Short-Term Load Forecasting In Thimphu and Phuentsholing Regions Using Machine Learning and Deep Learning Techniques

Jigme Namgyal*[1], Namgay Tenzin[2], Pema Galey[3], Tshering Denka[4]

[1, 4] Science and Humanities Department, College of Science and Technology, Royal University of Bhutan
[2] Electrical and Electronics Engineering Department, College of Science and Technology, Royal University of Bhutan
[3] Information Technology Department, College of Science and Technology, Royal University of Bhutan
E-mail: jigmenamgyal.cst@rub.edu.bt*[1], namgaytenzin.cst@rub.edu.bt[2], pemagaley.cst@rub.edu.bt[3], tsheringdenka.cst@rub.edu.bt[4]

Received: 14 April 2025; Revised: 9 June 2025; Accepted: 5 July 2025; Published: 15 July 2025

Abstract

Accurate short-term load forecasting is essential for efficient power system operation, energy management, and electricity pricing. Traditional statistical methods, such as seasonal autoregressive integrated moving averages with exogenous variables (SARIMAX), often fail to capture the intricate and dynamic patterns of electricity demand. Addressing the knowledge gap in developing countries, particularly Bhutan, this research explores advanced machine learning and deep learning techniques to enhance short-term load forecasting (STLF) accuracy in the Thimphu and Phuentsholing regions of Bhutan, characterized by unique electricity demand patterns due to population growth, industrial and commercial activities, and supply constraints. We evaluated SARIMAX, Support Vector Regression (SVR), Long Short-Term Memory Networks (LSTM), Convolutional Neural Networks (CNN), and hybrid CNN-LSTM architectures. On single-step STLF, we analyzed day-ahead aggregated load forecasts and 1-hour-ahead load forecasts based on historical load data over five years (2018-2022) for both Thimphu and Phuentsholing regions. In day-ahead aggregated load forecasting, the hybrid CNN-LSTM outperformed all other models with Mean Absolute Percentage Error (MAPE) values $2.332 \pm 0.075\%$ for Thimphu and $3.216 \pm 0.036\%$ for Phuentsholing, while also achieving the best MSE, RMSE, and R^2 metrics. For 1-hour-ahead forecasting, the CNN model achieved the lowest MAPE of $3.224 \pm 0.06\%$ in Thimphu and the hybrid CNN-LSTM model achieved a best MAPE of $3.687 \pm 0.027\%$ for Phuentsholing. Careful preprocessing, optimal feature engineering, and hyperparameter tuning were performed for all forecasting types. The findings demonstrate that data-driven approaches significantly enhance forecasting accuracy, providing valuable insights for energy planners to manage resources and maintain the power grid, preventing blackouts and other disruptions.

Key Words: *Short-Term Load Forecasting, Machine Learning, Deep Learning, Time-Series Analysis*

1. INTRODUCTION

Reliable short-term predictions of electricity demand are essential—they help keep power systems running smoothly, optimize energy use, and set fair electricity prices (Madrid & Antonio, 2021). Traditional statistical methods for load forecasting, such as autoregressive integrated moving average (ARIMA) and its advanced version, such as SARIMAX models, have limitations in capturing the complex and dynamic patterns of electricity demand, which may be affected by factors such as weather conditions, time of day, day of the week, and seasonality.

Machine Learning (ML) and Deep Learning (DL) techniques have shown great potential in improving the accuracy of short-term load forecasting by leveraging the power of data-driven approaches to capture non-linear and non-stationary patterns (Vanting et al., 2021). In the context of Thimphu and Phuentsholing regions, which have a unique electricity demand pattern due to population growth, its industrial and commercial activities and constraints on the electricity supply (Zam et al., 2021), there is a need to develop and evaluate machine learning and deep learning models for short-term load forecasting.

The coincidental peak load of Bhutan was recorded at 955.51 MW in 2023, which was an increase of 51.76% compared to 2022 when the coincidental peak load was 629.61 MW. During the winter months, when power generation is low, Bhutan imports electricity at a much higher tariff compared to the export tariff (BPSO, 2023). Bhutan's 13th Five-Year Plan (2024-2028) proposes to transform the country into a developed nation by 2030. This plan includes the development of Gelephu Mindfulness City and other developmental projects (Dema, 2024) as well as the development of 1 GW of solar photovoltaic systems (Lhaden, 2023). The increase in electrical load is evident, and load forecasting techniques will be crucial to match the growing load demand and power generation.

As Bhutan looks to diversify its energy sources with the introduction of large-scale solar photovoltaic power plants and wind energy conversion systems, which are inherently intermittent, accurate short-term load forecasting techniques are essential to ensure grid stability (Al-Ja'afreh et al., 2023). The forecasting should not only focus on estimating the energy resource, such as solar radiation and wind but also on estimating the possible coincidental load at the particular time when there is power generation from renewable energy sources. Additionally, there should be adequate ancillary services to cater to any deviations in power generation from renewable energy sources (Ray et al., 2024).

The existing research on short-term load forecasting using machine learning and deep learning techniques has mostly focused on developed countries with different load patterns and characteristics. Therefore, there is a knowledge gap in the application of these techniques to developing countries like Bhutan. This research aims to address this gap by developing and evaluating machine learning and deep learning models for STLF in Thimphu and Phuentsholing regions so that the energy planners can make necessary resources and maintenance planning on the power grid preventing blackouts and other disruptions. Furthermore, this study will contribute to a more sustainable and reliable global energy system by demonstrating the effectiveness of data-driven approaches in complex and dynamic electricity demand patterns, advancing our understanding of the potential benefits and limitations of these techniques.

The rest of the study is organized as follows: Section 2 presents reviews of key literature. Section 3 discusses the materials and methods, and Section 4 presents the results and discussions. Finally, Section 5 provides the conclusions and recommendations.

2. RELATED WORK

With the use of Artificial Intelligence, big data, and the Internet of Things for digitalization, smart grids are the future of power systems. He (2017) evaluated deep neural networks such as CNN and recurrent neural networks (RNN), along with traditional time series data analysis methods for load forecasting. The evaluation was conducted on a large dataset containing hourly loads for North China City for about 3 years. Performance evaluation metrics used were MAPE and mean average error (MAE). Their findings indicate that the parallel CNN-RNN model achieved the lowest MAPE and MAE of 1.04% and 104.24 MW, respectively. It is recommended that for the robust performance of the models, a large training dataset is needed with more relevant input features such as hourly temperature and humidity. A review of deep learning methods applied to smart grid load forecasting was done by (Almalaq & Edwards, 2017). Their survey explored the different applications of deep learning models that are used in power systems and smart grid forecasting. It was discovered that the use of CNN with k-means algorithm resulted in a great percentage reduction in root mean squared error (RSME) in comparison to other models. To improve the accuracy of STLF, Phyo et al. (2019), have applied LSTM and deep believe network models to their work of electricity load forecasting in Thailand. The historical data from the Electricity Generating Authority of Thailand was obtained from the period of January 2016 to January 2017. It was revealed that the LSTM model executed better than the DBN model. Ibrahim et al. (2022) conducted a complex case study on STLF in smart grids within Panama's power system. They identified key features such as the previous week's load, the previous day's load, and temperature. The deep learning regression model achieved the best performance with an R-squared (R^2) metric of 0.93 and an MAPE of about 2.9%. In addition, several performance analyses and comparisons of diverse ML and DL techniques have been deployed to study STLF in the study by Shahare et al. (2023). It is reported that the hybrid CNN-LSTM model revealed superior performance with a coefficient of correlation (R) value of 95.05%.

While these studies provide valuable insights into the application of ML and DL techniques for STLF, there are limitations that must be considered. For example, some studies may have used limited data which could affect the generalizability of their findings. In synthesizing the findings of these studies, it is evident that incorporating relevant input features such as temperature and humidity can improve the accuracy of the models. Additionally, the studies highlight the importance of selecting appropriate models for specific contexts and the need for large training datasets to ensure robust model performance.

3. MATERIALS AND METHODS

3.1 Tools and Dataset Description

All computations were performed in Jupyter Notebook IDE using the Python programming language and the necessary libraries for classical statistical and machine learning models. For DL model implementations, we used TensorFlow as the backend. The computations were carried out on a personal MacBook Air M1 with 8 GB of memory system.

Bhutan, for the same five-year period. However, we noticed significant missing values in the data consecutively, and it was recorded only on a daily basis. Since we aimed to obtain data at an hourly resolution, we obtained weather data for both Thimphu and Phuentsholing regions from the online weather API OpenMeteo, which included the following input parameters: maximum and minimum temperature recorded daily and hourly at 2 meters above ground, apparent temperature, precipitation hours recorded daily and hourly, precipitation duration recorded daily, snow depth in Thimphu during the winter season, sunlight and daylight duration recorded daily for the Phuentsholing region, and solar irradiance. Additionally, we developed calendar effects data, such as the day of the week,

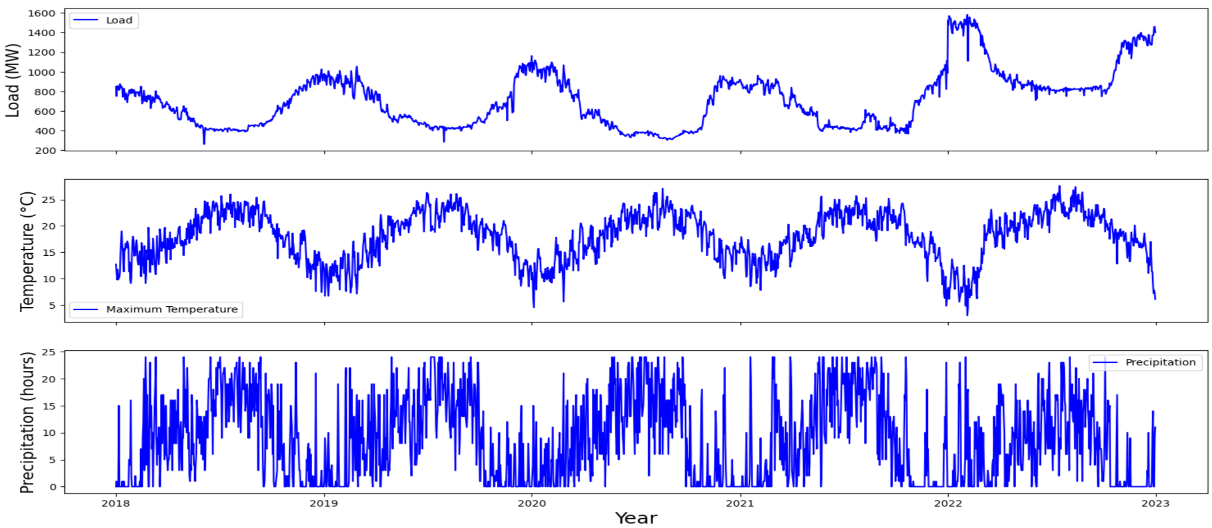

Fig.1: *Daily aggregated load consumption, maximum temperature, and precipitation duration in Thimphu from 2018-2022.*

To develop accurate and robust forecasting models, one must obtain endogenous variables inherent to the load forecasting system and exogenous variables that influence load consumption. In this research, we obtained time series historical load consumption data (endogenous variables) from the Bhutan Power System Operator (BPSO), Thimphu, Bhutan, for the period of five years from January 2018 to December 2022 for both Thimphu and Phuentsholing. All these time series data were recorded with hourly granularity in monthly Excel files, consisting of a total of 43,824 timestamps each for Thimphu and Phuentsholing. Regarding exogenous variables such as weather data, we initially obtained this data from the National Center for Hydrology and Meteorology Department, Royal Government of

time of the day, season, weekend, national and regional holidays, and working days.

3.2 Exploratory Data Analysis and Feature Engineering

The historical load consumption data at an hourly resolution, available in monthly Excel files, were loaded and converted to Data Frames using the pandas library. These data frames were then checked for missing values and outliers. In both regions, certain timestamps showed zero load consumption values. Upon verification with the Bhutan Power Corporation (BPC) Annual Report 2020, it was found that these instances were due to power blackouts and faulty recording equipment. However, in practical scenarios, zero load consumption is unrealistic for forecasting purposes. Therefore, we replaced the zero load consumption values with the average load consumption value.

For aggregated daily load consumption forecasting, we resampled the hourly resolution load values to a daily load consumption by summing up load values at 24-hour timestamps. Consequently, from the total of 43,824 hourly resolution load data records, the daily aggregated load was obtained as 1,826 samples for further analysis.

Regarding the weather data, there were no missing values in either the daily recorded data or the hourly recorded dataset. For other exogenous variables such as day of the week and time of the day, these features were derived from the time stamp of the historical load data. Monday is encoded as day 1, Tuesday as day 2 and so on. As for the seasonal feature, we have encoded winter month as 0, spring as 1, summer as 3 and autumn as 4. The holiday data which includes weekends, national holidays and local festivals for both the regions were manually recorded and the data is then transformed using a binary indicator where we assigned a binary value of 1 if the day is a working day and 0 if the day is a holiday.

Fig.1 depicts the daily aggregated load consumption, the maximum temperature recorded, and the duration of precipitation in hours for the Thimphu region. Upon examining the load consumption data, it is evident that consumption rises drastically during the winter season in Thimphu. This increase is due to the widespread use of heating systems in residential and office buildings in response to the cold, harsh weather. In contrast, the trend in Phuentsholing is different, with load consumption rising during the summer months. This increase is attributed to higher temperatures and the use of cooling systems throughout the region. Additionally, a notable pattern is observed in both regions: temperature rises during the monsoon season, accompanied by an increase in the duration of daily precipitation. For hourly load consumption, Thimphu showed a clear seasonal trend, whereas in Phuentsholing, the trend appears to be more stable, with a slight increase during the summer months. Additionally, the variation in load consumption patterns between weekdays and weekends was studied. It was evident that load consumption was relatively higher during weekdays compared to weekends. From Fig.2, the electrical load demand in Thimphu for the week of December 5-11, 2022, shows that the load demand on Saturday and Sunday is lower than on weekdays. This could be attributed to the closure of government offices. A similar trend is observed for Phuentsholing.

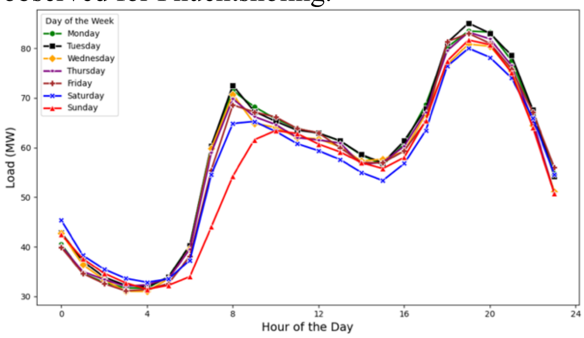

Fig.3: *Daily Electrical Load Demand for One Week in Thimphu (05-11 December 2022)*

Through box-plot analysis, Fig.3 reveals the hourly distribution of load consumption for a typical year in Phuentsholing. It demonstrates that the load consumption pattern varies throughout the day, with noticeable peak hours around the 10th hour until noon, at the 12th hour, and again in the evening between the 18th hour (6 PM) and the 21st hour (9 PM). Moreover, it was also discovered that load consumption patterns differ between regular working days and during public holidays and weekends.

A correlation heatmap is an essential tool in data exploration and analysis, providing a visual

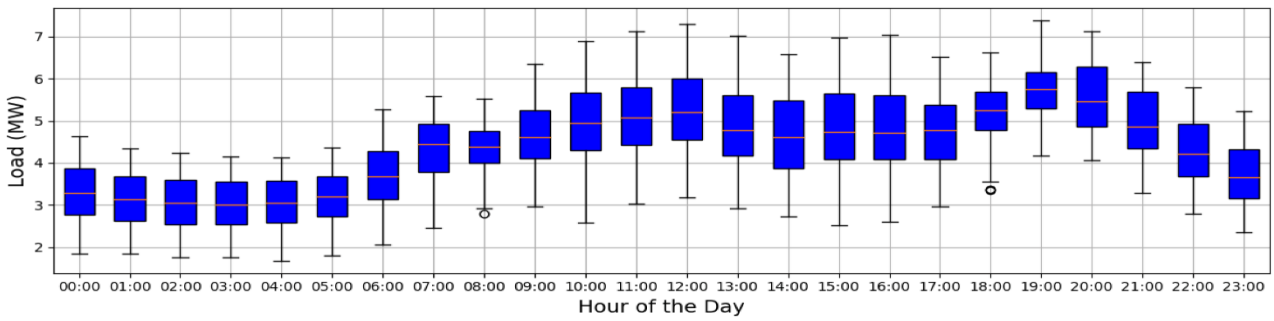

Fig.2: *Box Plot Analysis of Hourly Load Consumption in Phuentsholing in 2022.*

representation of the relationship between multiple variables. It uses colour gradients to indicate the strength of correlations between features, making it easy to identify patterns and dependencies. For instance, as shown in Fig.4, the load consumption of the same hour over the previous three days (Load_T-24, Load_T-48, Load_T-72) and the previous day's 24-hour average load (AvgLoad_T-24) exhibit high correlations of 0.98, 0.97, 0.97, and 0.75, respectively, with the current load demand. Moreover, a moderate relationship with temperature and relative humidity can be observed.

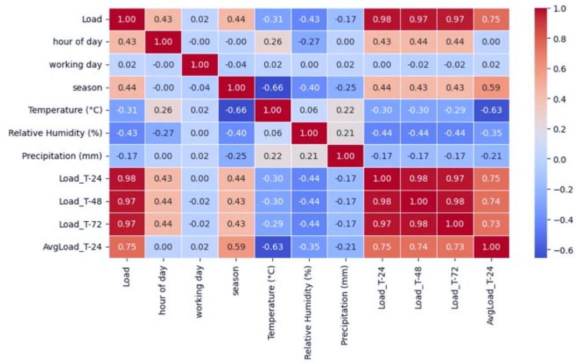

Fig.4: *Correlation Heatmap.*

In designing an accurate time series forecasting model, selecting important and relevant features is crucial. Feature engineering involves using domain knowledge to create features that enhance ML and DL algorithms, which is essential for improving model performance. It transforms raw data into meaningful inputs, capturing underlying patterns and relationships more effectively. To identify the most important features of our STLF models, we applied the random permutation technique and obtained relevant features for model training.

3.3 Overview of Machine Learning and Deep Learning Models

Seasonal Auto Regressive Integrated Moving Averages with Exogenous Regressors (SARIMAX)

The SARIMAX model is a powerful tool in time series analysis, particularly for forecasting tasks where the data exhibits seasonal patterns and external factors, such as calendar effects and weather-related data, influence the time series. SARIMAX extends the traditional ARIMA model by incorporating additional explanatory variables that enhance the model's forecasting accuracy. By integrating both autoregressive (AR) and moving average (MA) components with seasonality and exogenous variables, SARIMAX provides a robust framework for analyzing and forecasting time series data. The mathematical formulation of the model is:

$$y_t = \beta_t x_t + u_t, \quad (1)$$

$$\varphi_p(L)\tilde{\phi}_p(L^s)\Delta^d\Delta_s^D u_t = A(t) + \theta_q(L)\tilde{\theta}_Q(L^s)\zeta_t, \quad (2)$$

where the β in equation (1) represents the external input variables and y_t is the forecast variable. The remaining hyperparameters of the model are: p for the AR order, q for the MA order, I for the differencing order, P for the seasonal AR order, Q for the seasonal MA order, D for the seasonal differencing order, and s for the seasonal coefficients.

Since SARIMAX is one of the classical statistical models widely used in time series forecasting tasks, we chose this model as the baseline for evaluating the performance of more advanced ML and DL models.

Support Vector Regression (SVR)

Support Vector Regression (SVR) is an extension of the classification model Support Vector Machines introduced by Vapnik and colleagues in 1998. SVR finds a function that each prediction y deviates from the target value by no more than ϵ (Smola and Schölkopf, 2004). This approach helps to achieve robust regression by minimizing the margin of error within a specified tolerance level. The mathematical model is:

$$y = \boldsymbol{w}^T\phi(\boldsymbol{x}) + b,$$

where the weights $\boldsymbol{w}$ and the bias term b are obtained by solving the convex optimization problem

$$\min_{\boldsymbol{w},b,\xi_i,\xi_i^*} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$

Subject to

$$y_i - \boldsymbol{w}^T\phi(x_i) - b \le \epsilon + \xi_i^* \quad (i = 1,2,\dots,N)$$
$$\boldsymbol{w}^T\phi(x_i) - y_i + b \le \epsilon + \xi_i \quad (i = 1,2,\dots,N)$$
$$\xi_i, \xi_i^* \ge 0, \quad (i = 1,2,\dots,N)$$

$\phi(\cdot)$ is the transformation of the training data from the input space to a higher-dimensional kernel space, which enables SVR to capture complex patterns and trends that linear models might miss.

Long Short-Term Memory (LSTM) Network

LSTM networks are a specialized type of RNN that excels in capturing long-term dependencies in sequential data. They are particularly useful for time series forecasting, where understanding long-range temporal patterns is crucial. LSTM networks address the vanishing gradient problem, which are common in traditional RNNs, by using a unique architecture that includes a memory cell and three gating mechanisms, namely, the input gate, forget gate, and output gate. These gates control the flow of information, enabling the network to retain or discard information as needed over long sequences.

The mathematical architecture of LSTM is as follows:

Forget Gate: Decides what information from the previous cell state should be discarded.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Input Gate: Determines which new information will be stored in the cell state.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Cell State Update: Updates the cell state using the input and forget gates.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Output Gate: Determines the output from the current cell state.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

Here x_t represents the input at time t, h_{t-1} is the hidden state from the previous time step, σ is the activation function, tanh is the hyperbolic tangent function, W and b are the weights and biases learned during the training process of the network.

Convolutional Neural Networks (CNN)

CNNs have shown significant promise in the realm of load forecasting and time series forecasting. Originally developed for image processing tasks, CNNs excel at capturing spatial hierarchies in data. In the context of time series, this translates to effectively identifying temporal patterns and trends. The key to CNNs' effectiveness lies in their convolutional layers, which apply filters across the input data to extract relevant features. For load forecasting, these filters can uncover patterns such as daily or seasonal variations in electricity usage, allowing the model to make accurate predictions based on historical data.

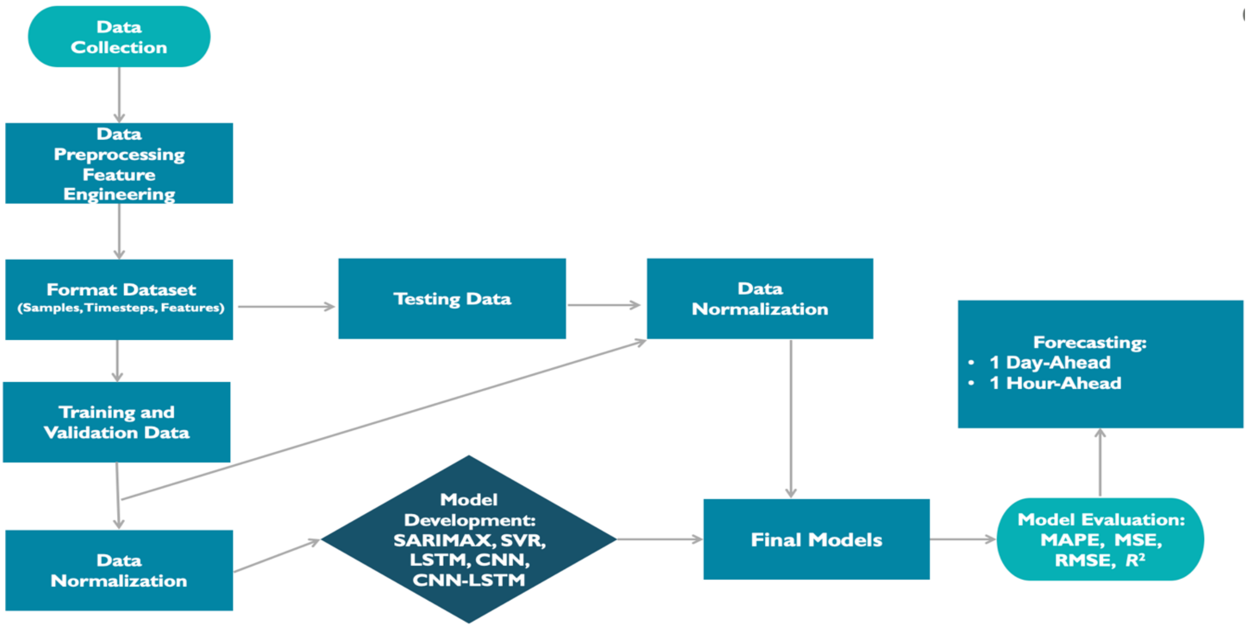

Fig.5: *Flowchart of the Proposed Study*

Hybrid CNN-LSTM

Hybrid CNN-LSTM models combine the strengths of both architectures to leverage the spatial feature extraction capabilities of CNNs and the temporal sequence learning abilities of LSTMs. This combination is particularly effective in capturing complex patterns in STLF, where both short-term fluctuations and long-term trends are important. In this model, CNN layers are first applied to the input data to extract local and hierarchical features. These features are then fed into LSTM layers, which capture the sequential dependencies over time. The convolutional layers focus on identifying significant patterns in the data, while the LSTM layers handle the temporal dependencies and long-term trends. This synergy allows the model to predict future load demand more accurately, considering both immediate changes and overall trends. The hybrid approach not only improves the forecasting accuracy but also enhances the model's ability to generalize across different scenarios and datasets.

3.4 Train/Validation/Test Split

In time series forecasting, the sequential nature of the data requires careful handling to ensure the model's ability to generalize to unseen data. In our study, we divided the single-step forecasting models into two types: one-day-ahead total power demand forecasting and one-hour-ahead load demand forecasting using an hourly dataset. In the single-step forecast of daily aggregated load demand, we split the dataset into 80%-10%-10% for training, validation, and testing, respectively. The data from January 1, 2018, to December 31, 2021, was used for training. The model was then validated on data from January 1, 2022, to June 30, 2022, and tested on data from July 1, 2022, to December 31, 2022. For the hourly resolution load dataset, we split the data into 85%-7.5%-7.5% for training, validation, and testing. Out of the total 43,824 samples, 37,250 timestamps were used for training, while 3,285 samples each were consecutively used for validation and testing. Fig.5 presents the overall framework of the proposed study, and the flow is consistent for developing forecasting models in both Thimphu and Phuentsholing.

3.5 Evaluation Metrics

Accurate evaluation metrics are essential in forecasting models to assess their predictive capabilities and ensure reliable performance. In each of the following evaluation metric, y_i refers to the $i-$th true value, $\hat{y}_i$ is the $i-$th predicted value, $\bar{y}$ is the mean value of true values in the test samples and n represents the total number test dataset.

- **Mean Absolute Percentage Error (MAPE)**: A relative measure of error between forecasted and actual values, expressed as a percentage, indicating the average magnitude of forecasting error. Lower MAPE values denote better forecasting accuracy.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100$$

- **Mean Squared Error (MSE)**: An average of the squared differences between forecasted and actual values, providing insight into the average squared magnitude of forecasting error.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- **Root Mean Squared Error (RMSE)**: A measure of the difference between forecasted and actual values, expressed in the same unit as the load variable. Smaller RMSE values indicate higher forecasting accuracy.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

- **Correlation Squared (R^2)**: Indicates the proportion of variance in the dependent variable (actual energy consumption) explained by the independent variable (forecasted energy consumption). R^2 values range between 0 and 1, with higher values indicating better fit between forecasted and actual values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

4. RESULTS AND DISCUSSIONS

We present the results in two sections: 4.1 Daily aggregated load forecasting, and 4.2 An hour-ahead load forecasting for both the regions.

4.1 Daily Aggregated 1-Day-Ahead Load Forecasting in Thimphu

Feature Selection

Effective model performance hinges on the careful selection of relevant features. In our study, feature selection was conducted individually for

Table 1: *1-Day-Ahead Single-Step Daily Aggregated Load Forecasting in Thimphu.*

Model	MAPE (%)	MSE (MW^2)	RMSE (MW)	R^2
SARIMAX	3.238	1647.061	40.584	0.969
SVR	2.401	1197.457	34.604	0.978
LSTM	3.101 ± 0.555	1550.863 ± 23.542	39.381 ± 4.852	0.971 ± 0.007
CNN	2.644 ± 0.186	1277.348 ± 1.578	35.74 ± 1.256	0.976 ± 0.002
CNN-LSTM	$\mathbf{2.332 \pm 0.075}$	$\mathbf{1110.889 \pm 0.272}$	$\mathbf{33.33 \pm 0.522}$	$\mathbf{0.98 \pm 0.001}$

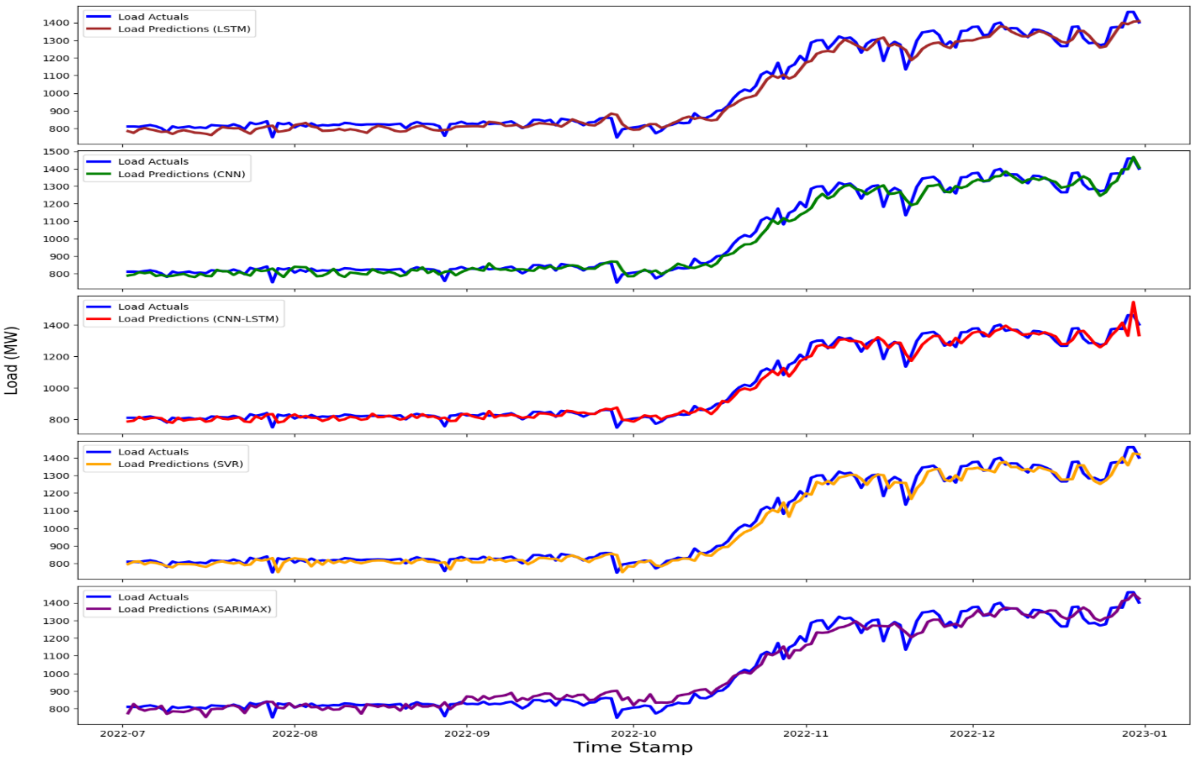

Fig.6: *Comparative analysis of single-step (1-day-ahead) predictions by ML and DL models in Thimphu from 1 July to 31 December 2022.*

each model, employing feature engineering through random permutation techniques for feature ranking. Subsequently, the least important features were discarded to enhance model efficacy.

For the SARIMAX and SVR models, 13 features were ultimately selected. These features cincluded the day of the week, working day, season, maximum temperature, snowfall depth, precipitation duration for the day, and the previous 7 days' lagged load consumption values.

In the case of the LSTM model, a look-back window size of 7 days was utilized to predict the next day's load consumption, incorporating 7 features: load, day of the week, working day, season, maximum temperature, snowfall depth,

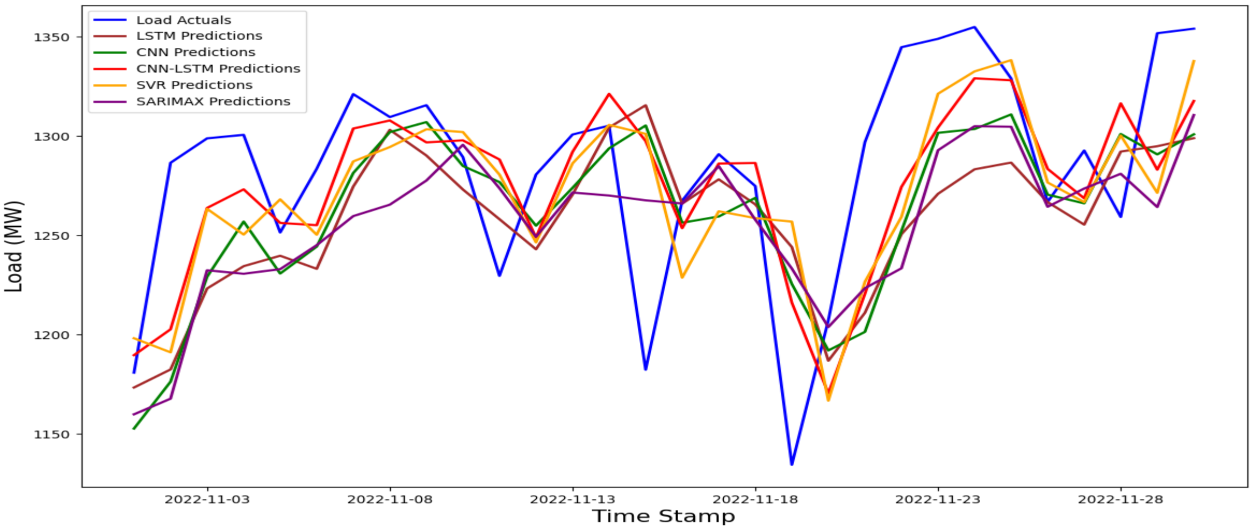

Fig.7: *Comparative analysis of single-step (1-day-ahead) predictions for the month of November 2022 in Thimphu using ML and DL models.*

and precipitation duration. Similarly, the CNN model employed the same look-back window size of 7 days and the same 7 features as the LSTM model. For the CNN-LSTM model, an expanded look-back window size of 8 days was adopted. This model included all the features used in the LSTM and CNN models, with the addition of minimum temperature, precipitation amount, and wind speed.

Parameter Selection

The forecasting models in our study rely significantly on the parameters selected in advance. For the SARIMAX model, we used the 'auto_arima' function from the 'statsmodels' library in Python. For the SVR model, we employed the Grid Search method, and for the deep learning models, Bayesian Optimization was utilized. While training SVR models, the RBF kernel showed superior performance compared to linear and polynomial kernels. Therefore, the RBF kernel was used throughout. This kernel is of the form

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

Where σ is the kernel parameter.

Moreover, since the features in the dataset that we have obtained have varying numerical ranges, they were normalized to the range of [0, 1] using the following min-max scaling formula:

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

where x_i is the $i-$th data, z_i is the $i-$th normalized data, $\min(x)$ and $\max(x)$ are the minimum and maximum values of the x feature.

Results

***Table 2:** 1-Day-Ahead Single-Step Daily Aggregated Load Forecasting in Phuentsholing.*

Model	MAPE (%)	MSE (MW^2)	RMSE (MW)	R^2
SARIMAX	3.802	32.004	5.657	0.829
SVR	3.619	28.564	5.344	0.848
LSTM	3.559 ± 0.1	28.998 ± 0.021	5.385 ± 0.144	0.846 ± 0.008
CNN	3.296 ± 0.021	26.061 ± 0.002	5.105 ± 0.045	0.861 ± 0.002
CNN-LSTM	$\mathbf{3.216 \pm 0.036}$	$\mathbf{25.18 \pm 0.0004}$	$\mathbf{5.018 \pm 0.019}$	$\mathbf{0.866 \pm 0.001}$

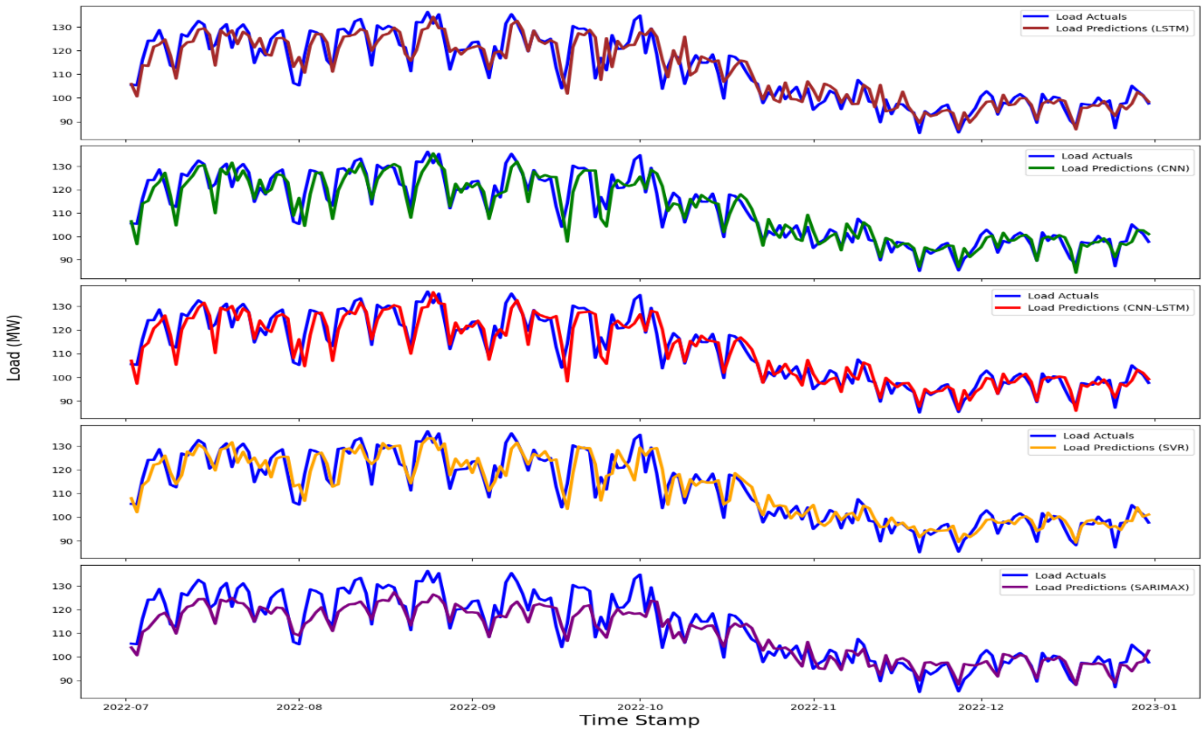

***Fig.8:** Comparative analysis of single-step (1-day-ahead) forecasting using ML and DL models in Phuentsholing from 01 July to 31 December 2022.*

For 1-day-ahead single-step daily aggregated load forecasting, we have employed all five models described previously with the evaluation metrics also outlined earlier. The overall summary of the results is tabulated in Table 1. For DL models, due to the random initialization of the weights before training, even though we use the optimized set of hyperparameters, the results vary slightly over different runs. Therefore, the results we report in the table are the average values over five runs along with the standard deviation.

From the tabulated results in Table 1, it is revealed that the hybrid CNN-LSTM model significantly outperforms the other models. It achieves the lowest MAPE at $2.332 \pm 0.075\%$, indicating its superior accuracy in forecasting. This validates the high accuracy of the model, as the MAPE is well below 10% (Lewis, 1984). In comparison, the SARIMAX model performs worst with the highest MAPE value, while the SVR, LSTM, and CNN models all show intermediate performance. The CNN-LSTM model also excels in terms of MSE, RMSE, and R^2, with values of $1110.889 \pm 0.271\ \mathrm{MW}^2, 33.33 \pm 0.522\ \mathrm{MW}$, and 0.98 ± 0.001 respectively, reinforce its overall effectiveness. This highlights the robustness of the hybrid CNN-LSTM model for this forecasting task. Moreover, the prediction capabilities of the models are demonstrated in Fig.6 and Fig.7.

Daily Aggregated 1-Day-Ahead Load Forecasting in Phuentsholing

Parameter selection was done similarly to Thimphu's case above.

Feature Selection

For the SARIMAX and SVR models, we selected 10 optimal features using random permutation feature engineering techniques. These features include the previous seven days' lagged loads, day of the week, working day indicator, and maximum temperature. For the DL models LSTM and CNN, we used a lookback window size of the past seven days of load consumption. In contrast, for the CNN-LSTM model, we employed an even-numbered lookback window size of eight days, allowing the generation of two subsequences during model training. Regarding features, the LSTM and CNN models utilized four features: load, day of the week, working day, and maximum temperature. In CNN-LSTM models, however, optimal performance was achieved using six features: load, day of the week, working day, minimum temperature, duration of sunshine, and precipitation.

Results

Table 2 shows the one-day-ahead load forecasting results for the Phuentsholing region, based on daily total electricity demand. It is evident that the hybrid CNN-LSTM model once again outperforms all other models, achieving the lowest MAPE at $3.216 \pm 0.036\%$. The MSE, RMSE, and R^2 values are also superior to those of the other models, with the CNN model performing closely behind with a MAPE of $3.296 \pm 0.021\%$. The baseline SARIMAX model has the lowest performance, while the SVR model performs slightly better. As observed, the ML and DL models significantly outperform the classical baseline method, suggesting their suitability for these forecasting tasks. Each model's prediction is also illustrated in Fig.8 and Fig.9.

Table 3: *1-Hour-Ahead Single-Step Load Forecasting in Thimphu.*

Model	MAPE (%)	MSE $(\mathbf{MW^2})$	RMSE (MW)	$\boldsymbol{R^2}$
LSTM	4.137 ± 0.241	6.477 ± 0.049	2.545 ± 0.222	0.974 ± 0.005
CNN	$\mathbf{3.224 \pm 0.06}$	$\mathbf{3.928 \pm 0.0006}$	$\mathbf{1.982 \pm 0.025}$	$\mathbf{0.984 \pm 0.0}$
CNN-LSTM	3.917 ± 0.119	5.063 ± 0.004	2.25 ± 0.066	0.98 ± 0.001

1-Hour-Ahead Load Forecasting

1-Hour-Ahead Load Forecasting in Thimphu

Model Selection

For 1-hour-ahead forecasting, we employed only the DL models: LSTM, CNN, and CNN-LSTM, as the dataset is large and training time is complex for SARIMAX and SVR models. The advantage of using DL is more pronounced when dealing with large datasets, as the computational time is not excessively long. As the features in our dataset have varying numerical ranges, we standardized them to have zero mean and unit standard deviation using the following formula:

$$z_i = \frac{x_i - \mu_i(x)}{\sigma_i(x)}$$

Where x_i is the $i-$th input data, z_i is the $i-$th standardized data, $\mu\_i(x)$ and $\sigma_i(x)$ are the mean and standard deviation of the feature x.

Feature Selection

For all DL models, our experiments revealed that a lookback window size of the past 48 hours provides optimal results for predicting the load

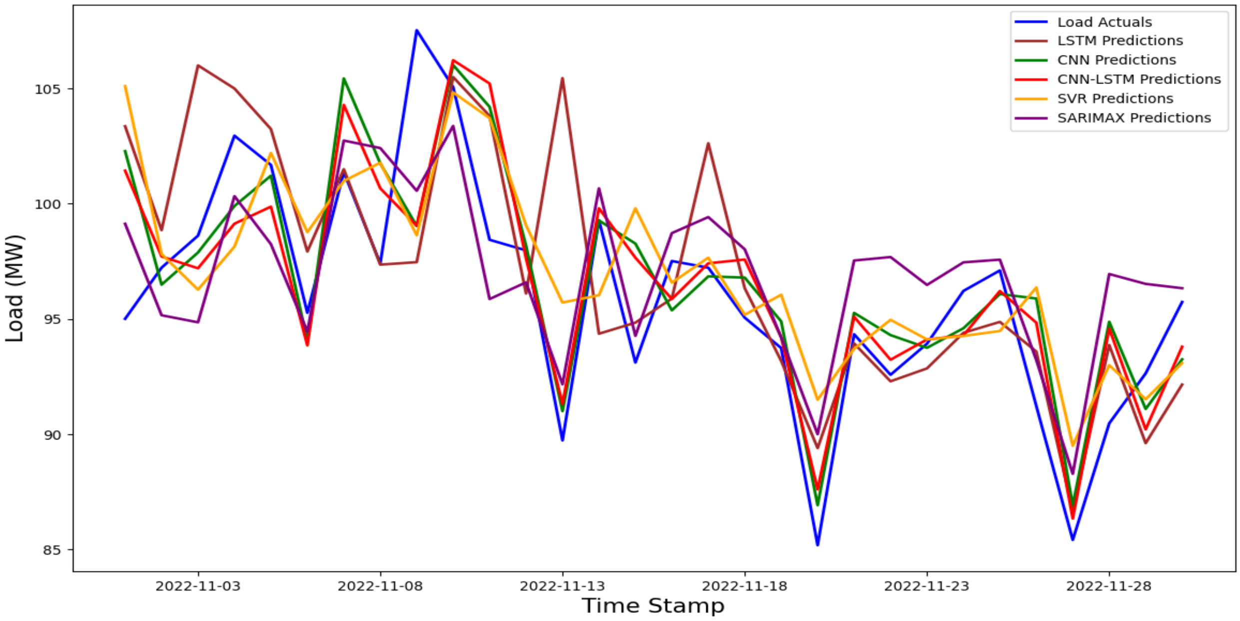

Fig.9: *Comparative analysis of step-ahead (1-day-ahead) load forecasting in Phuentsholing for the month of November 2022 using ML and DL models.*

Table 4 *1-Hour-Ahead Single-Step Load Forecasting in Phuentsholing.*

Model	MAPE (%)	MSE (MW^2)	RMSE (MW)	R^2
LSTM	4.02 ± 0.108	0.076 ± 0.00005	0.275 ± 0.007	0.939 ± 0.003
CNN	3.746 ± 0.036	$\mathbf{0.065 \pm 0.00}$	$\mathbf{0.254 \pm 0.002}$	$\mathbf{0.947 \pm 0.001}$
CNN-LSTM	$\mathbf{3.687 \pm 0.027}$	$\mathbf{0.065 \pm 0.00}$	0.255 ± 0.002	$\mathbf{0.947 \pm 0.001}$

one hour into the future. Using random permutation feature importance ratings, we identified nine features that result in effective model performance. These features are load, previous day's same hour load, two days prior same hour load, three days prior same hour load, previous 24 hours average load consumption, hour of the day, working day, temperature at 2 meters above ground, and apparent temperature.

Parameter Selection

A rigorous hyperparameter tuning was done by first searching two rounds for number of hidden layers and units, then searching for learning rate, dropout per cent, dense layers, choosing activation functions, optimizers, and then finally concluded

with the decision on the batch size. Bayesian Optimization techniques could not be done since the search took excessive time to execute.

Results

The 1-hour-ahead load forecasting in the Thimphu region shows that among the three DL models utilized, the CNN model significantly outperforms the LSTM and the hybrid CNN-LSTM model with the lowest MAPE value of $3.224 \pm 0.06\%$, an RMSE value of 1.982 ± 0.025 MW, and the highest R^2 value of 0.984. The hybrid CNN-LSTM model performs fairly with a MAPE of $3.917 \pm 0.119\%$, while the lowest performance was recorded with the LSTM, which has a MAPE of $4.137 \pm 0.241\%$. The overall results are shown in Table 3. Additionally, the results are obtained as the average of 5 different runs using the optimized set of hyperparameters. The accuracies of the forecasting models are seen on two randomly selected unseen test data as illustrated in Fig.10.

1-Hour-Ahead Load Forecasting in Phuentsholing

Similar to the 1-hour-ahead forecasting in Thimphu, we only use the DL models here as well.

Feature Selection

Using the same feature engineering techniques, it was identified that a total of 8 features, including load, previous day's same-hour load, two days prior same-hour load, three days prior same-hour load, past 24 hours average load, hour of the day,

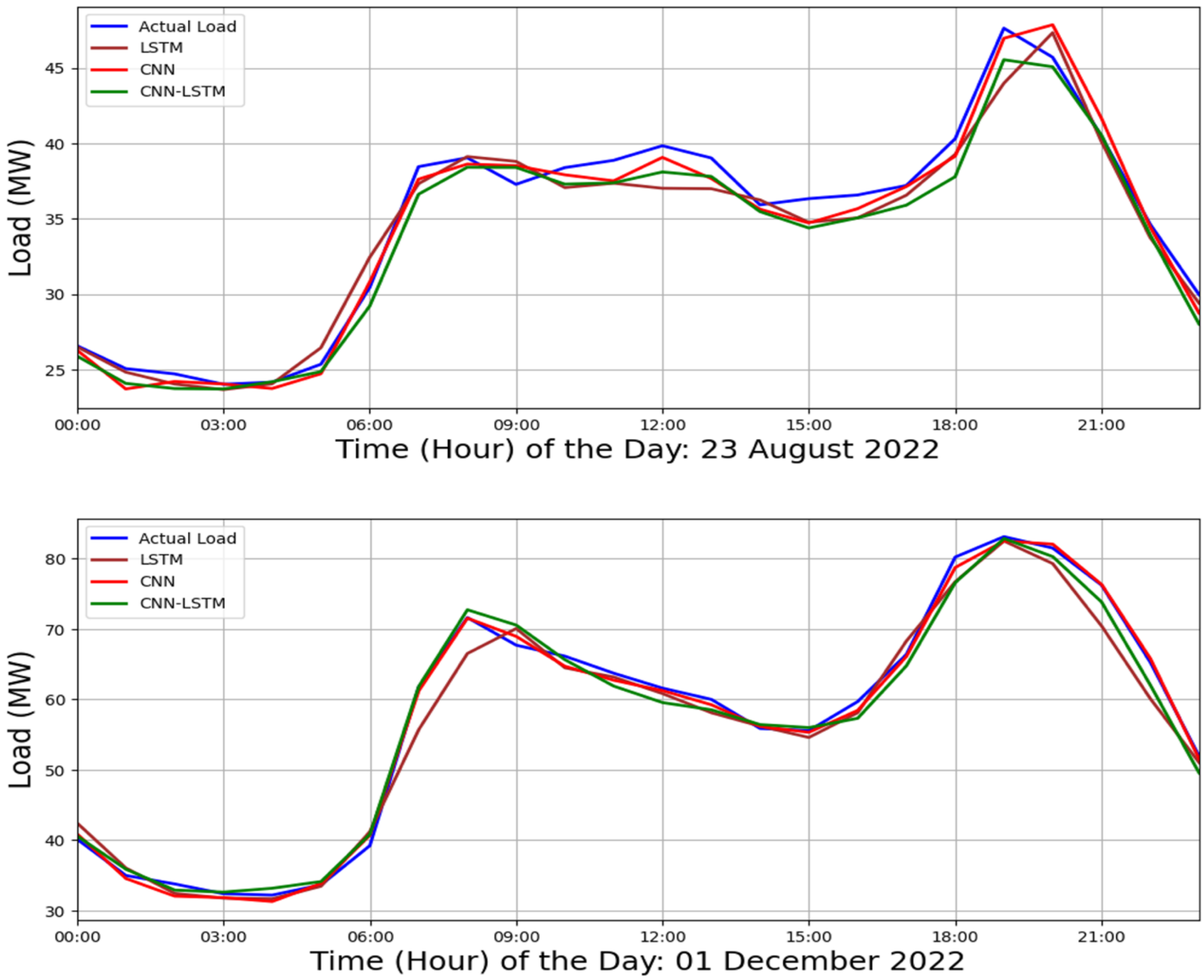

***Fig.10** Predicted 1-hour-ahead electrical load in Thimphu on two randomly chosen days using DL models.*

working day, and temperature, provided the best model performance.

Parameter Selection

The hyperparameter tuning was also rigorously conducted using the same procedure as in the case of model development in Thimphu.

The summary of the forecasting accuracy is reflected in Table 4.

Clearly, it is observed that the hybrid CNN-LSTM model outperforms the other two models in terms of MAPE value, achieving the lowest at $3.687 \pm 0.027\%$. The CNN model also performs well. In terms of MSE and R^2 values, both CNN and CNN-LSTM models show similar effectiveness. Additionally, CNN achieves the lowest RMSE value at 0.254 ± 0.002 MW, slightly lower than that of the CNN-LSTM model. Overall, CNN and CNN-LSTM models perform equally well, while the LSTM model lags in all evaluation metrics. These model performances can also be compared on two randomly chosen days, as shown in Fig.11.

Discussions

In the daily aggregated load forecasting for 1-day-ahead predictions in Thimphu and Phuentsholing regions, studied separately, ML and DL models demonstrated superior performance compared to the baseline classical model SARIMAX. Among the DL models, it is evident that the LSTM model did not exhibit any significant advantage over the CNN and CNN-LSTM models but performed closely behind. However, the hybrid CNN-LSTM model, which integrates CNN layers to extract important features before feeding them into LSTM layers, achieved the best performance in both regions across all evaluation metrics. Particularly noteworthy were the lowest MAPE scores of 2.332% and 3.216% achieved in Thimphu and Phuentsholing, respectively. Additionally, the SVR model showed remarkable performance and was compared favourably to the deep learning models.

Conversely, for the 1-hour-ahead forecasting task in Thimphu, it was observed that

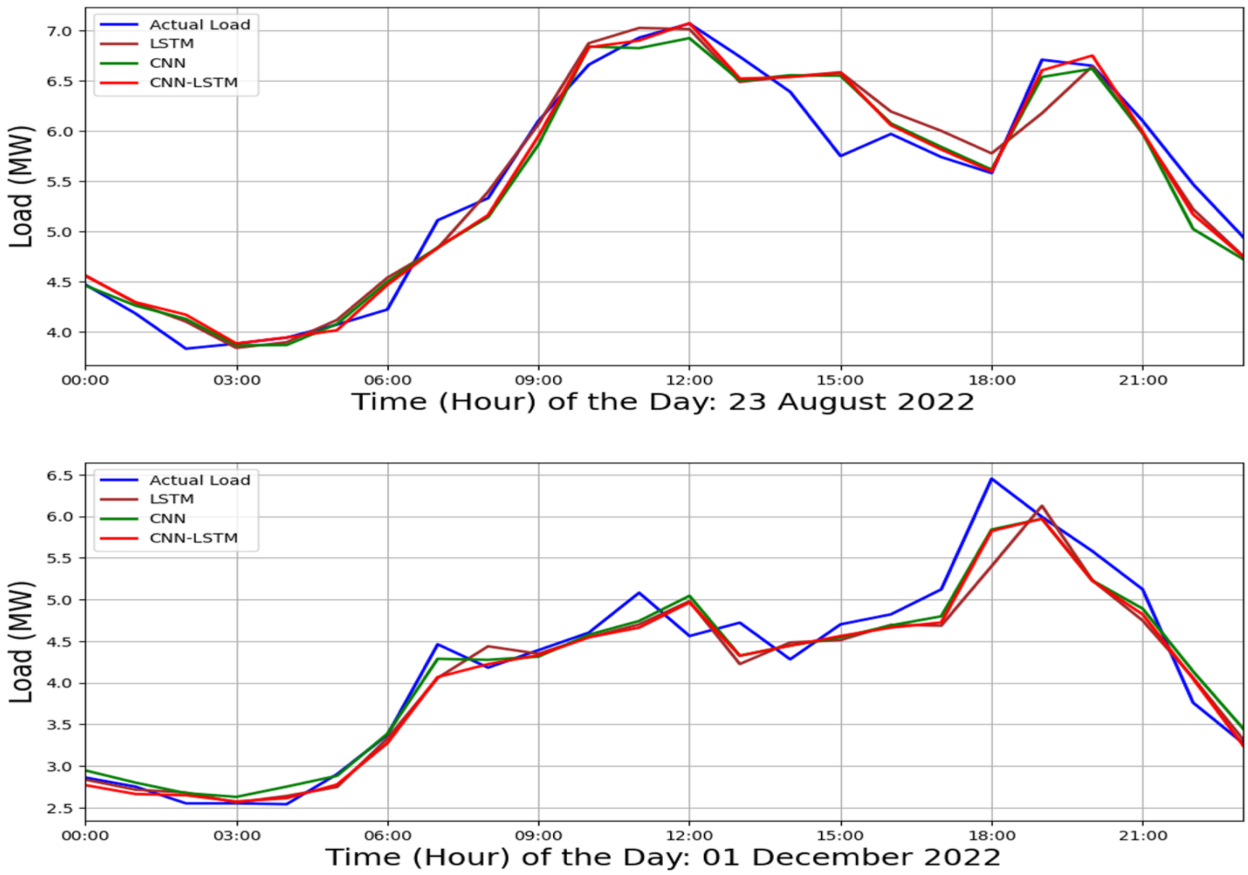

Fig.11 *Predicted 1-hour-ahead electrical load using DL models in Phuentsholing.*

the CNN model outperformed significantly in all evaluation metrics, a turn from the hybrid CNN-LSTM model's previous leadership in the performance board. The hybrid model maintained a slight advantage over the LSTM model. However, in the forecasting results for the Phuentsholing region, both the hybrid CNN-LSTM model and the CNN model demonstrated effective performance, while the LSTM model consistently lagged across all evaluation metrics.

Regarding the training and testing time complexity of these deep learning models, CNN exhibited the shortest time, followed by the hybrid CNN-LSTM, with LSTM requiring the longest duration for model training and testing. Therefore, for large datasets, the CNN model is highly recommended based on evaluation metrics and time efficiency perspectives.

An observable increase in evaluation metrics such as MAPE, MSE, and RMSE, alongside slightly reduced R^2 values in single-step 1-hour-ahead forecasting for both Thimphu and Phuentsholing regions compared to daily aggregated load demand forecasting, could be attributed to dataset size differences. Specifically, the 1-hour-ahead forecasting involved significantly more samples (43,826) compared to the aggregated demand samples (1,826). The complexity of model training time and the search for optimal hyperparameters also significantly influenced identifying the best model performances. In daily aggregated load demand forecasting, training was relatively straightforward, and hyperparameter tuning included advanced Bayesian optimization following careful manual adjustments. However, for 1-hour-ahead forecasting, all hyperparameter tuning had to be conducted manually due to the excessive time required for Bayesian optimization.

5. CONCLUSIONS

In this research, we have presented short-term load forecasting for Thimphu and Phuentsholing regions using machine learning and deep learning models, with the classical seasonal autoregressive integrated moving averages with explanatory variables method as the baseline model for performance comparisons.

In the first part, various popular techniques used for short-term electrical load forecasting were reviewed and experimented on the single-step (1-day-ahead) daily aggregated load demand in Thimphu and Phuentsholing regions. Thorough exploratory data analysis was also conducted, leading to the creation of temporal features, including calendar effect features that impact load consumption. Through random permutation feature engineering techniques, optimal and relevant features for each model were determined and evaluated.

Among the models employed for 1-day-ahead aggregated load forecasting, the hybrid convolutional neural network and long short-term memory network significantly outperformed all other models, achieving the lowest MAPE of 2.332% and 3.21% respectively in the Thimphu and Phuentsholing regions, while also maintaining the best values of MSE, RMSE, and R^2. This hybrid model used seven optimal features with a lookback window size of the previous seven days.

For 1-hour-ahead load forecasting, which involved a large dataset for training and testing, only deep learning models were evaluated due to their effectiveness in dealing with big data. The convolutional neural network had an edge over other models in the Thimphu region, achieving the lowest MAPE score of 3.224% using a lookback window of 24 hours with nine optimal features. In the Phuentsholing region, both the convolutional neural network model and the hybrid model showed almost equal performances, with the hybrid model recording the lowest MAPE of 3.687%, closely followed by the convolutional neural network.

Additionally, the convolutional neural networks are executed in the shortest span of time, making them especially suitable for big data applications.

The performance of the deep learning models could be further improved with more advanced computational resources and by deploying more effective hyperparameter optimization algorithms. It would be beneficial to explore hyperparameter tuning methods such as the tree-structured Parzen estimator.

However, this research has several limitations. The study was conducted only in the specific regions of Thimphu and Phuentsholing and does not represent the entire nation's picture of demand. Also, we have conducted only a forecast horizon of single-step load demand.

In the future, this research could be extended to other regions and the entire country for short-term load demand forecasting. In particular, multi-step forecasting such as 24-hours ahead into the future could also be taken up.

6. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Royal University of Bhutan for awarding the Annual University Research Grant (2023–2024), without which this study would not have been possible. We extend our sincere appreciation to the Bhutan Power System Operator for providing the historical load data and to the National Center for Hydrology and Meteorology for supplying the weather-related data, both of which were indispensable to this research.

REFERENCES

Al-Ja'afreh, M. A. A., Mokryani, G., & Amjad, B. (2023). An enhanced CNN-LSTM based multi-stage framework for PV and load short-term forecasting: DSO scenarios. *Energy Reports*, *10*, 1387–1408. https://doi.org/10.1016/j.egyr.2023.08.003

Almalaq, A., & Edwards, G. (2017). A Review of Deep Learning Methods Applied on Load Forecasting. *16th IEEE International Conference on Machine Learning and Applications (ICMLA),* 511-516. https://doi.org/10.1109/ICMLA.2017.0-110

BPSO. (2023). *Annual Transmission System Performance Report for the Year 2022 January-2023*. Bhutan Power System Operator. https://www.bpso.bt//uploads/reports/annualreport/Annual-Report-2023.pdf

Dema, N. (2024). *Prime Minister reviews ministries' proposals for 13th Five-Year Plan - BBSCL*. Bhutan Boardcasting Service. https://www.bbs.bt/news/?p=199347

Kainth, G. S. & Mehra, P. L. (1988). *Seasonality pattern of market arrival and prices of potato in Punjab.* Indian Journal of Agricultural Marketing. Vol. 2 , pp. 113-120.

Miles, J. W. (1957). *On the generation of surface waves by shear flows*, Journal of Fluid Mech., Vol. 3, pp. 185-204.

He, W. (2017). Load Forecasting via Deep Neural Networks. *Procedia Computer Science, 122*, 308-314. https://doi.org/10.1016/j.procs.2017.11.374

Ibrahim, B., Rabelo, Luis, Gutierrez-Franco, E., & Clavijo-Buritica, N. (2022). Machine Learning For Short-Term Load Forecasting in Smart Grids. *Energies, 15*(21). 10.3390/en15218079

Lewis, C.D. (1982), *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*, London, England; Boston, MA:Butterworth Scientific.

Lhaden, Y. (2023). *Sephu hosts Bhutan's first mega solar power plant*. Kuenselonline. https://kuenselonline.com/sephu-hosts-bhutans-first-mega-solar-power-plant/

Madrid, E. A., & Antonio, N. Short-Term Electricity Load Forecasting with Machine Learning. *Information 2021, 12*(12), 50. 10.3390/info12020050

Phyo, P. P., Jeenanunta, C., & Hashimoto, K. (2019). Electricity Load Forecasting in Thailand Using Deep Learning Models. *International Journal of Electrical and Electronic Engineering & Telecommunications, 8*(4). 10.18178/ijeetc.8.4.221-225

Ray, B., Lasantha, D., Beeravalli, V., Anwar, A., Nurun Nabi, M., Sheng, H., Rashid, F., & Muyeen, S. M. (2024). A comprehensive framework for effective long-short term solar yield forecasting. *Energy Conversion and Management: X*, *22*(December 2023), 100535. https://doi.org/10.1016/j.ecmx.2024.100535

Shahare, K., Mitra, A., Naware, D., Keshri, R., & Suryawanshi, H.M. (2023). Performance analysis and comparison of various techniques for short-term load forecasting. *Energy Reports, 9*, 799-808. https://doi.org/10.1016/j.egyr.2022.11.086

Vanting, N. B., Ma, Z. & Jørgensen, B. N. (2021). A scoping review of deep neural networks for electric load forecasting. *Energy Informatics, 4*(2), 49. 10.1186/s42162-021-00148-6

Zam, K., Gupta, M. K., & Uddin, S. M. N. (2021). The residential energy futures of Bhutan. *Energy Efficiency, 14*(4), 38. 10.1007/s12053-021-09948-x