter. The temperature impact on the performance of the different models was observed to depend on the machine learning /deep learning technique used and the customer type.

4. Introduction of a novel AI based process for determining distribution loading data integrity

A novel AI based process to determine power consumption data integrity in distribution networks was presented in Section 3.2. The system was developed and then tested on South African distribution loading data. The model falsely classified 11 % of the data as having low data integrity and 3 % as undeterminable, in comparison to a manual analysis which included feeder balancing and loads cross-checking.

The discussions in this section show that the novel contributions of this study were achieved. The system presented in this study can be extended to other distribution networks, such as power redistributors' reticulation networks or industrial customers' internal distribution networks, to improve maintenance planning. The application of machine learning/deep learning techniques in South African Dx load forecasting can lead to optimal planning of maintenance outages and thus achieve cost savings. There are areas that were identified during this study that presented opportunities in utility load forecasting and optimal maintenance planning. These areas were not pursued in this research, as they did not form part of the scope of this study and are presented in Section 7.4.

7.4. Recommended Future Work

There are opportunities related to this research work that were identified but not pursued in this research as they were not this study's focus. It is recommended that these identified opportunities be pursued further.

It is recommended that the machine learning and deep learning techniques studied here, be investigated for application in transmission networks load forecasting. The study should focus on integrating distribution networks' load forecasting with transmission network load forecasting to achieve improved load forecasting.

It was found that the adaptation of AI in load forecasting is almost non-existent in other African countries. It is therefore, also recommended that the feasibility of the techniques studied in this research be applied in other African countries and other developing countries.

The performance of machine learning and deep learning with the addition of other weather parameters, humidity, wind speed, rainfall, etc. should be investigated further for models that showed an improved performance with temperature in their development.

Lastly, the use of optimisation techniques such as particle swarm and genetic algorithms for optimal maintenance scheduling using load forecasting as an input, should be investigated.

7.5. Closing Remarks

This study presented a novel South African power distribution networks AI load forecasting system. The study introduced recent state of the art machine learning and deep learning techniques in South African distribution load forecasting. Deep learning techniques that were introduced to South African load forecasting were found to outperform machine learning techniques in forecasting three different distribution substation's loads. This study also investigated the impact of data 'clean-up' or modification during pre-processing. It was found that models developed without cleaning the loading data generally achieved lower errors than those developed with cleaned-up loading data in two of the three case studies. The impact of weather parameters were also investigated. This impact was studied by using temperature as an input variable in the development of machine learning and deep learning models. It was found that the impact of temperature depends on the customer type and the technique being used. However, it was observed that in general, machine learning techniques achieved lower errors without the inclusion of temperature and deep learning techniques reached lower load forecasting errors with the inclusion of temperature in the development of their models. This research's findings and introduced system can be implemented by utilities to assist in maintenance planning and hence reduce cost.

References

[1] S. Motepe, A. N. Hasan and R. Stopforth, "South African distribution networks load forecasting using ANFIS," in *IEEE Power Electronics Drivers and Energy Systems (PEDES)*, Chennai, India, 2018.

[2] A. N. Hasan, Effective use of artificial intelligence in predicting energy consumption and underground dam levels in two gold mines in South Africa, Johannesburg: D.Ing. Thesis: University of Johannesburg, 2013.

[3] Eskom, "Fact sheet: Transmission and distribution of electricity," December 2017. [Online]. Available: http://www.eskom.co.za/AboutElectricity/FactsFigures/Documents/TD0003TransmissionDistributionRev9.pdf. [Accessed 5 December 2018].

[4] IPP Projects, "Independent power producer procurement programme," [Online]. Available: https://www.ipp-renewables.co.za/. [Accessed February 2019].

[5] GET.invest, "South Africa energy sector," [Online]. Available: https://www.africa-eu-renewables.org/market-information/south-africa/energy-sector/.[Accessed February 2019].

[6] M. Kornatka, "Distribution of SAIDI and SAIFI indices and the saturation of the MV network with remotely controlled switches," in *18th International Scientific Conference on Electric Power Engineering (EPE)*, Kouty nad Desnou, 2017.

[7] C. Xie, J. Wen, W. Liu and J. Wang, "Power grid maintenance scheduling intelligence arrangement supporting system based on power flow forecasting," *Physics Procedia,* Vols. 24, Part B, pp. 832-837, 2012.

[8] A. Froger, M. Gendreau, J. E. Mendoza and L.-M. R. É. Pinson, "Maintenance scheduling in the electricity industry: A literature review," *European Journal of Operational Research,* vol. 251, pp. 695-706, 2016.

[9] H. Xianchao, Z. Lizi, S. Jun and Z. Jingwei, "Distribution Maintenance Time scheduling Using a Genetic Algorithm," in *Joint International Conference on Power System Technology and IEEE Power India Conference*, New Delhi, India, 2008.

[10] N. Kumarappan and K. Suresh, "Combined SA PSO method for transmission constrained maintenance scheduling using levelized risk method," *Electrical Power and Energy Systems,* vol. 73, p. 1025–1034, 2015.

[11] W. W. Godard, "Electrical utility load forecasting [includes discussion]," *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems,* vol. 74, no. 3, pp. 1428-1440, 1955.

[12] D. K. A. Gillies, B. Bernholtz and P. J. Sandiford, "A new approach to forecasting daily peak loads [includes discussion]," *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems,* vol. 75, no. 3, pp. 382-387, 1956.

[13] L. Han, Y. Peng, Y. Li, B. Yong, Q. Zhou and L. Shu, "Enhanced deep networks for short-term and medium-term load forecasting," *IEEE Access,* vol. 7, pp. 4045-4055, 2019.

[14] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo and Y. Xu, "Short-term residential load forecasting based on resident behaviour learning," *IEEE Transactions on Power Systems,* vol. 33, no. 1, pp. 1087-1088, 2018.

[15] S. Welikala, C. Dinesh, M.P.B. Ekanayake, R. I. Godaliyadda and J. Ekanayake, "Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting," *IEEE Transactions on Smart Grid,* vol. 10, no. 1, pp. 448-461, 2019.

[16] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu and Y. Zhang, "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," *IEEE Transactions on Smart Grid,* vol. 10, no. 1, pp. 841-851, 2019.

[17] T. Ouyang, Y. He, H. Li, Z. Sun and S. Baek, "Modeling and forecasting short-term power load with copula model and deep belief network," *IEEE Transactions on Emerging Topics in Computational Intelligence,* vol. 3, no. 2, pp. 127-136, 2019.

[18] K. Park, S. Yoon and E. Hwang, "Hybrid load forecasting for mixed-use complex based on the characteristic load decomposition by pilot signals," *IEEE Access,* vol. 7, pp. 12297-12306, 2019.

[19] J. López, M. Rider and Q. Wu, "Parsimonious short-term load forecasting for optimal operation planning of electrical distribution systems," *IEEE Transactions on Power Systems,* vol. 34, no. 2, pp. 1427-1433, 2019.

[20] M. Zhou and M. Jin, "Holographic ensemble forecasting method for short-term power load," *IEEE Transactions on Smart Grid,* vol. 10, no. 1, pp. 425-434, 2019.

[21] L. Marwala and B. Twala, "Forecasting electricity consumption in South Africa: ARMA, neural networks and neurro-fuzzy," in *International Joint Conference on Neural Networks*, Beijing, 2014.

[22] N.M Ijumba and J.P. Hunsley, "Improved load forecasting techniques for the newly electrified areas," in *IEEE Africon Conference in Africa*, Cape Town, South Africa, 1999.

[23] W. Yuill, R. Kgokong, S. Chowdhury and S.P. Chowdhury, "Application of adaptive neuro fuzzy inference system (ANFIS) based short term load forecasting in South African power networks," in *45th International Universities Power Engineering Conference UPEC2010*, Cardiff, Wales, 2010.

[24] W. Yuill, R. Kgokong, S. P. Chowdhury and S. Chowdhury, "Management of short term load forecasting in South African power networks," in *2010 International Conference on Power System Technology*, Hangzhou, China, 2010.

[25] L. Marwala and B. Twala, "Univariate modelling of electricity consumption in South Africa: neural networks and neuro-fuzzy systems," in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, Manchester, UK, 2013.

[26] S. Heunis and M. Dekenah, "A load profile prediction model for residential consumers in South Africa," in *Twenty-Second Domestic Use of Energy*, Cape Town, South Africa, 2014.

[27] L. Marwala, Forecasting electricity demand in South Africa using artificial intelligence, Johannesburg: D.Phil. Thesis: University of Johannesburg, 2015.

[28] L.K. Tartibu and K.T. Kabengele, "Forecasting net energy consumption of South Africa using artificial neural network," in *2018 International Conference on the Industrial and Commercial Use of Energy (ICUE)*, Cape Town, South Africa, 2018.

[29] R. Inglesi, "Aggregate electricity demand in South Africa: Conditional forecasts to 2030," *Applied Energy,* vol. 87, no. 1, pp. 197-204, 2010.

[30] C. Nikolopoulos, Expert systems, New York, USA: Marcel Dekker, Inc, 1997.

[31] E. Kumar, Artificial intelligence, New Delhi, India: I.K. International Publishing House Pvt. Ltd, 2008.

[32] S. Russel and P. Norvig, Artificial intelligence: A modern approach, Pearson Education Limited, 2016.

[33] F. Iafrate, Artificial intelligence and big data, Great Britain and USA: ISTE Ltd and Wiley & Sons, Inc, 2018.

[34] J.M. Martín-Doña, A.M. Gomezs, J.A. Gonzalez and A.M. Peinado, "A deep learning loss function based on the perceptual evaluation of the speech quality," *IEEE Signal Processing Letters,* vol. 25, no. 11, pp. 1680-1684, 2018.

[35] F. Fahiman, S.M. Erfani, S. Rajasegarar, M. Palaniswami and C. Leckie, "Improving load forecasting based on deep and k-shape clustering," in *International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA, 2017.

[36] X-W. Chen and X. Lin, "Big data deep learning: challenges and perspectives," *IEEE Access,* vol. 2, pp. 514-625, 2014.

[37] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014.

[38] WAYMO, "WAYMO - We're building the world's most experienced driver," WAYMO LCC, 2019. [Online]. Available: https://waymo.com/. [Accessed April 2019].

[39] TESLA, "TESLA - The future of driving," TESLA, 2019. [Online]. Available: https://www.tesla.com/autopilot. [Accessed April 2019].

[40] B. Marr, "Forbes - The amazing ways chinese tech giant alibaba uses artificial intelligence and machine learning," Forbes, 23 July 2018. [Online]. Available: https://www.forbes.com/sites/bernardmarr/2018/07/23/the-amazing-ways-chinese-tech-giant-alibaba-uses-artificial-intelligence-and-machine-learning/#53df7cc4117a. [Accessed April 2019].

[41] AWS AI, "AWS AI GE Healthcare drives better outcomes with machine learning," Amazon Web Services, [Online]. Available: https://aws.amazon.com/machine-learning/customers/innovators/ge_healthcare/. [Accessed April 2019].

[42] M. K. J. Han, Data mining: Concepts and techniques, San Francisco, CA, USA: Morgan Kaufmann Publishers, 2001.

[43] M. Tegmark, Life 3.0 being human in the age of artificial intelligence, United Kingdom: Allen Lane, 2017.

[44] E. Rich, K. Knight and S. B. Nair, Artificial intelligence third edition, New Delhi, India: Tata McGraw-Hill, 2009.

[45] E. Hurwitz, A. N. Hasan and C. Orji, "Soft biometric thermal face recognition using FWT and LDA feature extraction method with RBM DBN and FFNN classifier algorithms," in *Fourth International Conference on Image Information Processing (ICIIP)*, Shimla, India, 2017.

[46] N. Majumder, S. Poria, A. Gelbukh and E. Cambria, "Deep learning-based document modeling for personality detection from text," *IEEE Intelligent Systems,* vol. 32, no. 2, pp. 74-79, 2017.

[47] O. Abedinia, N. Amjady, M. Shafie-khah and J.P.S. Catalão, "Electricity price forecast using combinatorial neural network trained by a new stochastic search method," *Energy Conversion and Management,* vol. 105, pp. 642-654, 2015.

[48] J. Lagoa, F. D. Ridder and B. D. Schuttera, "Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms," *Applied Energy,* vol. 221, pp. 386-405, 2018.

[49] K. Chen, Z. He, K. Chen, J. Hu and J. He, "Solar energy forecasting with numerical weather predictions on a grid and convolutional networks," in *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*, Beijing, China, 2017.

[50] A. Narayan and K.W. Hipel, "Long short term memory networks for short-term electric load forecasting," in *IEEE International Conference On Systems, Man And Cybernetics (SMC)*, Banff, AB, Canada, 2017.

[51] R.F. Berriel, A.T. Lopes, A. Rodrigues, F.M. Varejão and T. Oliveira-Santos, "Monthly energy consumption forecast: A deep learning approach," in *International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA, 2017.

[52] K. Amarasinghe, D. L. Marino and M. Manic, "Deep neural networks for energy load forecasting," in *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, Edinburgh, UK, 2017.

[53] X. Peng, L. Xiong, J. Wen, Y. Xu, W. Fan, S. Feng and B. Wang, "A very short term wind power prediction approach based on multilayer restricted boltzmann machine," in *2016 IEEE PES Asia-Pacific Power and Energy Conference*, Xi'an, China, 2016.

[54] C-Y. Zhang, C. L. P. Chen, M. Gan and L.Chen, "Predictive deep Boltzmann machine for multiperiod wind speed forecasting," *IEEE Transactions on Sustainable Energy,* vol. 6, no. 4, pp. 1416-1425, 2015.

[55] Y.Q. Neo, T.T.Teo, W.L. Woo, T. Logenthiran and A. Sharma, "Forecasting of photovoltaic power using deep belief network," in *Proceeding of the 2017 IEEE Region 10 Conference (TENCON)*, Penang, Malaysia, 2017.

[56] J. Wang, J. Zhang and X. Wang, "Bilateral LSTM: A two-dimensional long short-term memory model with multiply memory units for short-term cycle time forecasting in re-entrant manufacturing systems," *IEEE Transactions on Industrial Informatics,* vol. 14, no. 2, pp. 748 - 758, 2018.

[57] T. Fujita, W. Bai and C. Quan, "Long short-term memory networks for automatic generation of conversations," in *2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Kanazawa, Japan, 2017.

[58] D. Tao, X. Lin, L. Jin and X. Li, "Principal component 2-d long short-term memory for font recognition on single chinese characters," *IEEE Transactions on Cybernetics,* vol. 46, no. 3, pp. 756 - 765, 2016.

[59] T.A. Jemimma and Y. J. V. Raj, "Brain tumor segmentation and classification using deep belief network," in *Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018)*, Madurai, India, 2018.

[60] Y-Y. Wu and C-M. Tsai, "Pedestrian, bike, motorcycle, and vehicle classification via deep learning: Deep belief network and small training set," in *2016 International Conference on Applied System Innovation (ICASI)*, Okinawa, Japan, 2016.

[61] L. Xiuli, Z. Xueying and W. Liyong, "Fault diagnosis method of wind turbine gearbox based on deep belief network and vibration signal," in *2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, Nara, Japan, 2018.

[62] G. Zhao, C. Zhang and L. Zheng, "Intrusion detection using deep belief network and probabilistic neural network," in *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, Guangzhou, China, 2017.

[63] S. Subramani, H. Wang, H. Q. Vu and G. Li, "Domestic violence crisis identification from Facebook posts based on deep learning," *IEEE Access,* vol. 6, pp. 54075-54085, 2018.

[64] K. Choi, G. Fazekas, K. Cho and M. Sandler, "The effects of noisy labels on deep convolutional neural networks for music tagging," *IEEE Transactions on Emerging Topics in Computational Intelligence,* vol. 2, no. 2, pp. 139-149, 2018.

[65] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci and N. Navab, "AggNet: Deep learning from crowds for mitosis detection in breast cancer histology images," *IEEE Transactions on Medical Imaging,* vol. 35, no. 5, pp. 1313-1321, 2016.

[66] V.H. Hinojosa and A. Hoese, "Short-term load forecasting using fuzzy inductive reasoning and evolutionary algorithms," *IEEE Transactions on Power System,* vol. 25, no. 1, pp. 565-574, 2010.

[67] B.A Høverstad, A. Tidemann, H. Langseth and P. Öztürk, "Short-term load forecasting with seasonal decomposition using evolution for parameter tuning," *IEEE Transactions on Smart Grid,* vol. 6, no. 4, pp. 911-918, 2015.

[68] J.W. Taylor and P.E. McSharry, "Short-term load forecasting methods: An evaluation based on european data," *IEEE Transactions on Power System,* vol. 22, no. 4, pp. 2213-2219, 2007.

[69] H. K. Alfares and M. Nazeeruddin, "Electric load forecasting: Literature survey and classification of methods," *International Journal of Systems Science,* vol. 33, no. 1, pp. 23-34, 2002.

[70] Z. Yun, Z. Quan, S. Caixin, L. Shaolan, L. Yuming and S. Yang, "RBF neural network and ANFIS-based short-term load forecasting approach in real-time price environment," *IEEE Transactions on Power System,* vol. 28, no. 3, pp. 853-858, 2008.

[71] T. Konjic, V. Miranda and I. Kapetanovic, "Fuzzy inference systems applied to LV substation load estimation," *IEEE Transactions on Power System,* vol. 20, no. 2, pp. 742-749, 2005.

[72] Suhartono, R. Faulina, D. A. Lusia, B. W. Otok, Sutikno and H. Kuswanto, "Ensemble method based on ANFIS-ARIMA for rainfall prediction," in *2012 International*

Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, Malaysia, 2012.

[73] K. Hsia, S. Lien, J. Su and W. Ciou, "ANFIS-based aided stabilization control and image tracking for mobile robot," in *2012 International conference on Fuzzy Theory and Its Applications (iFUZZY2012)*, Taichung, Taiwan, 2012.

[74] H. Chaudhary, S. Khatoon and R. Singh, "ANFIS based speed control of DC motor," in *2016 Second International Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity (CIPECH)*, Ghaziabad, India, 2016.

[75] N. Sivasankari and M. Malleswaran, "Evolutionary optimization in ANFIS for intelligent navigation system," in *2013 International Conference on Recent Trends in Information Technology (ICRTIT)*, Chennai, India, 2013.

[76] T. Kamel, Y. Biletskiy and L. Chang, "Open circuit fault diagnosis for the power electronic converter stages using ANFIS algorithm," in *2015 IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE)*, Halifax, NS, Canada, 2015.

[77] Y. Miche, A. Sorjamaa and A. Lendasse, "OP-ELM: Theory, experiments and a toolbox," in *18th International Conference on Artificial Neural Networks (ICANN)*, Prague, Czech Republic, 2008.

[78] G. Huang, Q. Zhu and C.K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing,* vol. 70, no. 1-3, pp. 489-501, 2006.

[79] F. Wang, Z. Zhao, X. Li, F. Yu and H. Zhang, "Stock volatility prediction using multi-kernel learning based extreme learning machine," in *2014 International Joint Conference on Neural Networks (IJCNN)*, Beijing, China, 2014.

[80] C. Wan, Z. Xu, P. Pinson., Z. Y. Dong and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Transactions on Power Systems,* vol. 29, no. 3, pp. 1033-1044, 2014.

[81] A. H. Nizar, Z. Y. Dong and Y. Wang, "Power utility nontechnical loss analysis with extreme learning machine method," *IEEE Transactions on Power Systems,* vol. 23, no. 3, pp. 946-955, 2008.

[82] X. Chen, Z. Y. Dong, K. Meng, Y. Xu, K. P. Wong and H. W. Ngan, "Electricity price forecasting with extreme learning machine and bootstrapping," *IEEE Transactions on Power Systems,* vol. 27, no. 4, pp. 2055-2062, 2012.

[83] M. Khoza and T. Marwala, "A rough set theory based predictive model for stock prices," in *2011 IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI)*, Budapest, Hungary, 2011.

[84] M. Khoza and T. Marwala, "Computational intelligence techniques for modelling an economic system," in *The 2012 International Joint Conference on Neural Networks (IJCNN)*, Brisbane, Australia, 2012.

[85] A. M. Farayola, A. N. Hasan and A. Ali, "Optimization of PV systems using data mining and regression learner MPPT techniques," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 10, no. 3, pp. 1080 - 1089, 2018.

[86] E. v. d. Walt and J. Eloff, "Using machine learning to detect fake identities: Bots vs humans," *IEEE Access,* vol. 6, pp. 6540 - 6549, 2018.

[87] G. Omer, O. Mutanga, E. M. Abdel-Rahman and E. Adam, "Performance of support vector machines and artificial neural network for mapping endangered tree species using worldview-2 data in dukuduku forest, South Africa," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 8, no. 10, pp. 4825 - 4840, 2015.

[88] C.-T. Sun and J-S. R. Jang, "Fuzzy modeling based on generalized neural networks and fuzzy clustering objective functions," in *Proceedings of the 30th IEEE Conference on Decision and Control*, Brighton, UK, 1991.

[89] J-S. R. Jang, "Fuzzy modeling using generalized neural networks and Kalman filter algorithm," in *Conference: Proceedings of the 9th National Conference on Artificial Intelligence*, Anaheim, CA, USA, 1991.

[90] J.-S.R. Jang and C.-T. Sun, "Predicting chaotic time series with fuzzy if-then rules," in *Second IEEE International Conference on Fuzzy Systems*, San Francisco, CA, USA, 1993.

[91] M. Negnevitsky, Artificial intelligence: A guide to intelligent systems 2nd edition, United Kingdom: Addison-Wesley, 2005.

[92] S. Motepe, B. Twala and R. Stopforth, "Determining South African distribution power system big data integrity using fuzzy logic: Power measurements data application," in *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics*, Bloemfontein, South Africa, 2017.

[93] L. Marwala and B. Twala, "Causality tests using basic and optimally-pruned extreme learning machines," in *2nd International Conference on Control and Robotics Engineering*, Bangkok, Thailand, 2017.

[94] Y. He, J. Deng and H. Li, "Short-term power load forecasting with deep belief network and copula models," in *9th International Conference on Intelligent Human-Machine Systems and Cybernetics*, Hangzhou, China, 2017.

[95] S. Motepe, A. N. Hasan, B. Twala and R. Stopforth, "Power distribution networks load forecasting using deep belief networks: The South African case," in *IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology*, Amman, Jordan, 2019.

[96] G. Hinton, S. Osindero and Y.W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation,* vol. 18, no. 7, pp. 1527-1554, 2006.

[97] X. Zhang, R. Wang, T. Zhang and Y. Zha, "Short-term load forecasting based on a improved deep belief network," in *2016 International Conference on Smart Grid and Clean Energy Technologies (ICSGCE)*, Chengdu, China, 2016.

[98] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences,* vol. 79, no. 8, pp. 2554-2558, 1982.

[99] H. Bourlard and C. J. Wellekens, "Speech dynamics and recurrent neural networks," in *International Conference on Acoustics, Speech, and Signal Processing*, Glasgow, UK,, 1989.

[100] Y. Bengio, R. Cardin, R. De Mori and Y. Normandin, "A hybrid coder for hidden Markov models using a recurrent neural networks," in *International Conference on Acoustics, Speech, and Signal Processing*, Albuquerque, NM, USA, 1990.

[101] S. Motepe, A. N. Hasan and R. Stopforth, "Improving load forecasting process for a power distribution network using hybrid ai and deep learning algorithms," *IEEE Access,* vol. 7, pp. 82584 - 82598, 2019.

[102] E. Dumbill, "Making sense of big data," *Big Data,* vol. 1, pp. 1-2, 2013.

[103] Gartner, "IT glossary: Big data," Gartner, 2012. [Online]. Available: http://www.gartner.com/it-glossary/big-data.

[104] S-J. Chen, T-S. Zhan, C-H. Huang, J-L. Chen and C-H. Lin, "Nontechnical loss and outage detection using fractional-order self-synchronization error-based fuzzy petri nets in micro-distribution systems," *IEEE Transactions on Smart Grid,* vol. 6, no. 1, pp. 411-420, 2015.

[105] M. Wang and Y. Sun, "A practical method to improve phasor and power measurements accuracy of DFT algorithm," *IEEE Transactions on Power Delivery,* vol. 21, no. 3, pp. 1054-1062, 2006.

[106] A. Domijan Jr., D. Czarkowski and J. Johnson, "Nonsinusoidal electrical measurement accuracy in adjustable-speed motors and drives," *IEEE Transactions on Industry Applications,* vol. 34, no. 6, pp. 1225-1233, 1998.

[107] G. Pasini, L. Peretto and R. Tinarelli, "Study of the accuracy requirements of the instrumentation for efficiency measurements in power conversion systems," *IEEE Transactions on Instrumentation and Measurements,* vol. 62, no. 8, pp. 2154-2160, 2013.

[108] D. Gallo, C. Landi and M. Luiso, "Accuracy analysis of algorithms adopted in voltage dip measurement," *IEEE Transactions on Instrumentation and Measurement,* vol. 59, no. 10, pp. 2652-2659, 2010.

[109] Q. Yang, D. Li, W. Yu and Y. Liu, "Towards data integrity attacks against optimal power flow in smart grids," *IEEE Internet of Things Journal,* vol. 4, no. 5, pp. 1726-1738, 2017.

[110] D. Supadu and R. Gaikward, "A survey on real time error detection service using MapReduce," *International Journal of Computer Engineering and Applications,* vol. 10, no. 2, pp. 47-53, 2016.

[111] Z. Gao, N. DeSalvo, P.D. Khoa, S. Kim, L. Xu, W. W. Ro, R. Verma and W. Shi, "Integrity protection for big data processing with dynamic redundancy computation," in *IEEE 12th International conference on automation computing*, Grenoble, France, 2015.

[112] I. Lebdaoui, S. El Hajji and G. Orhanou, "Managing big data integrity," in *2016 International Conference on Engineering & MIS (ICEMIS)*, Agadir, Morocco, 2016.

[113] A. N. Rukavitsyn, K. A. Borisenko, I. I. Holod and A. V. Shorov, "The method of ensuring confidentiality and integrity data in cloud computing," in *2017 IEEE International Conference on Soft Computing and Measurements (SCM)*, St. Petersburg, Russia, 2017.

[114] S. Motepe, B. Twala, Q-G Wang and R. Stopforth, "Determining distribution power system loading measurements accuracy using fuzzy logic," *Procedia Manufacturing,* vol. 7, pp. 435-439, 2017.

[115] X. Wang, M. Wang and X. Guo, "Thermal performance analysis of solar steam aided coal-fired power generation," in *2011 International Conference on Materials for Renewable Energy & Environment*, Shanghai, China, 2011.

[116] A. Almalaq and G. Edwards, "A review of deep learning methods applied on load forecasting," in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Cancun, Mexico, 2017.

[117] X. Dong, L. Qian and L. Huang, "Short-term load forecasting in smart grid: A combined CNN and k-means clustering approach," in *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Jeju, South Korea, 2017.

[118] D. L. Marino, K. Amarasinghe and M. Manic, "Building energy load forecasting using deep neural networks," in *42nd Annual Conference of the IEEE Industrial Electronics Society (IECON)*, Florence, France, 2016.

[119] T. Jarabek, P. Laurinec and M. Lucka, "Energy load forecast using S2S deep neural networks with k-shape clustering," in *2017 IEEE 14th International Scientific Conference on Informatics*, Poprad, Slovakia, 2017.

[120] S. Hosein and P. Hosein, "Load forecasting using deep neural networks," in *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Washington, DC, USA, 2017.

[121] J. Lago, F. D. Ridder and B. D. Schutter, "Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms," *Applied Energy ,* vol. 221, p. 386–405, 2018.

[122] S. Makridakis, "Accuracy measures: Theoretical and practical concerns," *International Journal of Forecasting,* vol. 9, no. 4, pp. 527-529, 1993.