Journal of Advances in Management, Engineering and Science (JAMES), 2025; 1(1): 27-35

Research Article

Hybrid Machine Learning–Time Series Forecasting Framework for Energy Demand Prediction under Weather Variations.

S. Monika, Research Scholar, Anna University, Chennai, India, monikasengottaiyan@gmail.com

Dr. M. Jawahar, Associate Professor, Department of Information Technology, Panimalar Engineering College, Chennai, India, mjawahar@gmail.com

Received 1st February 2025; Accepted 17th July 2025; https://doi.org/10.65470/james.3

Abstract – Precise energy consumption forecasting is essential in the rapidly evolving electric power market that is adjusting to a future without regulation and with intense competition. Establishing the Scene: Utilities, energy dealers, and system operators have come to rely more and more on short-term forecasts due to the deregulation of energy markets, while accurate demand projections have always been crucial for efficient power system planning and operation. To tackle this issue, this study investigates Energy Demand Prediction in the presence of weather changes by the application of data smoothing techniques, specifically the Savitzky-Golay filter and Gaussian kernel density estimation, both of which are optimised with PSO. A new hybrid BiLSTM-FCN design that takes use of feature extraction and temporal dependencies is also suggested after extensive investigation into sophisticated deep learning models like BiLSTM and FCN. Findings: The suggested models surpass the current state-of-the-art methods in experimental evaluation, which includes testing for classification error and comparison with other top-tier approaches. Finally, some thoughts: The BiLSTM-FCN model showcases its potential as a strong instrument for efficient power system planning and management under weather-induced changes with performance metrics of RMSE 2.50, MAE 2.10, and MSE 6.24. It is extremely effective for accurate energy demand forecasts.

Keywords— *Partial Swarm Optimization (PSO), Fully Convolutional Network (FCN), Energy Demand Prediction (EDP).*

INTRODUCTION

The significance of constructing energy-efficient buildings has grown in tandem with the worries surrounding climate change and energy security. Greenhouse gas emissions and energy consumption are significantly increased by buildings around the world. Improving the energy efficiency of buildings is an important step in lowering energy use and lowering the impact of climate change. There are a number of ways in which energy projections can be utilized to enhance the energy efficiency of buildings. One strategy is to optimize the operation of HVAC systems and other building systems in response to actual demand[1]. Energy performance predictions and analysis allows building operators to find strategies to lower energy usage without sacrificing comfort or safety. The forecast models could also be used to priorities energy efficiency modifications after they have been identified. According to the World Energy Outlook Report, there is a clear pattern that suggests a significant and noticeable increase in the demand for energy around the world. It is now very difficult to balance energy use with production. Energy forecasting is an essential activity that entails estimating and projecting future energy consumption and production. For effective energy resource management, future energy demand forecasting, and policy decisions, it is a critical instrument for energy management. Investments in new energy infrastructure, such as power plants and transmission lines, can be guided by national forecasts[2]. The importance of the energy supply, particularly the electrical supply, in economic operations and everyday life is growing as a result of societal and technological advancements. In order to ensure the stability and economy of the power system operation, a reasonable power dispatching scheme needs to be formulated. This is necessary due to the inability to store large quantities of electric energy and the synchronization of power generation, transmission, and utilization processes. The development of a smarter and more efficient power grid depends critically on accurate power load forecast.

The ability to foretell one's own behavior is fundamental to every facet of life and a major challenge in the engineering and scientific communities. In order to create a more effective plan, including the forecasting of prices, temperatures, and weather, all nations rely on their development projects and plans on the principles and current studies methodologies[3]. The ability to foretell the future of an area's atmosphere is known as weather prediction, and it is based on scientific understanding. Predictions of future weather are based on both qualitative data on the state of the climate and the application of scientific understanding processes in the atmosphere to foretell the causes of environmental change. Having the ability to swiftly analyze statistical data, identify patterns, and develop guidelines for future study and forecasting based on historical data requires a high level of intelligence. Every day, people rely on weather predictions to help them choose an appropriate outfit. With the help of weather forecasts, they can plan our activities around the likelihood of inclement weather, such as snow or freezing rain, and get through these periods unaffected. Forecasting has captured the interest of numerous scholars from other domains because to its impact on public social life.

The remaining sections of the document are organised as follows: Methodology is presented in Section 2. Experiments are summarised in Section 3. Results and discussions are shown in Section 4. Finally, our work is concluded in Section 5.

LITERATURE SURVEY

After establishing the FCM-BP model, which has higher accuracy, they utilize the improved Fuzzy C-Means clustering algorithm to filter out comparable daily data. Nevertheless, the BP neural network-based forecasting model's convergence speed is slow, its reliance on samples is large, and its generalizability is weak. While SVR is sensitive to parameter settings, it avoids falling into local optimal solutions and has quick convergence time as compared to BP neural networks[4]. Traditional forecasting models, however, have their limits because of environmental complexity and the intrinsic unpredictability of climate forecasting. Since ML techniques, such as DL, can quantify weather data and manage complex nonlinear relationships, they are well-suited to handle multidimensional complex weather characteristics and big renewable energy datasets, thus explaining reasons for lots of investigators are now utilizing them to forecast energy consumption[5]. As an example, they discovered that hybrid methods are more suited for predicting energy consumption when they used artificial intelligence methods like support vector and ANN techniques to forecast changes in electricity. They discovered that a mix of a single RF and XGBoost could increase the accuracy of electricity consumption predictions in cities[6]. The use of RL in energy systems has been the subject of an ever-expanding corpus of study. When it comes to complicated and dynamic sequential decision-making problems, RL approaches, and DQN in particular, have demonstrated encouraging results[7]. This Study has shown that they can be useful for managing energy on microgrids, optimizing demand response, and dispatching renewable energy sources. While these applications have shown promise, they have mostly focused on specific operational domains, ignoring the system-wide resilience and adaptability issues that arise in harsh environments. When dealing with high-dimensional state and action spaces, the scalability needed for large-scale power grids is also lacking in most RL implementations.

GNNs are an incredibly promising method among these advanced approaches. Weather forecasting is only one of several sectors where GNNs have proven to be highly effective, due to their ability to process graph-structured data. They excel at forecasting because they can learn from large datasets and understand complex correlations between meteorological factors[8]. Using weather and building usage rate data, they constructed a deep NN to predict energy demand[9]. Using data on energy usage and other temporal variables, they calculated environmental consumption using a model similar to RNN. One of the ways to represent data is using autoencoder, and they suggested using it to forecast energy demand[10]. Unfortunately, the model only included fully-connected layers, that meant that temporal features were not taken into account. Additionally, the latent space when the data features are represented is not described in that model, making it difficult to manage the conditions.

In addition, to better illustrate our point, we researched the dropout approach and tested both its absence and presence in suggested models.

In this paper, we primarily offer three things:

- They offer a BiLSTM FCN hybrid model as a deep learning option.
- The work demonstrates the use of cutting-edge approaches, such as BiLSTM and FCN, in hybrid models, specifically the BiLSTM-FCN EDP, to address the issue of weather variability. Additionally, they examine into the application of the dropout method.
- The experimental results show that the suggested models are end-to-end and don't need extensive data pre-processing, feature extraction, or refinement, and they also compare their performance to that of the current literature.

PROPOSED SYSTEM

Residential buildings can enhance their energy efficiency and achieve significant savings through energy demand prediction. Residential buildings account for the lion's share of building energy consumption, and this problem is only going to get worse as the social economy continues to expand at a rapid pace. Given the current dire circumstances, the focus of national energy saving efforts and the scientific community has shifted to the issue of building energy savings as a primary objective.

To improve their short-term supply planning, energy grid operators can benefit from improved data on community electricity demand and generation forecasts. To top it all off, homes will have to have a better grasp of their consumption habits if they want to make informed choices regarding energy-trading programs and appliance use[11]. Volatility in solar production affected by solar cell specifications and weather conditions, as well as volatility in load consumption caused by consuming behaviour, are the primary concerns to be addressed in this context. Analysis such as Gaussian Process Classifier, Bagger classifier, Gradient Boosting Classifier, and Quadratic Discriminant Analysis can be applied to this dataset, which includes behavioural and solar energy elements of communities.

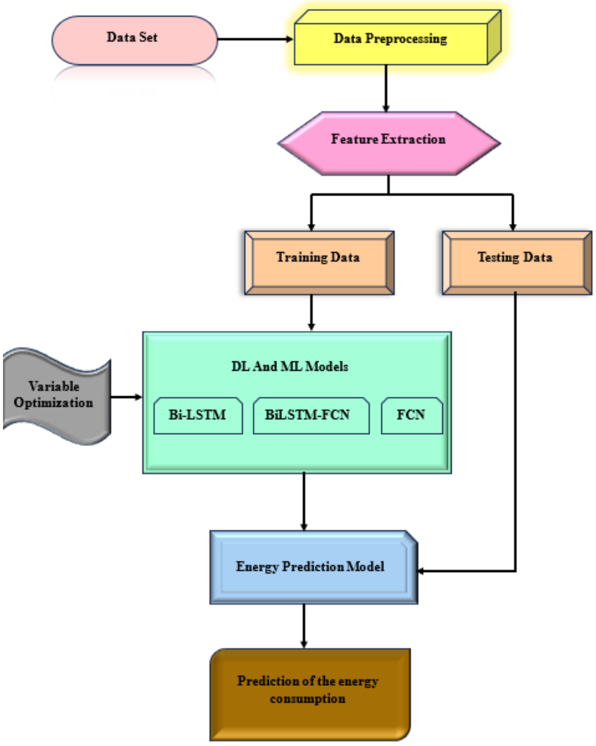

Fig. 1.The Framework for the Energy Demand Prediction

Figure 1 shows the structure of the model for making predictions[12]. They trained the prediction model with the first 60% of the data and tested it with the remaining 40%. Feature selection The significance of each parameter was determined using PSO, and the number of variables utilised to construct the prediction model was determined using RF. They might assume of the post's ensemble learning as having two layers. On the bottom, you'll see that the article employed a variety of deep learning algorithms to build basic models. These models can produce new data sets with the exact same size as the originals. The second-level Deep learning model was trained using the new data sets, and then the ensemble learning model was put up.

A. Data Preprocessing

It is necessary to preprocess data from smart meters before feeding it into models of DL to improve model performance, as the data is typically of low quality. Data normalisation, cleaning, and smoothing are examples of typical data preprocessing procedures. Data cleansing is the process of identifying and eliminating anomalies and missing values. One method for identifying outliers is by using the local outlier factor, which is dependent on distance. To compensate for missing data or extreme values, the average energy consumption of the same hour the previous two times is utilised[13]. The research delves into the theory and practice of two smoothing methods—the Savitzky-Golay filter and the Gaussian kernel density estimation—and their respective applications. To normalise data is to convert it from its raw form into a predetermined interval. Because models trained with neural networks are sensitive to the magnitude of their inputs, the suggested model's Equations (1) and (2) demonstrate the need to normalise the inputs and outputs to a range from 0 to 1.

$$C_r' = \frac{C_r - C_{min}}{C_{max} - C_{min}} \tag{1}$$

$$B_r' = \frac{B_r - B_{min}}{B_{max} - B_{min}} \tag{2}$$

C_r' and B_r' are the normalised versions of the input and output variables, respectively, whereas C_{max} , C_{min} , B_{min} and B_{max} are the appropriate minimum and maximum values of C_r and B_r , respectively.

B. Feature Selection:

The PSO optimisation algorithm is a well-known bio-inspired approach that can handle both continuous and discontinuous functions. The characteristics of birds that forage for food on a daily basis form the basis of its primary concept. One of its features is a mechanism for mining a population for logical answers represented by an individual inside that population[14]. Collaborative effort is crucial if the populace is to attain the necessary degree of intellect. Each member of the group must occupy a specific location whenever they look for an ideal spot.

Here is the full procedure based on the PSO steps:

1. Set the inertia weight, maximum speed, population size, and finishing criterion as initial values for the algorithm.
2. A new, identically sized random population should be generated using the data provided in Step I. Every part or member has to have all of the given values.
3. Next, for every piece of the population, find the objective function. In the first iteration, the population's best answer is represented by the $Pbest$.
4. It checks each particle's velocity against the updated maximum velocity. If the velocity is higher than the maximum velocity, it is reduced to that value.
5. Particles should be repositioned such that they fall between the specified upper and lower limits.
6. $Pbest$ is replaced following a comparison of the optimal solution for each particle. The optimal solution $Gbest$ is the highest achiever within the population.
7. The procedures from step IV are repeated until the termination requirements are met.

C. *Model Training:*

1) *BiLSTM:*

The BiLSTM is the result of a BRNN with an LSTM network architecture. RNN based on BiLSTM were created by combining the benefits of BRNN and LSTM.

One of the initial uses of BRNN was to showcase a structure that, when unfolded, became a bidirectional neural network. It permits data transmission in line with the data's intrinsic temporal sequences and data reversal to prior time steps when used to time series data[15]. Two hidden layers make up BRNN, and they are both linked to the input and output. One layer employs recurrent connections from prior time steps, whilst the other is inverted and transmits activation in reverse along the sequence. This illustrates the variation among various levels. Normal BP after temporal unfolding can train BRNN. A BRNN is defined in Equation $(3) - (5)$ by the three equations that follow.

$$s^{(g)} = \tau\left(D^{sc}c^{(g)} + D^{ss}s^{(g-1)} + y_s\right) \quad (3)$$

$$a^{(g)} = \tau\left(D^{ac}c^{(g)} + D^{aa}a^{(g+1)} + y_a\right) \quad (4)$$

$$\hat{b}^{(g)} = softmax(D^{bc}s^{(g)} + D^{bc}a^{(g)} + y_b \quad (5)$$

where $s^{(g)}$ represents the forward layers that are hidden value and $a^{(g)}$ represents the backward hidden layer value. During the present time step g, the input for recurrent edges comes from the current data point $c^{(g)}$ and the prior state $s^{(g-1)}$. The weight matrix and bias vectors are represented by D and y, respectively. The output layer $\hat{b}^{(g)}$ uses the τ sigmoid function as an activation function, while the is the normal function.

To address the issue of vanishing gradients, Hochreiter and Schmidhuber initially proposed the LSTM. A kind of RNN, LSTM shares the same input and output types as its parent network. A gate for forget, an gate for output, and an input gate are the three components that distinguish LSTM from RNN. So, it has control over what needs to be remembered and what can be erased. This is the reason why LSTM can remember previous data whereas RNN can't.

To put it more precisely, the results of the LSTM model's computations are proportional to the results of the subsequent calculations carried out at each time step. The full method for a state-of-the-art LSTM with forget gates in Equation $(6) - (11)$ is provided by these computations.

$$t^{(g)} = tans\left(D^{(tc)}c^{(g)} + D^{ts}s^{g-1} + y_t\right) \quad (6)$$

$$r^{(g)} = \tau\left(D^{rc}c^{(g)} + D^{rs}s^{(g-1)} + y_r\right) \quad (7)$$

$$u^{(g)} = \tau\left(D^{uc}c^{(g)} + D^{us}s^{(g-1)} + y_u\right) \quad (8)$$

$$l^{(g)} = \tau\left(D^{lc}c^{(g)} + D^{ls}s^{(g-1)} + y_l\right) \quad (9)$$

$$h^{(g)} = t^{(g)} \odot r^{(g)} h^{(g-1)} \odot u^{(g)} \quad (10)$$

$$s^{(g)} = tans\left(h^{(g)}\right) \odot l^{(g)} \quad (11)$$

In where $c^{(g)}$ represents the layer of input at the present time step g, $s^{(g)}$ stands for the hidden layer value of the LSTM, and $s^{(g-1)}$ signifies the output values produced by every memory cell in the layer of hidden at the prior time. The symbol denotes a sigmoid function $s^{(g-1)}$, whereas the symbol $\odot$ stands for element-wise multiplication and the symbol tanh for the hyperbolic tangent function.

In order to generate the final output, BiLSTM processes sequence data in both the forward and backward directions. It does this by using two hidden layers, one for capturing past information and the other for capturing future information. It has been proved that bidirectional networks are considerably better than unidirectional ones in many fields. They choose BiLSTM for our problem-specific learning needs because it allows us to access the long-range context in both input and output directions. Data was processed by unidirectional LSTM using just historical information that had been saved. Instead of training just one LSTM on the input sequence, BiLSTM uses all available time steps to solve the problem. In numerous areas, including phoneme classification, handwriting recognition, NLP, and speech recognition, BiLSTM has attained state-of-the-art results.

2) *FCN:*

An important class of graphical models, CNN may produce feature hierarchies. Wang et al. primarily proposed FCN for EDP and validated it under weather fluctuations; it is an extension of traditional CNNs. When it comes to EDP, FCNs are great for handling the temporal dimension without a tonne of data preprocessing or feature engineering, which is why they're mostly used in the temporal domain. In the first branch of both models, they employ FCN as a feature extractor in the suggested models.

Equations $(12) - (14)$ describe FCN for univariate WDP:

$$g = d \odot c + y \tag{12}$$

$$z = BN(g) \tag{13}$$

$$b = ReLU(z) \tag{14}$$

where the convolution operator is denoted by $\odot$ and the tensor, input vector, and bias vector at time step g are represented by $d, c, and\ y$ respectively.

Figure 2 shows the two branches that make up the proposed model architecture: FCN and BiLSTM. Three 1D convolutional kernels—numbering 8, 5, and 3—with filter widths of 128, 256, and 128—one in each block—make up the FCN architecture, which does not stride. After every block, there is an activation layer that uses ReLU and batch normalisation (BN). A global average pooling layer receives information after the convolutional blocks, and a SoftMax layer generates the last label, b.

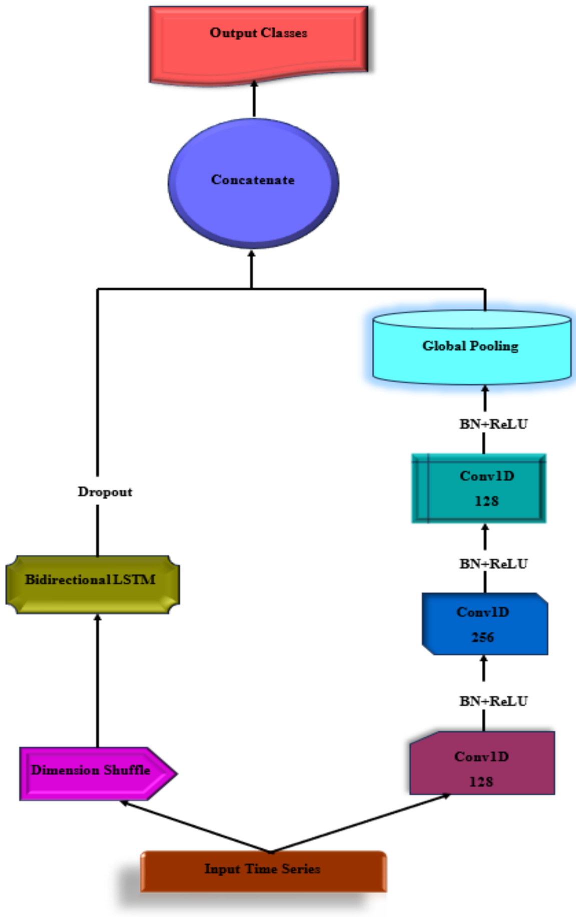

Fig. 2.BiLSTM-FCN Model

RESULT AND DISCUSSION

The severe difficulties caused by climate change and the increasing expansion of global energy demand have led to a common goal among industries: green operations and sustainable development of firms. Businesses can do their part to combat climate change and live up to their social obligations by accurately predicting and satisfying their demand for renewable energy, which is essential for running green operations. Yet, conventional approaches to predicting energy use often have poor accuracy and limited processing capability for complicated data, leaving businesses ill-equipped to handle the ever-changing demands of their management in today's market.

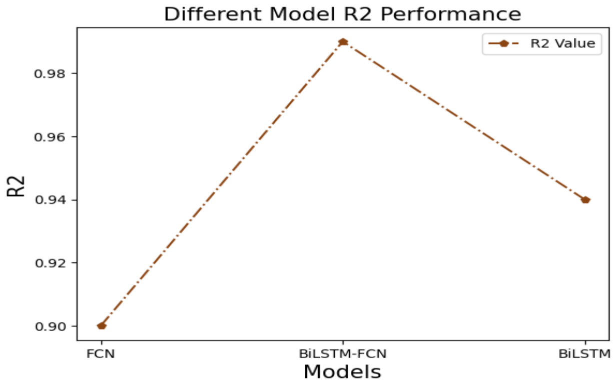

Fig. 3. Performance of R^2

Figure 3 illustrates a line graph that compares the R² scores of three models: FCN, BiLSTM-FCN, and BiLSTM. The BiLSTM-FCN model is better than the others since its R² value is around 0.99, which is substantially higher than the others' 0.90 and 0.94. This makes it evident that the BiLSTM-FCN model is better at making predictions.

TABLE I. PERFORMANCE RESULT(%)

Models	R^2	MAE	RMSE	MSE	Training Time
BiLSTM	0.94	6.27	8.64	9.30	1.1s
BiLSTM -FCN	0.99	2.10	2.50	6.24	1.5s
FCN	0.90	12.43	14.41	15.26	864s

Table 1 shows the results of computing the four performance indicators used to evaluate the three models. The FCN model was the only one that could not provide good results. Although the BiLSTM-FCN based model achieved somewhat better results in terms of MAE, RMSE, MSE, and R2, the BiLSTM based model took less time to train. While the BiLSTM-FCN and BiLSTM models trained more quickly, the FCN model trained quicker but fared poorly.

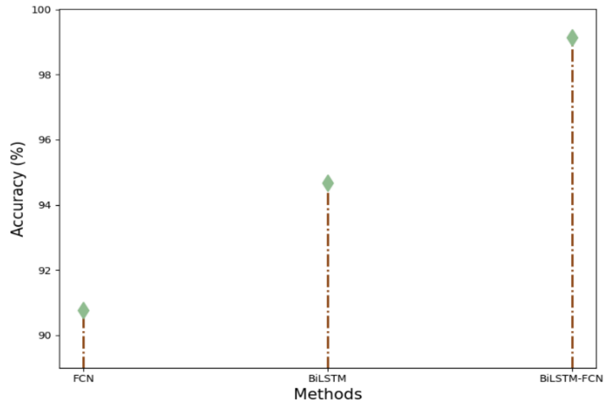

Fig. 4. Accuracy Comparison for Different Models

Figure 4 shows how well the three deep learning models with enhanced data performed in terms of making predictions. It is clear from the analysis that the Proposed BiLSTM-FCN models outperform the other models in terms of prediction performance.

TABLE II. COMPARISON RESULT (%)

Model	Accuracy	Miss Rate
Training BiLSTM-FCN	99.14	2.42
Testing BiLSTM-FCN	96.28	2.56

Table 2 displays the overall results of the training and testing phases for the proposed BiLSTM-FCN model. During training, the suggested BiLSTM-FCN model reached 99.14 percent accuracy with a 2.42 percent miss rate; during testing, it reached 96.28 percent accuracy with a 2.56% miss rate.

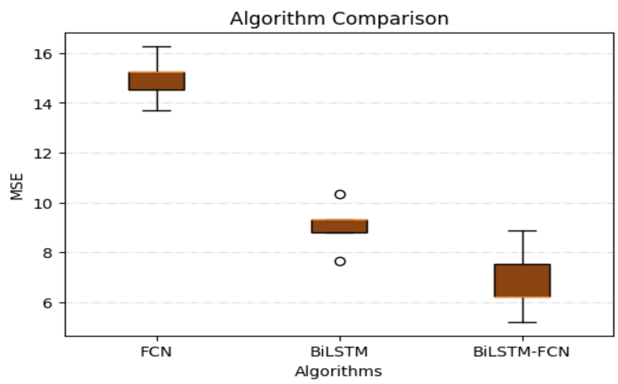

Fig. 5. Comparison of the MSE

Additionally, Figure 5 provides a visual representation of the absolute MSE error. It evaluates the BiLSTM, FCN, and BiLSTM-FCN models by comparing their MSE values. With the lowest MSE Value, they find the Proposed BiLSTM-FCN Model.

TABLE III. COMPARATIVE PERFORMANCE (%)

Metrics	BiLSTM		BiLSTM-FCN		FCN	
	Train	Test	Train	Test	Train	Test
MAE	6.27	13.56	**2.10**	**9.38**	12.43	19.32
RMSE	8.64	15.71	**2.50**	**10.68**	14.41	21.22

Table 3 displays the outcomes of energy demand prediction using several DL algorithms. Three different networks—BiLSTM, BiLSTM-FCN, and FCN—create these models. When compared to existing models, the proposed BiLSTM-FCN model performs well in both the train and testing sets.

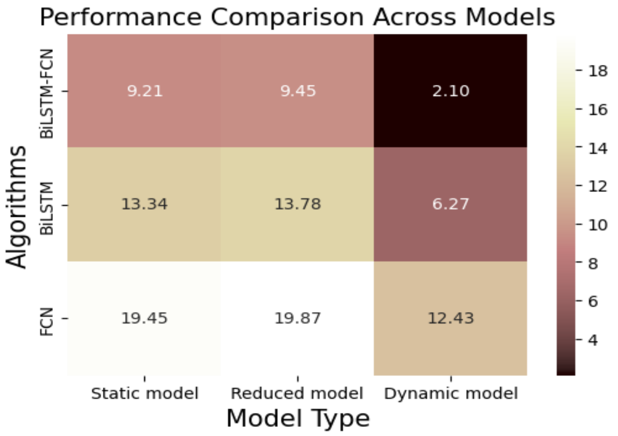

Fig. 6. Performance Comparison

All three models—static, reduced, and dynamic—have their MAE values displayed in Figure 6, a heatmap. In comparison to static and reduced models, dynamic reduced models that incorporate lag data on energy usage show better MAE predictive performance. After BiLSTM (8.64%) and FCN (14.41%), the most efficient model for training is BiLSTM-FCN (2.50). Using an RMSE of 10.68%, the model BiLSTM-FCN once again demonstrated its skills as the most efficient model. This is the sequence in which the other models appeared: BiLSTM (15.71%), FCN (21.22%). In terms of validation, the suggested model ranked first with an RMSE of 7.89%, followed by BiLSTM at 11.43%, and FCN at 18.55%. It shown graphically in Figure 7 are the RMSE results.

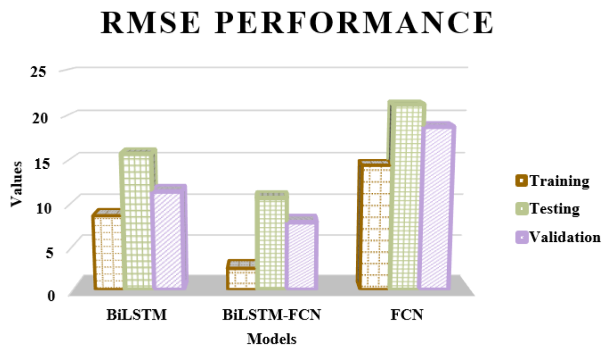

Fig. 7. RMSE Performance

CONCLUSION

In this study, they provide new methods that use cutting-edge DL algorithms to forecast energy use in response to weather changes. The prediction accuracy can be enhanced by these systems' ability to learn and discover the correlation of complicated hidden features. One sustainable way to replace traditional transportation systems with ones that consume less oil, are more energy efficient, and produce less petrol emissions is to leverage weather variability. Data smoothing isn't always the way to go, so use it with care. How the Savitzky-Golay filter and the Gaussian kernel density estimation, two types of smoothing, affected the accuracy of the model's predictions. PSO was used to optimise it for energy demand prediction. For end-to-end EDP in the face of weather changes, they have introduced a hybrid deep learning model called BiLSTM-FCN. As an EDP, FCN has shown to be effective at feature extraction in the face of weather changes, and a BiLSTM takes both types of dependencies into account. According to the findings of the experiments, our suggested models outperform the current state-of-the-art methodologies in every way. Compared to the alternatives, the suggested model outperforms them in terms of electricity demand forecasting, with RMSE of 2.50%, MAE of 2.10%, MSE of 6.24%.

REFERENCES

[1] H. S. Oliveira and H. P. Oliveira, "Transformers for Energy Forecast," *Sensors*, vol. 23, no. 15, pp. 1–21, 2023, doi: 10.3390/s23156840.

[2] V. D. Juyal and S. Kakran, "CNN-LSTM Based Weather Forecasting and Its Application in A Residential Home Energy Management System," *Evergreen*, vol. 11, no. 4, pp. 3243–3253, 2024, doi: 10.5109/7326959.

[3] S. Kareem, Z. J. Hamad, and S. Askar, "An evaluation of CNN and ANN in prediction weather forecasting: A review," *Sustain. Eng. Innov.*, vol. 3, no. 2, pp. 148–159, 2021, doi: 10.37868/sei.v3i2.id146.

[4] X. Deng *et al.*, "Bagging–XGBoost algorithm based extreme weather identification and short-term load forecasting model," *Energy Reports*, vol. 8, pp. 8661–8674, 2022, doi: 10.1016/j.egyr.2022.06.072.

[5] R. Qu, R. Kou, and T. Zhang, "The Impact of Weather Variability on Renewable Energy Consumption: Insights from Explainable Machine Learning Models," *Sustain.*, vol. 17, no. 1, pp. 1–25, 2025, doi: 10.3390/su17010087.

[6] B. Neeraja, W. Biswas, M. Sreelaxmi, B. S. Firake, J. S. Wasnik, and R. A. Kamde, "Advanced Energy Consumption Prediction for Smart Cities Using Meta-Learning Layers," *Proc. Int. Conf. Vis. Anal. Data Vis. ICVADV 2025*, pp. 317–322, 2025, doi: 10.1109/ICVADV63329.2025.10960914.

[7] S. Dong *et al.*, "Hierarchical deep Q-network-based optimization of resilient grids under multi-dimensional uncertainties from extreme weather," *Sci. Rep.*, vol. 15, no. 1, pp. 1–17, 2025, doi: 10.1038/s41598-025-09868-1.

[8] Y. Varshney, V. Kumar, D. K. Dubey, and S. Sharma, "Forecasting Precision : The Role of Graph Neural Networks and Dynamic GNNs in Weather Prediction," *J. Big Data Technol. Bus. Anal.*, vol. 3, no. 1, pp. 28–33, 2024.

[9] S. R. Madhukar, K. Singh, S. P. Kanniyappan, T. Krishnan, G. C. Sarode, and D. Suganthi, "Towards Efficient Energy Management of Smart Buildings: A LSTM-AE Based Model," *1st Int. Conf. Electron. Comput. Commun. Control Technol. ICECCC 2024*, pp. 1–6, 2024, doi: 10.1109/ICECCC61767.2024.10593988.

[10] J. Y. Kim and S. B. Cho, "Electric energy consumption prediction by deep learning with state explainable autoencoder," *Energies*, vol. 12, no. 4, 2019, doi: 10.3390/en12040739.

[11] "Dataset for Forecasting Electricity Demand." Accessed: Sep. 29, 2025. [Online]. Available: https://www.kaggle.com/datasets/itspot/dataset-for-forecasting-electricity-demand

[12] Y. Huang, Y. Yuan, H. Chen, J. Wang, Y. Guo, and T. Ahmad, "A novel energy demand prediction strategy for residential buildings based on ensemble learning," *Energy Procedia*, vol. 158, pp. 3411–3416, 2019, doi: 10.1016/j.egypro.2019.01.935.

[13] Z. Xiao *et al.*, "Impacts of data preprocessing and selection on energy consumption prediction model of HVAC systems based on deep learning," *Energy Build.*, vol. 258, 2022, doi: 10.1016/j.enbuild.2022.111832.

[14] E. Ofori-Ntow Jnr and Y. Y. Ziggah, "Electricity demand forecasting based on feature extraction and optimized backpropagation neural network," *e-Prime - Adv. Electr. Eng. Electron. Energy*, vol. 6, no. September, p. 100293, 2023, doi: 10.1016/j.prime.2023.100293.

[15] M. Khan, H. Wang, A. Riaz, A. Elfatyany, and S. Karim, "Bidirectional LSTM-RNN-based hybrid deep learning frameworks for univariate time series classification," *J. Supercomput.*, vol. 77, no. 7, pp. 7021–7045, 2021, doi: 10.1007/s11227-020-03560-z.