AIMS Energy

https://www.aimspress.com/journal/energy

AIMS Energy, 14(2): 310–334.
DOI: 10.3934/energy.2026014
Received: 10 October 2025
Revised: 22 January 2026
Accepted: 30 January 2026
Published: 10 March 2026

Research article

Real-time electricity load forecasting in South Africa using SOM-enriched deep learning ensembles

Katleho Makatjane[1,*], Caston Sigauke[2], Claris Shoko[3] and Ntebogang Moroke[4]

[1] Department of Statistics and Population Studies, University of the Western Cape, South Africa
[2] Department of Mathematical and Computational Sciences, University of Venda, Thohoyandou, South Africa
[3] Department of Statistics, University of Botswana, Gaborone, Botswana
[4] Department of Business Statistics and Operations Research, North West University, Mahikeng, South Africa

* **Correspondence:** Email: kmakatjane@uwc.ac.za; Tel: +27219593406.

Abstract: Accurate short-term electrical demand forecasting is critical for maintaining operational efficiency and energy security, especially in power-constrained systems like South Africa's Eskom. Statistical methods like autoregressive integrated moving average (ARIMA) and exponential smoothing often fail to represent nonlinear and regime-dependent trends in power demand. This study presents a dynamic ensemble that combines deep neural networks (DNN) and long short-term memory (LSTM) architectures, which are both augmented by self-organising maps (SOM)-based clustering. The proposed method divides historical hourly load data from the Drakensberg generation plant into discrete temporal regimes using SOM, then trains the DNN and LSTM architectures within each regime, and dynamically combines their predictions. Shapley additive explanations (SHAP) are used to improve the interpretability of the impact of each cluster and time hierarchies, while mean square error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) measures are used to assess prediction performance. The ensemble architecture delivers a higher accuracy, lowering MAPE to 2.20% while consistently outperforming individual benchmark architectures. The deployment on Amazon Web Services (AWS) proves the model's scalability and appropriateness for real-time applications. Although performance degrades in irregular demand clusters, adaptive re-clustering may alleviate this constraint. Overall, the combined DNN-LSTM-SOM strategy is a reliable, interpretable, and scalable solution for short-term load forecasting, enabling better operational planning and grid dependability in developing energy systems.

Keywords: deep neural networks; energy security; electricity load forecasting LSTM networks; nonlinear ensemble learning; self-organising maps

1. Introduction

Population growth, rising living standards, and continued industrialisation have driven sustained increases in electricity demand, resulting in highly intricate hourly load dynamics. Accurate short-term electricity load forecasting is therefore critical for ensuring power system security, reliability, and economic efficiency [1]. Improvements in forecasting accuracy directly enhance operational planning, reduce reserve margins, and support informed decisions regarding infrastructure investment [2]. Even marginal gains are economically significant; for example, a one per cent increase in forecasting accuracy can yield savings of several million dollars [3]. Furthermore, reliable anticipation of peak-load periods enables utilities to deploy effective demand-side management strategies and mitigate blackout risks.

Despite their long-standing use, classical time-series techniques such as autoregressive integrated moving average (ARIMA) and exponential smoothing models, among others, struggle to capture the nonlinear, volatile, and regime-dependent nature of electricity demand. These methods typically assume linearity and stationarity, assumptions that are frequently violated in hourly load data characterised by long-memory effects, abrupt structural changes, and evolving consumption patterns. In response, deep learning approaches, particularly long short-term memory (LSTM) networks, have gained prominence owing to their ability to model nonlinear relationships and long-range temporal dependencies, as indicated by [4, 5]. Nevertheless, both conventional artificial neural networks and standalone LSTM models may exhibit limited generalisation when confronted with pronounced regime shifts or highly heterogeneous operating conditions [6, 7]. To mitigate these limitations, recent studies, such as that of Reference [8], have increasingly adopted hybrid and ensemble forecasting strategies that combine complementary modelling paradigms, often yielding improved robustness and predictive accuracy relative to single-model approaches. However, Reference [9] emphasised that many existing deep learning ensembles implicitly assume homogeneous temporal behaviour and fail to model hierarchical or evolving demand regimes explicitly. Although self-organising maps (SOMs) have been employed to cluster load patterns before supervised learning, most SOM-deep neural network (DNN) and SOM-LSTM ensemble methods rely on static clustering structures and fixed or weakly adaptive ensemble weights (see, for instance, [10–12]). Such assumptions limit adaptability and raise concerns regarding sensitivity and operational robustness in nonstationary environments.

Motivated by these gaps, this study proposes a regime-aware, dynamically adaptive ensemble forecasting approach for short-term electricity load prediction. The core methodological contribution lies not in the individual use of SOM, DNN, or LSTM architectures, which are well established, but in their integrated and adaptive interaction are that designed explicitly to address nonstationarity. Specifically, a periodically updated SOM is employed to identify evolving load regimes, enabling the clustering structure to adjust as demand patterns change and facilitating the detection of emerging or transient behaviours rather than assuming fixed historical regimes. Within each identified regime, deep neural networks and long short-term memory architectures are trained separately to capture complementary nonlinear and temporal characteristics of electricity demand. Their forecasts are subsequently combined using a performance-driven dynamic ensemble weighting mechanism, in which model contributions are continuously recalibrated based on recent, regime-specific predictive accuracy. This dual adaptivity, operating at both the regime identification and ensemble weighting levels, constitutes the principal innovation of the proposed modelling approach. Our proposed

methodology is fundamentally distinct from existing machine learning-based load forecasting studies that focus on comparative model evaluation or static ensemble construction. For example, Reference [13] conducted a comparative analysis of k-nearest neighbour (KNN), support vector regression (SVR), random forest (RF), and extreme gradient boosting models without incorporating adaptive regime modelling or dynamic ensemble integration. Similarly, Reference [14] provided a comprehensive review of explainability and interpretability techniques in machine learning-based load forecasting rather than proposing a new adaptive forecasting methodology. Other recent contributions, such as that of Reference [15], introduced deep learning architectures for anomaly detection and load forecasting; however, these approaches rely on fixed ensemble structures and do not employ performance-driven dynamic ensemble weighting or evolving regime identification. In contrast, our proposed forecasting approach explicitly integrates unsupervised regime discovery, regime-specific deep learning, and dynamically adaptive ensemble weighting, thereby addressing a critical methodological gap in short-term electricity load forecasting under nonstationary operating conditions.

The approach is implemented as a three-stage forecasting application and validated using operational data from Eskom's Drakensberg pumped storage plant in South Africa, a system characterised by high volatility and frequent structural change. Deployment within a cloud-based environment enables scalable training, real-time inference, and continuous performance monitoring, demonstrating the practical feasibility of the proposed method. By integrating evolving regime discovery, regime-specific deep learning, and adaptive ensemble weighting within an operational forecasting architecture, we contribute a robust and practically relevant solution to short-term electricity load forecasting in complex, nonstationary power systems.

Research highlights and key findings

This study introduces and empirically validates a regime-aware, dynamically adaptive ensemble for short-term electricity load forecasting in South Africa. The core methodological and empirical highlights of the study are summarised in Table 1, with practical outcomes noted separately. This restructuring ensures that the methodological contributions and key empirical findings are immediately clear to the reader.

The remaining parts of the paper are organised as follows: Section 2 gives the detailed modelling framework, Section 3 presents the findings from the study, including the discussion of the findings, and finally Section 4 concludes and gives recommendations for further studies.

Table 1. Methodological and empirical highlights of the SOM-DNN-LSTM dynamic ensemble.

Aspect	Methodological highlight	Empirical finding	Contribution
Dynamic ensemble integration	Introduces a cluster-conditioned, uncertainty-aware ensemble combining SOM–DNN and SOM–LSTM forecasts	Achieves superior predictive performance relative to individual models (lowest MSE, MAE, MAPE)	Establishes a fully adaptive approach for nonstationary electricity demand
SOM-based regime identification	Utilises SOMs to dynamically identify temporal and behavioural load regimes	Reveals distinct weekday, weekend, and seasonal demand patterns	Enables regime-specific modelling and improved interpretability of demand dynamics
Regime-specific deep learning	Trains DNN and LSTM architectures within each SOM-identified regime cluster	Enhances generalisation during both stable and volatile demand periods	Demonstrates the value of conditional deep learning under heterogeneous load behaviour
Uncertainty-aware weighting	Computes weights using rolling prediction interval widths to quantify model reliability in real time	Automatically downweights less reliable models during volatile periods	Improves robustness and adaptivity without fixed or heuristic weights
Empirical performance validation	Systematically evaluates forecasting accuracy using multiple metrics (MSE, MAE, MAPE)	Confirms consistent performance gains across real-world Eskom load data	Provides rigorous empirical evidence for the effectiveness of regime-aware dynamic ensembling
Model interpretability	Applies SHAP analysis to quantify feature importance across regimes	Highlights the dominance of SOM-derived regime labels over raw temporal features	Ensures transparency and operational interpretability of model outputs
Implementation feasibility (secondary outcome)	Deploys the approach using AWS EC2 and Lambda for scalable training and inference	Demonstrates low-latency real-time prediction	Confirms practical applicability without affecting methodological contribution

2. Methodology

This part of the study covers the applied tools and techniques. By analysing incoming load demand, the suggested system generates a combined forecasting model for the Drakensberg power plant Unit in South Africa, operating in real-time. There are two components to this approach for hourly load forecasting, which are described in Section 2.1 and Section 2.2.

2.1. Data pre-processing architecture

This component regulates the lawful and efficient use of big data by sending it to an analytical server for predictive modelling. The architecture incorporates high-performance computing (HPC) for the development and training of initial models, utilising cloud-based services from Amazon Web Services to facilitate scalable and real-time deployments. TensorFlow is utilised for model development, reflecting the methodology in Reference [16], owing to its compatibility with various programming languages such as Python, JavaScript, C++, and Java, thereby ensuring versatility across sectors. Raw data files, usually in comma-separated format, are streamed into the system, parsed, and pre-processed prior to ingestion by the analytical model. The AWS Lambda functions enable real-time predictions, whereas large-scale model training and testing occur on AWS EC2 instances. This integration facilitates ongoing updates and adjustments to the model with new data, thereby enhancing the real-time evolution of a resilient, self-learning predictive model. The system is adaptable and

scalable, supporting both batch and streaming data pipelines for short-term electricity load forecasting, which improves operational efficiency and decision-making.

2.2. *Combined forecasting for hourly electricity load in real-time*

Electricity load forecasting at the hourly level is crucial for maintaining real-time grid stability and ensuring energy market operations. Single-model forecasts (e.g., DNN, LSTM) may fail under regime shifts or complex nonlinear dynamics shifts. Therefore, a combined ensemble modelling structure is proposed to enhance accuracy, robustness, and adaptability in this study. Figure 1 presents the schematic flowchart of our proposed methodology in this study.

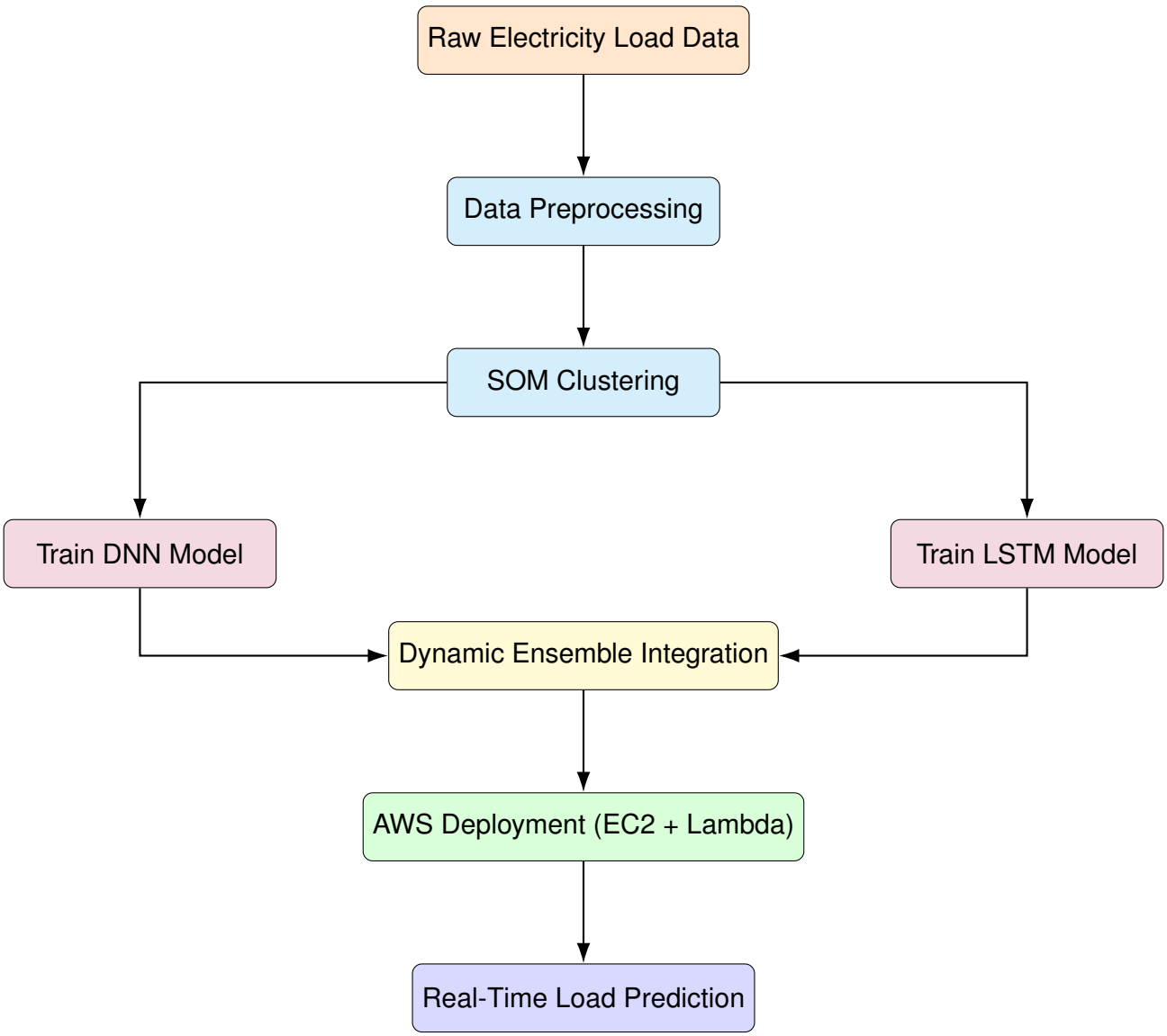

Figure 1. Flowchart of the proposed dynamic ensemble model for real-time road prediction. **Source:** authors' own computation.

2.2.1. Self-organising map network

This clustering approach is a well-known unsupervised competitive learning approach based on artificial neural networks, which was introduced by Reference [17] from a continuous high-dimensional input space Φ to a discretised low-dimensional output space ψ, that generates a map. The low-dimensional output space ψ comprises q neurons organised according to a fixed topology, as stated by Reference [18]. A neuron index is assigned to an input vector $x(t)$ by the mapping funcution $c(X) : \vartheta \to \chi$ which is determined by the weight vectors $Z = \left[z_1, \cdots, z_q\right]$, and this is computed by Eq 2.1 as follows

$$i^* = \arg\min_{\forall_i} \{\|x(t) - Z_i(t)\|\}, \tag{2.1}$$

where the current iteration is denoted by t and the Euclidean distance is shown by $\|\bullet\|$. We now follow Reference [17] and their competitive learning rule for training the weight vectors; this means that when an input vector is fed into the network, the neuron whose weight vector is the most similar to the input will have its weight and updated vector trained, and all of its neighbours as well are given by Eq 2.2.

$$Z_i(t+1) = Z_i(t) + \alpha(t)\, h(i^*, i; t)\, [x(t) - Z_i(t)]. \tag{2.2}$$

As a result, the input vector gains a little support from the weight vectors' modified neurons. The amount of movement is controlled by the learning rate α, which declines exponentially with time, as stated by Reference [18]. This change affects a specific number of neurons, called h, and the Gaussian function in Eq 2.3, which is a popular option for the neighbourhood function, and this function is given by Eq 2.3.

$$h(i^*, i; t) = \exp\left(\frac{\|s_i(t) - s_{i^*}\|^2}{2\sigma^2(t)}\right). \tag{2.3}$$

The distance between neurons i and i^* in the distance output space is represented by $\|s_i(t) - s_{i^*}\|$, while $\sigma(t)$ is the radius of the neighbourhood function at time t. This radius decreases exponentially to guarantee that the learning process begins with global adjustments and ends with local refinements, as stated by Reference [17].

2.3. Deep neural networks

Following SOM-based clustering, a deep neural network is constructed to perform supervised learning for forecasting log-transformed electricity demand. The DNN is designed to model complex nonlinear relationships among temporal features and regime-based patterns identified by the self-organising map [19]. The input feature set consists of both numerical and categorical predictors. Numerical features include the hour of the day, day of the week, month, quarter, and half year, while categorical features derived from the SOM clusters `SOM_Label` are one-hot encoded and appended to the feature matrix. This hybrid feature representation enables the model to account for both granular temporal variations and broader contextual regimes, such as "Afternoon Weekday" or "Early Morning Weekend"; they are also used as inputs for the LSTM network.

Let $x_t \in \Re^d$ denote the input feature vector at time t and y_t be the corresponding log-transformed electricity demand. The DNN seeks to learn a mapping function $f(\cdot)$ such that

$$\hat{y}_t = f(\mathbf{x}_t; \Theta), \tag{2.4}$$

where Θ represents the set of trainable parameters (weights and biases). The objective is to minimise the prediction error across all time steps using the mean squared error (MSE) loss, with mean absolute error (MAE) monitored as an auxiliary performance metric. Our proposed DNN architecture consists of an input layer corresponding to the number of features, followed by two hidden layers. The first hidden layer contains 64 neurons with ReLU activation and a dropout rate of 0.3, while the second hidden layer has 32 neurons with ReLU activation and a dropout rate of 0.2. The output layer consists of a single neuron with linear activation, making it suitable for continuous regression tasks. Figure 2 illustrates this DNN structure.

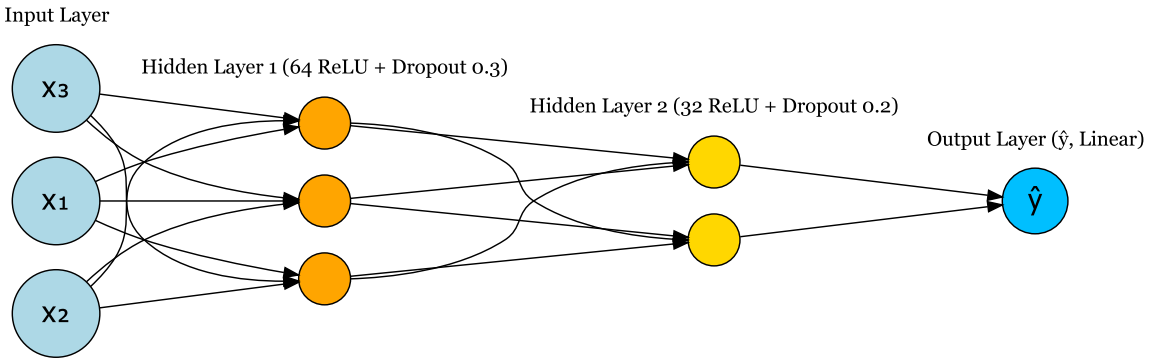

Figure 2. Deep neural network architecture. **Source:** authors' own computation.

For a DNN with L hidden layers, the forward pass is computed as follows:

$$h^{(1)} = \text{ReLU}(W^{(1)} x_t + b^{(1)}), \tag{2.5}$$

$$h^{(l)} = \text{ReLU}(W^{(l)} h^{(l-1)} + b^{(l)}), \quad l = 2, \ldots, L, \tag{2.6}$$

$$\hat{y}_t = W^{(L+1)} h^{(L)} + b^{(L+1)}, \tag{2.7}$$

where $W^{(l)}$ and $b^{(l)}$ denote the weight matrix and bias vector for layer l, and $\text{ReLU}(x) = \max(0, x)$ is applied element-wise. Model parameters are optimised using stochastic gradient descent using a learning rate of 0.01 and momentum of 0.9 as emphasised by Reference [20]. Early stopping is employed as a regularisation mechanism, with the patience parameter varied between 25 and 75 epochs and a minimum improvement threshold from 10^{-8} to 10^{-6} [21]. Training proceeds for a maximum of 1200 epochs, with a mini-batch size of 32 selected based on validation experiments, balancing convergence stability and computational efficiency. Key hyperparameters, including learning rate, momentum coefficient, number of hidden neurons, dropout rates, batch size, and early-stopping patience, are selected through structured validation using chronologically ordered data to preserve temporal dependencies [22].

2.3.1. Long-term-short memory network

Following the clustering of observations via the SOM procedure, each identified cluster in the SOM procedure is subsequently used to train a separate LSTM architecture, allowing it to specialise in the

unique temporal dynamics of that specific cluster. Once the input sequence from a given cluster is fed into the LSTM, the network processes it step-by-step; at step t, and, we compute the expected value of the memory cells $\tilde{C}_t$ and the input gate value i_t as shown in Eq 2.8 and Eq 2.9 .

$$i_t = \sigma\left(\omega_i\left[h_{t-1}, x_t\right] + b_i\right), \tag{2.8}$$

$$\tilde{C}_t = \tanh\left(\omega_c\left[h_{t-1}, x_t\right] + b_c\right), \tag{2.9}$$

as indicated by Reference [22], where $\omega_c\left[h_{t-1}, x_t\right]$ and b represent the weight matrices and the bias, respectively. The forget gate controls the weight of the state cell unit, and the value of the forget gate is computed by Eq 2.10.

$$f_t = \sigma\left(\omega_f\left[h_{t-1}, x_t\right] + b_f\right). \tag{2.10}$$

The process in Eq 2.10 updates the new state of the memory cell by Eq 2.11.

$$\tilde{C}_t = i_t\tilde{C}_t + f_t\tilde{C}_{t-1} \tag{2.11}$$

and with this new state of the memory cell, the output of the gate is computed as shown in Eq 2.12.

$$O_t = \sigma\left(\omega_x\left[h_{t-1}, xX_t\right] + b_0\right); \tag{2.12}$$

hence, the final output value of the cell is then explained by Eq 2.13.

$$h_t = O_t * \tanh\left(c_t\right), \tag{2.13}$$

where i, f, o, and c are the input, forget, output gate, and cell activation vectors, respectively. The sigmoid, denoted as ϕ and tanh functions, are used to train all of the features of the LSTM network design, where Figure 3 shows what the LSTM's basic structure looks like.

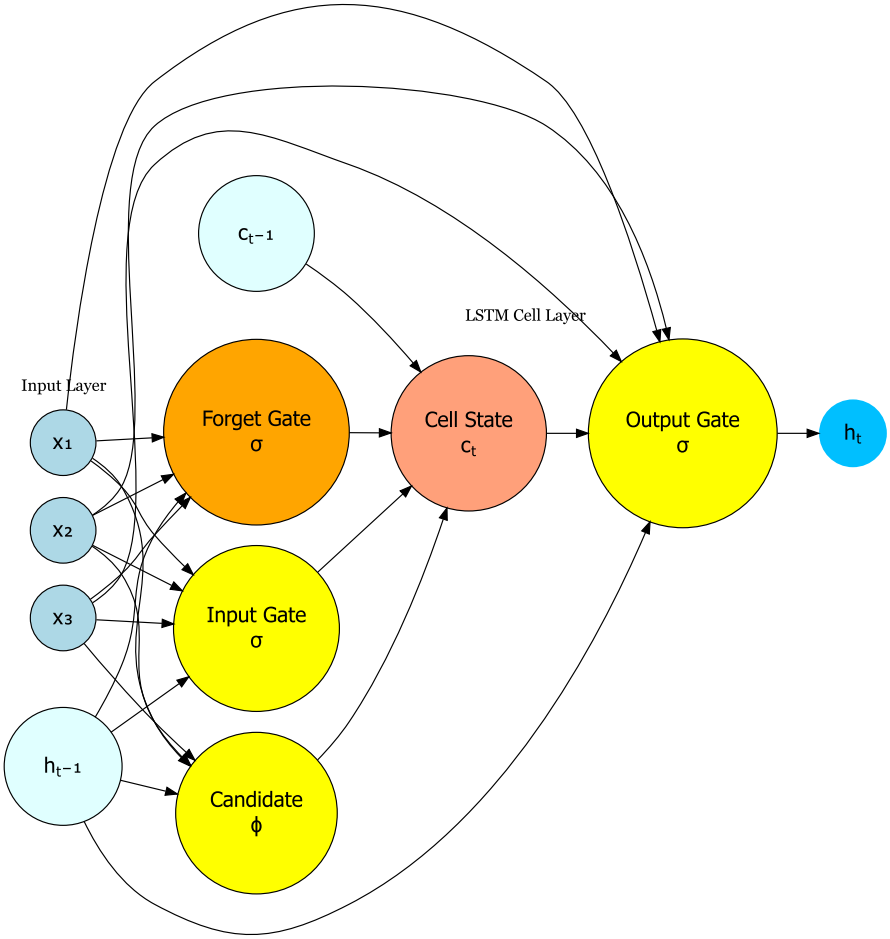

Figure 3. The LSTM architecture. **Source:** authors' own computation.

By letting $x_t \in \Re^d$ be the vector of exogenous features of the SOM clusters at time t, and $y_t \in \Re$ the corresponding log-transformed electricity demand, we therefore capture temporal dependencies by creating sequences of fixed length $T = 24$ as follows:

$$X_t = [x_{t-T+1}, x_{t-T+2}, \ldots, x_{24}], \tag{2.14}$$

with the associated prediction target y_{t+1}. This sliding-window approach preserves chronological order while enabling the network to learn intra-day load patterns [23]. The final hidden state h_T is passed through a fully connected linear layer to produce the forecast $\hat{y}_{t+1}$, where model parameters are optimised by minimising the mean squared error as indicated by Reference [24], and this is computed by Eq 2.15 as follows:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{2.15}$$

with mean absolute error (MAE) monitored as an auxiliary metric to assess robustness against extreme deviations. Optimisation is performed using the Adam optimiser, which adaptively adjusts learning rates based on first- and second-order gradient moments, ensuring stable convergence in the presence of noisy, nonstationary load data such as hourly load demand data under study. In

addition, hyperparameters are selected via validation-based sensitivity analysis [25]. We now let $\Theta = \{H, T, B, E\}$ denote the principal hyperparameters: H is the number of LSTM units, T the sequence length, B the mini-batch size, and E the number of epochs; therefore, the candidate configurations Θ_k are evaluated on a chronologically ordered validation set, with the optimal set Θ^* by minimising the validation error while maintaining stable convergence. Hence, the final configuration uses $H = 64$, $T = 24$, $B = 32$, and $E = 2000$, balancing forecasting accuracy, generalisation, and computational efficiency. Finally, we set the learning rate to 0.01 with the momentum of 0.9 via a stochastic gradient descent algorithm, where patience ranges from 25 to 75, and a minimum absolute difference is set between 10^{-8} and 10^{-6} to enable early stopping.

2.4. Dynamic ensemble integration of SOM-DNN and SOM-LSTM models

The suggested dynamic ensemble amalgamates predictions produced by the SOM-DNN and SOM-LSTM architectures to provide an adaptable and resilient short-term electrical load demand forecasting system. The suggested ensemble uses a completely adaptive, uncertainty-aware weighting mechanism that continually develops in real time, in contrast to static ensemble methods that depend on fixed or globally optimal weights. This design facilitates efficient adaptability to regime transitions, structural alterations, and phases of increased volatility in electricity consumption [26]. Incoming data are first categorised into temporal and behavioural regimes with the self-organising map, therefore encapsulating distinctive patterns such as weekdays, weekends, and seasonal fluctuations. Cluster-specific representations are concurrently supplied to two complementary base learners: a deep neural network, which captures intricate nonlinear interactions among exogenous variables, and a long short-term memory network, which explicitly models temporal dependencies within each regime.

Let $e_{DNN,t}$ and $e_{LSTM,t}$ denote recent forecast errors, for example, rolling root mean square error (RMSE) or mean absolute error for the SOM-DNN and SOM-LSTM architectures, respectively. The dynamic ensemble weights are then defined as

$$w_{DNN,t} = \frac{e_{LSTM,t}}{e_{DNN,t} + e_{LSTM,t}}, \qquad w_{LSTM,t} = \frac{e_{DNN,t}}{e_{DNN,t} + e_{LSTM,t}}. \tag{2.16}$$

In this formulation, the architecture with smaller error contributes more to the ensemble forecast. By construction, the weights always satisfy the following convexity constraint $w_{DNN,t} + w_{LSTM,t} = 1$. We therefore finally define the resulting ensemble forecast as

$$\hat{y}_{Ens,t} = w_{DNN,t}\,\hat{y}_{DNN,t} + w_{LSTM,t}\,\hat{y}_{LSTM,t}. \tag{2.17}$$

This approach ensures robust, adaptive forecasts for South Africa's electricity load demand, effectively balancing short-term spikes captured by the LSTM with cluster-based patterns captured by the SOM-DNN architecture. Importantly, it is algebraically equivalent to the original specification of the weighting mechanism and does not alter the estimation procedure, adaptive weight dynamics, or uncertainty quantification strategy. Consequently, this reformulation has no impact on the model's forecasting performance, empirical results, or study conclusions. Unlike error-based ensembles of Reference [27] or exponential-smoothing ensembles of Reference [28], our weighting mechanism is cluster-conditioned and uncertainty-aware. For each k SOM cluster and $i \in \{DNN, LSTM\}$, base learner, a cluster-specific prediction interval is constructed by

$$\left[\hat{y}_{i,t}^{(k),L},\ \hat{y}_{i,t}^{(k),U}\right], \tag{2.18}$$

with the width being

$$\mathrm{PIW}_{i,t}^{(k)} = \hat{y}_{i,t}^{(k),U} - \hat{y}_{i,t}^{(k),L}, \tag{2.19}$$

which serves as a direct measure of predictive uncertainty. To enhance robustness under transient fluctuations, extreme load events, and noisy observations, a rolling aggregation over a window of length W is applied as

$$\overline{\mathrm{PIW}}_{i,t}^{(k)} = \frac{1}{W} \sum_{s=t-W+1}^{t} \mathrm{PIW}_{i,s}^{(k)}. \tag{2.20}$$

The ensemble weights are then computed using an inverse-reliability formulation given by

$$w_{i,t}^{(k)} = \frac{\left(\overline{\mathrm{PIW}}_{i,t}^{(k)}\right)^{-1}}{\sum_{j\in\{DNN,LSTM\}} \left(\overline{\mathrm{PIW}}_{j,t}^{(k)}\right)^{-1}}, \tag{2.21}$$

which ensures non-negativity and convexity as follows:

$$0 \le w_{i,t}^{(k)} \le 1, \qquad \sum_{i\in\{DNN,LSTM\}} w_{i,t}^{(k)} = 1. \tag{2.22}$$

By substituting these dynamic, cluster- and uncertainty-aware weights into Eq 2.17, the ensemble forecast automatically emphasises the most reliable architecture within each temporal regime. During periods of heightened volatility or regime transitions, models exhibiting increased uncertainty (wider prediction intervals) are downweighted, while models with narrower intervals contribute more strongly to the final forecast. This error- and uncertainty-adaptive weighting strategy enhances forecasting accuracy and robustness without relying on ad hoc tuning parameters, fixed forgetting factors, or manual recalibration. Finally, the ensemble is implemented on Amazon Web Services (AWS), where AWS Lambda functions perform real-time inference and dynamic weight computation, and AWS EC2 instances handle periodic model training and retraining. Both components access a shared Amazon S3 repository for model artefacts and data storage, while Amazon CloudWatch provides continuous monitoring of logs and performance metrics.

3. Data analysis and discussion

This study uses data on the hourly electricity load from the Drakensberg generation plant unit operated by Eskom, covering the period from April 1, 2021, to May 4, 2025, comprising 35,880 observations. We split the data into two sets: the training and testing sets. Figure 4 presents a univariate data analysis for the variable of interest: the hourly electricity power generation for the Drakensberg plant in South Africa. This figure presents the kernel density (a nonparametric method used for estimating probability distributions), normal Q-Q plot, box and whisker plot, and time series. The kernel density plot shows that the hourly electricity data is not normally distributed and skewed to the left. Normal Q-Q plots also show slight deviations from normality. Time series plots further show complex nonlinear patterns.

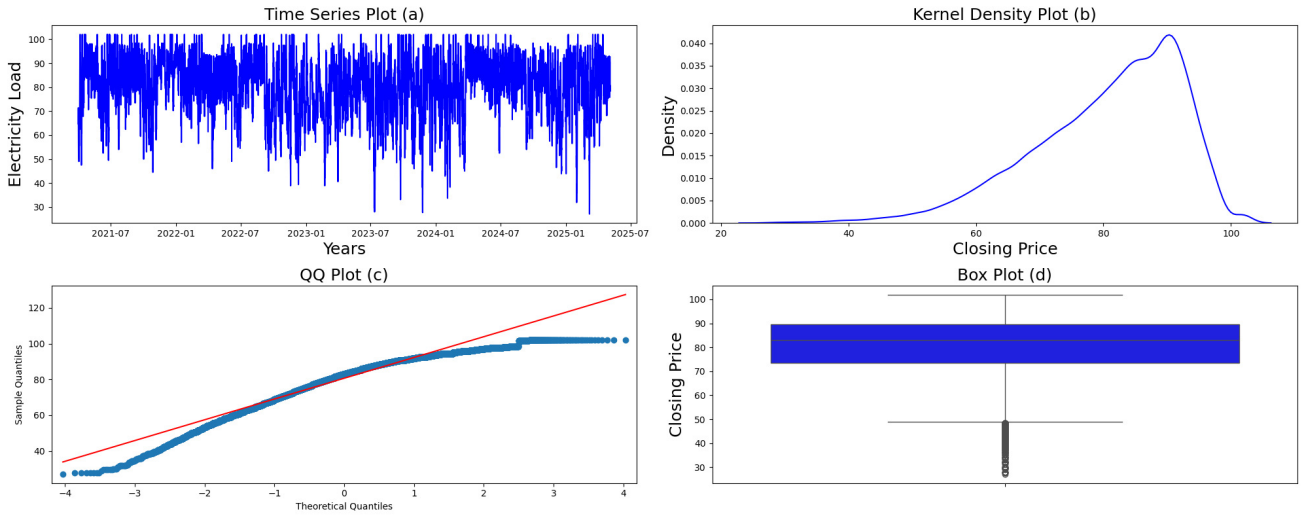

Figure 4. Time series plots.

To further confirm the departure from normality observed in Figure 4, we compute descriptive measures of central tendency, dispersion, asymmetry, and tail behaviour. The results indicate that the Drakensberg pumped storage plant consistently contributes to the national grid, with an average output of 80.67 MW and relatively low variability, as reflected by a standard deviation of 11.60 MW per hour. This limited dispersion suggests stable operational performance under normal conditions. The interquartile range, defined by the 25% and 75% quantiles at 73.4 MW and 89.7 MW, respectively, shows that half of the observed generation levels lie within a narrow and operationally reliable band. This concentration of output around the median reinforces the plant's role as a dependable short-term balancing resource. The maximum observed output of 102 MW further highlights the plant's capacity to respond effectively during periods of elevated demand or emergency dispatch, demonstrating its operational flexibility within system constraints.

Despite this overall stability, the distribution exhibits pronounced negative skewness of -0.88, together with a low minimum output of 27.1 MW, indicating that low-output events occur more frequently or with greater severity than high-output deviations. Such reductions in generation may be attributed to outages, water availability constraints, typical of pumped storage systems, or scheduled maintenance activities. These downside deviations complicate Eskom's ability to satisfy peak demand, particularly during periods of heightened system stress. According to Reference [29], maintaining sufficient emergency reserves is critical for preserving system stability, especially during winter peak periods. This concern is emphasised by the observed increase in the unscheduled capacity loss factor (UCLF), which rose by 1.2% relative to the same period in 2024, reaching 27.99% between 1 April and 8 May 2025. The negative skewness implies that low-generation events contribute disproportionately to supply volatility when considered alongside Eskom's broader generation portfolio. In contrast, the relatively low kurtosis value of 0.65 suggests light-tailed behaviour, indicating that extreme deviations from the mean are infrequent. This characteristic enhances short-term predictability and supports day-ahead operational planning. Nevertheless, the combination of a low minimum output and a negatively skewed distribution indicates the necessity of robust backup strategies and reserve planning to mitigate the risk of worsening load shedding during adverse operational conditions.

3.1. Temporal patterns and clustering analysis

This study employs classical decomposition to remove seasonal components from the electricity demand time series, yielding a deseasonalised dataset in line with the methodology proposed by Reference [16]. By isolating the underlying trend and irregular components, this transformation facilitates a more accurate representation of intrinsic demand dynamics. Following deseasonalisation, temporal hierarchies are constructed across multiple aggregation levels, including hour, day, week, month, quarter, half year, and year. These hierarchical representations serve as informative features that enable the extraction of latent temporal patterns critical for both clustering and forecasting tasks.

To uncover these patterns, a self-organising map is trained on the hierarchical feature set. Consistent with the findings of Reference [30], the SOM effectively groups observations according to similarities in their temporal structure, producing a topology-preserving representation of electricity demand behaviour. This structure enables the identification of dominant demand regimes while retaining nuanced variations across time, thereby supporting both interpretability and downstream modelling. The optimal number of SOM clusters is determined using silhouette analysis, shown in Figure 5. This metric jointly evaluates intra-cluster cohesion and inter cluster separation, providing a robust quantitative basis for cluster selection. The results indicate that $k = 3$ yields the highest silhouette score of 0.423, substantially outperforming alternative configurations [31]. Beyond this point, silhouette values decline sharply and stabilise between 0.22 and 0.26 as k increases, suggesting that additional clusters introduce overlap and reduce interpretability. These findings indicate that, although the SOM consists of numerous nodes, the intrinsic structure of the electricity demand data is most coherently represented by three principal clusters.

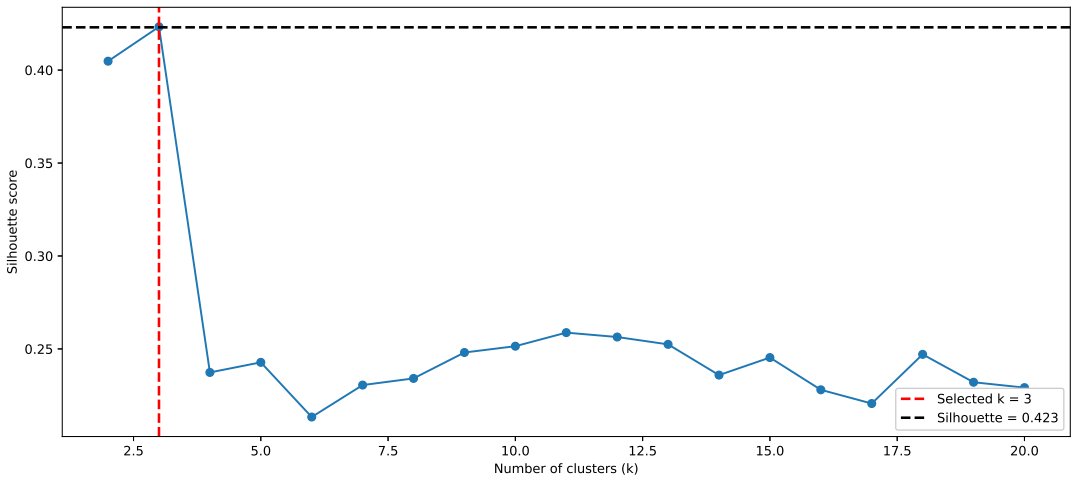

Figure 5. Silhouette analysis for selection of clusters in the SOM.

To further assess the internal structure and separability of the identified clusters, principal component analysis (PCA) is applied to the scaled SOM outputs, reducing the feature space to three principal components for visual inspection. Figure 6 presents the resulting three-dimensional projection, applied in the reduced space, following the approach advocated by Reference [32]. The figure reveals compact and well-separated clusters with minimal overlap, indicating a high degree of

internal homogeneity and meaningful separation between demand regimes. The alignment between the dominant variance directions and the clustering structure suggests that the SOM-derived features effectively capture the principal modes of electricity demand variability.

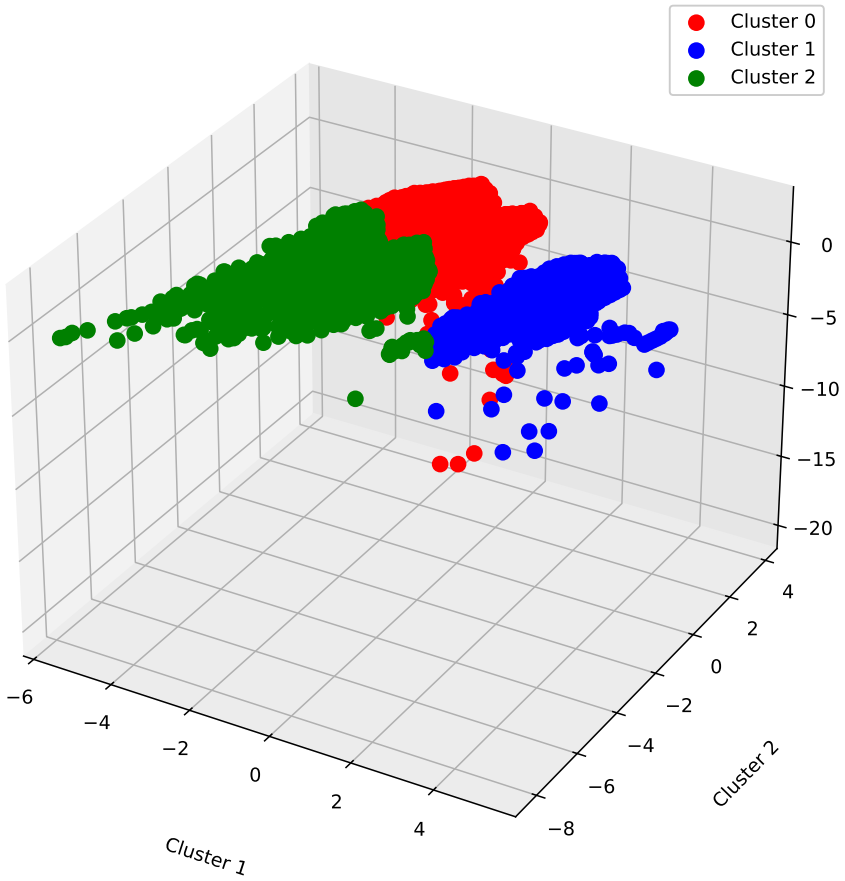

Figure 6. Three-dimensional PCA projection of SOM-derived features with K-means clustering ($k = 3$).

Together, the silhouette analysis and the PCA-SO visualisation provide complementary validation of the three-cluster solution. While the silhouette score offers quantitative justification for the selected partition, the three-dimensional projection provides intuitive confirmation that the identified demand regimes form distinct and coherent structures in the reduced feature space. Following this validation, each SOM cluster is examined in greater detail to characterise its underlying temporal behaviour.

Rather than representing additional clusters, Table 2 reports a comprehensive statistical summary of SOM-derived load patterns observed within the three validated clusters. These patterns are identified through the temporal attributes of the observations assigned to each cluster and are labelled according to dominant time-of-day, day-of-week, and seasonal characteristics. The table summarises key distributional properties alongside median temporal indicators, thereby providing a substantive interpretation of the geometrically distinct clusters observed in Figure 6. In doing so, the analysis explicitly links the reduced feature space representation to meaningful temporal electricity demand behaviour.

Table 2. SOM-derived load pattern summary statistics with temporal hierarchy labels.

Load Pattern Label	Count	Mean	Max	Min	Std Dev	Median Hour	Median DayOfWeek	Median Month
Morning Weekday	11281	4.437676	4.624973	3.475067	0.127446	7.0	1.0	6.0
Afternoon Weekend	2195	4.403711	4.579852	3.703768	0.142717	16.0	6.0	7.0
Evening Weekday	4542	4.379547	4.576771	3.384390	0.124927	19.0	0.0	7.0
Early Morning Weekend	5941	4.372700	4.624973	3.384390	0.172288	5.0	5.0	6.0
Night Weekend	2089	4.359824	4.622027	3.499533	0.142661	21.0	6.0	7.0
Holiday Spike	3548	4.351757	4.624973	3.443618	0.166067	6.0	4.0	7.0
Afternoon Weekday	1187	4.320236	4.567468	3.517498	0.175645	14.0	4.0	3.0
Summer Low	3547	4.300841	4.622027	3.332205	0.166691	19.0	3.0	6.0
Winter Spike	1550	4.242698	4.545420	3.299534	0.199758	20.0	4.0	7.0

Among the identified patterns, the Morning Weekday regime emerges as the dominant demand pattern, accounting for the largest proportion of observations and exhibiting relatively low variability. Its early-morning timing and stable load levels are indicative of routine weekday industrial and residential activity. This regularity aligns closely with the compact structure associated with this regime in the PCA-SOM projection, reinforcing its interpretation as a core component of weekday electricity demand. In contrast, the Afternoon Weekend and Night Weekend patterns capture distinct weekend-specific behaviours, characterised by later median hours and increased variability. The SOM's ability to differentiate between weekday and weekend consumption dynamics is demonstrated by its separation from weekday-oriented patterns, which is in line with well-established results in electricity demand analysis, such as those published by [33]. Furthermore, the presence of multiple evening-related weekday patterns reveals finer-grained substructures within evening demand behaviour. These patterns suggest gradual transitions across time-of-day and day-of-week dimensions, reflecting heterogeneous consumption dynamics that would likely be obscured under more aggregated modelling approaches.

Building on these clustered demand representations, two forecasting architectures, a deep neural network and a long short-term memory network, are trained using SOM-enriched temporal and cluster-based features. A dynamic ensemble structure is subsequently employed to integrate the predictions of both architectures, leveraging their complementary strengths in capturing nonlinear relationships and temporal dependencies. Our modelling approach is different from that of [34], who used a weighted combination of the deterministic exponential smoothing recurrent neural network (ESRNN), which gets 60% of the weight, and the stochastic variational autoencoder (VAE), which gets 40%, and that of Reference [35], the weighted sum ensemble, where the weights of each model are determined by how well it performs. The forecasting results, presented in Table 3, indicate that while the SOM-LSTM and SOM-DNN models achieve broadly comparable accuracy, the dynamic SOM-DNN-LSTM ensemble consistently outperforms both individual architectures. Specifically, the ensemble achieves the lowest MSE, MAE, and MAPE values, demonstrating superior predictive accuracy and generalisability in electricity load forecasting at the Drakensberg generation plant.

Table 3. Forecasting performance.

Model	MSE	MAE	MAPE
SOM-LSTM	0.018994	0.100915	2.36%
SOM-DNN	0.019168	0.102238	2.39%
SOM-DNN-LSTM	0.010913	0.081750	2.20%

These findings carry significant implications for both short-term electricity forecasting and broader grid management. The superior performance of the SOM-DNN-LSTM ensemble, as evidenced by its lower MSE, MAE, and MAPE, emphasises the effectiveness of integrating deep learning models with clustering techniques. The enhanced predictive accuracy offered by this ensemble is critical for operational planning in power utilities such as Eskom, enabling more reliable unit commitment, load balancing, and demand response strategies, which collectively contribute to more efficient and cost-effective grid operations. By incorporating SOM-derived features within a dynamic ensemble structure, the model demonstrates improved generalisation compared with individual architectures, enhancing its ability to capture complex load patterns. This capability is particularly valuable for utility companies, where accurate demand forecasting informs operational efficiency, optimises reserve utilisation, and facilitates the integration of renewable energy sources. Furthermore, the application has potential applicability beyond the Drakensberg generation plant, offering a robust structure for other facilities and regions facing similar forecasting challenges. The effectiveness of the ensemble architecture is also illustrated in Figure 7, which shows its ability to replicate observed electricity load patterns closely. For comparison, the predictive performance of the SOM-DNN and SOM-LSTM models is presented in Figures 9 and 10 in the Appendix, respectively.

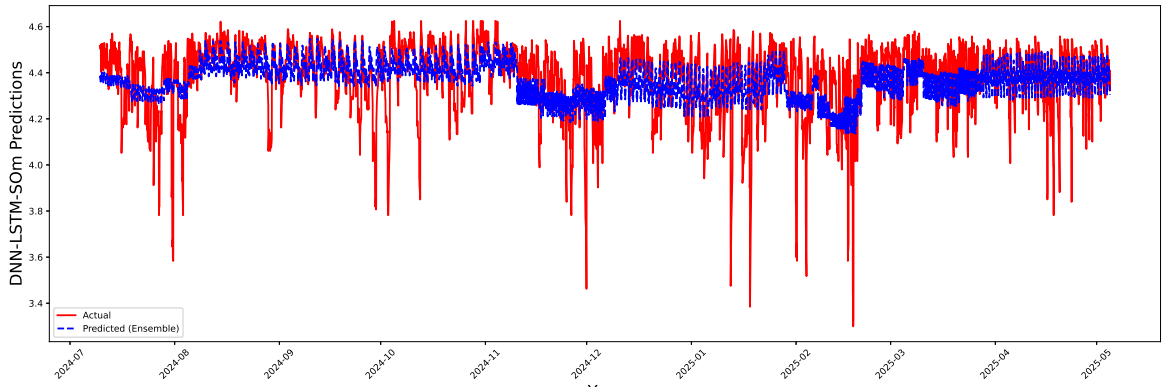

Figure 7. Original Vs predicted time series using combined ensemble architecture.

Having found that the ensemble outperformed the SOM-LSTM and SOM-DNN, we further assess its stability in order to assess its capacity to accurately predict electricity load patterns and to ascertain whether it accurately estimates potential demand loads. This is done by following the approach of Reference [36], and we found that the mean width across the entire test set is 0.0594, indicating that, on average, the true load values are expected to lie within approximately ±0.03 of the predicted mean with 99% probability. The narrowest intervals, with a minimum width of 0.0279, correspond to regions where the SOM-DNN-LSTM is particularly confident, likely due to well-represented patterns in the training data. Conversely, the widest interval of 0.1323 highlights instances of increased uncertainty, which may correspond to atypical feature combinations or periods of higher volatility in the data. The relatively low standard deviation of 0.0161 confirms that most intervals are tightly clustered around the mean, implying consistent predictive reliability across the majority of the test set. Overall, these results demonstrate that the architecture provides robust point predictions while appropriately reflecting varying degrees of epistemic uncertainty, which is crucial for informed decision-making in practical applications.

3.2. Feature importance and temporal interaction for forecasting the Drakensberg generation plant

The dynamic ensemble model exhibits heterogeneous predictive performance across demand patterns, as summarised in Table 4. Relatively low MSE, MAE, and MAPE values are observed for stable and frequently occurring regimes, notably Weekend High, Summer Low, and High Demand (Evening). These periods are characterised by regular and persistent consumption patterns, enabling the model to capture underlying dynamics with high precision. In contrast, the Winter Peak and Weekday Morning patterns display comparatively larger prediction errors, reflecting greater volatility and structural irregularity. Such patterns are inherently more difficult to forecast due to heightened sensitivity to exogenous factors and weaker intraday regularity. Overall, the ensemble benefits from large sample sizes within dominant patterns, strong temporal regularities, and the aggregation of multiple learners, which collectively smooth individual model biases and enhance robustness.

Table 4. Dynamic ensemble performance per cluster patterns.

Patterns	MSE	MAE	MAPE	Count
Weekday Morning	0.035982	0.140580	3.421130	310
Winter Peak	0.039748	0.140144	3.401033	238
Low Demand (Night)	0.023169	0.115899	2.731857	710
Weekday Evening	0.023190	0.107347	2.545095	710
Mid Demand (Morning)	0.020666	0.105548	2.488495	1189
Holiday Spike	0.011182	0.090645	2.053775	439
Weekend High	0.012763	0.078990	1.821690	2257
Summer Low	0.009488	0.079329	1.815871	418
High Demand (Evening)	0.009768	0.075379	1.738220	909

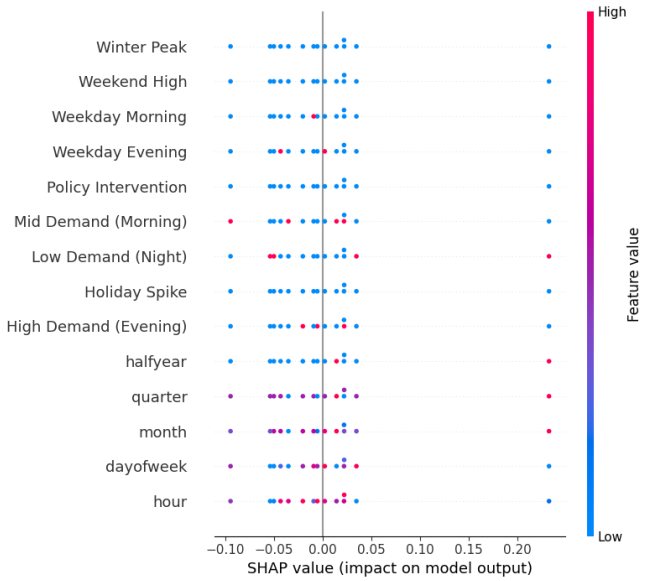

Figure 8. SHAP values for feature importance.

Figure 8 presents the SHAP values, which quantify both the relative importance of features and the

direction of their influence on the ensemble model's predictions. The results reveal a clear hierarchical structure in the way information is utilised within the forecasting structure. The most influential predictors are the SOM-derived cluster labels, which encode distinct demand regimes identified by the self-organising map. Among these, Winter Peak exhibits the widest dispersion of SHAP values, indicating that observations classified within this regime substantially increase or decrease predicted demand depending on prevailing conditions. This behaviour is consistent with the strong seasonal and weather-driven variability typically associated with winter electricity consumption. Similarly, Weekend High, Weekday Morning, and Weekday Evening exert pronounced influence, emphasising the importance of recurrent daily and weekly operational patterns. These findings confirm that the SOM effectively captures the dominant structural regimes governing demand behaviour.

In contrast, the original temporal variables, hour, day of week, month, quarter, and half year, exhibit SHAP values clustered close to zero once SOM-based regime classification is applied. This indicates that their marginal contribution to prediction is limited when considered independently. Nevertheless, quarters and half years display slightly greater dispersion, suggesting the persistence of residual seasonal effects beyond regime classification. Importantly, this does not imply irrelevance of temporal features; rather, their influence is largely conditional and mediated through the demand regimes identified by the SOM. The SHAP plot further reveals pronounced nonlinear effects. Later hours may either increase or suppress predicted demand depending on the associated regime, while later months and broader seasonal indicators generally increase predictions, particularly within winter-dominated regimes. Demand-related indicators also exhibit asymmetric behaviour: high values of Low Demand (Night) may either raise or lower predictions, whereas elevated Mid Demand (Morning) values typically reduce forecasted demand. These patterns justify the use of nonlinear ensemble learning.

Table 5. Interaction of demand clusters and GMM regimes with temporal correlations.

Cluster / Regime	corr(hour)	corr(day)	corr(month)
Weekday Morning	-0.302	-0.256	-0.001
Weekday Evening	0.519	-0.258	0.002
Weekend High	0.092	0.561	-0.002
Low Demand (Night)	-0.751	0.001	0.002
Mid Demand (Morning)	-0.040	-0.254	-0.005
Winter Peak	0.002	-0.402	0.573
Holiday Spike	0.009	0.018	0.037
High Demand (Evening)	0.597	-0.249	0.004
Policy Intervention	0.682	0.000	-0.004
Half Year	0.530	-0.493	0.246
Quarter	0.461	0.571	0.129
GMM Regime 1	-0.523	-0.194	0.476
GMM Regime 2	-0.459	0.585	-0.089

While individual SHAP values describe marginal effects, interaction and correlation analysis provide insight into how features jointly influence predictions. Table 5 summarises the relationship between demand clusters, Gaussian mixture model (GMM) regimes, and key temporal variables. Although the absolute magnitude of interactions is modest, consistent and interpretable patterns emerge that align closely with the SHAP importance results.

The temporal correlations reinforce the SHAP-based interpretation. The Weekday Evening and High Demand (Evening) show strong positive associations with hour, confirming that increasing demand later in the day defines these regimes. Conversely, Low Demand (Night) is strongly negatively correlated with hour, reflecting systematically reduced demand during late-night periods. Weekend High is predominantly influenced by the day of the week rather than intraday or seasonal effects, highlighting the dominance of calendar-driven behaviour. Seasonality is most pronounced for the Winter Peak pattern, which is strongly correlated with month and largely independent of hour. This explains its elevated forecasting uncertainty and confirms that winter demand is driven primarily by seasonal rather than intraday dynamics. Holiday spikes and Policy Intervention regimes display weaker and more diffuse temporal relationships, consistent with episodic and externally driven influences.

Taken together, the SHAP feature importance, interaction structure, and temporal correlations demonstrate a coherent hierarchical modelling strategy. The SOM-derived clusters capture the dominant operational regimes of the Drakensberg generation plant, while temporal variables act as complementary signals that refine predictions within each regime. This layered structure explains both the strong performance observed for stable, high-demand periods and the increased uncertainty associated with more volatile regimes, particularly Winter Peak.

3.3. Operational and economic implications of short-term electricity forecasting

The empirical findings demonstrate that short-term electricity demand forecasting using self-organising maps integrated with deep learning architectures offers substantial operational and economic value for Eskom and electricity-dependent industries in South Africa. By accurately capturing regime-dependent demand dynamics, the proposed SOM-DNN-LSTM ensemble enables more informed generation scheduling, reduces reliance on high-cost peaking resources such as diesel-fired plants, and enhances overall load balancing across the national grid. These improvements directly contribute to lower operating costs, reduced system stress, and improved supply reliability, which are critical in a power system characterised by constrained capacity and increasing demand volatility. From an operational perspective, the ability to distinguish stable demand regimes from volatile or event-driven patterns, such as Winter Peak or Holiday Spike, allows Eskom to anticipate periods of heightened uncertainty and allocate reserves more efficiently; this capability is particularly important for mitigating the risk of load shedding, as improved short-term forecasts support proactive rather than reactive system management. Furthermore, enhanced forecast precision facilitates better coordination between baseload, mid-merit, and peaking generation assets, thereby improving overall system efficiency and operational resilience.

The benefits of accurate short-term forecasting extend beyond the utility sector to electricity-intensive industries, including mining, manufacturing, and data centre operations. Reduced uncertainty in demand projections enables firms to optimise production schedules, manage electricity procurement more strategically, and minimise costly disruptions arising from supply instability. Moreover, reliable forecasts support participation in demand response and load-shifting programmes, allowing businesses

to reduce energy costs while contributing to broader grid stability. In this regard, improved forecasting accuracy strengthens both firm-level competitiveness and system-wide efficiency. These findings are consistent with recent studies in the South African context. Reference [37] demonstrate that hybrid deep learning frameworks combining long short-term memory networks with ensemble learners significantly improve forecasting accuracy during peak demand periods, reinforcing the effectiveness of integrated, nonlinear modelling approaches. Similarly, Reference [38] emphasise that reliable short-term electricity demand forecasts are essential for national power system planning and for informing strategic instruments such as the integrated resource plan, particularly under conditions of structural change and increasing renewable energy penetration. In addition, Reference [39] highlight the operational importance of accurate load forecasting in maintaining grid stability amid growing variability from intermittent energy sources.

Taken together, these results confirm that the proposed SOM-enhanced dynamic ensemble framework is not only methodologically robust but also operationally relevant. By hierarchically combining regime identification with deep learning-based temporal forecasting, the model achieves improved generalisation and interpretability relative to single-architecture approaches. This makes it a practical and scalable solution for short-term electricity demand forecasting at the Drakensberg generation plant and offers a transferable modelling charter for other power system contexts facing similar challenges related to demand variability, capacity constraints, and the ongoing energy transition.

4. Limitations, recommendations, and conclusions

This section examines the principal limitations of the proposed forecasting framework, outlines recommendations to enhance its performance and generalisability, and summarises the key findings and their implications for real-time electricity demand management in South Africa.

4.1. Limitations

While the proposed dynamic ensemble model integrating DNN and LSTM architectures demonstrates significant advances in short-term electricity demand forecasting, several limitations are noteworthy. The model's sophistication, particularly the combination of SOM-based cluster conditioning and dynamic ensemble weighting, imposes substantial computational demands, requiring high-performance hardware and technical expertise, which may not be universally accessible. The framework was developed and validated using data from the Drakensberg generation plant, and its predictive performance in other regions with distinct demand patterns, influenced by regional, seasonal, or socio-economic factors, remains untested. Predictive accuracy decreases for irregular or highly variable load patterns, particularly within clusters like Winter Peak and Weekday Morning, which show random and less predictable behaviours, leading to higher error metrics. Although SOM clustering enhances interpretability and facilitates regime-based forecasting, the relevance of cluster structures is dependent on historical data, and without periodic updating, these clusters may become outdated, potentially diminishing forecast reliability. Furthermore, while global feature importance has been assessed via SHAP analysis, the model does not explore local explanations for individual predictions, anomalous days, or extreme events, limiting detailed interpretability under atypical conditions.

4.2. Recommendations

To address these limitations and improve operational robustness, the model could be extended to include data from multiple Eskom generating units and regions, thereby testing its generalisability across diverse demand profiles. The incorporation of exogenous variables such as weather conditions, public holidays, industrial activity, and economic indicators is likely to enhance forecasting performance, particularly for clusters characterised by high variability or irregular demand spikes. Implementing adaptive SOM frameworks that update cluster boundaries in response to evolving consumption patterns would maintain the temporal relevance of clusters and strengthen long-term reliability. Integrating probabilistic forecasts or prediction intervals would provide valuable information regarding uncertainty and operational risk, supporting informed grid management decisions. Additionally, optimising model efficiency through techniques such as pruning, quantisation, or simplified ensemble configurations could reduce computational overhead while maintaining accuracy. Finally, expanding interpretability by including local SHAP explanations for individual predictions and outlier events would improve transparency and stakeholder confidence.

4.3. Conclusions

This study introduces a dynamic, cluster-conditioned ensemble framework for short-term electricity demand forecasting in South Africa, combining the pattern recognition capabilities of DNNs with the temporal modelling strengths of LSTMs, conditioned on SOM-derived clusters. The ensemble dynamically integrates the outputs of these architectures based on cluster-specific predictive reliability, enabling adaptation to evolving demand patterns and regime shifts. The model demonstrates superior performance in high-frequency and recurring demand clusters such as "Weekend High" and "High Demand (Evening)," while SOM-derived labels enhance interpretability by capturing regime-specific behaviour. SHAP analysis confirms that predictions are predominantly driven by these clustered behavioural features rather than raw temporal inputs. The system is deployed on Amazon Web Services, utilising AWS Lambda for low-latency inference and EC2 instances for model training and retraining, allowing continuous adaptation and robust performance under non-stationary conditions. Despite reduced accuracy in highly stochastic or structurally changing demand clusters, the ensemble provides a scalable, interpretable, and flexible solution for short-term load forecasting, with significant potential to improve grid stability, reduce forecasting errors, and support data-driven energy management across South Africa's electricity infrastructure. Future extensions incorporating multi-region data, adaptive clustering, exogenous predictors, and uncertainty quantification will further enhance model robustness and operational applicability in real-time settings.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare that they have no conflicts of interest. The authors confirm that there are no financial, personal, professional, or institutional relationships that could be perceived as influencing the research reported in this manuscript.

Authors contribution

Katleho Makatjane: Formal data analysis. Caston Sigauke: Conceptualisation, Writing—original draft. Claris Shoko: Methodology, Writing—review and editing. Ntebogang Moroke: Conceptualisation, review and editing of the final draft.

References

1. Rostum M, Zamel AA, Moustafa H, et al. (2020) Electrical load forecasting: A methodological overview. *Int J Eng Technol* 9: 842–869. https://doi.org/10.14419/ijet.v9i3.30706
2. Bunn DW (2000) Forecasting loads and prices in competitive power markets. *Proc IEEE* 88: 163–169. https://doi.org/10.1109/5.823996
3. Hobbs BF, Jitprapaikulsarn S, Konda S, et al. (1999) Analysis of the value for unit commitment of improved load forecasts. *IEEE Trans Power Syst* 14: 1342–1348. https://doi.org/10.1109/59.801894
4. Dudek G (2016) Pattern-based local linear regression models for short-term load forecasting. *Electr Power Syst Res* 130: 139–147. https://doi.org/10.1016/j.epsr.2015.09.001
5. Percuku A, Minkovska D, Hinov N (2025) Enhancing electricity load forecasting with machine learning and deep learning. *Technologies* 13: 59. https://doi.org/10.3390/technologies13020059
6. Al-Ani BRK, Erkan TE (2022) A study of load demand forecasting models in electricity using artificial neural networks and fuzzy logic models. *Int J Eng* 35: 1111–1118. https://doi.org/10.5829/ije.2022.35.06c.02
7. Eberhart, Shi Y (2001) Particle swarm optimisation: Developments, applications and resources. *Proceedings of the 2001 Congress on Evolutionary Computation, IEEE* 1: 1–86. https://doi.org/10.1109/CEC.2001.934374
8. Kim J, Kim H, Kim H, et al. (2025) A comprehensive survey of deep learning for time series forecasting: Architectural diversity and open challenges. *Artif Intell Rev* 58: 1–95. https://doi.org/10.1007/s10462-025-11223-9
9. Islam BU, Ahmed SF (2022) Short-term electrical load demand forecasting based on LSTM and RNN deep neural networks. *Math Probl Eng* 2022: 2316474. https://doi.org/10.1155/2022/2316474
10. Mohan P, Patil KK (2018) Deep learning-based weighted SOM to forecast weather and crop prediction for agriculture applications. *Int J Intell Eng Syst* 11: 167–176. https://doi.org/10.22266/ijies2018.0831.17
11. Guo F, Li S, Zhao G, et al. (2024) A SOM-LSTM combined model for groundwater level prediction in karst critical zone aquifers considering connectivity characteristics. *Environ Earth Sci* 83: 267. https://doi.org/10.1007/s12665-024-11567-5
12. Raj RA, Sarathkumar D, Andrews LJB, et al. (2023) Key gases in transformer oil—An analysis using self-organising map (SOM) neural networks. *2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT)*, 642–647. https://doi.org/10.1109/CSNT57126.2023.10134597

13. Aguilar Madrid E, Antonio N (2021) Short-term electricity load forecasting with machine learning. *Information* 12: 50. https://doi.org/10.3390/info12020050

14. Baur L, Ditschuneit K, Schambach M, et al. (2024) Explainability and interpretability in electric load forecasting using machine learning techniques: A review. *Energy AI* 16: 100358. https://doi.org/10.1016/j.egyai.2024.100358

15. Wang A, Yu Q, Wang J, et al. (2023) Electric load forecasting based on deep ensemble learning. *Appl Sci* 13: 9706. https://doi.org/10.3390/app13179706

16. Shoko C, Moroke ND, Makatjane K (2024) A deep learning framework for modelling temporal dependencies and hierarchies in hourly electricity demand load. In: Acharjya P, Koley S, Barman S (eds) Machine Learning and Computer Vision for Renewable Energy. *IGI Global*, 42–65. https://doi.org/10.4018/979-8-3693-2355-7.ch003

17. Baptista ML, Henriques EM, Goebel K (2021) A self-organising map and a normalising self-organising multi-layer perceptron approach to baselining in prognostics under dynamic regimes. *Neurocomputing* 456: 268–287. https://doi.org/10.1016/j.neucom.2021.05.031

18. Sun S, Wang S, Wei Y, et al. (2020) A clustering-based nonlinear ensemble approach for exchange rates forecasting. *IEEE Trans Syst Man Cybern Syst* 50: 2284–2292. https://doi.org/10.1109/TSMC.2018.2799869

19. Alotaibi MA (2022) Machine learning approach for short-term load forecasting using deep neural network. *Energies* 15: 6261. https://doi.org/10.3390/en15176261

20. Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press, Cambridge. Available from: `https://virtualmmx.ddns.net/gbooks/DeepLearning.pdf`.

21. Wang H, Li T, Zhuang Z, et al. (2023) Early stopping for deep image prior. *Trans Mach Learn Res*. Available from: `https://openreview.net/forum?id=231ZzrLC8X`.

22. Oruh J, Viriri S, Adegun A (2022) Long short-term memory recurrent neural network for automatic speech recognition. *IEEE Access* 10: 30069–30079. https://doi.org/10.1109/ACCESS.2022.3159339

23. Toba AL, Kulkarni S, Khallouli W, et al. (2025) Long-term traffic prediction using deep learning long short-term memory. *Smart Cities* 8: 126. https://doi.org/10.3390/smartcities8040126

24. Echrigui R, Hamiche M (2023) Optimising LSTM models for EUR/USD prediction in the context of reducing energy consumption. *E3S Web Conf* 412: 01069. https://doi.org/10.1051/e3sconf/202341201069

25. Kim HS, Choi D, Yoo DG, et al. (2022) Hyperparameter sensitivity analysis of deep learning-based pipe burst detection model for multiregional water supply networks. *Sustainability* 14: 13788. https://doi.org/10.3390/su142113788

26. He X, Zhao W, Zhang L, et al. (2024) A novel ensemble deep reinforcement learning model for short-term load forecasting based on Q-learning dynamic model selection. *J Eng* 2024: 1–12. https://doi.org/10.1049/tje2.12409

27. Murthy R, Padmalatha NA (2025) Iterative error-based ensemble forecasting model: A novel framework for enhanced forecast accuracy. *IUP J Appl Econ* 24: 85–95. https://doi.org/10.71329/IUPJAE/2025.16.1.85-95

28. Dantas TM, Oliveira FLC (2018) Improving time series forecasting: An approach combining bootstrap aggregation, clusters and exponential smoothing. *Int J Forecast* 34: 748–761. https://doi.org/10.1016/j.ijforecast.2018.05.006

29. Eskom (2025) Power system remains stable with a sustained reduction in unplanned outages. Available from: https://www.eskom.co.za/power-system-remains-stable-with-a-sustained-reduction-in-unplanned-outages/.

30. Xaba D, Makatjane K, Senosi A (2024) Prediction accuracy of SARIMA-STAR-CNE. In: Vasant P, Panchenko V, Munapo E, et al (eds) Intelligent Computing and Optimisation. *Lect Notes Netw Syst*, 1169. Springer, Cham. https://doi.org/10.1007/978-3-031-73324-6_31

31. Januzaj Y, Beqiri E, Luma A (2023) Determining the optimal number of clusters using silhouette score as a data mining technique. *Int J Online Biomed Eng*, 19. https://doi.org/10.3991/ijoe.v19i04.37059

32. Majidi F (2023) A hybrid SOM and K-means model for clustering time series energy consumption clustering. *arXiv*. https://doi.org/10.48550/arXiv.2312.11475

33. Toussaint W, Moodley D (2020) Clustering residential electricity consumption data to create archetypes that capture household behaviour in South Africa. *S Afr Comput J* 32: 1–34. Available from: https://hdl.handle.net/10520/ejc-comp-v32-n2-a3

34. Sigauke C, Moroke N, Makatjane K, et al. (2025) A deep learning forecasting of downside risk: Application of a combined ESRNN-VAE. *Front Appl Math Stat* 11: 1662252. https://doi.org/10.3389/fams.2025.1662252

35. Meswal H, Kumar D, Gupta A, et al. (2024) A weighted ensemble transfer learning approach for melanoma classification from skin lesion images. *Multimed Tools Appl* 83: 33615–33637. https://doi.org/10.1007/s11042-023-16783-y

36. Amnuaypongsa W, Wangdee W, Songsiri J (2025) Neural network-based prediction interval estimation with large width penalisation for renewable energy forecasting and system applications. *Energy Convers Manag X* 27: 101119. https://doi.org/10.1016/j.ecmx.2025.101119

37. Mazibuko T, Akindeji K (2025) Hybrid forecasting for energy consumption in South Africa: LSTM and XGBoost approach. *Energies* 18: 4285. https://doi.org/10.3390/en18164285

38. SANEDI (2023) Short-term electricity demand forecasting for South Africa: Implications for grid planning and policy. *South African National Energy Development Institute*. Available from: https://sanedi.org.za/.

39. Eskom (2023) Medium-Term System Adequacy Outlook 2024–2028. Eskom Holdings SOC Ltd, Pretoria. Available from: https://www.eskom.co.za/wp-content/uploads/2023/11/Medium_Term_System_Adequacy_Outlook_2024-2028.pdf.

Appendix

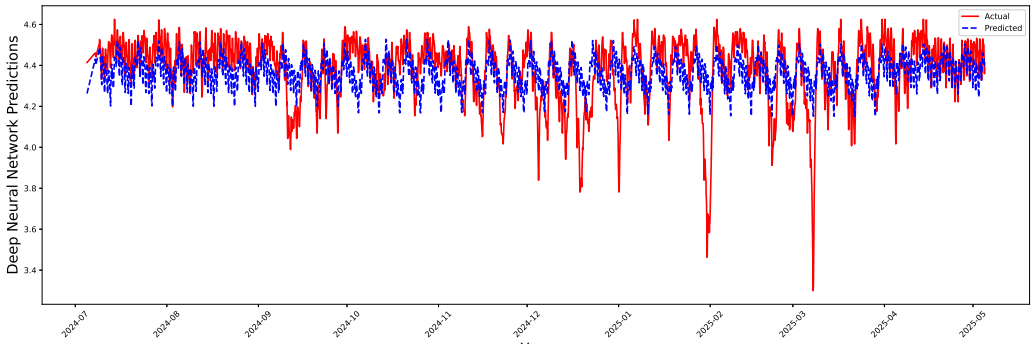

Figure 9. Original Vs Predicted Time Series using Deep Neural Network Architecture.

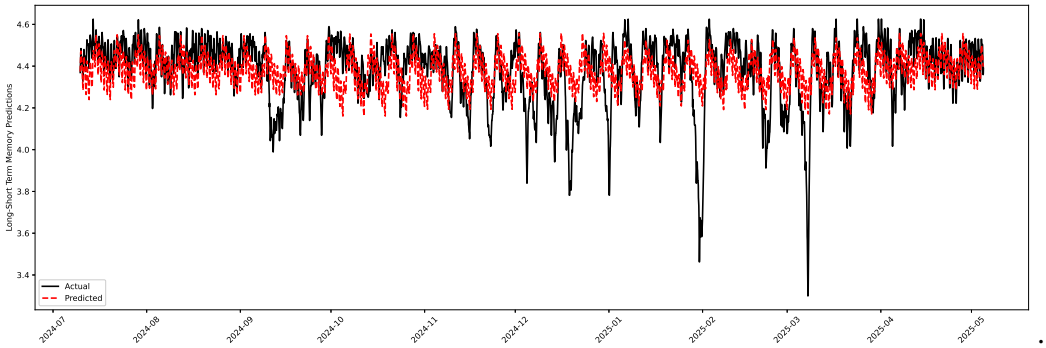

Figure 10. Original Vs Predicted Time Series using Long-term Short Memory Architecture.