Short-term Power Load Forecasting for a 33/11 KV Sub-Station by Utilizing Attention-Based Hybrid Deep Learning Architectures

Mukkamala R., Sunku V.S.,
School of Energy & Clean Technology, NICMAR University of Construction Studies, Hyderabad, Telangana, India

Abstract. Estimating electric power load at substations is a fundamental task for system operators, as it is essential for the reliable and optimal operation of the power system. Effective load forecasting is critical for optimal power generation, as precise predictions facilitate the economical use of electrical infrastructure. The primary objective of this study is to develop advanced deep learning (DL) attention-based models aimed at improving the accuracy of short-term electric power load forecasting at substations. This enhancement is essential for ensuring the reliable and efficient operation of power systems. To accomplish this objective, a comprehensive evaluation of various machine learning (ML) and deep learning (DL) architectures was conducted. This evaluation included the following models: Autoregressive Integrated Moving Average (ARIMA), Multi-Layer Perceptron (MLP), Random Forest (RF), Gradient Boosting (GB), Long Short-Term Memory (LSTM) networks with Attention mechanisms, Double Attention mechanisms, LSTM-Convolutional Neural Network (CNN) Attention, Recurrent Neural Networks (RNN) with Input Attention, Bidirectional LSTM (BiLSTM) with Attention, and CNN-BiLSTM Attention mechanism. These models applied to hourly estimated energy consumption data (in kilowatts) sourced from the 33/11 KV substation in Telangana, India. The performance of these models measured using several key metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2). The most important result is that the CNN-BiLSTM attention model significantly outperforms the other models, achieving an MSE of 0.0079, an RMSE of 0.0889, and an R^2 value of 0.8547. that underscores that the CNN-BiLSTM attention model represents an effective and practical tool for accurate power load forecasting.
Keywords: short-term load, forecasting, deep learning, machine learning, attention-based mechanisms. performance metrics.

UDC:620.91
DOI: https://doi.org/10.52254/1857-0070.2025.3-67.02

Prognoza pe termen scurt a sarcinii electrice pentru o substație de 33/11 KV prin utilizarea arhitecturilor hibride de învățare profundă bazate pe atenție

Mukkamala R., Sunku V.S.,
School of Energy & Clean Technology, NICMAR University of Construction Studies, Hyderabad, Telangana, India.
Rezumat. Estimarea sarcinii energiei electrice la substații este o sarcină fundamentală pentru operatorii de sistem, deoarece este esențială pentru funcționarea fiabilă și optimă a sistemului energetic. Prognoza eficientă a sarcinii este esențială pentru generarea optimă de energie, deoarece predicțiile precise facilitează utilizarea economică a infrastructurii electrice. Obiectivul principal al acestui studiu este de a dezvolta modele avansate de învățare profundă (DL) bazate pe atenție, care vizează îmbunătățirea preciziei prognozei pe termen scurt a sarcinii energiei electrice la substații. Această îmbunătățire este esențială pentru asigurarea funcționării fiabile și eficiente a sistemelor energetice. Pentru a atinge acest obiectiv, a fost efectuată o evaluare cuprinzătoare a diferitelor arhitecturi de învățare automată (ML) și învățare profundă (DL). Această evaluare a inclus următoarele modele: Autoregressive Integrated Moving Average (ARIMA), Multi-Layer Perceptron (MLP), Random Forest (RF), Gradient Boosting (GB), rețele Long-Short-Term Memory (LSTM) cu mecanisme de atenție, LSTM-Convolutional Neural Network (CNN) Attention etc. Aceste modele s-au aplicat datelor de consum de energie estimate orar (în kilowați) provenite de la substația de 33/11 kV din Telangana, India. Performanța acestor modele a fost măsurată utilizând mai mulți indicatori cheie, inclusiv Eroarea Medie Pătratică (MSE), Eroarea Medie Pătratică Root (RMSE) și R-pătrat (R^2). Cel mai important rezultat este că modelul de atenție CNN-BiLSTM depășește semnificativ celelalte modele, atingând un MSE de 0.0079, un RMSE de 0.0889 și o valoare R^2 de 0.8547, ceea ce subliniază faptul că modelul de atenție CNN-BiLSTM reprezintă un instrument eficient și practic pentru prognozarea precisă a sarcinii energetice.

Cuvinte-cheie: sarcină pe termen scurt, prognoză, învățare profundă, învățare automată, mecanisme bazate pe atenție, indicatori de performanță.

Краткосрочное прогнозирование нагрузки на подстанции 33/11 кВ с использованием гибридных архитектур глубокого обучения на основе внимания

Муккамала Р., Сунку В.С.
Факультет энергетики и чистых технологий, Университет строительных исследований NICMAR

Аннотация. Оценка электрической нагрузки на подстанциях является фундаментальной задачей для системных операторов, поскольку она необходима для надежной и оптимальной работы энергосистемы. Эффективное прогнозирование нагрузки критически важно для оптимальной выработки электроэнергии, поскольку точные прогнозы способствуют экономичному использованию электрической инфраструктуры. Основная цель данного исследования — разработка усовершенствованных моделей глубокого обучения (ГО), основанных на анализе внимания, для повышения точности краткосрочного прогнозирования электрической нагрузки на подстанциях. Это усовершенствование необходимо для обеспечения надежной и эффективной работы энергосистем. Для достижения этой цели была проведена комплексная оценка различных архитектур машинного обучения (МО) и глубокого обучения (ГО). Эта оценка включала следующие модели: авторегрессионное интегрированное скользящее среднее (ARIMA), многослойный персептрон (MLP), случайный лес (RF), градиентное усиление (GB), сети с длинной краткосрочной памятью (LSTM) с механизмами внимания, механизмы двойного внимания, LSTM-свёрточная нейронная сеть (CNN) с механизмом внимания, рекуррентные нейронные сети (RNN) с входным вниманием, двунаправленная LSTM (BiLSTM) с вниманием и механизм внимания CNN-BiLSTM. Эти модели применялись к почасовым данным о потреблении энергии (в киловаттах), полученным с подстанции 33/11 кВ в Телангане, Индия. Производительность этих моделей измерялась с использованием нескольких ключевых метрик, включая среднеквадратичную ошибку (MSE), среднеквадратичную ошибку (RMSE) и коэффициент детерминации (R^2). Наиболее важным результатом является то, что модель внимания CNN-BiLSTM значительно превосходит другие модели, достигая среднеквадратичной ошибки (СКО) 0.0079, среднеквадратичной ошибки (СКО) 0.0889 и значения R^2 0.8547. Это подчёркивает, что модель внимания CNN-BiLSTM представляет собой эффективный и практичный инструмент для точного прогнозирования нагрузки. Эта возможность не только обеспечивает экономичное использование электроэнергетической инфраструктуры, но и поддерживает надёжные процессы принятия решений на основе данных в рамках эксплуатации энергосистемы.

Ключевые слова: кратковременная нагрузка, прогнозирование, глубокое обучение, машинное обучение, механизмы, основанные на внимании, показатели производительности.

INTRODUCTION

Power Load forecasting is a fundamental component of efficient power system operation and strategic planning, directly influencing the allocation of energy resources, grid stability, and infrastructure development. The deregulation of energy markets, coupled with the increasing integration of renewable energy sources and evolving consumer behavior, has introduced considerable complexity to load patterns. Traditional statistical models and earlier machine learning (ML) techniques, while useful in straightforward scenarios, often fail to adequately capture the complex temporal, spatial, and contextual relationships inherent in contemporary electricity demand data. This complexity necessitates the adoption of advanced forecasting methodologies to ensure both accuracy and reliability. Recent advancements in DL, particularly through the integration of attention mechanisms, have created significant opportunities for addressing the challenges associated with electricity demand forecasting. LSTM frameworks that integrate attention mechanisms have demonstrated significant effectiveness in identifying critical temporal patterns. This capability enhances prediction accuracy by prioritizing pertinent data and minimizing the influence of irrelevant noise [1]. Double attention frameworks enhance analytical capabilities by simultaneously capturing both temporal and spatial dependencies, making them particularly well-suited for analyzing complex datasets. Furthermore, hybrid models, such as LSTM-CNN with attention, effectively combine the strengths of sequence modeling and convolutional feature extraction to further enhance forecasting performance. RNNs that utilize input attention mechanisms are proficient in prioritizing the most pertinent input variables, thereby enhancing the interpretability and predictive accuracy of the models. The BiLSTM framework, when integrated with attention mechanisms, improves sequential modeling by analyzing data in both forward and backward directions.

This approach facilitates a comprehensive understanding of the dependencies inherent in the data. Furthermore, CNN-BiLSTM attention frameworks adeptly combine convolutional feature extraction with bidirectional sequential processing, thereby offering a robust solution for forecasting under dynamic and nonlinear load conditions. To evaluate the efficacy of these advanced modeling techniques, this study employs a range of performance metrics critical for assessing forecasting accuracy and reliability. Metrics such as MSE, RMSE and R^2 are utilized to quantify model performance. These metrics provide valuable insights into the precision of predictions, the magnitude of errors, and the capacity of the models to capture variability in the underlying load dataset.

The objective of this study is to investigate and validate advanced DL architectures through the analysis of real-time power load datasets. By employing attention-driven mechanisms and hybrid architectures, this research aims to enhance forecasting accuracy for short-term predictions. It seeks to address the complexities and variability inherent in demand patterns while establishing a scalable framework that is compatible with contemporary power systems. The findings underscore the transformative potential of these models in bolstering grid reliability, optimizing resource allocation, and advancing sustainable energy management practices. ML and DL frameworks have emerged as promising alternatives for forecasting power load, owing to their capacity to adapt and learn autonomously from data. These models efficiently handle the dynamic, nonlinear, and complex characteristics of power load data, surpassing the limitations of traditional statistical approaches [2]. LSTM networks, for instance, have been effectively utilized to enhance the reliability and accuracy of power load forecasting through multi-step predictive models [3]. BiLSTM framework further extend this capability by capturing dependencies in both forward and backward directions, making them particularly suitable for sequential forecasting tasks [4]. Attention mechanisms have been incorporated into DL frameworks to overcome the limitations of sequential models by enabling them to concentrate on the most critical time steps or features in the data. The LSTM Attention mechanism enhances time-series forecasting by dynamically assigning weights to relevant inputs [5]. Similarly, Double Attention models provide a dual focus—both temporal and feature-based enabling more precise predictions for complex datasets [6]. RNN with attention mechanisms have been employed to selectively emphasize significant past events, further improving forecasting performance [7]. Hybrid models that combine attention mechanisms with advanced architectures have shown promise in overcoming the challenges of power load forecasting. BiLSTM with an attention mechanism leverages bidirectional information flow while focusing on the most relevant features and time steps. Additionally, hybrid models such as CNN - BiLSTM attention integrate spatial feature extraction with temporal dependencies to achieve superior predictive performance in power systems [8]. While notable advancements have been made in the field, research on CNN-BiLSTM attention mechanisms for power load forecasting remains is limited. This gap presents a significant opportunity for further investigation that could improve forecasting accuracy and efficiency in energy management. Recent studies have explored CNN, BiLSTM, and attention-based architectures for time series forecasting, including power load prediction; however, key differences persist in how these models are integrated and the scope of their applications. Many of these models utilize mechanisms like CNN, BiLSTM, and attention but are either structured in parallel pipelines, combined with auxiliary techniques, or lack a focus on fully integrated CNN-BiLSTM-attention frameworks optimized for short-term power load forecasting [9,10,11]. Additionally, these approaches have often been validated using real-time datasets from Indian substations, emphasizing their practical applicability. This work seeks to advance the field by addressing the current gap in fully integrated, end-to-end CNN-BiLSTM-attention models for localized short-term load forecasting, aiming for demonstrably superior accuracy.

METHODOLOGY and DATA COLLECTION

The historical hourly active power load data from a 33/11 KV substation in Telangana, India, was collected for the period from January 1, 2021, to December 31, 2021.

The time series dataset comprises a total of 8,760 data points, of which 66 are identified as missing values during the preparation phase.

The absence of these data points is attributed to substation shutdowns for maintenance or instances of power outages [https://data.mendeley.com/datasets/tj54nv46hj/1].

This dataset was employed for the purpose of training and evaluating the DL architectures. This study focuses on a univariate forecasting approach, utilizing only historical load data to assess the intrinsic performance of the proposed models.

LSTM Attention:

The LSTM attention model is a powerful mechanism for sequential data modeling, particularly for time series forecasting in power systems. LSTM networks are designed to overcome the vanishing gradient problem in standard RNNs by incorporating memory cells that can store information over long time intervals [12]. The attention mechanism further enhances LSTM architecture by allowing the model to focus on the most relevant past time steps, thus improving the prediction of future load values [13]. This is especially crucial in power load forecasting, where past consumption patterns, seasonality, and abrupt changes due to external factors need to be effectively captured. The attention mechanism computes a weighted average of all hidden states, assigning higher weights to the most influential time steps [14]. By integrating attention, the model dynamically adjusts its focus, leading to better handling of non-linear and non-stationary power load data.

LSTM Double Attention:

Recent advancements in power load forecasting indicate that deep learning models employing LSTM with double attention mechanisms result in notable enhancements in prediction accuracy and reliability. The LSTM double attention mechanism integrates attention layers at both the encoder and decoder stages of the model, thereby enabling a more comprehensive analysis of input features and temporal dependencies. A recent study has effectively integrated feature attention and sequence attention to capture the complex patterns present in power load data, resulting in enhanced forecasting performance compared to traditional models [15]. The double attention mechanism is particularly advantageous in multi-step forecasting scenarios, where it is crucial to model the dependencies among various past and future time steps. In the realm of power load forecasting, this model adeptly captures intricate temporal dependencies as well as sudden fluctuations in demand [16]. Research has indicated that double attention models significantly outperform single attention mechanisms when it comes to capturing long-range dependencies in time series data.

RNN Input Attention:

Traditional Recurrent Neural Networks (RNNs) encounter significant challenges when addressing long-term dependencies, primarily due to the phenomenon known as vanishing gradients. The implementation of input attention mechanisms has emerged as a solution to this limitation, enabling the network to concentrate on pertinent historical inputs. The RNN Input Attention model effectively emphasizes critical features within the input sequence, thereby filtering out extraneous noise and enhancing overall prediction accuracy [17]. In the domain of power load forecasting, this model proficiently captures vital consumption patterns along with external factors such as temperature variations and holidays, which greatly influence load demand [18]. The model dynamically adjusts the attention weights assigned to different input time steps, ensuring that significant events are appropriately weighted during forecasting [19]. The RNN Input Attention model in Figure 1 represents a considerable advancement in overcoming the limitations associated with traditional recurrent neural networks (RNNs).

LSTM-CNN Attention:

The LSTM-CNN Attention model in Figure 2 synergizes CNNs and LSTM networks with an attention mechanism. Convolutional Neural Network (CNN) layers are proficient in capturing local spatial features in data, such as short-term fluctuations in load demand, while Long Short-Term Memory (LSTM) networks are predominantly operative in modeling long-term dependencies [20]. The incorporation of the attention mechanism further enhances this hybrid model by directing the LSTM to concentrate on significant temporal patterns. This synergy proves especially beneficial for datasets characterized by high variability and noise, as it enables the model to extract pertinent features while disregarding extraneous data [21]. Empirical studies have demonstrated the effectiveness of CNN-LSTM hybrid models in intricate forecasting tasks, including predictions of energy consumption [22]. The LSTM-CNN Attention model adeptly integrates the advantages of CNNs, LSTMs, and attention mechanisms, thereby facilitating the efficient management of complex and noisy datasets.

Bi-LSTM Attention:

The Bi-LSTM attention model, illustrated in Figure 3, represents an advancement over standard LSTM networks by processing input sequences in both forward and backward directions.

This bidirectional method allows the model to effectively capture dependencies from both past and future data, a capability that is particularly crucial in power load forecasting, where future trends may be influenced by preceding and succeeding time intervals. The integration of an attention mechanism further enhances the model by directing focus toward the most pertinent segments of the data, thereby improving prediction accuracy [23].

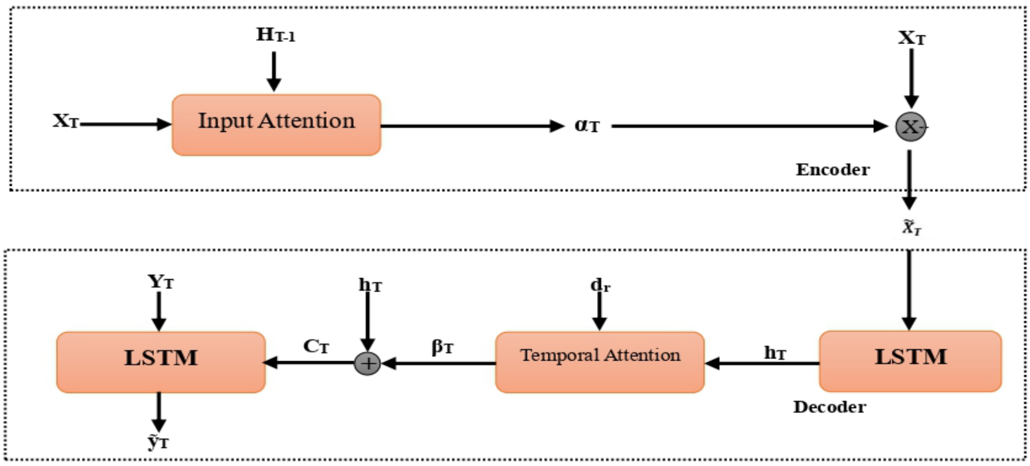

Fig. 1. RNN Input with Attention Mechanism.

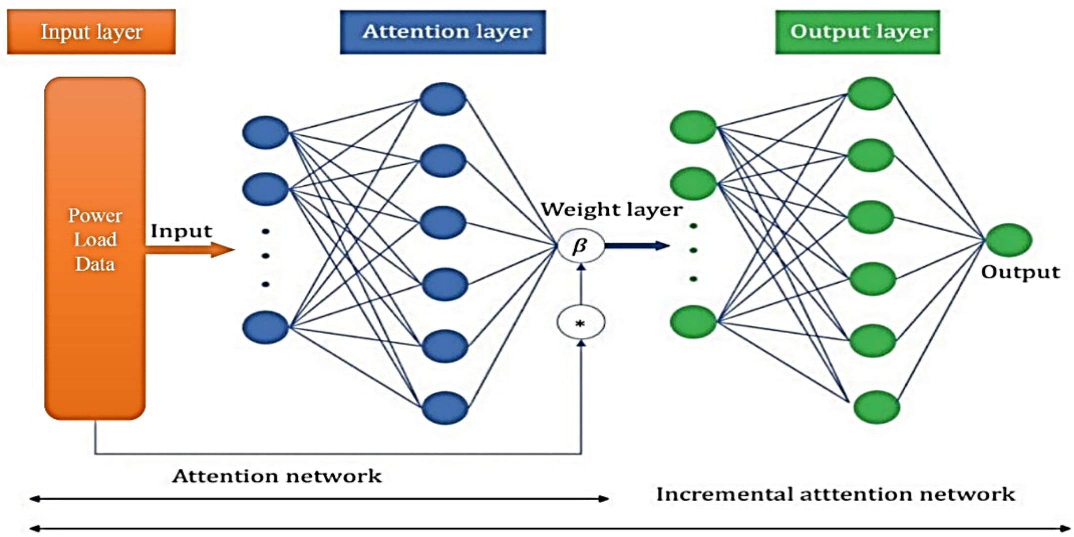

Fig. 2. LSTM CNN with Attention Mechanism.

Bi-LSTM models have gained widespread acceptance in time series forecasting tasks due to their proficiency in capturing bidirectional dependencies and their robustness in managing complex datasets [24]. These Bi-LSTM neural networks, combined with an attention mechanism, systematically analyze sequential load data for each load pattern while minimizing their objective function to enhance load forecasting performance.

CNN-BiLSTM Attention:

The CNN-BiLSTM attention model combines the spatial feature extraction capabilities of CNNs with the sequential learning strengths of BiLSTMs, integrated with an attention mechanism for refined focus on crucial data points [25]. CNN layers first extract important local features from the input data, which are then passed to BiLSTM layers that model the temporal dependencies in both forward and backward directions [24].

The attention mechanism allows the model to prioritize significant patterns, leading to superior forecasting performance. This hybrid approach is highly effective in power load forecasting, as it captures both short-term variations and long-term trends.

Studies have shown that CNN-BiLSTM Attention models outperform other architectures in complex forecasting tasks due to their ability to handle multidimensional data and capture intricate relationships [26].

The CNN-BiLSTM-Attention architecture is a sophisticated model that integrates convolutional neural networks (CNN) with bidirectional long short-term memory networks (Bi-LSTM) and an attention mechanism, thereby enhancing feature extraction and sequence modeling capabilities. [27,28]

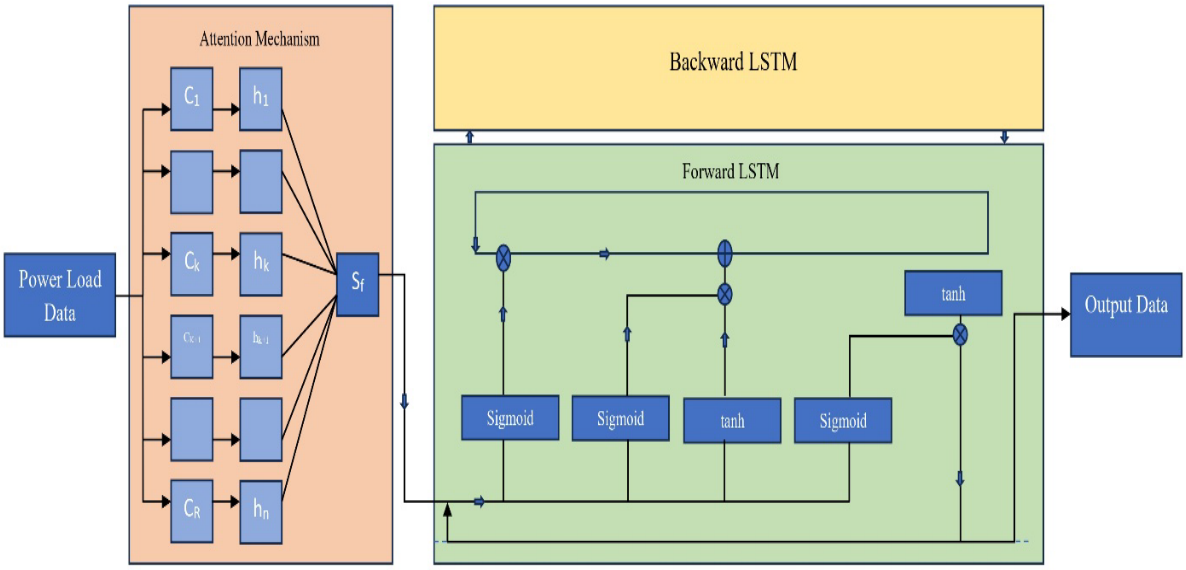

Fig. 3. BiLSTM with Attention mechanism.

The architecture in Figure 4 commences with an Input Layer, which receives the input sequence and transforms it into numerical embeddings via an Embedding Layer. Subsequently, a CNN Layer is employed, utilizing multiple convolutional filters of varying kernel sizes to capture local features within the sequence data. This phase is essential for identifying patterns in smaller segments of the data. The resulting feature maps undergo pooling operations, such as max pooling or global average pooling, to minimize dimensionality while preserving the most salient features. Following this, the pooled feature maps are directed to a Bi-LSTM Layer. This layer processes data in both forward and backward directions, effectively capturing sequential dependencies from both preceding and subsequent contexts. The concatenated hidden states generated by the Bidirectional Long Short-Term Memory (Bi-LSTM) network are subsequently fed into an Attention Layer. In this layer, the attention mechanism assigns greater importance to the most relevant hidden states, thereby enabling the model to focus on critical features within the sequence. Ultimately, the Output Layer processes the weighted context vector derived from the attention mechanism through a fully connected dense layer, leading to the final classification or prediction [29].

This hybrid architecture adeptly integrates Convolutional Neural Networks (CNNs) for local feature extraction, Bi-LSTM for sequential modeling, and attention for enhanced interpretability, rendering it highly suitable for applications such as power load forecasting.

I. PERFORMANCE METRICS

The evaluation of the models' performance was conducted using three main statistical metrics: mean square error (MSE), root mean square error (RMSE), and coefficient of determination (R^2), as outlined in Equations (1) to (3).

$$MSE = 1/n \cdot \sum_{k=1}^{n} \left| F_k - O_k \right|^2, \quad (1)$$

$$RMSE = 1/n \sqrt{\sum_{k=1}^{n} \left| F_k - O_k \right|^2}, \quad (2)$$

$$R^2 = 1 - \frac{\sum_{k=1}^{n} \left| F_k - O_k \right|^2}{\sum_{k=1}^{n} \left| O_k - \bar{O} \right|^2}. \quad (3)$$

The formula for calculating the error between the predicted and observed power generation values is as follows: F_k denotes the expected power generation for a specific time interval of (One hour), and O_k represents the observed power generation during the same interval. To enhance the analysis of the accuracy of the predictions, the average predicted power generation ($\bar{F}_k$) and the average observed power generation ($\bar{O}$) are calculated. Furthermore, the total number of observations (n) is also considered.

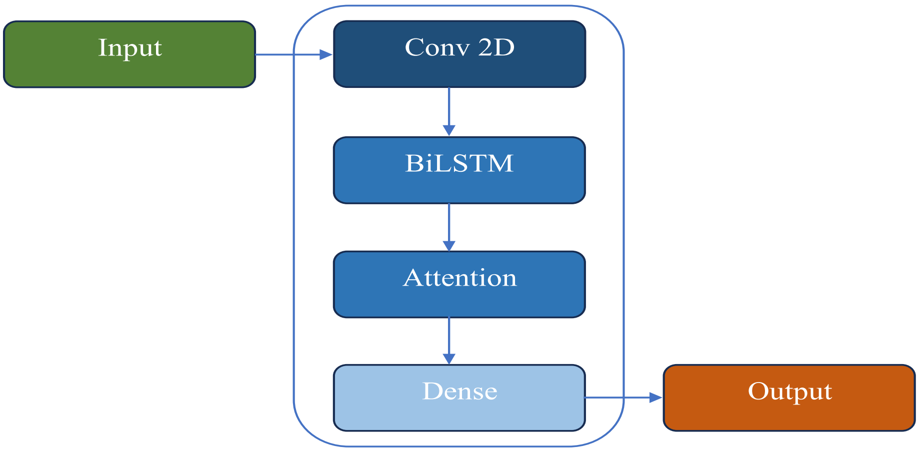

Fig. 4. CNN Bi-LSTM Attention.

Additionally, the Coefficient of Determination (R^2) is calculated to measure how well the predicted values explain the variance in the observed data. This metric requires the computation of the average observed power generation $\bar{O}$ across all time intervals. The value of R^2 ranges from 0 to 1, where values closer to 1 indicate a better fit of the model to the actual data.

III. RESULTS AND DISCUSSION

DATA PREPROCESSING

Data preprocessing is a critical phase in the development of power load predictions.
Table 1 presents the statistical characteristics (KW) of the dataset utilized. This phase employs several essential methodologies aimed at enhancing the quality of input data, which has a direct impact on the accuracy of predictions. To maintain data continuity, any missing values are systematically removed.

Furthermore, the exclusion of outliers, which represent extreme variations within the data, significantly enhances model performance by minimizing undesirable noise during the training process. Normalization is a vital component of data preprocessing, as it standardizes the variables to a uniform range between zero and one. This procedure is crucial in preventing any single variable from exerting disproportionate influence on the model's learning, thereby facilitating more efficient convergence and improving overall accuracy. Upon completion of the preprocessing phase, it was confirmed that the dataset contained no missing values. The data was subsequently normalized utilizing the Minmax Scaler to achieve a range of 0 to 1. Following normalization, the preprocessed dataset was divided into training and testing subsets. Of the 7342 input points, 5873 (80%) were designated for training, while the remaining 1468 (20%) were allocated for testing purposes to assess the predictive performance of the model.

EXPERIMENTAL RESULTS OF BILSTM ATTENTION AND CNN BILSTM ATTENTION MODELS

The enhanced performance of the CNN-BiLSTM Attention model is largely due to its dual functionality in extracting spatial features through Convolutional Neural Networks (CNN) and in capturing long-term dependencies using Bidirectional Long Short-Term Memory (BiLSTM) networks. While models such as Double Attention and LSTM Attention also demonstrated commendable performance and exhibited minor limitations in addressing complex data variations.

The study also compared results with ARIMA, MLP, Random Forest, and Gradient Boosting models, in addition to DL Attention Models. The findings underscore the critical role of hybrid models in augmenting forecasting accuracy batch size of 32 and a total of 75 epochs are selected for the experiment.

Figure 5 presents the actual versus predicted values across three-time segments: Time-Step 1, Time-Step 2, and Time-Step 3, using various deep learning attention models. The CNN-BiLSTM Attention model outperforms the others by closely aligning with the actual values, showing minimal deviation across all segments. In contrast, the LSTM and BiLSTM Attention models exhibit moderate accuracy, characterized by slight lags and amplitude errors. The CNN component effectively captures local features, while the BiLSTM and attention layers enhance sequence learning and focus. These findings underscore the effectiveness of hybrid attention-based architectures for short-term time series forecasting. To conduct a thorough analysis and comparison of the outcomes produced by these models, several fundamental evaluation metrics are employed, specifically MSE, RMSE, and R^2. The proposed model, CNN-BiLSTM with attention mechanism, is assessed against other models including LSTM attention, double attention, RNN Input attention, and BiLSTM. These metrics serve to delineate the differences between the predicted and actual values in power load forecasting. The corresponding values of these indicators are available in Table2. Significantly, the R^2 score of the CNN-BiLSTM model exceeds that of all other attention mechanisms and simpler models, signifying a superior level of accuracy in forecasting power load. Furthermore, this research introduces a cutting-edge forecasting model utilizing the CNN-BiLSTM Attention framework.

Table 1

Details of the Dataset.

Mean	SD	Min	Peak
2130	1302	412	8.435

Table 2

Performance metrics of various models

Model	MSE[KW]	RMSE[KW]	R^2 [-]	R
ARIMA	3127.38	1.77	-1.6879	0
MLP	291.40	0.54	0.7496	0.8658
RANDOM Forest	316.08	0.562	0.7283	0.8534
Gradient Boosting	324.98	0.570	0.7207	0.8489
LSTM Attention	0.008433	0.09183	0.84519	0.91934
Double Attention	0.008399	0.91647	0.84581	0.91968
LSTM CNN Attention	0.009040	0.95129	0.83387	0.91316
RNN Input Attention	0.010290	0.10143	0.81110	0.90061

Bi-LSTM Attention	0.010333	0.10165	0.81031	0.90017
CNN-BiLSTM Attention	0.007912	0.08894	0.85476	0.92453

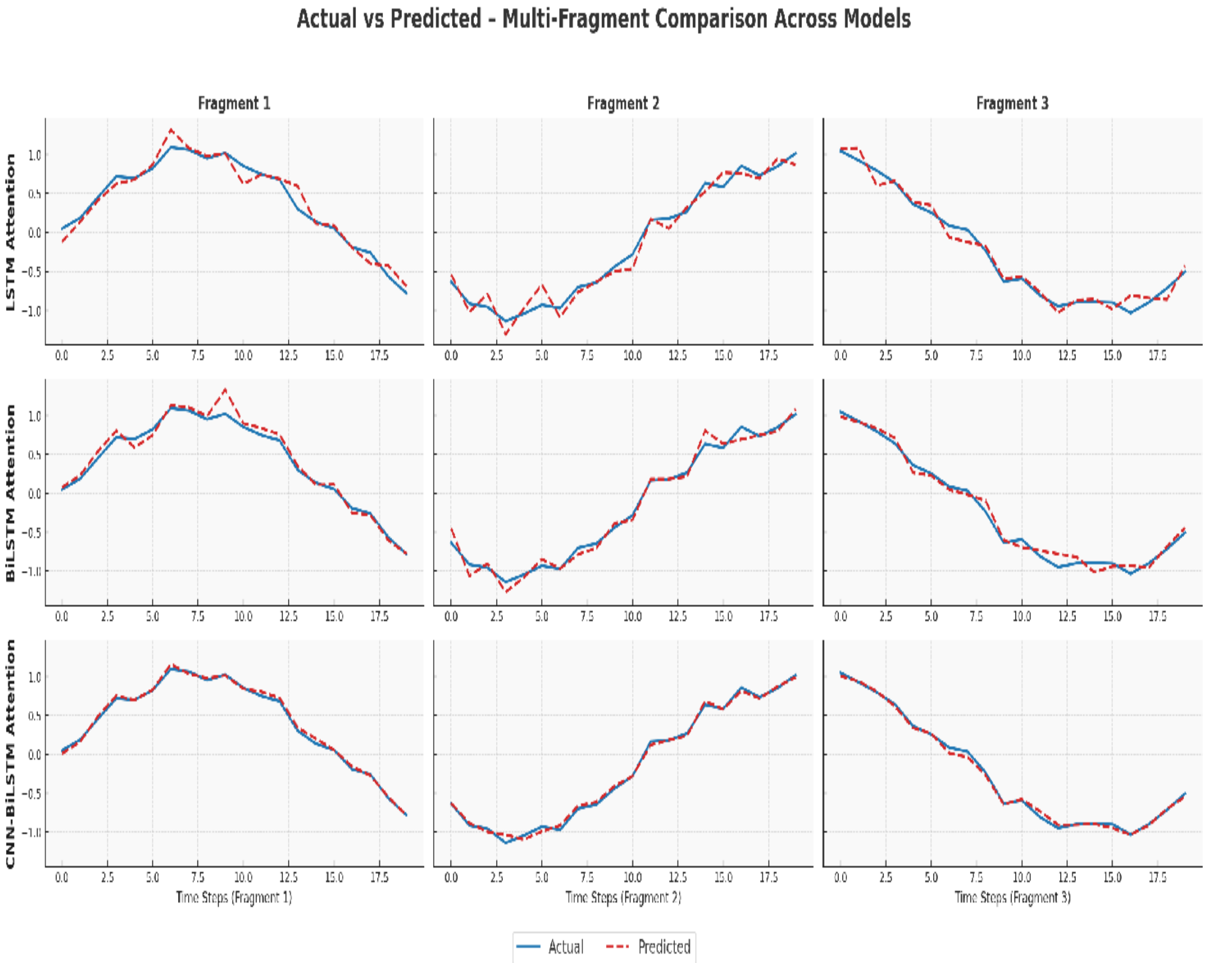

Fig. 5. Fragment Analysis of Predicted vs Actual Values Using Attention Mechanisms.

Table 3
Diebold-Mariano Test Results.

Metric/Comparison	**Bi-LSTM Attention**	**CNN Bi-LSTM Attention**	**P-VALUE**	**REMARK**
MSE	0.0099	0.0079	0.015	Lower is Better
RMSE	0.0996	0.0889	0.020	Lower is Better
R^2	0.8176	0.8547	0.012	Lower is Better
R	0.8176	0.8547	0.018	Lower is Better
Diebold-Mariano Test (p-value)			0.001	Statistically Significant($p<0.05$)

A Diebold-Mariano test was conducted to compare the CNN Bi-LSTM Attention model and the Bi-LSTM Attention model, resulting in a p-value of 0.0028 given in Table 3. This finding indicates that the performance of the CNN Bi-LSTM Attention model is statistically significant when compared to that of the Bi-LSTM Attention model.

This innovative model has demonstrated a notable enhancement in the accuracy of power load forecasting predictions. Its exceptional predictive capabilities signify a meaningful advancement in the realm of renewable energy. In comparison to alternative machine learning models, the CNN Bi-LSTM Attention model excelled in power load forecasting by achieving the lowest error rate, with a mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R^2) and R of 0.0079, 0.0889, and 0.8547 and 0.9245, respectively. The learning curve associated with the CNN Bi-LSTM Attention model indicates a satisfactory fit, which suggests the absence of both overfitting and underfitting. Furthermore, the performance of the CNN Bi-LSTM Attention model was evaluated against five other deep learning attention models. These metrics highlight the model's outstanding forecasting capabilities. When evaluated against advanced models such as the LSTM-CNN-based Self Attention Mechanism (SAM) and the LSTM model employing a double attention mechanism, the proposed CNN-BiLSTM Attention model exhibits either competitive or superior performance. While the LSTM-CNN SAM model demonstrated efficacy in managing noisy and volatile data, achieving over a 10% improvement in Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) across various datasets, the CNN-BiLSTM Attention model revealed lower error metrics, affirming its robustness and reliability in dynamic power load forecasting scenarios [30]. Moreover, the CNN-BiLSTM Attention model adeptly addresses non-stationary time-series data, paralleling the convolution layer's function within the LSTM-CNN SAM model.However, incorporating attention mechanisms in the proposed model significantly enhances its capability to capture both local and global dependencies, thereby facilitating improved generalization across diverse datasets.

The CNN Bi-LSTM model exhibited superior performance, achieving the lowest values across the established performance metrics. Moreover, the CNN Bi-LSTM Attention model possesses the capability to be trained and utilized in conjunction with various additional factors that influence power load forecasting.

CONCLUSION

The effectiveness of the CNN-BiLSTM Attention model was substantiated through a series of rigorous experiments, which identified significant improvements in critical performance metrics. The proposed methodology exhibited exceptional accuracy in forecasting, exceeding not only the CNN-BiLSTM Attention model itself but also other state-of-the-art models. This assertion is reinforced by the recorded values of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R^2, and the square root of R^2, which were documented as 0.0079, 0.0889, 0.85476, and 0.9245, respectively. In summary, the CNN-BiLSTM Attention model is characterized by its remarkable predictive accuracy, resilience in the face of data volatility, and effective management of complex time-series data. This attribute position it as a premier solution for applications in power load forecasting. The study may be further developed by integrating multiple variables as inputs to enhance the prediction of power load. This can be accomplished by obtaining datasets from the energy sector.

REFERENCES

[1] Pan Wu, Zilin Huang, Yuzhuang Pian, Lunhui Xu, Jinlong Li, Kaixun Chen. . A Combined Deep Learning Method with Attention-Based LSTM Model for Short-Term Traffic Speed Forecasting,2020, https://doi.org/10.1155/2020/8863724

[2] Sheraz Aslam, Herodotos Herodotou , Syed Muhammad Mohsin , Nadeem Javaid , Nouman Ashraf and Shahzad Aslam. A Survey on Deep Learning Methods for Power Load and Renewable Energy Forecasting in Smart Microgrids,2021, https://doi.org/10.1016/j.rser.2021.110992

[3] Dedong Tang, Chen Li, Xiaohui Ji, Zhenyu Chen, Fangchun Di, "Power Load Forecasting Using a Refined LSTM",2019.https://doi: 10.1145/3318299.3318353

[4] Sima Siami-Namini, Neda Tavakoli, Akbar Siami Namin. The Performance of LSTM and BiLSTM in Forecasting Time Series,2010.https://doi: 10.1109/ICMLA.2019.00119

[5] Xianyun Wen and Weibang Li. Time Series Prediction Based on LSTM-Attention-LSTM Model,2023. https://doi: 10.1109/ACCESS.2023.3276628

[6] Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., & Cottrell, G. A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction,2017. https://doi: 10.24963/ijcai.2017/366

[7] Yeqi Liu, Chuanyang Gong, Ling Yang, Yingyi Chen. DSTP-RNN: A Dual-Stage Two-Phase Attention-Based Recurrent Neural Network for

Long-Term and Multivariate Time Series Prediction,2017. https://doi: 10.48550/arXiv.1904.07464
[8] Zhangliang Wang, Li Jia, Chang Ren. Attention-Bidirectional LSTM Based Short Term Power Load Forecasting,2021. https://doi: 10.1109/PSGEC51302.2021.9541781
[9] Firuz Kamalov, Zicmane , Murodbek, Safara live, Linda Smail, Mihail Senyuk and Pavel Matrenin ,Attention-Based Load Forecasting with Bidirectional Finetuning,2024. Energies 2024, 17, 4699. https://doi.org/10.3390/en17184699
[10] Syed Muhammad Hasanat , Rehmana Younis ,Saad Alahmari ,Muhammad Talha Ejaz ,Muhammad Haris , Hamza Yousaf , Sadia Watara ,Kaleem Ullah , and Zahid Ullah, Enhancing Load Forecasting Accuracy in Smart Grids: A NovelParallel Multichannel Network Approach Using 1D CNN andBi-LSTM Models,2024,International Journal of Energy ResearchVolume 2024, Article ID 2403847, 13 pages https://doi.org/10.1155/2024/24038
[11] Sujan Ghimire , Ravinesh C. Deo , David Casillas-Pérez and Sancho Salcedo-Sanz,Electricity demand error corrections with attention bi-directional neural networks,2024,Volume 291, 129938, https://doi.org/10.1016/j.energy.2023.129938
[12] Guanlin Ding , Andrew Plummer & Ioannis Georgilas, Deep learning with an attention mechanism for continuous biomechanical motion estimation across varied activities, 2022, https://doi: 10.3389/fbioe.2022.1021505
[13] Sotiris Pelekis, Evangelos Karakolis, Francisco Silva, Vasileios Schoinas, Spiros Mouzakitis, Georgios Kormpakis, Nuno Amaro, John Psarras. In Search of Deep Learning Architectures for Load Forecasting: A Comparative Analysis and the Impact of the Covid-19 Pandemic on Model Performance ,2023. https://doi.org/10.48550/arXiv.2302.13046
[14] Grzegorz Dudek, Paweł Pełka, Slawek Smyl. A Hybrid Residual Dilated LSTM end Exponential Smoothing Model for Mid-Term Electric Load Forecasting, 2020, https://doi.org/10.48550/arXiv.2004.0050 8
[15] Xiaorui, Shao, Chang Soo Kim. Multi-Step Short-Term Power Consumption Forecasting Using Multi-Channel LSTM With Time Location Considering Customer Behavior,2020, https://ieeeplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9133384
[16]Xian yen Wen and Weibang L, Time Series Prediction Based on LSTM-Attention-LSTM Model ,2023, https://10.1109/ACCESS.2023.3276628
[17] Zengxi FENG, Xun GE, Yaojia ZHOU and Jiale LIA, Power Load Prediction by LSTM Model Based on the Double Attention Mechanism for Hospital Building,2023, https://doi.org/10.1051/wujns/2023283223
[18] Ozcan, A.; Catal, C.; Kasif, A. Energy Load Forecasting Using Dual-Stage Attention-Based Recurrent Neural Network. Sensors 2021, 21, 7115. https://doi.org/10.3390/s21217115
[19] Ying Chen, Zhiqiang Li, Xiaoyang Liu, and Yuhan Wang. A Hybrid CNN-LSTM-Attention Model for Electric Energy Consumption Forecasting,2023 https://10.1145/3588155.3588182
[20] Jonghoon Kim and Woojin Cho. Predicting the Household Power Consumption Using CNN-LSTM Hybrid Networks , 2018. https://doi.org/10.1007/978-3-030-03493-1_50
[21] Aya Abdelsalam Ismail, Mohamed Gunady, Luiz Pessoa, Héctor Corrada Bravo, Soheil Feizi, Input-Cell Attention Reduces Vanishing Saliency of Recurrent Neural Networks,2019, https://10.48550/arXiv.1910.12370
[22] J Son, J Cha, H Kim and YM Wi. Day-Ahead Short-Term Load Forecasting for Holidays Based on a Hybrid Model, 2022,10.1109/ACCESS.2022.3150344
[23] Mingfei Zhang, Zhoutao Yu, and Zhenghua XU.Short-Term Load Forecasting Using Recurr Recurrent Neural Networks With Input Attention Mechanism and Hidden,2020, 10.1109/ACCESS.2020.3029224
[24]Yushan Liu , Zhouchi Liang, and Xiao Li Enhancing Short-Term Power Load Forecasting for Industrial and Commercial Buildings: A Hybrid Approach Using TimeGAN, CNN, 2023, DOI 10.1109/OJIES.2023.3319040
[25]Jie Du, Yingying Cheng, Quan Zhou, Jiaming Zhang, Xiaoyong Zhang, Gang Li. Power Load Forecasting Using BiLSTM-Attention ,2020, doi:10.1088/1755-1315/440/3/032115
[26] Zhang J., Zhang Y., and Liu Y.s. Short-Term Smart Grid Load Forecasting Based on CNN-BiLSTM with Attention Mechanism,**2021,** 10.1109/ACCESS.2021.3079071
[27] Je Case, Load Electricity Forecasting using CNN-BiLSTM with Attention Mechanism,2023, https://github.com/JeCase/LoadElectricity_Forecasting_CNN-BiLSTM-Attention.

About authors.

Mukkamala Ram Babu, Assistant Professor, NICMAR University of Construction Studies, Hyderabad, Telangana. Places of work: NUCS,Hyderabad ,Telanagana. Research interests:Semantic Web,AI/ML. E-mail: rmukkamala@nicmar.ac.in

Sunku Venkata Siva Raja Prasad, Associate Professor, NICMAR University of Construction Studies, Hyderabad, Telangana. Research interests: AI/ML Safety, Risk Management
E-ail:rajaprasad@nicmar.ac.in