

Article

Forecasting the Total South African Unplanned Capability Loss Factor Using an Ensemble of Deep Learning Techniques

Sibonelo Motepe ^{1,*} , Ali N. Hasan ²  and Thokozani Shongwe ¹ 

¹ Department of Electrical and Electronic Engineering Technology, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg 2092, South Africa; tshongwe@uj.ac.za

² Department of Electrical Engineering, Faculty of Engineering Science and Technology, Higher Colleges of Technology, Abu Dhabi 25026, United Arab Emirates; alinabeal99@gmail.com

* Correspondence: djscvii@gmail.com

Abstract: Unplanned power plant failures have been seen to be a major cause of power shortages, and thus customer power cuts, in the South African power grid. These failures are measured as the unplanned capability loss factor (UCLF). The study of South Africa's UCLF is almost non-existent. Parameters that affect the future UCLF are, thus, still not well understood, making it challenging to forecast when power shortages may be experienced. This paper presents a novel study of South African UCLF forecasting using state-of-the-art deep learning techniques. The study further introduces a novel deep learning ensemble South African UCLF forecasting system. The performance of three of the best recent forecasting techniques, namely, long short-term memory recurrent neural network (LSTM-RNN), deep belief network (DBN), and optimally pruned extreme learning machines (OP-ELM), as well as their aggregated ensembles, are investigated for South African UCLF forecasting. The impact of three key parameters (installed capacity, demand, and planned capability loss factor) on the future UCLF is investigated. The results showed that the exclusion of installed capacity in the LSTM-RNN, DBN, OP-ELM, and ensemble models doubled the UCLF forecasting error. It was also found that an ensemble model of two LSTM-RNN models achieved the lowest errors with a symmetric mean absolute percentage error (sMAPE) of 6.43%, mean absolute error (MAE) of 7.36%, and root-mean-square error (RMSE) of 9.21%. LSTM-RNN also achieved the lowest errors amongst the individual models.

Keywords: deep learning; forecasting; power outages; coal power plants; recurrent neural networks; ensemble techniques



Citation: Motepe, S.; Hasan, A.N.; Shongwe, T. Forecasting the Total South African Unplanned Capability Loss Factor Using an Ensemble of Deep Learning Techniques. *Energies* **2022**, *15*, 2546. <https://doi.org/10.3390/en15072546>

Academic Editors: Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 31 January 2022

Accepted: 7 March 2022

Published: 31 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

South Africa has been seen to be a late participant in the three key industrial revolutions [1]. The use of artificial intelligence (AI) and data is on the rise in South Africa [2–4]. This rise means that South Africa might not be a late participant in the fourth industrial revolution. In 2007, 2013, 2018, and 2019, South Africa experienced a shortage in power supply due to various challenges, leading to load shedding [1]. South Africa's public power utility, Eskom, has on several occasions stated its inability to accurately predict/forecast the unplanned capability loss factor (UCLF) as one of the major factors leading to an unreliable power supply and unpredictable load shedding [5,6]. UCLF is a term that refers to the measure of unplanned plant breakdown. The behavior of South African UCLF has not been well studied. Pretorius et al. studied the impact of the South African energy crisis on emissions [7]. This study only talks about an increase in UCLF due to maintenance deferral. The study does not talk about how to forecast UCLF, nor the major factors that contribute to UCLF that can help in the forecasting of UCLF. The UCLF, planned capability loss factor (PCLF), and other capability loss factor (OCLF), together with the installed capacity, determine the power available to supply customers. The PCLF is the planned plant

outages for the maintenance or refurbishments of the plant. This is typically a planned, set value set by the utility. The utility can decide to change their planned outage/PCLF depending on different factors. The OCLF accounts for other or random losses and is usually significantly smaller than the UCLF [8]. The installed capacity gives the number of megawatts of the installed power plant units. Micali studied the prediction of new coal power plants' availability in the absence of data in South Africa [8]. The author mentions that the work is a precursor to predicting UCLF in new plants. The author proposes using expert opinion with some data from stations where data are available. However, the work in [8] did not focus on the total UCLF, assumed limited availability of data, did not use AI techniques, and depended on expert knowledge. In [9], the authors state that expert knowledge can change from one expert to the next, and thus expert results can be different from the same data. The author, in addition, did not investigate factors that affect power supply and may influence the UCLF [8]. There is, thus, a gap in South Africa in terms of accurately forecasting UCLF. In addition, the study of the total South African UCLF behavior is a gap as only precursor work exists, and the precursor work is focused on new plants. Another gap is the use of intelligent systems that are not reliant on human experts in UCLF forecasting.

To add to the previous paragraph, the knowledge of when the power system might experience a power shortage is still a topic of interest and is not only important for the utility, but also customers. Knowing when there may be a power shortage, and hence a requirement to reduce consumption, helps customers plan their operations. Unplanned failures have been studied before. In [10], real-time prediction of distribution system outage duration using historical outage records to train neural networks was studied. The Netherlands collects information on unplanned outages from its utilities to inform its maintenance and investment policies [11].

South Africa is the highest producer of electricity in Africa and is in the top 25 producers of power in the world [12,13]. Over 80% of South Africa's power is produced by coal-fired power stations and a nuclear power station. The total South African power grid UCLF can, thus, be modeled as that of the coal and nuclear power stations. Despite the recent move towards cleaner energy, the largest power-producing countries, such as India and China, still rely heavily on coal-fired power stations [12]. The study of coal thermal power plants and behavior is, thus, still of interest [14–17]. The study of the South African coal-fired power station UCLF is, therefore, important as coal power plants are still highly used and are still a research topic of interest.

Forecasting and prediction have been topics of interest for many researchers [10,18]. This is mainly due to an interest in understanding and predicting the future behavior of certain variables. Artificial intelligence (AI) techniques have become popular in these forecasting/prediction tasks. One of the reasons for this popularity is their ability to model non-linearity with high accuracy. Khoza and Marwala used an ensemble of the multi-layer perceptron and rough set theory to predict the direction that the South African gross domestic product (GDP) would take [18]. Galius proposed a probabilistic model for modeling power distribution network blackouts [19]. In Egypt, power cable failures were analyzed to help prevent future power outages [20]. In [21], bilateral long short-term memory (LSTM) was used to forecast the short-term cycle of wafer lots for the planning and control of wafer manufacturing. The rise of computational power and access to labeled data has led to an increase in the utilization of deep learning techniques [22]. Deep learning techniques have been seen to have an excellent performance in multiple areas, such as language and speech processing, as well as computer vision [23,24]. Alhussein et al. used a hybrid of convolutional neural networks (CNN) and long short-term memory (LSTM) to forecast individual house loads [25]. Here, the researchers use CNN to select features from the input data and LSTM to learn the sequence. The authors stated a mean absolute percentage error (MAPE) improvement greater than 4% in comparison to LSTM-based models. Kong et al. also combined CNN and LSTM for short-term load forecasting in Singapore [26]. Pandit et al. compared LSTM and Markov chain models in weather

forecasting for German offshore wind farms to improve their wind turbine availability and maintenance [27]. Deep learning has also been used to forecast wind speeds at turbine locations [28]. The authors combine CNN and the gated recurrent unit (GRU) to achieve satisfactory results in comparison to existing models. Deep learning techniques have also been used to forecast the Korean postal delivery service demand [29]. This observed performance of deep learning techniques has also led to their adoption in recent load forecasting studies [30,31]. A gap still exists in the application of the state-of-the-art techniques in forecasting UCLF (and South African UCLF), as applied in forecasting in the different engineering areas.

As observed, a number of studies have used a combination of techniques to achieve improved performance [25–29]. This combination of techniques is usually termed ensemble or hybrid techniques. Ensemble techniques have also been used for classification in different engineering applications. Ramotsoela et al. used an ensemble of five artificial intelligence techniques to detect intrusion in water distribution systems [32]. The ensemble model used here combined an artificial neural network (ANN), RNN (recurrent neural network), LSTM, GRU, and CNN in a voting system. The ensemble model classified its output as an anomaly if at least two constituent models classified their outputs as an anomaly. CNN models have been combined to determine driver behavior from multiple data streams [33]. The proposed ensemble model incorporated a voting system to enhance the classification accuracy. A double ensemble model of semi-supervised gated stacked auto-encoders has been used to predict industrial key performance indicators [34]. Drif et al. proposed an ensemble of auto-encoders for recommendations [35]. The authors used an aggression method to combine outputs from the sub-models to form the ensemble model output. Bibi et al. used an ensemble-based technique to forecast electricity spot prices in the Italian electricity market [36]. The authors estimated deterministic components using semi-parametric techniques and then determined stochastic components using time series, and machine learning algorithms. The final forecast is obtained from the estimates of both components [36]. Shah et al. used a similar approach to Bibi et al. in short-term electricity demand forecasting for the Nordic electricity market [37]. The similarity is that the authors separated their approach into a deterministic and a stochastic component and then combined the estimates from them to obtain the final forecast. None of the literature covers the use of ensemble techniques in forecasting UCLF. The use of ensemble techniques in UCLF forecasting is, thus, an existing research gap.

This paper introduces the following contributions: (i) A novel study of the South African UCLF behavior using state-of-the-art AI (deep learning and ensemble) techniques. (ii) An investigation of the impact of the installed capacity, historic demand, and PCLF on the UCLF forecasting accuracy. (iii) An introduction of a novel deep-learning ensemble total South African UCLF forecasting system.

The remainder of this paper is arranged as follows: Section 2 presents the techniques used in this research. Section 3 presents the experimental setup. The proposed UCLF forecasting system is presented in Section 4. Section 5 then presents the experimental results and the discussion of the results. The paper conclusions are presented in Section 6. Section 7 presents the limitations of the study as well as future work. The paper flow chart is shown in Figure 1.

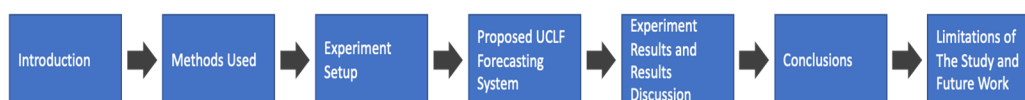


Figure 1. The paper arrangement flow chart.

2. Methods Used

This section presents the four techniques used in this research.

2.1. OP-ELM

The optimally pruned extreme learning machine (OP-ELM) is the improved version of the extreme learning machine. This improved technique, introduced by Miche et al., uses the leave-one-out (LOO) method to select the optimal number of neurons [38]. LOO marginalizes the irrelevant neurons built into ELM's network. This marginalization helps overcome the shortfall in the approximation of the training dataset's correlated and irrelevant variables. Given a training set x_i , with a target vector t_i , the OP-ELM's objective is to obtain the minimum possible error function. The OP-ELM equation is given by (1). If there exists an input weight vector connecting the k th hidden neuron and the input (w_k), a k th hidden node's bias (b_k), and an output weight connecting the output and the k th hidden neuron (β_k), such that $\sum_{k=1}^j f(w_k, b_k, x_i) \beta_k = y_i$, (1) can be re-written as (2).

$$\sum_{k=1}^j f(w_k, b_k, x_i) \beta_k = t_i \quad (1)$$

$$H\beta = T \quad (2)$$

$$H = \begin{bmatrix} f(w_1, b_1, x_1) & \cdots & f(w_k, b_k, x_1) \\ \vdots & \cdots & \vdots \\ f(w_1, b_1, x_m) & \cdots & f(w_k, b_k, x_m) \end{bmatrix}_{m \times j} \quad (3)$$

$$\beta = H^*T = (HH^T)^{-1}HT^T \quad (4)$$

where y_i is the output vector, t_i is the output target vector, H is the hidden layer's output matrix, and $k = 1, 2 \dots j$. The input weights and biases are assigned at random and do not require tuning. The hidden layer's output matrix parameters are also assigned random values. If H is a square matrix, matrix inversion can be used to determine the output weights. In a case where H is not a square matrix, the Moore–Penrose Equation (4) is used to determine the output weights. The neurons are ranked using multi-response sparse regression, and the LOO is then applied.

2.2. LSTM-RNN

The fading of previously learned patterns is a challenge experienced in standard RNN architectures. The LSTM-RNN has a memory cell to overcome this shortcoming. The memory cell is managed by non-linear gating units. The gated units of an LSTM-RNN unit can be seen in Figure 2. These gated units, the forget gate (f_n), input gate (i_n), and output gate (o_n), are presented by Equations (5)–(7), respectively. Equations (8)–(10), respectively, present the input node (g_n), the state (s_n), and the cell state (h_n). Here, n is the time step, \varnothing is the tanh function, σ is the sigmoid function, and the W matrices are the respective network activation functions' corresponding input weights. The LSTM-RNN cells are stacked after each other to achieve a deep layered LSTM-RNN. The memory cells give the models the ability to sustain memory.

$$f_n = \sigma(W_{fz}z_n + W_{fh}h_{n-1} + b_f) \quad (5)$$

$$i_n = \sigma(W_{iz}z_n + W_{ih}h_{n-1} + b_i) \quad (6)$$

$$o_n = \sigma(W_{oz}z_n + W_{oh}h_{n-1} + b_o) \quad (7)$$

$$g_n = \varnothing(W_{gz}z_n + W_{gh}h_{n-1} + b_g) \quad (8)$$

$$s_n = g_n \odot i_n + s_{n-1} \odot f_n \quad (9)$$

$$h_n = \varnothing(s_n) \odot o_n \quad (10)$$

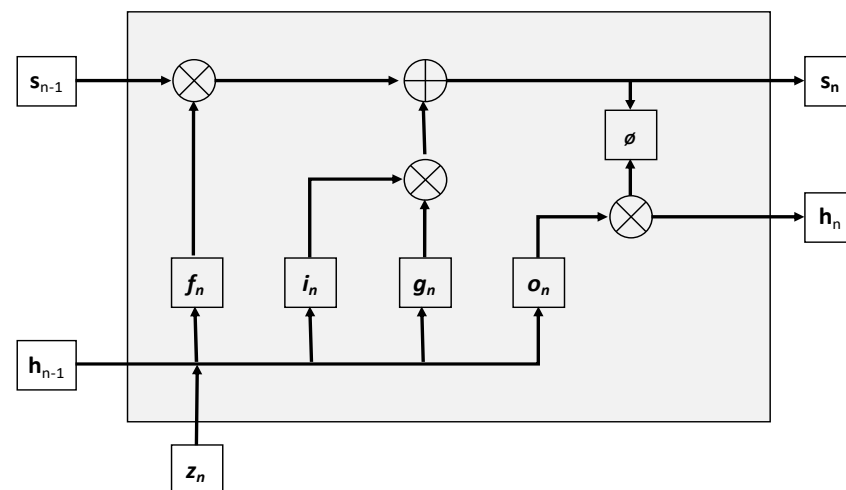


Figure 2. An LSTM–RNN cell with gated units.

2.3. DBN

The deep belief network (DBN) is built by stacking restricted Boltzmann machines (RBM). The technique was introduced in the mid-2000s by Geoffrey Hinton. There are no connections between the neurons on the same layer. There is a symmetrical and bi-directional connection between the layers. The model determines the hidden state, visible state, initial weight, and biases in the first step using unsupervised learning. Supervised learning, using back-propagation, is used to append the unsupervised learning pre-trained model. The joint distribution over the visible and hidden units is given by (11) [39].

$$P(m, h) = \frac{e^{-E(m, h)}}{\sum_n \sum_h e^{-E(m, h)}} \quad (11)$$

where $E(m, h)$ is the energy function. The conditionally independent conditional probabilities are given by (12) and (13). If the values of the hidden and visible units are from 0 to 1, (12) and (13), respectively, become (14) and (15), with $i = 1, 2 \dots k_h$ and $j = 1, 2 \dots k_m$.

$$p(m|h) = \prod_j p(m_j|h) \quad (12)$$

$$p(h|m) = \prod_i p(h_i|m) \quad (13)$$

$$p(m_j = 1|h) = \text{sigmoid}\left(\alpha_j + \sum_{i=1}^{k_h} W_{ij}h_i\right) \quad (14)$$

$$p(h_j = 1|m) = \text{sigmoid}\left(\beta_j + \sum_{i=1}^{k_m} W_{ij}m_i\right) \quad (15)$$

2.4. Ensemble

Ensembles of models of the three techniques used in this study, LSTM-RNN, OP-ELM, and DBN, are investigated for UCLF forecasting. Ensemble models are a combination of multiple models to try to achieve better performance than that of the individual models. There is a number of different ways that models can be combined to form an ensemble [30]. Figure 3 shows a summary of the aggregate method, which is commonly used in regression problems. Here, models operate in parallel, and their outputs are aggregated to obtain the ensemble model's output. The aggregate ensemble model output, O_ϕ , can be written as (16). Here, O_{mk} is the ensemble model's k th output for models $m_1, m_2 \dots m_n$, and n is the number of models used to develop the assembly model. The equally weighted method was used, where each model's output into the ensemble model is given an equal weight.

$$O_\phi = \frac{1}{n} \sum_{k=1}^{k=1} O_{mk} \quad (16)$$

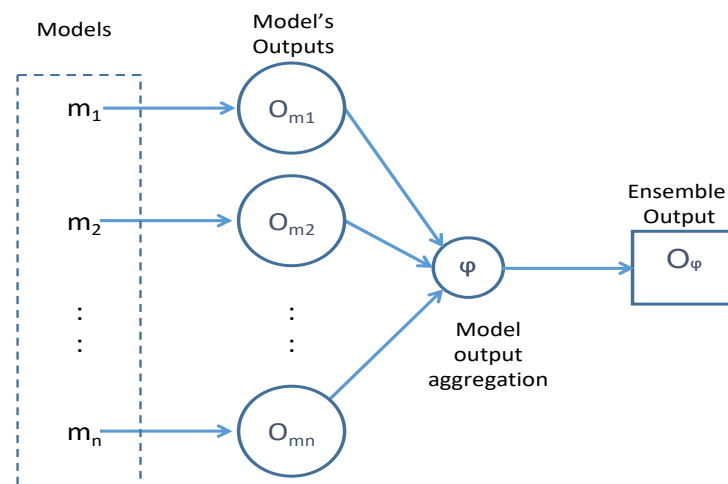


Figure 3. Summary of the aggregate ensemble method.

3. Experimental Setup

This section presents the experiment setup via two sub-sections. The first sub-section presents the South African coal generation plants overview. The second sub-section presents the experimental approach.

3.1. South African Key Coal Power Generation Plants Overview

South Africa has 15 key coal-powered thermal power stations. These stations are owned and operated by Eskom. Two of these stations are the new supercritical power stations, Medupi and Kusile, which are still under construction and at different stages of completion. The power stations are mostly concentrated in Mpumalanga Province, mainly due to the large availability of coal in this province. Twelve coal power stations are located in Mpumalanga, two in Limpopo Province, and one in Free State Province. Figure 4 shows the location of the South African coal-fired power stations [40]. South Africa also has one nuclear power generation station located in the Western Cape Province. This power station has an installed capacity of 1940 MW. This nuclear station and the coal-fired power stations contribute to over 80% of South Africa's installed capacity and supply the country's baseload. The PCLF and UCLF data used in this research are from these coal-fired powered stations and the nuclear power station, collected from a centralized database.

3.2. Data Description

The data used in this study were real utility data collected from January 2010 to December 2019. Figure 5 shows the different periodicities of the UCLF over time. Figure 5c shows the periodicity over weeks in parts of the South African winter (June–July) and summer (November–December) season in the year 2019.

The collected data were for four variables: the installed capacity, demand, PCLF, and UCLF. To investigate how these variables affect the UCLF forecast accuracy of the different techniques, the variables were arranged into five experiments, as shown in Figure 6. A tick indicates that a variable is used in the respective experiment and a cross indicates that the variable was not used in the experiment. The experiment with the best performance will, thus, indicate which variables should be used with which technique to achieve the lowest year-ahead UCLF forecasting error. The installed capacity is the total power that can be generated by the installed power generation plants in megawatts. The demand is the historic total national power demand in megawatts. The PCLF and UCLF are the respective historic variables in megawatts. The UCLF data used for the input in the training and testing of the models were split into the UCLF two years before the target UCLF, *UCLF T-2 Years*, and the UCLF a year before the target, *UCLF T-1 Year*. The UCLF data used was a daily peak value. A variable indicating if it is a weekend or a weekday, the

Weekend Index, was also used as an input. This variable was a 1 for weekends and a 0 for weekdays. This variable was included for the models to be able to differentiate the data for a weekday and the weekend, respectively. This resulted in six input variables. The training period was between 1 January 2012 and 31 December 2018. The testing period was between 1 January 2019 and 31 December 2019. Thus, the forecasts were a daily peak UCLF for the year-ahead forecast period. All the variables, except the weekend index, were normalized to be between 0 and 1. The training input data were, thus, a $2555 \times n$ matrix, where the 2555 is the daily input values over 7 years and n is the number of variables used in the respective experiment, as described next.

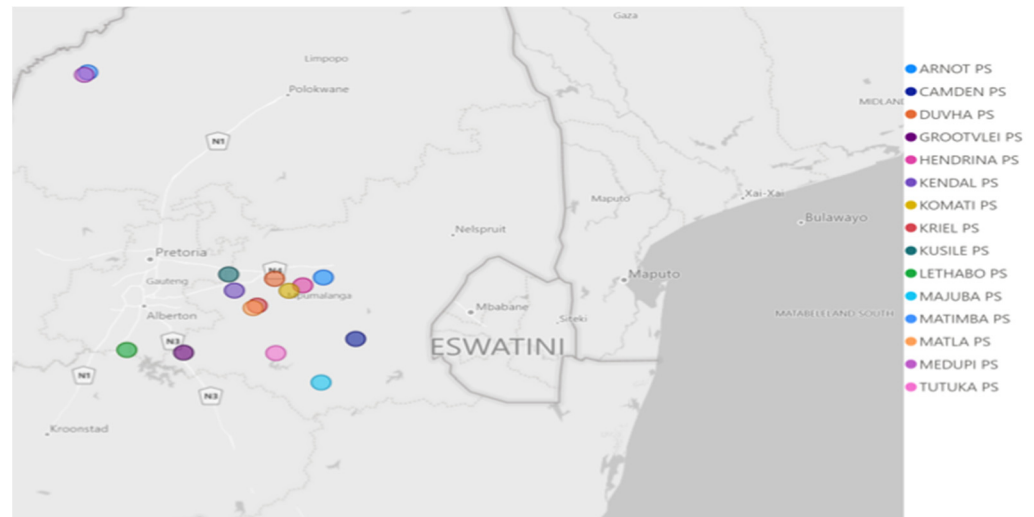


Figure 4. Location of 15 key South African coal-fired power stations.

The training input variable matrix sizes were, thus, 2555×6 for Exp 1, 2555×5 for Exp 2 to Exp 4, and 2555×3 for Exp 5.

3.3. Experimental Approach

The different techniques' models were, respectively, developed using various approaches.

The OP-ELM models were trained by tuning the model dimensions. A different number of hidden nodes were used to train the model in the respective experiments. Optimal pruning using the LOO method was key in determining the model's dimensions. Various dimensions were investigated and the model with the lowest errors in each experiment was captured and is presented in the results section.

LSTM-RNN models were trained with different numbers of stacked hidden LSTM units. The variation of the hidden units was consistent in all the different experiments. Similar to the OP-ELM, the performance results for the model with the lowest obtained UCLF forecast errors were captured.

Single layered DBN models were developed with the number of hidden units being varied for the respective models, the lowest number of hidden units used was four with the highest number of hidden units being sixteen.

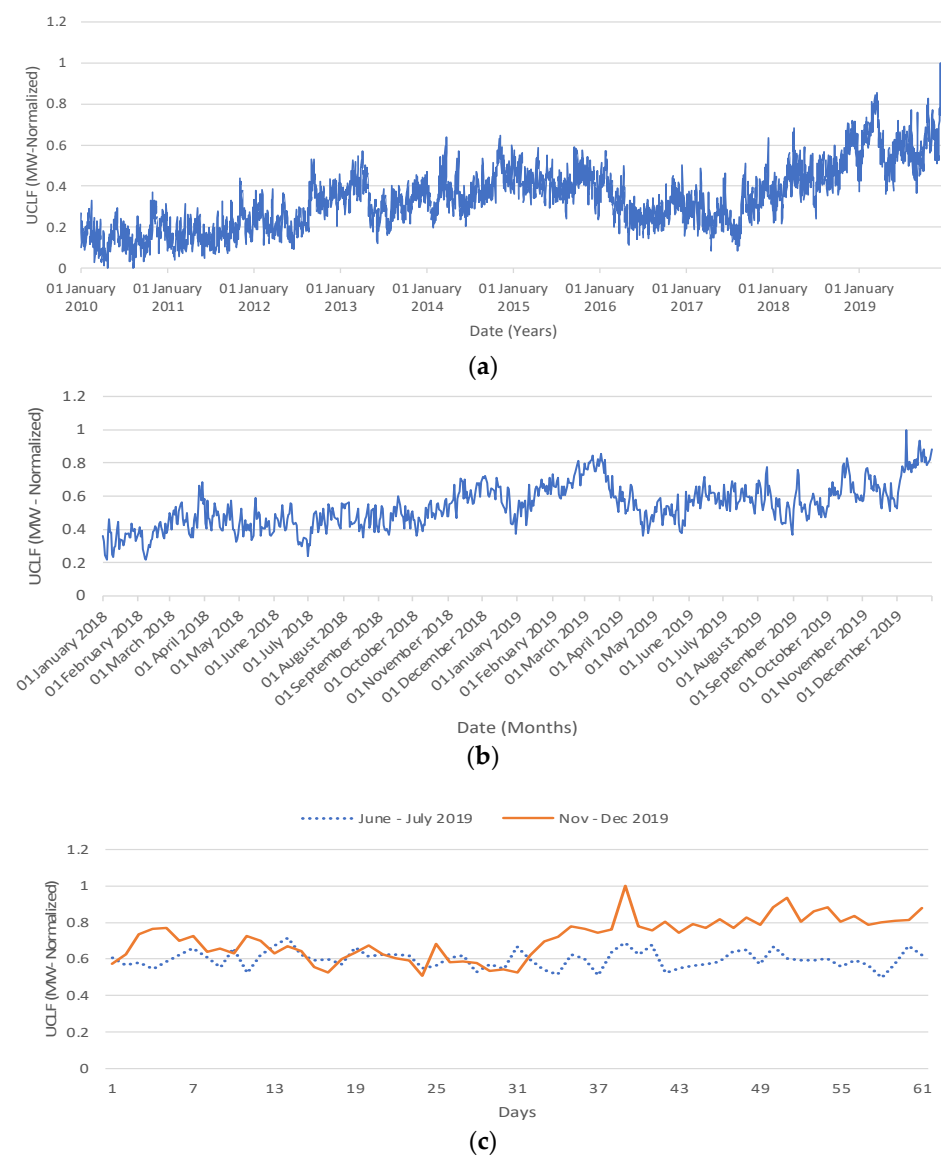


Figure 5. The South African UCLF (MW—normalized): (a) UCLF for a period between January 2010 and December 2019; (b) monthly periodicity of UCLF between January 2018 and December 2019; (c) weekly periodicity for June–July 2019 and November–December 2019.

Variable	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5
Weekend Index	✓	✓	✓	✓	✓
Installed Capacity	✓	✓	✓	✗	✗
Demand	✓	✗	✓	✓	✗
PCLF	✓	✓	✗	✓	✗
UCLF T-2 Years	✓	✓	✓	✓	✓
UCLF T-1 Year	✓	✓	✓	✓	✓

Figure 6. Variables used in the different experiments conducted per technique.

The aggregation ensemble approach was used for the ensemble of the three techniques. These ensembles were of two techniques at a time. Here, the various respective parameters per technique are tuned and combined to form different ensemble models. The performance results of the forecast results with the lowest errors are captured per experiment. For each technique and experiment, the other hyperparameters, such as training rate and the number of layers, were kept the same. In future work, the effect of optimizing the hyperparameters can be investigated.

3.4. Performance Measures Used

Each model's performance was measured using three key performance measures: symmetric mean absolute percentage error (sMAPE), mean absolute error (MAE), and root-mean-square error (RMSE). Motepe et al. state that the MAE, RMSE, MPE, MAPE, and sMAPE are common forecasting error measurements [30]. They further state the challenge that the MAPE faces when target values are too small, which leads to errors being too large. The three used performance measurements in this research are presented in (17)–(19).

$$sMAPE = \frac{2}{N} \sum_{k=1}^N \frac{|F_k - T_k|}{|F_k| + |T_k|} \quad (17)$$

$$MAE = \frac{\sum_{k=1}^N |F_k - T_k|}{N} \quad (18)$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^N (F_k - T_k)^2}{N}} \quad (19)$$

where F_k is the forecasted value, T_k is the target value, and N is the number of forecasted values.

3.5. Statistical Significance Test

After the model performance is measured, the model results can be found to not be statistically different from each other. This means that despite one model achieving results with a lower error in comparison to the next model's results, the model with the lower error does not necessarily outperform the model it is being compared to. A statistical test can be used to determine if model results are statistically significantly different. One such test is the t-test. The t-test uses the mean and the variance to check if two samples are from the same sample. The test calculates a significant value, also termed the p -value. A p -value less than the acceptable value means that the samples being compared have a significant difference, and vice versa. A p -value of 0.05, which is a commonly used value in scientific studies, was used in this study. The statistical significance test is performed, for each technique between the results with the lowest overall errors and results with the lowest errors from Exp 1, Exp 2, Exp 3, Exp 4, and/or Exp 5.

4. Proposed UCLF Forecasting System

Figure 7 presents the proposed UCLF forecasting system. The power stations monitor their plant's performance and report this locally at the station and centrally. These data are then stored in a central database. The UCLF data are part of these stored power station data. A record of the power station units that are on planned outages, PCLF, for maintenance or refurbishment is also stored centrally. These PCLF data are then provided by a central planning department in conjunction with the central operations department. The planning department also provides the installed capacity data to the central database. The system operator or an equivalent department would then provide the demand data. The data are pre-processed, and the variables are then consolidated for input into the deep learning (DL) ensemble UCLF forecasting module. The DL ensemble UCLF forecast module contains a DL ensemble model that forecasts the UCLF. The UCLF forecast is then stored and used by the planning, operations, and system operator. The DL ensemble model is developed and tested offline, and then deployed in the system. The UCLF forecast data together with the

actual UCLF data are then used by a model performance evaluation module to periodically check if the model's accuracy is still acceptable based on the utility's requirements.

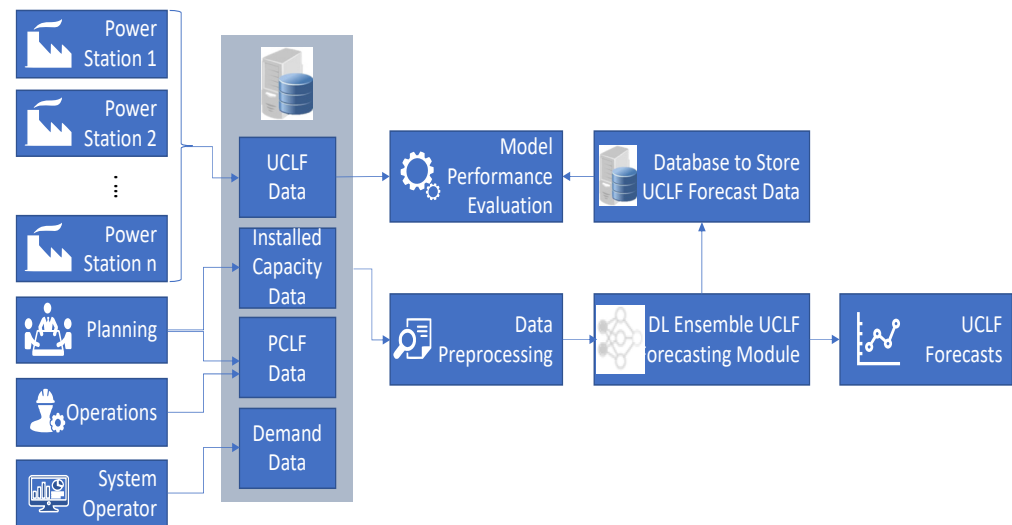


Figure 7. Proposed deep learning UCLF forecasting system.

5. Experiment Results and Results Discussion

This section presents the results of the five different experiments for the four techniques. The results are then discussed.

5.1. OP-ELM Results

The different experiments were conducted with different OP-ELM models, as described in Section 3. The lowest obtained errors per experiment are captured in Table 1. It was found that the OP-ELM model developed using variables for Experiment 2 and 50 hidden nodes achieved the lowest errors. This model achieved an sMAPE of 10.21%, MAE of 11.57%, and RMSE of 14.65%. These performance results are in bold in Table 1. This model was, therefore, developed without the demand as an input. Experiments 4 and 5's lowest obtained errors were higher than the lowest obtained errors in the other three experiments. The exclusion of the installed capacity, in Experiments 4 and 5, was observed to lead to an increase in the errors. In these experiments, the sMAPE increased by over 90% in comparison to the sMAPE in the other experiments. This increase in the errors was also observed to be approximately twice the observed errors in Experiment 2.

Table 1. OP-ELM experiments results.

Experiment	Hidden Nodes	Performance		
		sMAPE	MAE	RMSE
Exp 1	81	0.208919	0.124393	0.157294
Exp 2	50	0.204172	0.115727	0.146514
Exp 3	50	0.231884	0.134929	0.173026
Exp 4	125	0.405396	0.198972	0.228686
Exp 5	18	0.519556	0.246778	0.27476

A statistical significance test was conducted to determine if the results with the lowest errors from each experiment had a significant difference from the results with the overall lowest errors. The statistical significance test results are captured in Table 2. From the significance test results, a *p*-value of less than 0.05 was observed. Thus, the results are significantly different from each other. The exclusion of the demand, therefore, increases model forecasting error.

Table 2. OP-ELM models' lowest errors statistical significance test.

	Exp 1	Exp 3	Exp 4	Exp 5
<i>p</i> -value	0.020517	0.001810	2.0375×10^{-91}	7.4004×10^{-118}

5.2. LSTM-RNN Results

LSTM-RNN models were developed using the different variables per respective experiment. The performance of the different LSTM-RNN models was observed. The lowest obtained year-ahead UCLF forecast errors, per experiment, are captured in Table 3.

Table 3. LSTM-RNN experiments results.

Experiment	Hidden Units	Performance		
		sMAPE	MAE	RMSE
Exp 1	511	0.15897	0.091421	0.114164
Exp 2	64	0.173154	0.097143	0.117865
Exp 3	511	0.168273	0.09699	0.122862
Exp 4	256	0.343999	0.179548	0.214021
Exp 5	767	0.407081	0.206088	0.237777

A model with 511 hidden units and Experiment 1 variables had the lowest errors. Here, an sMAPE of 7.95%, MAE of 9.14%, and RMSE of 11.42% were achieved. Higher errors were observed in Experiments 4 and 5, where the installed capacity was excluded. These errors were approximately twice the errors in Experiment 1. A statistical significance test was conducted to determine if the results with the lowest errors in each experiment were significantly different from the results with the overall lowest errors. The results were found to be statistically different from each other as a *p*-value of less than 0.05 was observed in all four cases. The obtained *p*-values are captured in Table 4.

Table 4. LSTM-RNN models' lowest errors statistical significance test.

	Exp 2	Exp 3	Exp 4	Exp 5
<i>p</i> -value	0.022794	9.3999×10^{-14}	9.2587×10^{-211}	7.8709×10^{-256}

5.3. DBN Results

The DBN models were developed as discussed in Section 3. The errors for the models' year-ahead UCLF forecast results were observed and the lowest obtained errors per experiment are captured in Table 5. A model with nine hidden nodes developed using all the variables was found to achieve the lowest errors, with an sMAPE of 9.74%, MAE of 11.52%, and RMSE of 13.74%. Experiments 4 and 5 showed an increase that was approximately three times the errors observed in Experiment 1.

The statistical significance test was conducted as described in Section 3.5 and the test result showed that the forecasting results were significantly different. Table 6 shows the statistical significance test results. The *p*-value can be seen to be less than 0.05 in each case, indicating a significant difference in the respective cases.

Table 5. DBN experiments results.

Experiment	Hidden Nodes	Performance		
		sMAPE	MAE	RMSE
Exp 1	9	0.194736	0.115172	0.137397
Exp 2	8	0.328704	0.172461	0.172461
Exp 3	8	0.300888	0.159492	0.189725
Exp 4	4	0.608786	0.279951	0.304046
Exp 5	4	0.588584	0.273245	0.298614

Table 6. DBN models' lowest errors statistical significance test.

	Exp 2	Exp 3	Exp 4	Exp 5
<i>p-value</i>	7.3796×10^{-268}	1.7369×10^{-253}	9.7011×10^{-264}	3.9572×10^{-259}

5.4. Ensemble Results

Ensemble models of the three techniques were developed using the aggregate method with two individual developed models at a time; that is, from Equation (16), $n = 2$. All the individual models developed in this research were ensembled in this manner and their performance was observed. The performance parameters for the ensemble model whose year-ahead UCLF forecast achieved the lowest errors per experiment are presented in Table 7. Thus, not all results are included in Table 7, just the results with the lowest errors per experiment. The ensemble technique name is constructed by combining the name of the original technique used and the number of hidden nodes, for the OP-ELM and DBN, and the number of hidden units, for the LSTM, next to the name. The lowest obtained errors were achieved using an ensemble model of two LSTM models with 192 and 26 hidden units, respectively. This model achieved an sMAPE of 6.43%, MAE of 7.36%, and RMSE of 9.21%, which are bolded in Table 7. The respective errors in Experiments 4 and 5 were approximately twice the errors in Experiment 1. The accuracy of the model in Experiment 2 was higher than that for the models in Experiment 3. The models in Experiments 2 and 3 had lower accuracy than the model in Experiment 1, and higher accuracy than the models in Experiments 4 and 5.

Table 7. Ensemble experiments results.

Experiment	Ensemble Technique	Performance		
		sMAPE	MAE	RMSE
Exp 1	LSTM192-LSTM26	0.1286794	0.073588	0.092046
	LSTM192-DBN9	0.143504	0.080100	0.099055
	LSTM383-OPELM16	0.163263	0.093770	0.120741
	DBN9-DBN8	0.155915	0.089143	0.11129
	DBN9-OPELM16	0.168854	0.096859	0.122968
	OPELM81-OPELM16	0.198971	0.108824	0.140945
Exp 2	LSTM383-LSTM64	0.161214	0.092327	0.112328
	LSTM128-DBN8	0.170443	0.097098	0.124192
	LSTM64-OPELM50	0.167899	0.095670	0.118292
	DBN8-DBN8	0.328704	0.172461	0.205681
	DBN8-OPELM80	0.217165	0.118987	0.150817
	OPELM50-OPELM15	0.206886	0.114157	0.147775

Table 7. Cont.

Experiment	Ensemble Technique	Performance		
		sMAPE	MAE	RMSE
Exp 3	LSTM511-LSTM511	0.168272	0.096990	0.122861
	LSTM511-DBN8	0.210692	0.118462	0.148410
	LSTM511-OPELM50	0.185430	0.106055	0.130075
	DBN8-DBN8	0.300887	0.159491	0.189725
	DBN9-OPELM50	0.225704	0.123672	0.153060
	OPELM50-OPELM15	0.229431	0.127089	0.160191
Exp 4	LSTM256-LSTM256	0.343998	0.179547	0.214021
	LSTM256-DBN4	0.466254	0.228923	0.257488
	LSTM256-OPELM100	0.359248	0.184216	0.223038
	DBN4-DBN4	0.608786	0.279950	0.304045
	DBN4-OPELM100	0.480638	0.230885	0.263031
	OPELM100-OPELM125	0.392240	0.194914	0.232011
Exp 5	LSTM767-LSTM767	0.407080	0.206087	0.237776
	LSTM767-DBN4	0.4929245	0.239221	0.267542
	LSTM767-OPELM18	0.459199	0.225823	0.254908
	DBN4-DBN4	0.588584	0.273245	0.298614
	DBN4-OPELM18	0.551643	0.259683	0.285607
	OPELM18-OPELM18	0.519555	0.246777	0.274759

Table 8 presents the results for a statistical significance test conducted as discussed in Section 3.5. A p -value less than 0.05 was observed for each test conducted. This observation indicated that all the results being compared were significantly different from each other.

Table 8. Ensemble models' lowest errors statistical significance test.

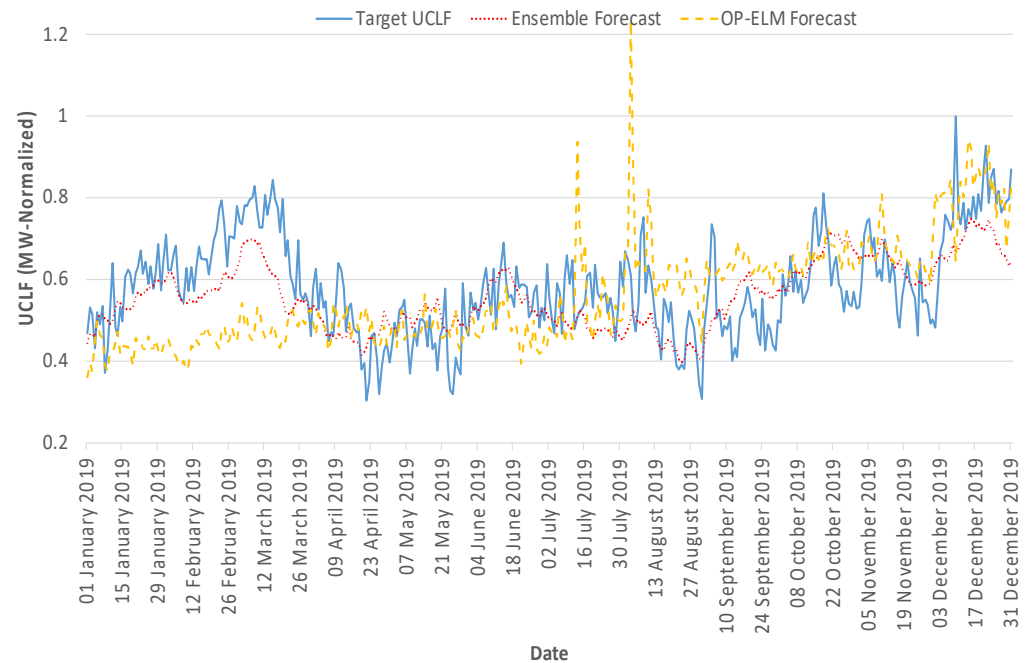
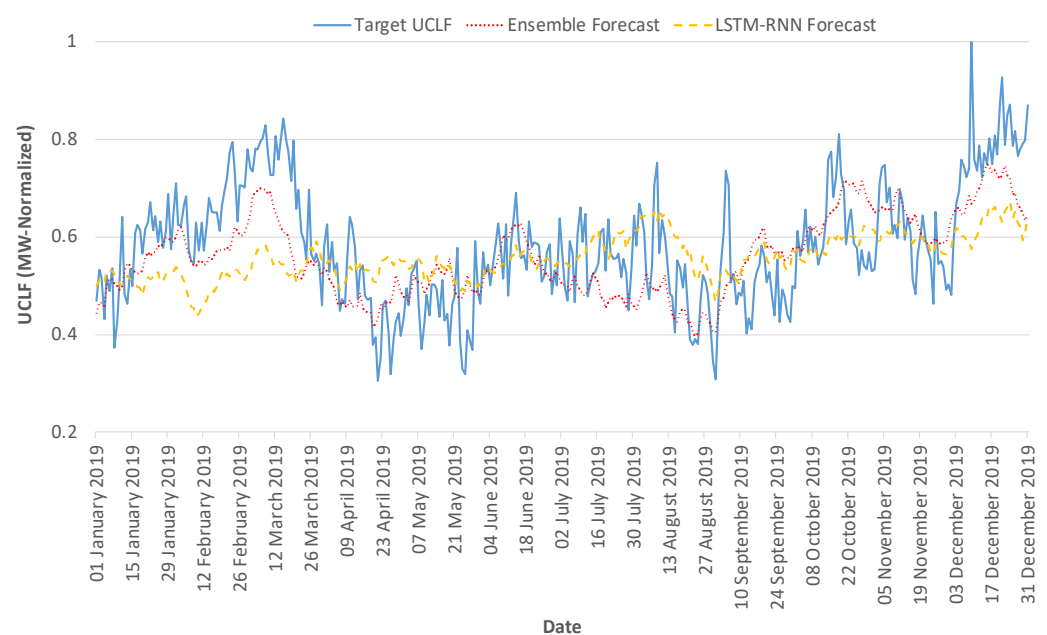
	Exp 2	Exp 3	Exp 4	Exp 5
p-value	7.3796×10^{-268}	1.7369×10^{-253}	9.7011×10^{-264}	3.9572×10^{-259}

5.5. Results Discussion

The lowest obtained year-ahead UCLF forecasting errors from each technique are summarized in Table 9. These results show that the lowest UCLF forecasting errors were obtained by the ensemble model. The ensemble model was then followed by the LSTM-RNN, DBN, and then OP-ELM. The two deep learning techniques, thus, achieved higher accuracies than the non-deep learning technique, OP-ELM. It was observed that with all techniques, apart from OP-ELM, the lowest errors were attained in Experiment 1. Experiments 4 and 5 showed a sharp increase in errors, relative to the rest of the experiments with all the techniques. Thus, the exclusion of the installed capacity as an input variable decreased the accuracy of the models of the techniques used. The plots of the target UCLF and the year-ahead forecasted UCLF for the models with the lowest errors per technique are presented in Figures 8–11. These plots are plotted for the period of 1 January 2019 to 31 December 2019. Each plot of the individual models also includes the ensemble model with the lowest forecasting error. The plots of the UCLF forecast by the models that make up the ensemble model are plotted in Figure 9.

Table 9. Summary of lowest obtained errors per used technique.

Technique	Experiment	Performance		
		sMAPE	MAE	RMSE
OP-ELM	Exp 2	0.204172	0.115727	0.146514
LSTM-RNN	Exp 1	0.15897	0.091421	0.114164
DBN	Exp 1	0.194736	0.115172	0.137397
Ensemble	Exp 1	0.128679	0.073588	0.092046

**Figure 8.** A plot of the OP-ELM and ensemble lowest error model year-ahead UCLF forecast against the target UCLF.**Figure 9.** A plot of the LSTM-RNN and ensemble lowest error model year-ahead UCLF forecast against the target UCLF.

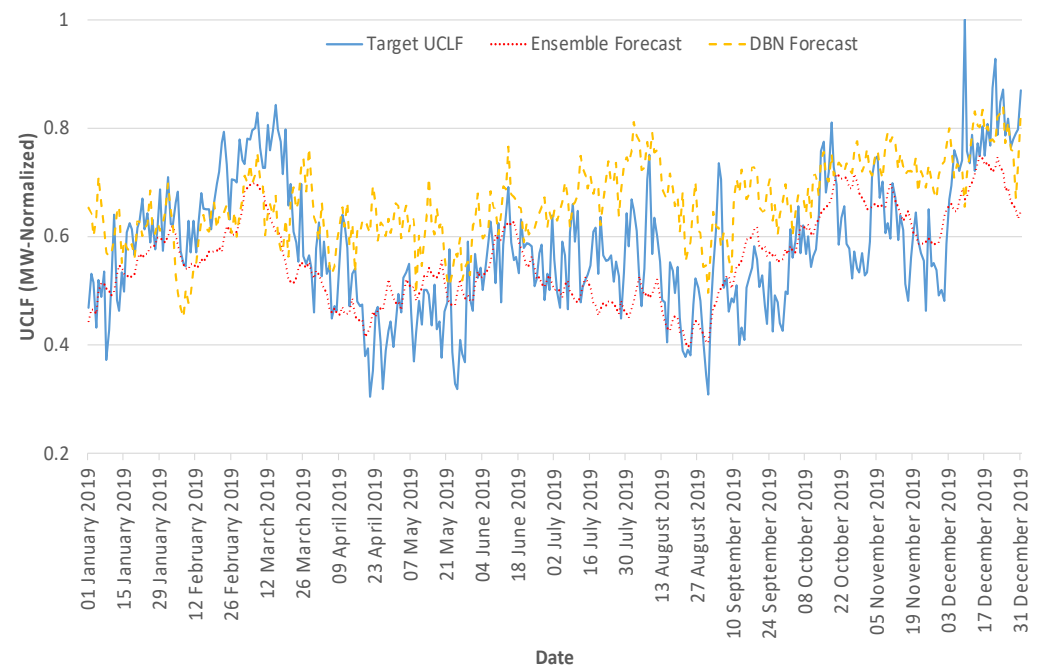


Figure 10. A plot of the DBN and ensemble lowest error model year-ahead UCLF forecast against the target UCLF.

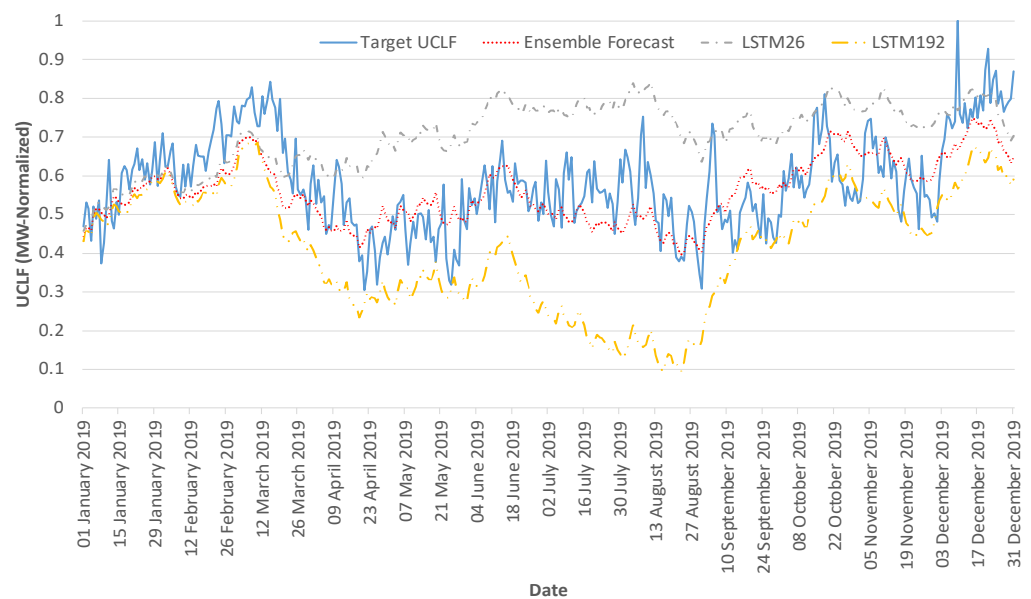


Figure 11. A plot of the ensemble lowest error model and the two aggregated models' year-ahead UCLF forecast against the target UCLF.

6. Conclusions

This paper contributed to the body of knowledge about South African UCLF forecasting. (i) A novel study of the South African UCLF behavior using state-of-the-art AI (deep learning and ensemble) techniques was presented. LSTM-RNN, DBN, OP-ELM, and ensembles of these three techniques' models were investigated in South African UCLF forecasting. (ii) An investigation of the impact of the installed capacity, historic demand, and PCLF on the UCLF forecasting accuracy was presented. It was found that the installed capacity had the biggest impact on the UCLF forecasting error, with the exclusion of this variable doubling the errors with the respective techniques used. (iii) A novel deep-learning ensemble total South African UCLF forecasting system was introduced. It was found that

an ensemble of LSTM models achieved the lowest errors with an sMAPE of 6.43%, MAE of 7.36%, and RMSE of 9.21%. The lowest achieved LSTM model UCLF forecast errors were an sMAPE of 7.95%, MAE of 9.14%, and RMSE of 11.42%. The lowest achieved DBN model UCLF forecast errors were an sMAPE of 9.74%, MAE of 11.52%, and RMSE of 13.74%. The lowest achieved OP-ELM model UCLF forecast errors were an sMAPE of 10.21%, MAE of 11.57%, and RMSE of 14.65%. The lowest attained error was, thus, given by the ensemble model, followed by LSTM-RNN. The non-deep learning techniques' lowest achieved error was higher than that of the lowest errors achieved by the other techniques. Thus, ensemble deep learning techniques can be used to effectively forecast the total South African UCLF and, thus, load shedding.

7. Limitations of the Study and Future Work

This section presents the limitation of this study. As with most research, not all research-related aspects can be covered in a single study. As mentioned in Section 1, the study of South African UCLF behavior and UCLF forecasting is a new research area. This study does not focus on the speed of training the models, but rather on how well the models forecast the UCLF. Future work can include looking at the model training performance from the training speed perspective. The study forecast period is a year. This period was selected as it gives a wide enough window for the utility, at a daily resolution, to understand the UCLF behavior for the year. This understanding allows the utility company to plan over the year. The study does not research the performance of the models in shorter-term forecast windows, e.g., hourly, daily, weekly, etc. The performance of the models can, in the future, be studied for different forecast windows. Future research work should also consider looking at recent state-of-the-art techniques, such as temporal convolutional networks (TCN), gated recurrent units (GRU), and quasi-recurrent neural networks (QRNN). Given the performance of the equally weighted ensemble techniques in this paper, weighted ensemble techniques should be considered in future work. This future work can also investigate the ensemble models' performance when combining more than two models. Other benchmark techniques, such as naïve and multilayer perceptron, can be considered in future work.

Author Contributions: Conceptualization, S.M.; methodology, S.M.; software, S.M.; validation, S.M.; formal analysis, S.M.; investigation, S.M.; resources, S.M., A.N.H. and T.S.; data curation, S.M.; writing—original draft preparation, S.M.; writing—review and editing, S.M., A.N.H. and T.S.; visualization, S.M.; supervision, A.N.H. and T.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The authors would like to acknowledge the University of Johannesburg's Global Excellence and Stature (GES) 4.0 for funding of this research.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is not available on any public platform.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Motepe, S.; Hasan, A.N.; Twala, B.; Stopforth, R.; Alajarmeh, N. South African Power Distribution Network Load Forecasting Using Hybrid AI Techniques: ANFIS and OP-ELM. In Proceedings of the Aegean Conference on Electrical Machines and Power Electronics, and Optimization of Electrical & Electronic Equipment Conference (ACEMP-OPTIM), Istanbul, Turkey, 2–4 September 2019.
2. Hasan, A.; Twala, B.; Ouahada, K.; Marwala, T. Energy Usage Optimisation in South African Mines. *Arch. Min. Sci.* **2014**, *59*, 53–69. [[CrossRef](#)]
3. Malatji, E.M.; Zhang, J.; Xia, X. A multiple objective optimisation model for building energy efficiency investment decision. *Energy Build.* **2013**, *61*, 81–87. [[CrossRef](#)]
4. Motepe, S.; Hassan, A.N.; Stopforth, R. South African Distribution Networks Load Forecasting Using ANFIS. In Proceedings of the IEEE Power Electronics Drivers and Energy Systems (PEDES), Chennai, India, 18–21 December 2018.

5. ESKOM and The Department of Public Enterprise. Update: ESKOM Electricity Supply. Available online: http://www.eskom.co.za/news/Documents/20190403ESKOM_BriefingFINAL.pdf (accessed on 1 April 2020).
6. Pombo-van Zyl, N. Warning: Stage 2 Loadshedding Returns States Eskom. ESI Africa, Africa's Power Journal, February 2020. Available online: <https://www.esi-africa.com/industry-sectors/transmission-and-distribution/warning-high-risk-of-loadshedding-returns-states-eskom/> (accessed on 1 May 2020).
7. Pretorius, I.; Piketh, S.J.; Burger, R.P. The impact of the South African energy crisis on emissions. *WIT Trans. Ecol. Environ.* **2015**, *198*, 255–264.
8. Micali, V. Prediction of Availability for new power plant in the absence of data. In Proceedings of the 2012 9th Industrial and Commercial Use of Energy Conference, Stellenbosch, South Africa, 15–16 August 2012.
9. Motepe, S.; Hasan, A.N.; Stopforth, R. Improving Load Forecasting Process for a Power Distribution Network Using Hybrid AI and Deep Learning Algorithms. *IEEE Access* **2019**, *7*, 82584–82598. [\[CrossRef\]](#)
10. Jaech, A.; Zhang, B.; Ostendorf, M.; Kirschen, D.S. Real-Time Prediction of the Duration of Distribution System Outages. *IEEE Trans. Power Syst.* **2019**, *34*, 773–781. [\[CrossRef\]](#)
11. Wolse, H.; Geist, G.; Hoving, B.; Oosterlee, P.; Polman, H. Experience and tendencies after 40 years outage data registration in the Netherlands. *CIREN-Open Access Proc. J.* **2017**, *2017*, 2279–2282. [\[CrossRef\]](#)
12. BP. Statistical Review of World Energy. 69th Edition. 2020. Available online: <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2020-full-report.pdf> (accessed on 1 May 2020).
13. Index Mundi. Available online: <https://www.indexmundi.com/g/r.aspx?t=50&v=79&l=en> (accessed on 1 November 2020).
14. Smith, R.K. Analysis of hourly generation patterns at large coal-fired units and implications of transitioning from baseload to load-following electricity supplier. *J. Mod. Power Syst. Clean Energy* **2018**, *7*, 468–474. [\[CrossRef\]](#)
15. Fu, J.; Xiao, H.; Wang, H.; Zhou, J. Control Strategy for Denitrification Efficiency of Coal-Fired Power Plant Based on Deep Reinforcement Learning. *IEEE Access* **2020**, *8*, 65127–65136. [\[CrossRef\]](#)
16. Wang, Y.; Lou, S.; Wu, Y.; Wang, S. Flexible Operation of Retrofitted Coal-Fired Power Plants to Reduce Wind Curtailment Considering Thermal Energy Storage. *IEEE Trans. Power Syst.* **2019**, *35*, 1178–1187. [\[CrossRef\]](#)
17. Che, P.; Liu, Y.; Che, L.; Lang, J. Co-Optimization of Generation Self-Scheduling and Coal Supply for Coal-Fired Power Plants. *IEEE Access* **2020**, *8*, 110633–110642. [\[CrossRef\]](#)
18. Khoza, M.; Marwala, T. Computational intelligence techniques for modelling an economic system. In Proceedings of the 2012 International Joint Conference on Neural Networks (IJCNN), Brisbane, Australia, 10–15 June 2012.
19. Galias, Z. Probabilistic Model for Studying Blackouts in Power Networks. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2017**, *7*, 218–227. [\[CrossRef\]](#)
20. Attia, A. Analysis of failure in power cables for preventing power outage in Alexandria electricity distribution company in Egypt. *CIREN-Open Access Proc. J.* **2017**, *2017*, 20–24. [\[CrossRef\]](#)
21. Wang, J.; Zhang, J.; Wang, X. 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 Trans. Ind. Inform.* **2017**, *14*, 748–758. [\[CrossRef\]](#)
22. Chen, X.-W.; Lin, X. Big Data Deep Learning: Challenges and Perspectives. *IEEE Access* **2014**, *2*, 514–525. [\[CrossRef\]](#)
23. Martín-Doña, J.M.; Gomez, A.M.; Gonzalez, J.A.; Peinado, A.M. A deep learning loss function based on the perceptual evaluation of the speech quality. *IEEE Signal Process. Lett.* **2018**, *25*, 1680–1684. [\[CrossRef\]](#)
24. Masita, K.L.; Hasan, A.N.; Paul, S. Pedestrian Detection Using R-CNN Object Detector. In Proceedings of the IEEE Latin American Conference on Computational Intelligence (LA-CCI), Guadalajara, Mexico, 7–9 November 2018.
25. Alhussein, M.; Aurangzeb, K.; Haider, S.I. Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting. *IEEE Access* **2020**, *8*, 180544–180557. [\[CrossRef\]](#)
26. Kong, Z.; Zhang, C.; Lv, H.; Xiong, F.; Fu, Z. Multimodal Feature Extraction and Fusion Deep Neural Networks for Short-Term Load Forecasting. *IEEE Access* **2020**, *8*, 185373–185383. [\[CrossRef\]](#)
27. Pandit, R.K.; Kolios, A.; Infield, D. Data-driven weather forecasting models performance comparison for improving offshore wind turbine availability and maintenance. *IET Renew. Power Gener.* **2020**, *14*, 2386–2394. [\[CrossRef\]](#)
28. Kou, P.; Wang, C.; Liang, D.; Cheng, S.; Gao, L. Deep learning approach for wind speed forecasts at turbine locations in a wind farm. *IET Renew. Power Gener.* **2020**, *14*, 2416–2428. [\[CrossRef\]](#)
29. Munkhdalai, L.; Park, K.H.; Batbaatar, E.; Theera-Umpon, N.; Ryu, K.H. Deep Learning-Based Demand Forecasting for Korean Postal Delivery Service. *IEEE Access* **2020**, *8*, 188135–188145. [\[CrossRef\]](#)
30. Motepe, S.; Hasan, A.N.; Twala, B.; Stopforth, R. Effective load forecasting for large power consuming industrial customers using long short-term memory recurrent neural networks. *J. Intell. Fuzzy Syst.* **2019**, *37*, 8219–8235. [\[CrossRef\]](#)
31. Han, L.; Peng, Y.; Li, Y.; Yong, B.; Zhou, Q.; Shu, L. Enhanced Deep Networks for Short-Term and Medium-Term Load Forecasting. *IEEE Access* **2019**, *7*, 4045–4055. [\[CrossRef\]](#)
32. Ramotsoela, T.D.; Hancke, G.P.; Abu-Mahfouz, A.M. Behavioural Intrusion Detection in Water Distribution Systems Using Neural Networks. *IEEE Access* **2020**, *8*, 190403–190416. [\[CrossRef\]](#)
33. Zhang, C.; Li, R.; Kim, W.; Yoon, D.; Patras, P. Driver Behavior Recognition via Interwoven Deep Convolutional Neural Nets with Multi-Stream Inputs. *IEEE Access* **2020**, *8*, 191138–191151. [\[CrossRef\]](#)

34. Sun, Q.; Ge, Z. Deep Learning for Industrial KPI Prediction: When Ensemble Learning Meets Semi-Supervised Data. *IEEE Trans. Ind. Inform.* **2021**, *17*, 260–269. [[CrossRef](#)]
35. Drif, A.; Zerrad, H.E.; Cherifi, H. EnsVAE: Ensemble Variational Autoencoders for Recommendations. *IEEE Access* **2020**, *8*, 188335–188351. [[CrossRef](#)]
36. Bibi, N.; Shah, I.; Alsubie, A.; Ali, S.; Lone, S.A. Electricity Spot Prices Forecasting Based on Ensemble Learning. *IEEE Access* **2021**, *9*, 150984–150992. [[CrossRef](#)]
37. Shah, I.; Iftikhar, H.; Ali, S.; Wang, D. Short-Term Electricity Demand Forecasting Using Components Estimation Technique. *Energies* **2019**, *12*, 2532. [[CrossRef](#)]
38. Miche, Y.; Sorjamaa, A.; Lendasse, A. OP-ELM: Theory, Experiments and a Toolbox. In Proceedings of the 18th International Conference on Artificial Neural Networks (ICANN), Prague, Czech Republic, 3–6 September 2008.
39. Motepe, S.; Hasan, A.N.; Twala, B.; Stopforth, R. Power Distribution Networks Load Forecasting Using Deep Belief Networks: The South African Case. In Proceedings of the IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology, Amman, Jordan, 9–11 April 2019.
40. ESKOM. Power station GPS Coordinates. Eskom. Available online: http://www.eskom.co.za/Whatweredoing/ElectricityGeneration/PowerStations/Pages/Power_Station_GPS_Coordinates.aspx (accessed on 6 November 2020).