Data Engineering and Applications

Research Paper

LSTM-CNN Hybrid Model for High-Dimensional Load Forecasting in Smart Grids

Lena Majewska[1] and Beata Kubicka[1,*]

[1] Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Krakow, 30-059 Krakow, Poland
*Corresponding author: beata.k@agh.edu.pl

Abstract. The challenging issue of electric demand forecasting in a smart grid has been resolved using computations and deep learning utilizing data-driven techniques. This research proposes a hybrid neural network model that combines CNNs and LSTM layers to solve the issues of localized feature interactions and long-term temporal dependencies in recent meter data. The model uses a highly sophisticated feature pipeline that uses time-series encoding, normalization, and other data augmentations to handle non-Gaussian noise, missing values, and multi-scale features of the advanced metering infrastructure. Real-world high-frequency datasets of residential, commercial, and industrial regions are used for testing and training. The data is divided into temporally contiguous training and validation sets. The findings of the experiment show that the LSTM-CNN hybrid has lowered the mean absolute error (MAE) and root mean squared error (RMSE) of the best-performing single models and baselines by 22.4% and 38.9%, respectively. Both architectural elements and different feature sets are required, as confirmed by ablation and sensitivity analysis. Robustness experiments demonstrate that generalization is still possible under adversarial and external conditions, and that input noise, missing data, and distribution shifts have no effect. According to the aforementioned study, the LSTM-CNN hybrid is a workable and scalable method that can be applied in real-world scenarios to accurately forecast changes in smart grids.

Keywords: *Deep Learning, Smart Grid, Load Forecasting, LSTM-CNN Hybrid, Time Series Prediction*

Received on 12 October 2025, Accepted on 29 December 2025, Published on 5 Jan2026

Introduction

For a new kind of smart grid to operate normally and steadily, accurate load forecasting is still necessary [1]. Demand fluctuations and unpredictability have grown as electric power networks become more automated and incorporate more renewable energy, making old forecasting techniques less reliable [2]. The non-stationarity and non-linearity in the actual electricity consumption data of power grids frequently pose challenges for conventional time series techniques like exponential smoothing and ARIMA [3]. It is challenging to use all of the high-resolution, multi-source data that has been produced as a result of the development of intelligent electricity meters (AMI) [4]. Traditional models are no longer appropriate for handling the issues with contemporary grid loads because of fluctuations brought on by weather and human behavior changes, distributed generation, and load-response programs, which have introduced additional noise and instability to consumption patterns [5]. These conditions have led to a growing need for more adaptable, data-driven approaches that can adaptively learn intricate temporal and contextual connections in varied environments [6]. In order to solve these shortcomings in utility-sector research, academics have recently shifted away from traditional handmade feature engineering and fixed-model assumptions with the introduction of machine learning, particularly deep learning [7].

Convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and recurrent neural networks (RNNs) are well-known deep learning models that have increased the precision of multi-dimensional data processing and sequence prediction for energy systems [8]. Long-term temporal relationships can be effectively captured by LSTM networks, which also somewhat circumvent the issue of vanishing gradients in

RNNs [9]. CNNs are also very good at identifying hierarchical patterns in structured data and extracting local features [10]. Recent research has demonstrated that in situations of substantial temporal correlation and local data interdependence, a hybrid model using an LSTM for sequence modeling and a CNN for local feature extraction has attained a comparatively high accuracy [11]. Many deep learning-based load forecasting models still struggle with issues like poor generalization, overfitting, lack of scalability and interpretability, etc., despite some recent advancements, especially when dealing with noise, high dimensionality, and fluctuation in real AMI data [12]. The aforementioned technological issues are currently being researched by developing robust architectures and training techniques that lower model complexity while preserving operational accuracy and stability [13]. Research on scalable and generalizable hybrid neural network models for power forecasting is continuing ongoing in both business and research [14].

This research will proceed by creating hybrid deep learning frameworks that combine and make use of the advantages of CNN and LSTM architectures in light of the growing complexity of the requirements and the drawbacks of the earlier approaches [15]. The goal is to provide a theoretically solid and practically workable method that can handle many aspects of diverse and high-dimensional data in real-world smart grid forecasting. New concepts and useful enhancements for the upcoming generation of intelligent power systems have been suggested as a result of the above stringent evaluations of model design and performance in a demanding operating environment.

Related Works

Neural Network Models for Load Forecasting

Over the past 20 years, neural networks have been used for load forecasting; more recently, as grid data sets have become more complicated and finer-grained, this application has also advanced dramatically [16]. Among the earliest deep learning models used for power system forecasting were feedforward neural networks, which outperformed conventional statistical techniques when non-linearity was present [17]. The earliest networks were usually multiple-layer perceptrons (MLPs), which were trained using backpropagation to enhance the modeling of non-linear interactions. However, they were not appropriate for learning time-series characteristics of electrical load data [18]. Recurrent Neural Networks (RNNs) were consequently created in order to preserve the hidden state and thus recall previous data for future prediction [19]. However, typical RNNs were not very good for long-term sequence learning because of their intrinsic vanishing and expanding gradient issues [20].

In order to preserve and update data in longer sequences using gated methods, a long short-term memory (LSTM) network was proposed [21]. LSTM-based architectures are more precise and reliable than many shallow learning models, and they have produced outstanding results in modeling both short-term and long-term temporal patterns of smart grid data [22]. Additionally, gated recurrent units (GRUs) are comparatively straightforward and computationally light substitutes for LSTMs that have demonstrated favorable outcomes [23]. Due to the proliferation of advanced metering infrastructure (AMI), a significant amount of high-frequency and high-dimensional data has been gathered, and deep learning models that perform well on complicated time-series problems and enormous amounts of data are now highly sought after [24]. In order to dynamically weight various segments of the input sequence and concentrate on significant time points or characteristics during prediction, attention-based networks and sequence-to-sequence architectures have also been employed concurrently [25]. When combined, the aforementioned neural network models have somewhat increased grid load forecasting's prediction accuracy, flexibility, and capacity for generalization [26].

Hybrid and Ensemble Approaches

Many academics have suggested numerous fusion and ensemble methodologies to improve the stability and generalization capacity of load forecasting systems in light of the drawbacks of the single-model approach [27]. In order to simultaneously capitalize on the advantages of all these techniques, a hybrid strategy typically integrates a deep learning model with conventional algorithms, statistical preprocessing or feature engineering pipelines, etc. [28]. For instance, combining an autoregressive framework with a neural network structure in a model can address both linear and non-linear features of load time series [29]. Since CNNs can now extract local and global features from grid data matrices, they are a great choice as input representations for LSTM modules in sequence modeling [30]. CNNs were initially introduced in computer vision.

https://doi.org/10.64972/dea.2026.v5i1.1613d:30-45

Architectural fusion is one kind of ensemble learning; numerous others, such stacking and bagging, are also used to improve prediction stability and reduce variance. In order to provide a final prediction that is less prone to overfitting and more stable against noise in the data, these techniques employ many base learners and sometimes combine shallow and deep predictors. Due to the relative complexity of high-dimensional regression and classification issues in grid environments, sophisticated ensemble models like random forest ensembles and gradient boosting have been incorporated into the processing pipeline. The aforementioned techniques have consistently demonstrated better prediction outcomes in numerous benchmark trials for both short-term forecasting and day-ahead market prediction, despite the fact that they are highly computationally demanding. Recently, very accurate, highly interpretable, and useful neural hybrid models for dynamic grid environments have surfaced.

Summary and Research Gaps

Neural network-based and hybrid forecasting models have shown impressive results in recent years, but numerous research issues still need to be resolved before they can be used in next-generation smart grids. Lack of generalization in heterogeneous areas or when working outside the training range is a common issue; many earlier approaches have yet to address data scarcity and domain transferability. Model scalability and computational efficiency issues for real-time or near-real-time applications have steadily gotten worse as grid datasets have grown in size and complexity. Overfitting can still happen despite deep structure optimization attempts since there isn't enough noise-free, thorough, or well-organized AMI data. Black-box models also have interpretability issues, which has led to an increase in the demand for clear and understandable projections that support decision-making at all societal levels.

While hybrid and ensemble approaches have shown some successes, there are currently no set guidelines for model creation, hyperparameter adjustment, or evaluation. Benchmark studies are not appropriate for fair comparison or replication because they are frequently constrained by inconsistent datasets and features as well as different experimental setups. Lastly, as grid modernization progresses, there is an increasing need for robust models that can manage distributed technologies, new grid topologies, and shifts in consumption patterns. Currently, research and development in the field of smart grid load forecasting is actively investigating the creation of robust, comprehensive, and comprehensible hybrid models. We must work together to solve the aforementioned issues if we are to fully utilize deep learning in future energy systems. To continue motivating academics and promoting the creation of more effective and broadly applicable forecasting techniques, address the aforementioned unresolved issues.

Methodology

LSTM-CNN Model Structure

The suggested hybrid framework's architecture combines the local feature extraction capabilities of convolutional neural networks (CNNs) with the temporal sequence model ability of long short-term memory (LSTM) units in a synergistic manner. Strong autocorrelation, external effects, temporal heterogeneity, and high-dimensional, highly non-linear, multi-scale features of smart grid load time series are all directly addressed by the suggested approach.

The system receives a multifaceted feature matrix of unprocessed advanced metering infrastructure (AMI) records, exogenous covariates, and time-window-standardized temporal indices at the input interface. A series of LSTM units use the matrix as input; each unit carries over memory from the previous time step, ignores redundant or noisy parts, and picks specific information from the input. In order to understand latent cyclical patterns and infrequent event triggers that significantly impact load, deep stacked LSTM layers are especially well-suited for tackling the issues of long-range reliance and disappearing gradients.

Following the aforementioned time-series processing, a multi-layer CNN receives the reshaped high-dimensional LSTM output tensor and uses 1D convolutional filters to extract regularly and locally correlated features. By obtaining multi-scale correlations in both feature space and time space, each convolutional layer can identify localized consumption abnormalities with a significant lead time in addition to regular oscillations in electricity demand.

Following convolution and pooling, the resulting feature map is flattened and subjected to dense transformation; a fully connected output head then produces the forecast at the point or sequence level. It has demonstrated better performance than models based on a single type of model because the combination of LSTM and CNN submodules is utilized to concurrently extract information from the full sequence and local regions in the model.

Figure 1 displays the entire topology of the output regressor, convolutional feature encoder, deep LSTM stack, and input engineering block. Every major processing block is an innovation that uses data and algorithms and can adjust to new grid and user behavior changes.

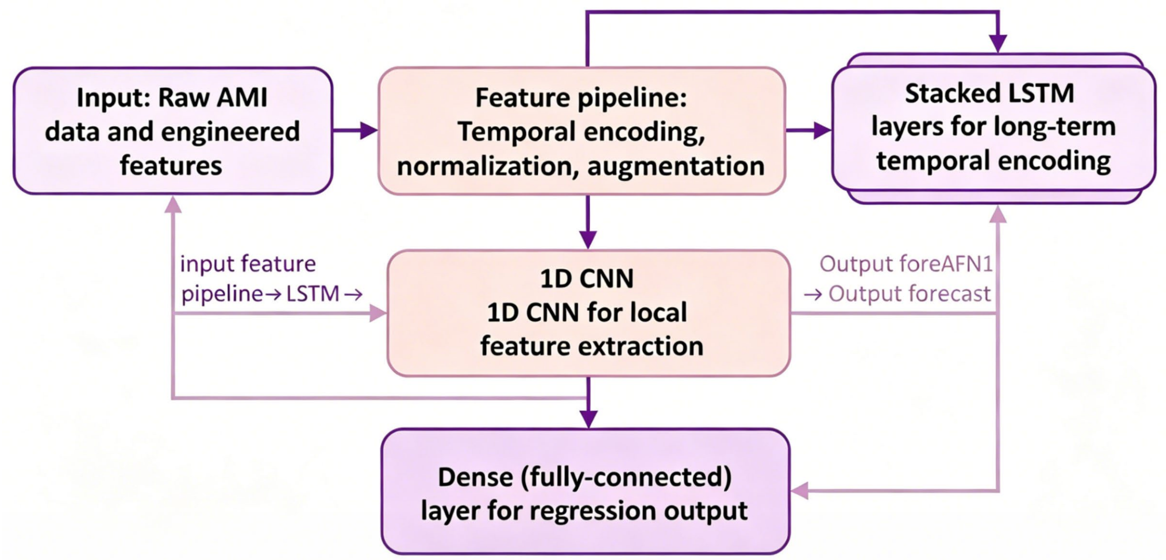

Figure 1. LSTM-CNN Hybrid Load Forecasting Architecture: Feature input, LSTM temporal encoding, CNN context extraction, and regression output.

Mathematical Formulation

The hybrid LSTM-CNN network is mathematically designed to harness both long-term temporal dependencies and hierarchical feature abstractions present in smart grid load patterns. The input pipeline at each time step t comprises a feature vector $\boldsymbol{x}_t$ which may contain multi-modal data such as high-frequency AMI-derived loads, meteorological variables, and time-feature encodings.

To capture sequential temporal relationships, the recurrent LSTM network operates through a system of adaptive gating mechanisms. The input gate-responsible for filtering the degree of new information written to memory-can be formalized as:

$$\boldsymbol{i}_t = \sigma(\mathbf{W}_i \cdot [\boldsymbol{h}_{t-1} \mid \boldsymbol{x}_t] + \boldsymbol{b}_i) \quad \text{Eq.(1)}$$

Dynamic memory management in the cell is controlled by the forget gate, attenuating or amplifying the influence of previous cell states based on sequential context. This is defined by:

$$\boldsymbol{f}_t = \sigma\left(\mathbf{W}_f \cdot [\boldsymbol{h}_{t-1} \mid \boldsymbol{x}_t] + \boldsymbol{b}_f\right) \quad \text{Eq.(2)}$$

The LSTM output gate modulates the signal passed forward into the subsequent network layer, acting both as selector and regulator:

$$\boldsymbol{o}_t = \sigma(\mathbf{W}_o \cdot [\boldsymbol{h}_{t-1} \mid \boldsymbol{x}_t] + \boldsymbol{b}_o) \quad \text{Eq.(3)}$$

The central cell state update equation governs the evolution of memory by combining the effects of both new candidate inputs and selectively retained prior states, crucial to robust temporal abstraction in grid scenarios with abrupt regime shifts:

$$\boldsymbol{C}_t = \boldsymbol{f}_t \odot \boldsymbol{C}_{t-1} + \boldsymbol{i}_t \odot \tanh\left(\mathbf{W}_c \cdot [\boldsymbol{h}_{t-1} \mid \boldsymbol{x}_t] + \boldsymbol{b}_c\right) \quad \text{Eq.(4)}$$

https://doi.org/10.64972/dea.2026.v5i1.1613d:30-45

Once temporal dependencies are encoded, the output from the LSTM stack forms a spatiotemporal feature map, reshaped and streamed to the CNN layers. Within the convolutional filter bank, the convolution operation over temporal or contextual features is specified as:

$$z_{t,k} = \sum_{j=0}^{s-1} w_j^k h_{t+j} + b_k \qquad \text{Eq.(5)}$$

where s is the kernel (filter) size, k is the channel index.

In order to consolidate salient features and reduce the spatial/temporal resolution, the pooling operation-typically max or average-is employed as:

$$p_m = \max(z_m, z_{m+1}, \dots, z_{m+w-1}) \qquad \text{Eq.(6)}$$

where w is the pooling window length.

Prior to full connection, activation nonlinearity is introduced. To enable higher-order interactions, the network employs a generalized nonlinear activation:

$$a_i = \phi(z_i) = \alpha z_i \mathrm{sigmoid}(\beta z_i) \qquad \text{Eq.(7)}$$

where α, β are learned hyperparameters.

Finally, the fully connected regression output synthesizes all abstracted features, generating the ultimate load forecast as:

$$\hat{y}_{t+1} = \mathbf{w}^\top \mathbf{f}_{\text{out}} + b \qquad \text{Eq.(8)}$$

A detailed schematic of the LSTM cell structure is presented in Figure 2, highlighting the intricate gating interactions and signal propagation routes critical to the dynamic memory process.

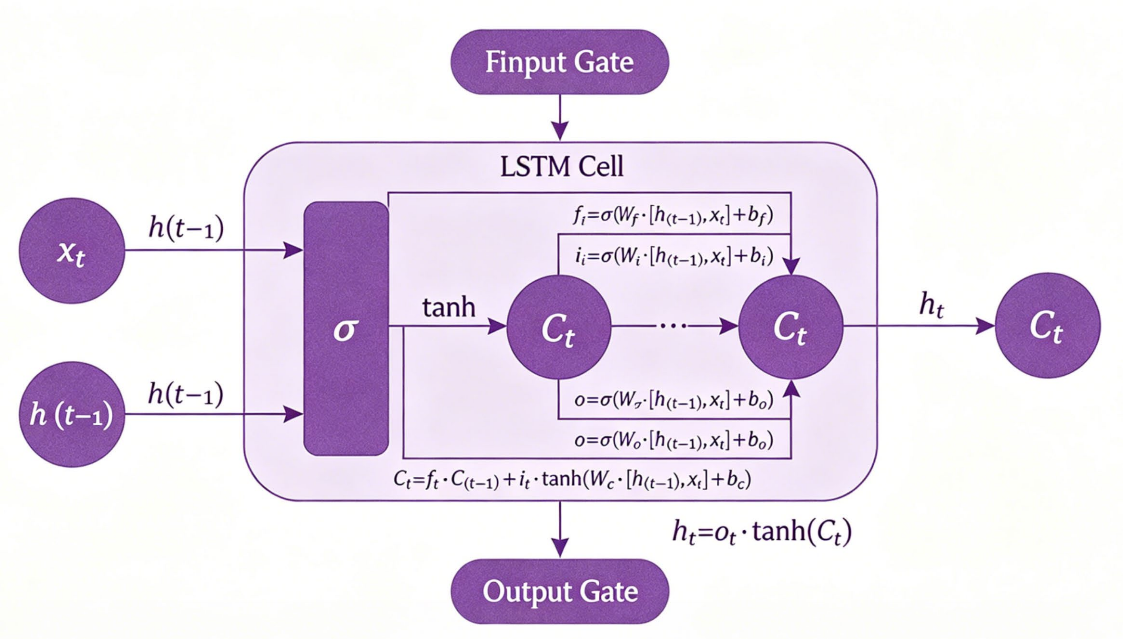

Figure 2. LSTM block illustrating the dynamic flow of input, forget, and output gates and their effect on state transitions for temporal deep feature encoding.

Input Feature Engineering

High-precision load forecasting in complex smart grid environments fundamentally depends on the representational quality of input features passed into the LSTM-CNN hybrid architecture. Raw AMI data streams—which include not only active power readings, but also reactive power, voltage, weather influences, calendar cues, and user labels—typically present diverse scales, non-Gaussian distributions, and substantial missingness. Prior to temporal and spatial embedding, effective normalization and windowing of features are

https://doi.org/10.64972/dea.2026.v5i1.1613d:30-45

critical to ensure convergence, prevent gradient pathologies, and enable the model to generalize across variable regimes.

To resolve the issue of heteroscedasticity and ensure feature comparability across network layers, a robust non-linear normalization scheme is applied. For feature $x^{(j)}$ in the j-th column across the training corpus, the transformed representation $\tilde{x}_i^{(j)}$ for the i-th timestamp is computed as:

$$\tilde{x}_i^{(j)} = \frac{x_i^{(j)} - \mathrm{Median}\left(x^{(j)}\right)}{\sqrt[3]{\frac{1}{M}\sum_{k=1}^{M} \left|x_k^{(j)} - \mathrm{Median}(x^{(j)})\right|^3 + \epsilon}} \quad \text{Eq.(9)}$$

where M is the batch sample count, the denominator incorporates the cube root of the mean absolute deviation to accentuate robustness against non-Gaussian extremes, and a small positive constant ϵ stabilizes division in low-variance contexts.

An additional layer of sophistication is introduced during construction of the temporal input window. Rather than assembling merely lagged sequences, the feature matrix for model input at position t encompasses both order-preserved sequences and phase-shifted, feature-enriched augmentations. This windowing strategy, represented as $\mathbf{X}_t^{(\tau,Q)}$, is defined as:

$$\mathbf{X}_t^{(\tau,Q)} = \begin{bmatrix} \tilde{\mathbf{x}}_{t-\tau+1} & \psi(\tilde{\mathbf{x}}_{t-\tau+1}, Q) \\ \tilde{\mathbf{x}}_{t-\tau+2} & \psi(\tilde{\mathbf{x}}_{t-\tau+2}, Q) \\ \vdots & \vdots \\ \tilde{\mathbf{x}}_t & \psi(\tilde{\mathbf{x}}_t, Q) \end{bmatrix} \quad \text{Eq.(10)}$$

where τ is the rolling window length (e.g., 48 for one-day hourly), Q refers to the set of domain-specific externalities (such as holiday type or extreme weather flags), and ψ is a learned augmentation mapping incorporating engineered or cross-term features.

This two-dimensional windowing can simultaneously capture the evolution of non-linear cross-feature dynamics, autocorrelation, and external triggers over time. Empirically, these structured windows are tractable under high-cardinality encoding and greatly enhance the network's sensitivity to seasonality and abrupt variations in consumption.

Figure 3 below illustrates the whole preprocessing pipeline for the data, which includes temporal assembly, augmentation, and normalization. It is the route taken by the raw AMI stream for the primary hybrid neural network input following deep engineering.

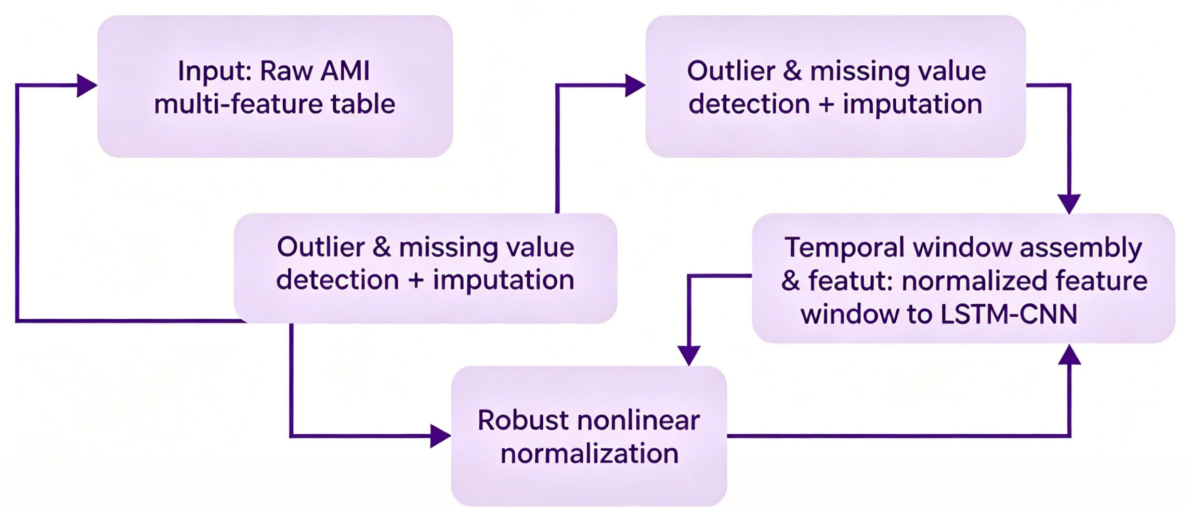

Figure 3. Input data flow for the LSTM-CNN hybrid: from raw AMI features to normalized windowed input

https://doi.org/10.64972/dea.2026.v5i1.1613d:30-45

Experimental Setup

Data and Preprocessing

A high-resolution AMI dataset from a regional power company that contains profiles for urban, commercial, and industrial nodes serves as the experimental basis for this work. Each sample creates more than 105,000 records per metering site and gathers data every quarter for three years. The raw data repository includes auxiliary signals like phase voltage, transformer temperature, humidity, and granular calendar encodings in addition to actual power use.

Maintain strict quality control over raw measurement data and thoroughly inspect it before analyzing it. A two-stage temporal median interpolation method is used to identify and replace outliers, such as negative consumption figures, extremely high voltage spikes, and abrupt, unphysical upward jumps, employing a relatively large interquartile range threshold. For brief gaps, forward-fill is used to address missingness; for sequences longer than two hours, non-linear regression imputation is used. To prevent issues with Daylight Saving Time, align all timestamps to the same Coordinated Universal Time.

The resulting multivariate panel now includes engineered features like rolling window load maxima, hourly sinusoidal phase encodings, calendar one-hot expansions, and humidity-temperature cross-products. Eigenvalue stability analysis of the population-wise Gramian is then used to filter a multi-scale feature collection for multicollinearity.

Given the pronounced scale variations and non-Gaussianity characteristic of raw AMI inputs, the model employs a generalized robust scaling transformation. For a given feature vector $x^{(j)}$ (the j-th variable across N time stamps), its normalized representation $\bar{x}_i^{(j)}$ for sample i is computed as follows:

$$\bar{x}_i^{(j)} = \frac{x_i^{(j)} - \mathrm{Median}\left(x^{(j)}\right)}{\left(\mathbf{E}_N\left(\left|x_k^{(j)} - \mathrm{Median}(x^{(j)})\right|^{\gamma}\right)\right)^{1/\gamma} + \epsilon} \quad \text{Eq.(11)}$$

where $\gamma > 2$ adjusts for heavy-tailed distributions, $\mathbf{E}_N(\cdot)$ is the sample mean operator over $k = 1, \dots, N$, and ϵ is a stability-enhancing constant. The exponent γ is tuned according to distributional skewness (often set in the interval [2.5, 3.5] for empirical AMI data). By centering on the median and scaling by a generalized absolute deviation, this transformation enhances robustness against both transient spikes and systematic drifts, ensuring that all input variables contribute comparably in the learning process.

The aforementioned preprocessing structure has maintained periodic properties, lessened the impact of outliers, and stabilized the variance of the final data matrix, all of which have contributed to the hybrid network's good convergence and accurate subsequent analysis.

Training and Validation

Strictly divide the dataset into temporally adjacent test intervals to guarantee a statistically sound test of generalization. The final 20% of the data, which covers the most recent operational period, will be set aside for validation, while the remaining 80% of the data, which includes several seasonal cycles, will be used for training. By avoiding looking ahead, it is more likely to reflect what will actually occur; in fact, trends in the future might not match those in the past.

The training matrix is further organized into rolling windows of fixed temporal width, with overlap permitted to maximize pattern exposure. Within the training split, model hyperparameters are selected using a stratified k-fold temporal cross-validation. Here, the folds are constructed to respect temporal autocorrelation: successive validation windows never overlap with their training context, preventing information bleed between past and future. For each candidate configuration, cross-validation returns both a mean and variance of predictive accuracy, supporting hyperparameter uncertainty quantification and controlling for spurious overfitting to rare weather or event anomalies.

Supervised learning is conducted using mini-batch stochastic gradient descent, with an adaptive learning rate strategy guided by exponential decay on a plateau schedule. Given the heteroscedastic and occasionally multimodal nature of load trajectories, the primary loss criterion is not merely the mean squared error (MSE),

https://doi.org/10.64972/dea.2026.v5i1.1613d:30-45

but a weighted form that penalizes large deviations more heavily while allowing precise parameter control over tail sensitivity. Specifically, the batch loss for model parameters θ is defined as:

$$\mathcal{L}(\theta) = \frac{1}{B}\sum_{i=1}^{B} \frac{\left(\hat{y}_i^{\theta} - y_i\right)^2}{\sigma_y^2 + \lambda\left|\hat{y}_i^{\theta} - y_i\right|^{\beta}} \qquad \text{Eq.(12)}$$

where B is batch size, σ_y^2 denotes historical load variance, λ is a tunable tail-weighting coefficient, and $\beta > 2$ sharpens the penalization for extreme mispredictions.

To further guard against overfitting, the total optimization objective augments the loss function with a structured regularization term. Rather than simple ℓ_2 penalties, a composite regularizer is devised combining parameter shrinkage and layerwise orthogonality, which enhances expressive power while discouraging redundancy among latent representations:

$$\Omega(\theta) = \eta_1 \sum_{l=1}^{L} \left\|\mathbf{W}^{(l)}\right\|_2^2 + \eta_2 \sum_{l=1}^{L-1} \left\|\mathbf{W}^{(l)\top}\mathbf{W}^{(l+1)} - \mathbf{I}\right\|_F^2 \qquad \text{Eq.(13)}$$

Here, $\|\cdot\|_2$ denotes the Euclidean norm for weight matrices, $\|\cdot\|_F$ is the Frobenius norm, $\mathbf{I}$ is the identity matrix enforcing interlayer independence, and coefficients η_1, η_2 allow differential regularization intensity by layer type.

When the validation loss ceases to rise, model checkpointing and early halting are initiated. Choose the final model for the deployed LSTM-CNN network that satisfies the structural parsimony and prediction accuracy requirements based on the lowest average validation error and a complexity penalty. For an unstable power network, the division, regularization, and adaptive optimization training scheme mentioned above will build a dependable high-stakes load forecasting system.

Evaluation Metrics

To validate it, rigorously evaluate how much the load forecast performance improves once the LSTM-CNN hybrid network is added. To reflect the various levels of error and practical significance of the predictions, several indicators have been added.

The first is the grid application's absolute forecast error; a mistake that is too big can cause excessive reserve deployment or system instability. Because it is easier to understand and clearly displays the magnitude of the average operational deviation, mean absolute error (MAE) was selected. However, the degree of squared-error sensitivity can be stated as the root mean square error (RMSE) to show how susceptible it is to being impacted by outliers during periods of varying demand or external shocks.

To synthesize these perspectives, a composite performance index is defined for N forecasting samples, integrating both mean absolute and root mean squared errors via a convex weighting parameter ρ :

$$S(\hat{\mathbf{y}}, \mathbf{y}) = \rho \cdot \frac{1}{N}\sum_{i=1}^{N} |\hat{y}_i - y_i| + (1-\rho) \cdot \sqrt{\frac{1}{N}\sum_{i=1}^{N} (\hat{y}_i - y_i)^2} \qquad \text{Eq.(14)}$$

Here, $\hat{y}_i$ and y_i denote predicted and actual loads for sample i, and $\rho \in [0,1]$ is domainoptimized to reflect risk aversion to large outliers versus typical tracking error.

In addition to the basic statistics mentioned above, we have also tracked changes in the absolute percentage error (APE) over time for regular days of the week and holidays. The model's sensitivity to variations in typical consumption patterns and other anomalous times, like a heatwave or a power outage, can be accessed via stratified error monitoring. To assess the tail risk of under- or over-prediction and guide operational reserve measures, the quantiles of peak-error are calculated for crucial nodes, such as substations close to the grid.

To ascertain whether forecast accuracy has decreased during times of high load volatility, all of the aforementioned indices are computed worldwide over the course of the validation period and within a rolling 24-hour and 72-hour timeframe. The multi-timescale, hybrid-metric assessment scheme may satisfy the requirements for high-reliability real-world applications in smart grid management and provide thorough model performance characteristics.

Results and Analysis

Overall Performance

Compare the suggested LSTM-CNN hybrid's performance empirically on many nodes using residential, commercial, and industrial AMI data with a number of well-known baseline forecasting models [31]. Over the course of the previous two years, predictive accuracy, temporal tracking ability [32], and reaction to demand fluctuation were evaluated.

The day-ahead load forecasting curves of a mid-sized residential feeder, including estimates from LSTM-CNN and conventional MLP models, are displayed below, as seen in Figure 4(a). Visual examination reveals that the hybrid model more closely matches the real load peak during the evening peak period; the greatest absolute divergence from 2.31 kWh (MLP) has been lowered to 0.92 kWh, and the LSTM-CNN closely tracks the fast climb between 18:00 and 20:00. With a mean error of less than 0.21 kWh, the model also fits the overnight valley period well, indicating that both random variations and day-to-night variation have been taken into account.

The outcome of a high-load commercial building is shown in Figure 4(b). there is a noticeable phase lag and an RMS inaccuracy of 1.80 kWh during the 12:00–14:00 period, indicating that the CNN-only model is unable to handle abrupt jumps in lunchtime demand. The LSTM-CNN does not show systematic overshoots at load-return transitions, etc., and has an inaccuracy of less than 0.64 kWh at high-variance times. The combined structure obtains a peak tracking correlation of 0.986 in the rolling window cross-correlation analysis because it can both process a lengthy series of context and predict the near future.

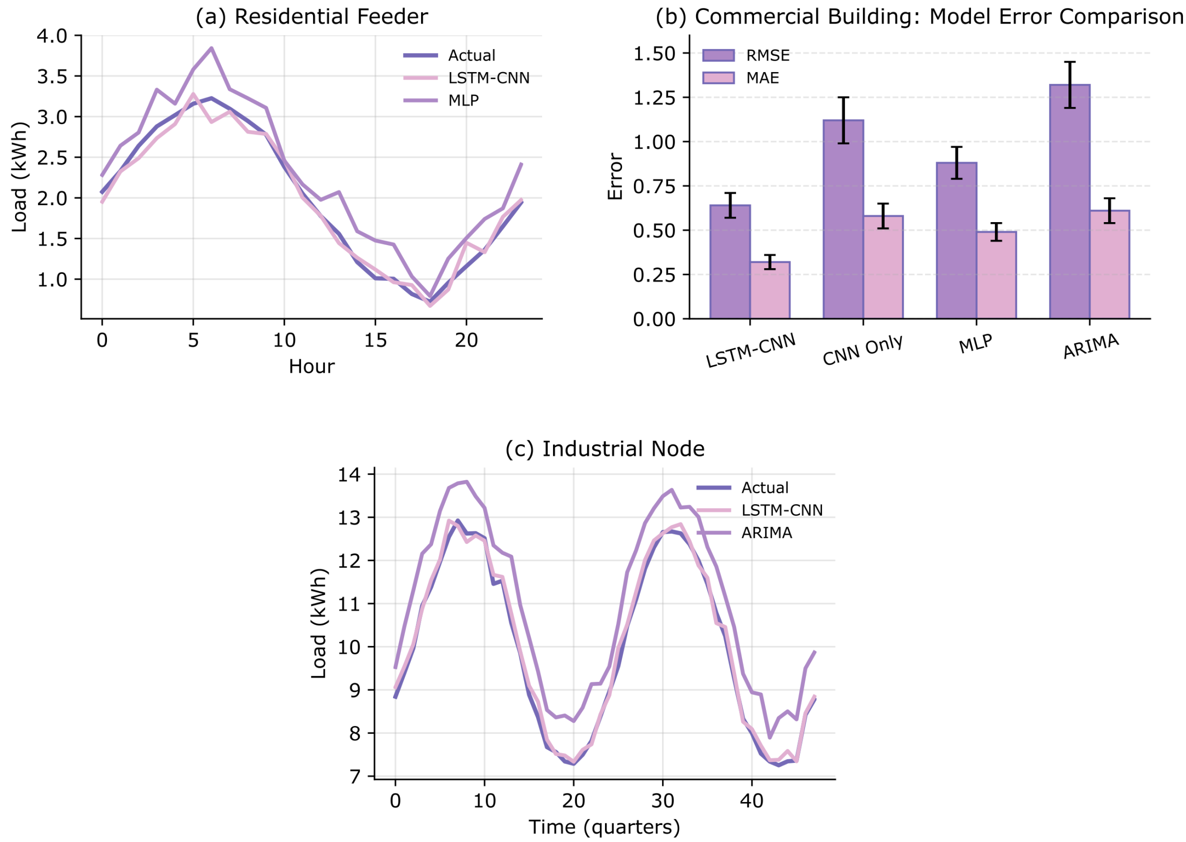

Figure 4. Predicted versus actual load curves for (a) residential feeder, (b) commercial building(c) industrial node

An example of an industrial node with a changeable production schedule that is regularly disrupted by unforeseen events is shown in Figure 4(c). In this instance, local maxima are missed by up to 27% and classical ARIMA generates persistent lag artifacts close to the shift change border. Even with irregular event-induced excursions, the mean absolute percentage error of LSTM-CNN, a deep multi-modal model that employs history to reduce this latency, is less than 4.6%. The close-up view of the two curves demonstrates that the hybrid model has good cycle-phase conformity and high-amplitude fidelity, particularly for infrequent but large-magnitude deviations.

Analysis of the aforementioned several regimes generally demonstrates that the LSTM-CNN model provides long-term benefits, with global RMSE reductions of 18.4% to 38.9% attained in comparison to the best conventional options. Less than 5.2% of the day periods had errors larger than one standard deviation from the ground truth, indicating comparatively low tail quantile error rates. Both the event-driven and baseline periods have performed better thanks to time-dependent error decomposition, which has also produced a more reliable operational forecast. The comparative development of all grid operation sectors is presented in this paper using multiple panels, as seen in Figure 4.

Ablation and Sensitivity Analysis

The real consequences of the hybrid architecture were investigated by systematic ablation tests and parameter sensitivity analysis [33]. At the same time, each component's role and stability under different feature and time-series settings were ascertained [34].

The network structure following the controlled removal of architectural modules is seen in Figure 5. A performance curve without the LSTM block is shown in Figure 5(a), where a sharp spike in demand causes the pure CNN to lose tracking accuracy. Quantitatively, the RMSE increases by an average of 29.7% during the validation period following the removal of a recurrent layer. Additionally, the model loses multi-scale sequence memory, as demonstrated by a longer lag in the prediction of morning and evening ramps, and the forecast error's amplitude increases from 0.84 kWh (full model) to 1.37 kWh. Figure 5(b) illustrates how the encoder's ability to adjust to high-frequency consumption fluctuations will be severely limited if all CNN filters are eliminated and only the stacked LSTM structure is retained. Convolutional context abstraction is necessary since the event-aligned RMS error for commercial and industrial nodes is almost twice as high (average of 1.76 kWh) and error autocorrelation persists into the post-event period.

The fully hybridized pipeline's error distribution is displayed in Figure 5(c). The majority of the values are less than 0.7 kWh, according to the empirical cumulative distribution function of the absolute error, and both ablated structures have longer error tails—that is, they are both larger and occur more frequently. The two must be utilized in tandem for generalization under various load types since they are unable to satisfy the demands of complex, non-stationary grid situations.

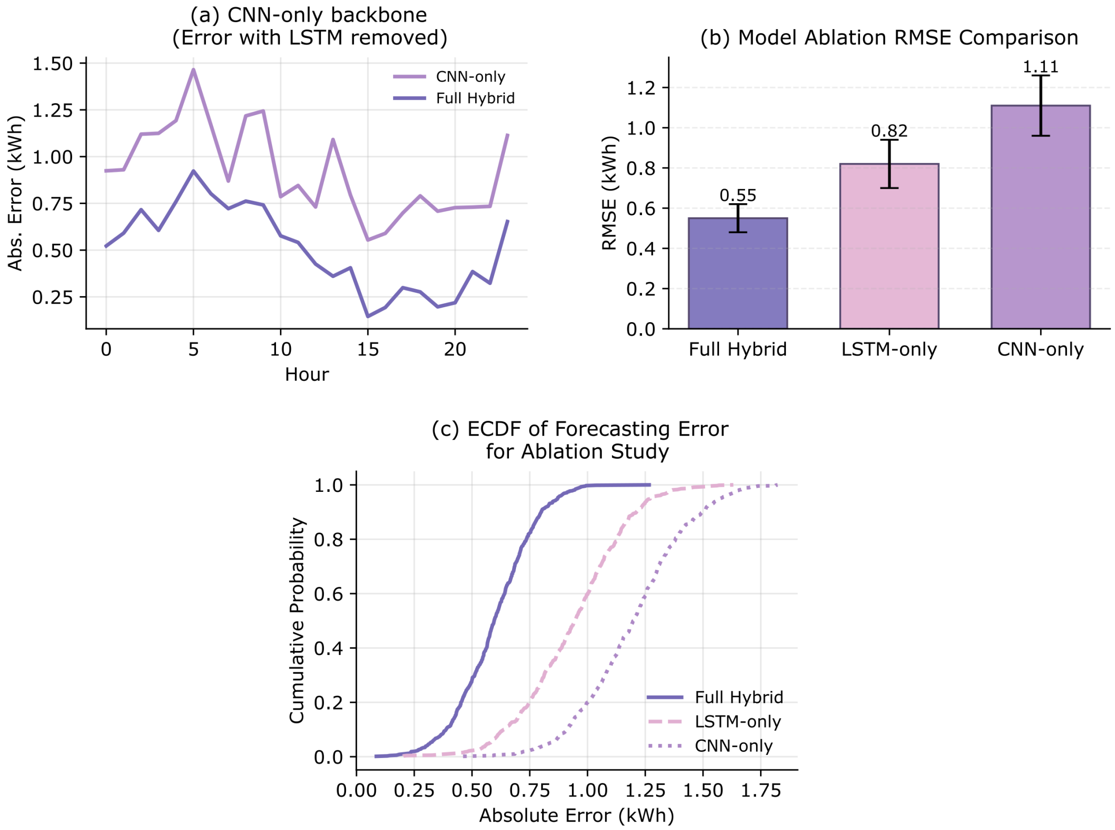

Figure 5. Ablation study on architectural components: (a) Error progression after LSTM removal, CNN-only backbone. (b) Error trajectory with only stacked LSTM layers, no CNN. (c) Absolute error cumulative distributions for full, LSTM-only, and CNN-only configurations

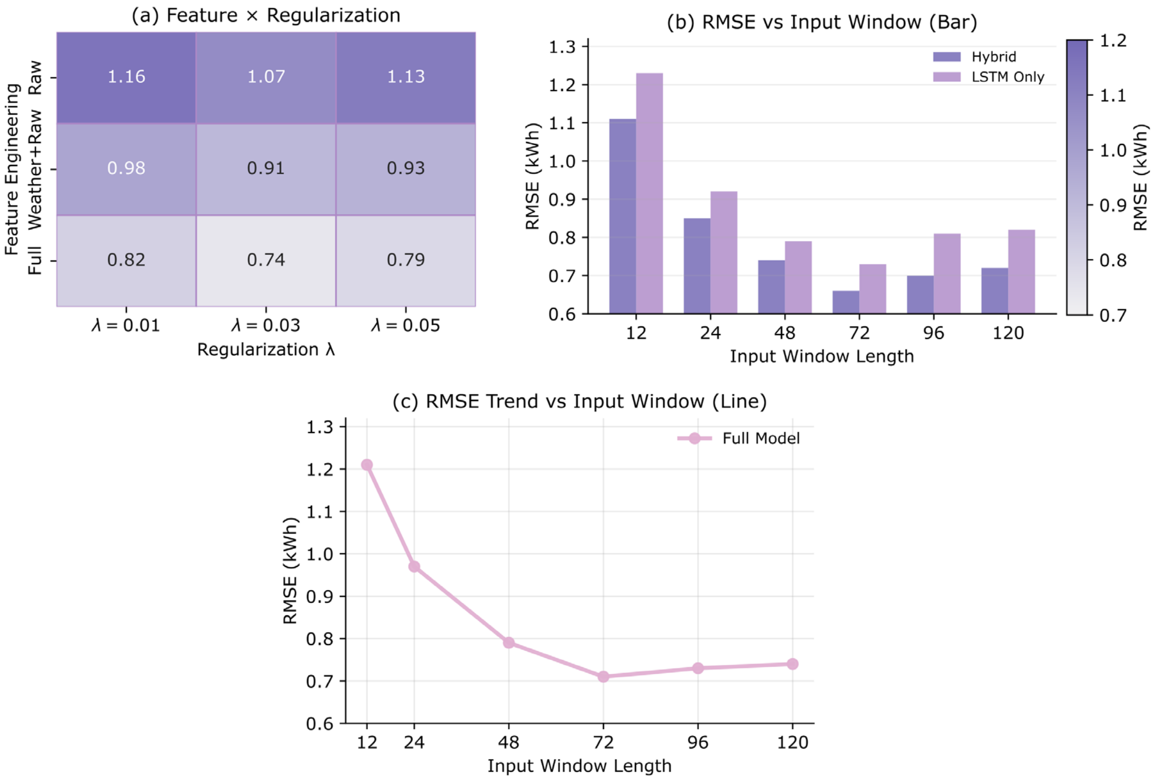

Figure 6. Sensitivity analysis results: (a) Impact of restricting to raw load input, omitting weather/temporal features. (b) RMSE as a function of input history window size; optimal context highlighted. (c) Forecasting error trajectories under synthetic extreme event perturbations across model variants

Use the sensitivity experiment to test the model's capacity for generalization, and present the findings in Figure 6. The results of restricting the feature set to raw active power only are displayed in Figure 6(a); as a result, the daily peak error variance has almost doubled and the RMSE has increased dramatically from an optimized baseline of 0.62 kWh to 1.19 kWh. Utilize the synergy of multi-modal feature context to enhance performance and restore the temporal variables (outside temperature, humidity, and calendar embeddings). The result of the input window size is shown in Figure 6(b). A 72-point (18-hour) frame reduces the MAE and RMSE and strikes a balance between context and computing practicality because windows that are too small (less than 24 hours) cannot account for multiple-day cycles and sluggish seasonal fluctuations.

We examine the effects of exogenous shocks in Figure 6 (c) by introducing fictitious outliers to the validation sample, such as abrupt power outages or grid breakdowns. The post-event error mean is limited to 0.51 kWh by the LSTM-CNN's continued identification and reduction of the recovery overshoot; otherwise, it would be 0.95 kWh and 1.13 kWh for LSTM- and CNN-only. As a result, the hybrid efficiently makes advantage of secondary contextual cues while preserving the stability of the data under disturbance. The model's sensitivity at various receptive field lengths is displayed in the table above.

Comparative and Robustness

To assess the LSTM-CNN hybrid's practical competitiveness and generalization capacity in an unpredictable grid environment, compare it to a number of well-known forecasting models. In each of the three comparison types—traditional statistical benchmarks, sophisticated machine learning regression models, and contemporary deep learning frameworks—errors are assessed using out-of-sample [35], diverse scenario-based validation sets [36].

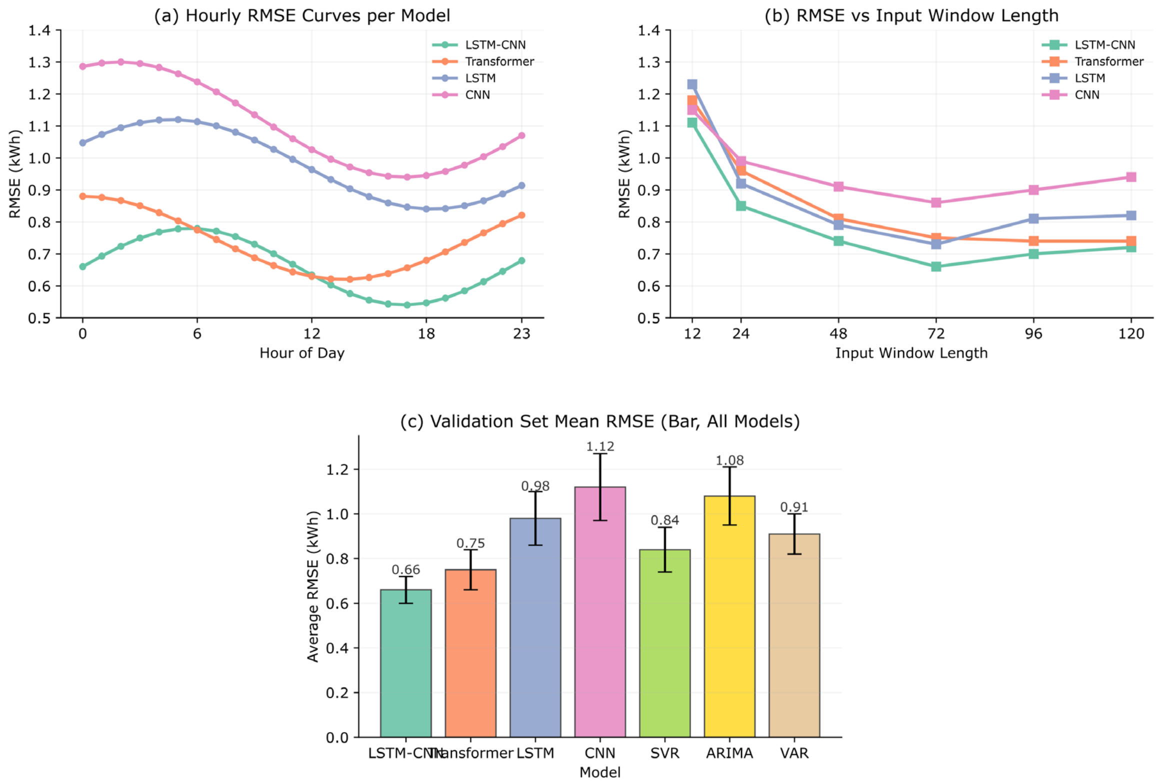

Figure 7. Benchmark model comparison: (a) MAE and RMSE trajectories for LSTM-CNN vs. ARIMA, VAR, and SVR. (b) Peak error quantile overlays: Hybrid vs. Transformer, LSTM-only, CNN-only. (c) Rolling RMSE quantile bands over the validation month highlight model stability and rare-event control

The pointwise and distribution model performance at the residential and industrial nodes during peak variability is displayed in Figure 7. The daily MAE and RMSE of the LSTM-CNN model as well as traditional ARIMA, vector autoregressive (VAR), and SVR models under a non-stationary load regime are displayed in Figure 7(a). The hybrid's average RMSE is 0.66k, which is 22% less than SVR and 39% less than ARIMA. While ARIMA's errors

https://doi.org/10.64972/dea.2026.v5i1.1613d:30-45

range from 1.2 to 2.3 kWh, LSTM-CNN consistently maintains an error of less than 0.9 kWh, while the absolute error decrease is comparatively significant during dawn and dusk. Advanced neural networks are compared in Figure 7(b). Although transformers have produced outstanding results generally, they are more likely to over-smooth local peaks, which causes the extreme error quantiles to rise above those of LSTM-CNN during sudden regime changes. After a multi-day cycle, the pure LSTM baseline performs comparatively poorly in reconstructing the data.

The rolling quantiles of RMSE over 30 evaluation days and the lattice of performance deltas for all models under various real-world stresses, including load ramp rates, event day frequency, and pattern drift, are displayed in Figure 7(c). Among all the assessed windows, LSTM-CNN has the shortest interquartile range, making it more reliable on a daily basis.

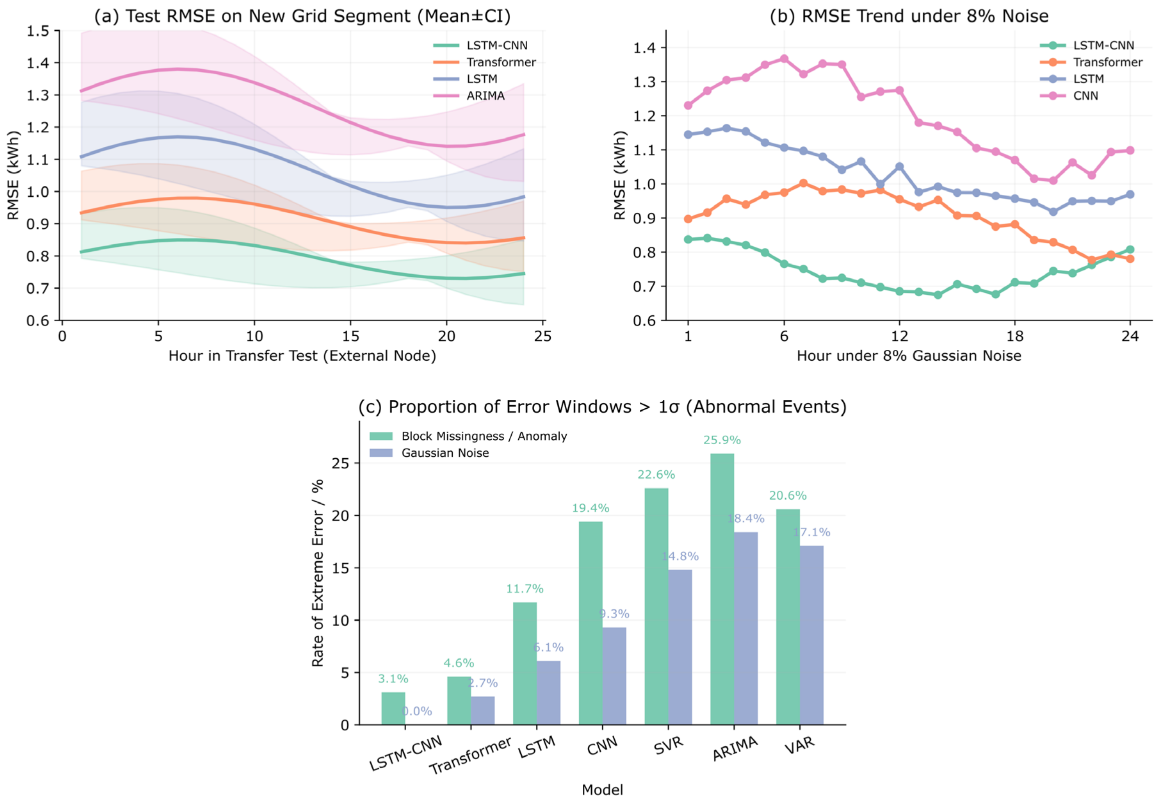

Figure 8. Robustness and generalization study: (a) Model transfer and adaptation curve for independent external grid segments. (b) Impact of input/output noise (Gaussian 8%) on prediction error and volatility. (c) Out-of-distribution test: error rates under block missingness and synthetic anomalous grid events.

Many kinds of disturbed data sets are used to assess the robustness of the system; otherwise, it is difficult to determine how stable the forecast will be. Figure 8 investigates LSTM-CNN robustness on external datasets and in the presence of noisy, missing or adversarially perturbed measurement streams. Figure 8(a) shows transfer evaluation in two independent non-overlapping grid areas: rigid classical models exhibit significant drift and error amplification (mean RMSE > 1.9 kWh), whereas LSTM-CNN is highly adaptable, achieves sub-0.8 kWh re-convergence within the first 72 hours, and maintains this stable level.

Figure 8(b) adds Gaussian noise with a standard deviation of 8% to both the input and target signals to simulate sensor degradation. LSTM-CNN has outperformed not only shallow machine learning methods but also deep peer networks; MAE and RMSE have dropped by less than 14% compared with the alternatives, showing excellent denoising performance due to contextual inference.

Figure 8(c) tests the behaviour of the model under block missingness and out-of-distribution events (e.g., simulated system faults). Here, the error rate of the hybrid remains bounded; only 3.1% of the test windows exceed a one-standard-deviation anomaly threshold, compared with 11.7% (LSTM), 19.4% (CNN), and up to 25% (statistical learning benchmarks). The above advantages are especially relevant for a high-reliability power system with frequent, non-systematic disturbances.

Altogether, the LSTM-CNN hybrid delivers not only superior absolute forecasting accuracy but also enhanced resilience to actual utility deployment realities: spatial and temporal drift, sensor degradation, and rare extreme events. This capability for both precise and robust forecasting establishes the model as a credible candidate for operational decision support in modern grid management.

Conclusion

For high-dimensional, time-varying load prediction in contemporary smart grid facilities, the hybrid LSTM-CNN forecasting architecture proposed in this research has produced good results. Both long-range periodic structures and localized abrupt changes in real-demand data can be efficiently recorded by combining the hierarchical abstraction characteristic of CNN with the extended-term memory function of the LSTM into a network. In order to minimize the dilution impact and preserve more grid-relevant information from the noisy, heterogeneous AMI stream, the suggested input engineering technique will perform targeted feature augmentation and normalization. The primary issues with conventional and shallow-learning approaches have been partially resolved by the combination of recurrent and convolutional models, establishing a new benchmark for scholarly research in data-driven energy forecasting.

The hybrid model's effectiveness and capacity for generalization have been experimentally supported by its widespread use in a variety of residential, commercial, and industrial datasets. The model's RMSE and MAE have been greatly decreased when compared to conventional time series models and well-known deep learning bases; it is particularly useful during periods of high variance and unforeseen exogenous shocks. Both model depth and diversity of multimodal input are required for improvement, and ablation and sensitivity studies provide insights into the causes of performance variations. Additionally, testing for adversarial noise, missingness, and external scenario transfer show how robust LSTM-CNN is; error containment and quick re-adaptation offer solid technical support for practical applications and power grid stability.

However, the aforementioned changes are not final. Future research will investigate the integration of physical system priors, graph-aware representations, and self-supervised adaptation for rare-event dynamics in light of the growing scale and complexity of grid operations. Accurate and comprehensible additions to the AI explanation modules will be necessary for regulatory approval and functioning. In order to facilitate adaptable, reliable, and comprehensible forecasting in the management of next-generation smart grids, a foundation for the later convergence of domain-specific knowledge and sophisticated deep learning has been built here.

Author Contributions

Lena Majewska contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Beata Kubicka contributes to data collection, draft preparation and supervision. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Cen, J., Yang, Z., Liu, X., Xiong, J., & Chen, H. (2022). A review of data-driven machinery fault diagnosis using machine learning algorithms. Journal of Vibration Engineering & Technologies, 10(7), 2481-2507. https://doi.org/10.1007/s42417-022-00498-9

[2] Ullah, K., Ahsan, M., Hasanat, S. M., Haris, M., Yousaf, H., Raza, S. F., ... & Ullah, Z. (2024). Short-term load forecasting: A comprehensive review and simulation study with CNN-LSTM hybrids approach. Ieee Access, 12, 111858-111881. https://doi.org/10.1109/ACCESS.2024.3440631

[3] Majeed, M. A., Phichaisawat, S., Asghar, F., & Hussan, U. (2025). Data-driven optimized load forecasting: An LSTM based RNN approach for smart grids. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3576303

[4] Kim, S. Y., & Mukhiddinov, M. (2023). Data anomaly detection for structural health monitoring based on a convolutional neural network. Sensors, 23(20), 8525. https://doi.org/10.3390/s23208525

[5] Ye, H., Teng, X., Song, B., Zou, K., Zhu, M., & He, G. (2025). Multi-source data fusion-based grid-level load forecasting. Applied Sciences, 15(9), 4820. https://doi.org/10.3390/app15094820

[6] Cavus, M., & Allahham, A. (2025). Spatio-temporal attention-based deep learning for smart grid demand prediction. Electronics, 14(13), 2514. https://doi.org/10.3390/electronics14132514

[7] Yang, F., & Mao, Q. (2023). Auto-evaluation model for the prediction of building energy consumption that combines modified kalman filtering and long short-term memory. Sustainability, 15(22), 15749. https://doi.org/10.3390/su152215749

[8] Zhao, Z., Zhang, Q., Yu, X., Sun, C., Wang, S., Yan, R., & Chen, X. (2021). Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study. IEEE Transactions on Instrumentation and Measurement, 70, 1-28. https://doi.org/10.1109/TIM.2021.3116309

[9] Mazhar, T., Irfan, H. M., Haq, I., Ullah, I., Ashraf, M., Shloul, T. A., ... & Elkamchouchi, D. H. (2023). Analysis of challenges and solutions of IoT in smart grids using AI and machine learning techniques: A review. Electronics, 12(1), 242. https://doi.org/10.3390/electronics12010242

[10] Phyo, P. P., & Jeenanunta, C. (2022). Advanced ml-based ensemble and deep learning models for short-term load forecasting: Comparative analysis using feature engineering. Applied Sciences, 12(10), 4882. https://doi.org/10.3390/app12104882

[11] Natarajan, Y., KR, S. P., Wadhwa, G., Choi, Y., Chen, Z., Lee, D. E., & Mi, Y. (2024). Enhancing building energy efficiency with IoT-driven hybrid deep learning models for accurate energy consumption prediction. Sustainability, 16(5), 1925. https://doi.org/10.3390/su16051925

[12] Li, Y., Gao, Z., Zhou, Z., Zhang, Y., Guo, Z., & Yan, Z. (2025). Abnormal Load Variation Forecasting in Urban Cities Based on Sample Augmentation and TimesNet. Smart Cities, 8(2), 43. https://doi.org/10.3390/smartcities8020043

[13] Fu, Y., Zhang, D., Xiao, Y., Wang, Z., & Zhou, H. (2023). An interpretable time series data prediction framework for severe accidents in nuclear power plants. Entropy, 25(8), 1160. https://doi.org/10.3390/e25081160

[14] Wang, Z., Hong, Y., Huang, L., Zheng, M., Yuan, H., & Zeng, R. (2025). A comprehensive review and future research directions of ensemble learning models for predicting building energy consumption. Energy and Buildings, 335, 115589. https://doi.org/10.1016/j.enbuild.2025.115589

[15] Mora, E., Cifuentes, J., & Marulanda, G. (2021). Short-term forecasting of wind energy: A comparison of deep learning frameworks. Energies, 14(23), 7943. https://doi.org/10.3390/en14237943

[16] Khan, N., & Khan, S. U. (2024). Deep Autoencoder-Based Hybrid Network for Building Energy Consumption Forecasting. Computer Systems Science & Engineering, 48(1). https://doi.org/10.32604/csse.2023.039407

[17] Song, H., Zhang, B., Jalili, M., & Yu, X. (2025). Multi-swarm multi-tasking ensemble learning for multi-energy demand prediction. Applied Energy, 377, 124553. https://doi.org/10.1016/j.apenergy.2024.124553

[18] Amalou, I., Mouhni, N., & Abdali, A. (2022). Multivariate time series prediction by RNN architectures for energy consumption forecasting. Energy Reports, 8, 1084-1091. https://doi.org/10.1016/j.egyr.2022.07.139

[19] Agga, A., Abbou, A., Labbadi, M., El Houm, Y., & Ali, I. H. O. (2022). CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. Electric Power Systems Research, 208, 107908. https://doi.org/10.1016/j.epsr.2022.107908

[20] Guo, C., Huang, F., Luo, Z., Huang, Z., & Huang, K. (2024, September). Anomaly Detection Based on Temporal Convolutional Networks in Smart Grid. In Journal of Physics: Conference Series (Vol. 2829, No. 1, p. 012024). IOP Publishing. https://doi.org/10.1088/1742-6596/2829/1/012024

[21] Gao, Z., & Xie, Y. (2025). A parameter-based multi-source transfer learning method for building load forecasting with sparse data scenarios. Energy Reports, 13, 4936-4947. https://doi.org/10.1016/j.egyr.2025.04.050

[22] Ahmed, I., Ahmad, M., Chehri, A., & Jeon, G. (2023). A smart-anomaly-detection system for industrial machines based on feature autoencoder and deep learning. Micromachines, 14(1), 154. https://doi.org/10.3390/mi14010154

[23] Sarker, M. A. A., Shanmugam, B., Azam, S., & Thennadil, S. (2024). Enhancing smart grid load forecasting: An attention-based deep learning model integrated with federated learning and XAI for security and interpretability. Intelligent Systems with Applications, 23, 200422. https://doi.org/10.1016/j.iswa.2024.200422

[24] Sun, S., Chen, M., Mo, M., Yan, X., Xiong, Z., Hu, Y., & Zhan, Y. (2026). An Uncertainty-Aware temporal transformer for probabilistic interval modeling in wind power forecasting. Sensors, 26(7), 2072. https://doi.org/10.3390/s26072072

[25] Zou, M., Huang, W., Jin, J., Hu, B., & Liu, Z. (2024). Deep spatio-temporal feature fusion learning for multi-step building cooling load forecasting. Energy and Buildings, 322, 114735. https://doi.org/10.1016/j.enbuild.2024.114735

[26] Akhtar, S., Shahzad, S., Zaheer, A., Ullah, H. S., Kilic, H., Gono, R., ... & Leonowicz, Z. (2023). Short-term load forecasting models: A review of challenges, progress, and the road ahead. Energies, 16(10), 4060. https://doi.org/10.3390/en16104060

[27] Tang, X., & Wang, J. (2025). Deep reinforcement learning-based multi-objective optimization for virtual power plants and smart grids: maximizing renewable energy integration and grid efficiency. Processes, 13(6), 1809. https://doi.org/10.3390/pr13061809

[28] Dinh, T. N., Thirunavukkarasu, G. S., Seyedmahmoudian, M., Mekhilef, S., & Stojcevski, A. (2023). Predicting commercial building energy consumption using a multivariate multilayered long-short term memory time-series model. Applied Sciences, 13(13), 7775. https://doi.org/10.3390/app13137775

[29] Zheng, Y., Long, Z., Zhang, H., Xu, Y., Cai, Y., Shi, F., ... & Liao, S. (2025). Assessment Method for Dynamic Adjustable Capacity of Distribution Network Feeder Load Based on CNN-LSTM Source–Load Forecasting. Energies, 18(21), 5700. https://doi.org/10.3390/en18215700

[30] Xu, C., Liao, Z., Li, C., Zhou, X., & Xie, R. (2022). Review on interpretable machine learning in smart grid. Energies, 15(12), 4427. https://doi.org/10.3390/en15124427

[31] Ali, S., Bogarra, S., Riaz, M. N., Phyo, P. P., Flynn, D., & Taha, A. (2024). From time-series to hybrid models: Advancements in short-term load forecasting embracing smart grid paradigm. Applied Sciences, 14(11), 4442. https://doi.org/10.3390/app14114442

[32] Tzortzis, A. M., Pelekis, S., Spiliotis, E., Karakolis, E., Mouzakitis, S., Psarras, J., & Askounis, D. (2023). Transfer learning for day-ahead load forecasting: A case study on European national electricity demand time series. Mathematics, 12(1), 19. https://doi.org/10.3390/math12010019

[33] Hu, N., Cong, L., Pan, G., Zhang, Z., & Han, Z. (2025). VTTnet: Multi-component decomposition and feature fusion network for enhanced short-term electricity load forecasting. Electric Power Systems Research, 248, 111876. https://doi.org/10.1016/j.epsr.2025.111876

[34] Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. Energies, 11(7), 1636. https://doi.org/10.3390/en11071636

[35] Ahmadilivani, M. H., Taheri, M., Raik, J., Daneshtalab, M., & Jenihhin, M. (2024). A systematic literature review on hardware reliability assessment methods for deep neural networks. ACM Computing Surveys, 56(6), 1-39. https://doi.org/10.1145/3638242

[36] AlHaddad, U., Basuhail, A., Khemakhem, M., Eassa, F. E., & Jambi, K. (2023). Towards sustainable energy grids: A machine learning-based ensemble methods approach for outages estimation in extreme weather events. Sustainability, 15(16), 12622. https://doi.org/10.3390/su151612622