

Performance Analysis of a Bayesian Regularization Neural Network for Short-Term Electricity Load Forecasting

Ahmad Fauzi*, Tarmizi Tarmizi, Khairun Saddami,
Maulisa Oktiana, and Ramzi Adriman

Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Indonesia
ringsigli2@gmail.com

Abstract

Short-term electricity load forecasting (STLF) is important for supporting reliable and efficient power system operation. Bayesian Regularization Neural Network (BRNN) can model nonlinear relationships in electricity load data while controlling model complexity. This study analyzes the performance of a previously developed BRNN model by examining the effects of network architecture and forecasting horizon. Four input variables, namely hour, air temperature, air humidity, and previous-day electricity load, were used to forecast the following-day electricity load. Five BRNN configurations with 5, 10, 15, 20, and 25 hidden units were evaluated using Mean Squared Error (MSE) and correlation coefficient (R), while forecasting performance was assessed using Mean Absolute Percentage Error (MAPE) for horizons $h+1$ to $h+7$. The 4-10-1 architecture achieved the lowest training MSE of 0.000933 and was selected as the reference configuration based on the predefined minimum-training-MSE criterion. The testing R values ranged from 0.930 to 0.952 across the evaluated architectures. For the selected configuration, MAPE values across $h+1$ to $h+7$ were 8.46%, 7.76%, 8.16%, 5.28%, 2.45%, 8.02%, and 6.11%, respectively. The lowest MAPE of 2.45% occurred at $h+5$. These results indicate that BRNN performance varies with network architecture and forecasting horizon, while the observed performance should be interpreted within the evaluated dataset and experimental conditions. Beyond reporting a single best forecasting result, this study provides a diagnostic analysis of how hidden-layer size affects training error and training-testing consistency, and how forecasting accuracy varies across multiple horizons under the evaluated experimental conditions.

Keywords: Bayesian Regularization Neural Network; Short-Term Electricity Load Forecasting; Neural Network; Forecast Horizon; Mean Absolute Percentage Error.

I. INTRODUCTION

Short-term Load Forecasting (STLF) is an important component of power system operation because accurate electricity demand forecasts are required to support generation scheduling, system operation management, maintenance planning, and the balance between electricity supply and demand [1]. STLF generally concerns the prediction of electricity demand over the next several hours to several days. Forecasting accuracy is increasingly important because electricity load characteristics can vary according to temporal patterns, consumer behavior, environmental conditions, and other external factors [2], [3]. Recent reviews also indicate that STLF remains a challenging forecasting problem because electricity demand exhibits nonlinear and dynamic characteristics, while forecasting performance is strongly influenced by the selected model, input variables, and characteristics of the dataset [4].

One of the approaches widely used for STLF is the Artificial Neural Network (ANN), which has demonstrated the ability to model nonlinear relationships between input and output variables [5], [6]. ANN-based models have been extensively investigated for electricity load forecasting, including conventional neural networks and enhanced architectures designed to improve

forecasting accuracy [7]. Recent studies have also demonstrated that ANN performance can be improved by incorporating relevant temporal and non-temporal factors into the forecasting model [8].

In addition to the neural network architecture, the selection of input variables is an important factor in STLF. Temporal variables and external factors such as temperature, humidity, and other environmental conditions can provide additional information for representing variations in electricity demand [9]. Historical electricity load is also an important source of information because previous consumption patterns can be used to capture temporal dependencies and autoregressive characteristics in electricity demand [10]. Therefore, the selection of input variables and network architecture needs to be considered together when developing an ANN-based STLF model.

One approach used to control the complexity of ANN models is Bayesian Regularization (BR). Bayesian regularization provides a mechanism for controlling network parameters and balancing the ability of the model to fit the training data with its ability to generalize to unseen data [11]. Bayesian neural networks have also been investigated for short-term load forecasting, demonstrating the applicability of Bayesian approaches to nonlinear

electricity demand prediction [12], [13]. More recent research has further investigated Bayesian neural networks for short-term electric net-load forecasting, indicating the continued relevance of Bayesian approaches in electricity forecasting applications [14].

In a previous study, a Bayesian Regularization Neural Network (BRNN) was applied to short-term electricity load forecasting in Langsa City using time, air temperature, humidity, and historical electricity load variables. Several hidden-layer configurations were evaluated to identify the network structure with the best training performance. The results showed that the configuration with four input units, 10 hidden-layer units, and one output unit achieved the lowest MSE during training, while the BRNN model obtained a MAPE of 2.45% at the h+5 forecasting horizon [15].

In recent years, STLF research has increasingly explored deep learning architectures to capture complex temporal relationships in electricity demand. Convolutional and recurrent neural network architectures have been investigated for short-term load forecasting [16], while CNN-based approaches have also been applied under changing electricity consumption conditions [17]. LSTM-based models have been used for short-term net-load forecasting, demonstrating the potential of recurrent architectures for capturing temporal dependencies [18]. Other data-driven approaches have also been investigated for STLF, reflecting the diversity of forecasting models and modeling strategies currently available [19], [20].

Although the previous results indicate a high level of forecasting accuracy, the performance of BRNN should not be evaluated solely based on a single best accuracy value. Changes in network configuration may result in different performance characteristics, while variations in the forecasting horizon may also affect prediction error. Therefore, a more structured analysis of the relationship between network architecture, training performance, generalization capability, and forecasting horizon is required to obtain a more comprehensive understanding of BRNN performance. This is consistent with the broader STLF literature, which emphasizes that forecasting performance should be evaluated with consideration of model configuration, input characteristics, and data properties [2]–[4].

This study analyzes the performance of BRNN for short-term electricity load forecasting. The analysis is conducted using the training and testing results of BRNN with several network architecture configurations and forecasting horizons. Model performance is evaluated using Mean Squared Error (MSE), the correlation coefficient (R), and Mean Absolute Percentage Error (MAPE). In addition to forecasting accuracy, this study analyzes indications of overfitting based

on the comparison between training and testing performance.

The main contribution of this study is a systematic diagnostic analysis of an existing BRNN forecasting model, with emphasis on multi-horizon generalization behavior rather than on proposing a new learning algorithm. Specifically, the study examines: (1) the effect of hidden-layer size on training performance; (2) the consistency between training and testing performance using a correlation-based generalization gap; (3) the trade-off between training error minimization and testing correlation performance; and (4) the variability of forecasting accuracy across seven forecasting horizons. The analysis therefore establishes an evidence-based diagnostic framework for identifying differences between training optimization and unseen-data correlation, as well as temporal error patterns that may be relevant to short-term operational forecasting. Because the underlying BRNN weights and experimental data originate from the previously reported Langsa City study [15], the contribution of this paper is explicitly positioned as a reanalysis and deeper performance characterization, rather than as a newly developed BRNN method.

II. METHOD

2.1 Research Design

This study employed a quantitative research design to analyze the performance of a BRNN for short-term electricity load forecasting. Figure 1 illustrates the research flow used to analyze the performance of the BRNN for short-term electricity load forecasting. The research begins with dataset collection, followed by data normalization to transform the variables into an appropriate numerical range for the neural network model. The normalized data are then divided into training and testing datasets. Next, the BRNN architecture is designed using five different configurations of hidden-layer units, namely 5, 10, 15, 20, and 25 units. Each configuration is trained using the Bayesian regularization algorithm. The training performance is subsequently evaluated using MSE and the correlation coefficient (R) to identify the best-performing configuration.

The selected BRNN model is then applied to the testing dataset. Forecasting accuracy is evaluated using MAPE for seven forecasting horizons, from h+1 to h+7. Finally, the results are analyzed to examine the influence of network architecture and forecasting horizon on BRNN performance, including indications of overfitting and the model's generalization capability.

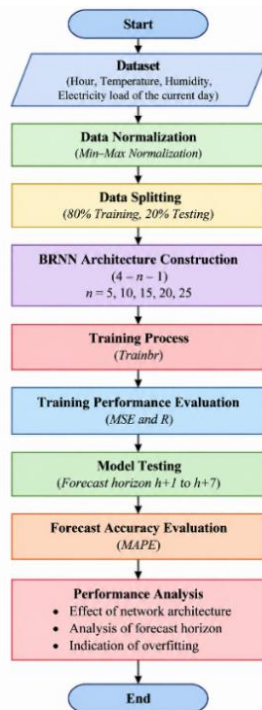


Figure 1. Research Flowchart

2.2 Dataset and Research Variables

The dataset used in this study consists of electricity load and weather-related data used to develop the BRNN model for short-term electricity load forecasting. The input variables consist of hour, air temperature, humidity, and electricity load from the previous day, while the electricity load of the following day is used as the target variable.

The selection of these variables is intended to represent temporal and environmental factors that may influence electricity demand [8], [9]. The hour variable represents the temporal pattern of electricity consumption, while air temperature and humidity provide information on environmental conditions. The previous day's electricity load serves as historical information that can help the model learn the temporal pattern of electricity demand [10].

The relationship between the input and target variables can be represented as follows:

$$D = \{H_t, T_t, RH_t, L_t\} \quad (1)$$

where H_t denotes the hour, T_t represents air temperature, RH_t represents humidity, and L_t represents the electricity load on day t . The target variable is defined as:

$$Y_t = L_t + 1 \quad (2)$$

where $L_t + 1$ represents the electricity load on the following day. Thus, the BRNN model uses four input variables to predict the electricity load corresponding to the subsequent forecasting period.

2.3 Data Normalization

Before being used in the BRNN modeling process, the input and target data were normalized to ensure that all variables had comparable numerical ranges. Min-Max normalization was applied to transform the original data into a normalized range. This process helps maintain numerical stability during the neural network training process. The Min-Max normalization is expressed by Equation (3):

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Where X represents the original data value, X_{min} and X_{max} represent the minimum and maximum values of the corresponding variable, respectively, and X' denotes the normalized value.

The same normalization procedure was applied to the variables used in the BRNN model. After the normalization process, the resulting data were used for the training and testing stages. The normalization process ensures that differences in the numerical scales of the input variables do not disproportionately affect the learning process of the BRNN model.

2.4 Data Splitting

The normalized dataset was divided into two subsets, namely training data and testing data. In this study, 80% of the dataset was used for training, while the remaining 20% was used for testing.

The training dataset was used to develop the BRNN model and determine the network parameters through the Bayesian regularization training process. Meanwhile, the testing dataset was not used during the training process and was used to evaluate the model's forecasting performance on previously unseen data.

The separation of the training and testing datasets enables the performance of each BRNN architecture to be evaluated objectively. The testing results were subsequently analyzed using Mean Absolute Percentage Error (MAPE) for forecasting horizons from $h+1$ to $h+7$. In addition, the comparison between training and testing performance was used to assess the model's generalization capability and identify potential indications of overfitting.

2.5 BRNN Architecture

The BRNN architecture was designed using a multilayer feed-forward neural network consisting of an input layer, a hidden layer, and an output layer. Bayesian regularization was employed to control model complexity and improve the generalization capability of the neural network [11], [12]. The model uses four input units corresponding to the four input variables, namely hour, air temperature, humidity, and electricity

load from the previous day. The output layer consists of one output unit representing the predicted electricity load. To analyze the effect of hidden-layer size on model performance, five BRNN configurations were evaluated. The number of units in the hidden layer was varied from 5 to 25 units, resulting in the following architectures: 4-5-1, 4-10-1, 4-15-1, 4-20-1, and 4-25-1. The notation represents the number of input units, hidden-layer units, and output units, respectively. The five network configurations used in this study are presented in Table 1. These configurations were used to investigate the effect of the number of hidden-layer units on the model learning process.

2.6 Training Process

The BRNN training process uses the Bayesian regularization function through the Trainbr function in MATLAB R2015a. Bayesian regularization is used to control model complexity during the learning process and to improve the balance between model fitting and generalization [11], [13].

The BRNN was trained using the Trainbr function with a performance goal of 1×10^{-3} and a maximum of 500 epochs. The training parameters used in the BRNN model are summarized in Table 2. In this study, the model's generalization capability was analyzed by comparing its training and testing performance.

Table 1. BRNN Architecture Configurations

Architecture	Input unit	Hidden Unit	Output unit
BRNN 4-5-1	4	5	1
BRNN 4-10-1	4	10	1
BRNN 4-15-1	4	15	1
BRNN 4-20-1	4	20	1
BRNN 4-25-1	4	25	1

Table 2. BRNN Training Parameters

Parameters	Value
Training function	Trainbr (Bayesian Regularization)
Performance goal	1×10^{-3}
Performance function	MSE
Maximum epoch	500

2.7 Model Performance Evaluation

Model performance was evaluated in two stages: training performance evaluation and testing performance evaluation. The training evaluation was conducted using the MSE and R, while the testing evaluation used the MAPE to measure the accuracy of the forecasting results. MSE was used to measure the average squared difference between the actual and predicted values during the training process. A lower MSE indicates a smaller prediction error. In this study, MSE was used as

the primary criterion for comparing the five BRNN architecture configurations and selecting the configuration with the lowest training error [19]. The R was used to evaluate the relationship between the predicted output and the actual target values. The R value provides an indication of how well the model captures the pattern of the target data. A value closer to 1 indicates a stronger positive correlation between the predicted and actual values [21]. The correlation coefficient was evaluated for both the training and testing datasets. The comparison of R values between the two datasets was also used as an additional indicator of the model's generalization capability [21].

MAPE was used to evaluate the forecasting accuracy of the BRNN model during the testing stage. MAPE measures the average absolute percentage difference between the actual and predicted values. In this study, MAPE was calculated for seven forecasting horizons, namely h+1, h+2, h+3, h+4, h+5, h+6, and h+7. The MAPE values obtained from each horizon were compared to determine the forecasting horizon with the best BRNN performance [22]. This analysis also provides an overview of how forecasting accuracy changes as the prediction horizon increases. The MSE, correlation coefficient R, and MAPE are calculated using Equations (4), (5), and (6), respectively [20], [21]:

$$MSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad (5)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (6)$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the actual values, $\bar{\hat{y}}$ is the mean of the predicted values, and n is the number of observations.

To further assess the consistency between training and testing performance, the generalization gap was calculated as the difference between the training and testing correlation coefficients [21]:

$$\Delta R = R_{Training} - R_{Testing} \quad (7)$$

The absolute value of ΔR was used to evaluate the magnitude of the difference between training and testing correlation performance, where a smaller absolute value indicates greater consistency between the two datasets.

III. RESULTS AND DISCUSSION

3.1 Performance of BRNN Configurations

This study analyzes the performance of five BRNN configurations by varying the number of hidden units. The evaluated architectures were 4-5-1, 4-10-1, 4-15-1, 4-20-1, and 4-25-1. The performance of each configuration was evaluated

using the Mean Squared Error (MSE) and correlation coefficient (R) obtained from the training and testing datasets. The results are presented in Table 3. The MSE values obtained from the five configurations were relatively close, ranging from 0.000933 to 0.000998. The 4-10-1 architecture produced the lowest training MSE of 0.000933, followed by the 4-25-1 architecture with an MSE of 0.000950. The 4-15-1, 4-20-1, and 4-5-1 configurations produced MSE values of 0.000986, 0.000995, and 0.000998, respectively. Based on the predefined minimum training MSE criterion, the 4-10-1 architecture provided the best training performance among the evaluated configurations.

The correlation coefficients also show a strong relationship between the actual and predicted load values. The training R values ranged from 0.939 to 0.947, while the testing R values ranged from 0.930 to 0.952. The 4-20-1 configuration achieved the highest testing R of 0.952, whereas the 4-25-1 configuration obtained the lowest testing R of 0.930. These results indicate that increasing the number of hidden units does not necessarily produce a consistent improvement in model performance. The lowest training MSE was obtained with 10 hidden units, whereas the highest testing correlation was obtained with 20 hidden units. Therefore, the effect of network complexity needs to be evaluated using multiple performance indicators.

Table 3. BRNN Performance and Generalization Gap for Different Network Configurations

Architecture	Epoch	MSE	R	R	ΔR
			Training	Testing	
4-5-1	5	0.000998	0.939	0.948	-0.009
4-10-1	6	0.000933	0.947	0.945	0.002
4-15-1	9	0.000986	0.941	0.950	-0.009
4-20-1	12	0.000995	0.941	0.952	-0.011
4-25-1	6	0.000950	0.947	0.930	0.017

3.2 Analysis of Architecture and Generalization Performance

The 4-10-1 architecture was selected as the reference configuration because the study defined the minimum training MSE as the primary criterion for architecture selection. The testing R was subsequently used as an independent indicator of generalization performance rather than as the selection criterion.

To further assess the generalization behavior of the evaluated architectures, the difference between the training and testing correlation coefficients (ΔR) was calculated as an additional indicator. The 4-10-1 architecture produced the smallest absolute ΔR of 0.002, indicating the highest consistency between its training and testing correlation performance among the evaluated configurations. In contrast, the 4-25-1 architecture showed the largest absolute ΔR of 0.017, indicating a greater difference between training and testing performance. The 4-5-1 and 4-

15-1 configurations exhibited negative ΔR values because their testing correlation coefficients were slightly higher than their corresponding training values.

The combined analysis of training MSE, testing correlation, and the correlation-based generalization gap reveals a trade-off among the evaluated BRNN architectures. The 4-10-1 configuration provided the lowest training MSE and the smallest absolute difference between training and testing R, whereas the 4-20-1 configuration achieved the highest testing R. This indicates that minimizing training error does not necessarily result in the highest testing correlation. Therefore, BRNN architecture selection should consider both predictive performance and consistency between training and testing behavior.

The differences between training and testing correlation coefficients provide an indication of the behavior of the BRNN models when applied to previously unseen data. For the 4-10-1 configuration, R decreased slightly from 0.947 during training to 0.945 during testing. The relatively small difference suggests consistent correlation performance between the two datasets. In contrast, the 4-25-1 configuration showed a larger decrease from 0.947 to 0.930. From a performance-analysis perspective, these results suggest that the evaluated BRNN configurations generally maintained a strong correlation between actual and predicted values on the testing data.

However, these results should not be interpreted as definitive evidence that overfitting is completely absent. The correlation-based generalization gap provides a descriptive indication of consistency between training and testing performance, but a more comprehensive assessment of generalization would require additional validation procedures, such as repeated experiments or rolling-origin time-series validation.

3.3 Forecasting Performance Across Different Horizons

After identifying the reference BRNN configuration, the forecasting performance was analyzed across seven forecasting horizons, from $h+1$ to $h+7$. MAPE was used as the primary measure because it expresses forecasting error in percentage form and facilitates comparison across different forecasting horizons. The results are summarized in Table 4. The results demonstrate that forecasting performance varies across the seven horizons. The MAPE values range from 2.45% to 8.46%. The lowest MAPE was obtained at $h+5$, with a value of 2.45%, while the highest MAPE was obtained at $h+1$, with a value of 8.46%. An important observation is that the forecasting error does not increase monotonically as the forecasting horizon becomes longer. For example, the MAPE decreases from 8.16% at $h+3$

to 5.28% at h+4 and further to 2.45% at h+5. It subsequently increases to 8.02% at h+6 before decreasing again to 6.11% at h+7. This pattern indicates that the forecasting performance in the evaluated dataset is not determined solely by the length of the forecasting horizon.

Characteristics of the load profile and the temporal variability represented in the corresponding observations may also contribute to the differences in forecasting error. Therefore, the h+5 result should be interpreted as the best-performing horizon within the evaluated dataset and experimental setting, rather than as evidence that BRNN generally performs best at a five-step forecasting horizon under all conditions.

The non-monotonic error pattern is also relevant from a power-system perspective. The sharp reduction in MAPE at h+5 followed by an increase to 8.02% at h+6 indicates that forecast difficulty is not determined by horizon length alone. Such variation can occur when the target observations contain changes in load trajectory that are not fully represented by the four input variables used in this study. In particular, the larger h+6 error should be treated as an indication of sensitivity to the local load profile, rather than as evidence of a specific consumer-behavior mechanism, because consumer composition and causal demand drivers were not directly analyzed. This interpretation is consistent with the broader finding that short-term load predictability can vary across system levels and temporal conditions [28]. Further validation using additional exogenous variables, load-profile decomposition, and rolling-origin evaluation would be required to establish the physical causes of the observed error transition.

Table 4. BRNN Forecasting Performance Across Different Horizons

Forecasting Horizon	MAPE (%)	Maximum Error (%)	Minimum Error (%)
h+1	8.46	12.68	3.95
h+2	7.76	12.60	0.83
h+3	8.16	13.74	0.07
h+4	5.28	13.89	0.54
h+5	2.45	9.13	0.46
h+6	8.02	23.49	2.70
h+7	6.11	10.08	0.14

To further characterize the variation in forecasting accuracy across the evaluated horizons, a statistical summary of the MAPE values was calculated. The results are presented in Table 5. The seven forecasting horizons produced a mean MAPE of 6.61%, with a standard deviation of 2.02%. The MAPE ranged from 2.45% to 8.46%, indicating noticeable variation in forecasting accuracy across horizons. The relatively low MAPE at h+5 contrasts with the higher errors observed at h+1 and h+6, confirming that the forecasting error does not change monotonically with the forecasting horizon. Therefore, the h+5

result should be interpreted as the best-performing horizon within the evaluated dataset rather than as a general characteristic of BRNN.

It should also be noted that the present study reports MAPE, together with maximum and minimum percentage errors, for each forecasting horizon. Additional horizon-specific metrics such as MAE or RMSE would require the complete set of actual and predicted observations, which is not reconstructed in this analysis; therefore, no additional numerical error metrics are introduced without the underlying prediction series.

Table 5. Statistical Summary of Forecasting Performance

Statistic	MAPE (%)
Mean	6.61
Median	7.76
Standard deviation across the seven evaluated horizons	2.02
Minimum	2.45
Maximum	8.46
Range	6.01

3.4 Detailed Analysis of the h+5 Forecasting Horizon

The h+5 forecasting horizon produced the lowest MAPE, namely 2.45%. This result indicates a relatively small average percentage difference between the predicted and actual electricity load values for this particular forecasting horizon. The maximum and minimum percentage errors at h+5 were 9.13% and 0.46%, respectively. The largest error occurred at approximately 18:00 WIB, when the actual electricity load was 67.95 MW while the predicted value was 74.78 MW. In contrast, the smallest error occurred at approximately 21:00 WIB, when the actual load was 84.09 MW, and the predicted value was 84.48 MW.

The difference in prediction performance between these two periods illustrates the sensitivity of forecasting accuracy to variations in the load profile. The relatively larger error around 18:00 WIB may be associated with a period of substantial changes in electricity consumption. However, because the present analysis does not include a dedicated causal analysis of consumer behavior or load composition, this observation should be considered an interpretation of the prediction pattern rather than a definitive explanation of its cause.

The MAPE value of 2.45% also indicates that the BRNN model can produce predictions that are relatively close to the actual values under the evaluated h+5 condition. Nevertheless, the result is specific to the dataset, preprocessing procedure, network configuration, and experimental conditions used in this study. From an operational forecasting perspective, this result suggests that the h+5 horizon may be useful as a tactical forecasting window under conditions similar to those evaluated in this study. However, a 2.45%

MAPE should not by itself be interpreted as a guaranteed operational safety margin, because the analysis is based on a single dataset and does not quantify the consequences of forecast error for dispatch, reserve allocation, or system security.

3.5 Discussion

The analysis provides several observations regarding the performance of the BRNN for short-term electricity load forecasting. First, network architecture influences the training performance of BRNN. Among the five configurations evaluated, the 4-10-1 architecture achieved the lowest MSE. However, the 4-20-1 architecture achieved the highest testing correlation coefficient. This difference demonstrates that model selection based solely on one performance indicator may not fully represent the behavior of the model on unseen data.

Second, the results indicate that increasing the number of hidden units does not automatically improve forecasting performance. The MSE decreased when the number of hidden units increased from 5 to 10, but increased again at 15 and 20 hidden units before decreasing slightly at 25 hidden units. This non-monotonic behavior suggests that network complexity needs to be evaluated empirically for the specific forecasting problem. This observation is consistent with the broader STLFL literature, where forecasting performance depends not only on model complexity but also on the characteristics of the input data and the selected model configuration [17], [18], [23].

Third, forecasting accuracy varies considerably across forecasting horizons. The lowest MAPE was obtained at $h+5$, while the largest average MAPE occurred at $h+1$. Consequently, the relationship between forecasting horizon and forecasting error in the evaluated data cannot be characterized simply as an increasing error with increasing horizon.

The findings are consistent with the general consideration in neural-network-based load forecasting that model configuration and input characteristics can affect forecasting performance [17], [18]. Previous research has investigated Bayesian neural-network approaches for short-term load forecasting, including Bayesian methods for controlling model complexity and improving forecasting performance [11]–[14]. Importantly, the results presented in this study should be interpreted as a performance analysis of the previously developed BRNN forecasting model, rather than as a claim of a newly developed forecasting method. The BRNN modeling framework and its underlying experimental results have previously been reported in a study focusing on short-term electricity load forecasting in Langsa City [15]. The present work reorganizes and analyzes the available results specifically from the

perspectives of network architecture, generalization behavior, and forecasting horizon.

This distinction matters because the present article's main contribution is systematically analyzing and interpreting model performance, not proposing a new BRNN algorithm. The analysis highlights how the number of hidden units and forecasting horizon influence the observed performance and provides a more focused interpretation of the strengths and limitations of the evaluated BRNN configuration. For future work, the performance analysis can be extended through additional datasets covering different months or seasons, rolling-origin time-series validation, repeated experiments, and comparisons with contemporary forecasting methods. Recent studies have investigated GRU-based architectures for short-term load forecasting [23], as well as LSTM and RNN-based deep neural networks [24]. More advanced architectures, including deep convolutional LSTM and stacked GRU models, have also been proposed for short-term power load forecasting [25]. Hybrid forecasting approaches also provide potential alternatives for comparison with BRNN. VMD-GRU-TCN and other decomposition-based deep learning approaches have been investigated for short-term electricity load forecasting [26], [27], while ensemble learning has also been explored as an approach for improving forecasting performance [22]. STLFL applications also extend across different levels of the electricity system, from system-level forecasting to secondary substations [28], as well as microgrid energy management applications [29]. Recent reviews highlight the continued development of deep learning approaches for STLFL [4], while multimodal fusion represents an emerging direction for short-term electric load forecasting [30].

IV. CONCLUSION

This study analyzed the performance of a BRNN for short-term electricity load forecasting by examining the effects of network architecture and forecasting horizon. Five BRNN configurations with 5, 10, 15, 20, and 25 hidden units were evaluated using MSE and correlation coefficient as training and testing performance indicators. The analysis showed that the 4-10-1 configuration achieved the lowest MSE of 0.000933, indicating the best training performance among the evaluated network configurations. The testing correlation coefficients ranged from 0.930 to 0.952, indicating a strong relationship between the predicted and actual load values across the evaluated architectures. These results also indicate that increasing the number of hidden units does not necessarily improve model performance. Forecasting performance across the $h+1$ to $h+7$ horizons showed MAPE values ranging from

2.45% to 8.46%. The lowest MAPE of 2.45% was obtained at the h+5 horizon, while the highest MAPE of 8.46% occurred at h+1. The variation in MAPE across horizons indicates that forecasting performance does not necessarily deteriorate monotonically as the forecasting horizon increases.

The analysis indicates that BRNN can provide relatively accurate short-term electricity load forecasts under the evaluated experimental conditions. However, the findings should be interpreted within the scope of the dataset, network configurations, and evaluation procedure used in this study. Since the underlying BRNN modeling results have been previously reported, the contribution of this study is positioned as a systematic performance analysis of the previously developed BRNN model, particularly with respect to network architecture, generalization behavior, and forecasting horizon.

From a practical standpoint, the results suggest that the BRNN 4-10-1 configuration can be considered for short-term tactical forecasting when the operating conditions resemble the evaluated dataset, particularly around the h+5 horizon where the observed MAPE was 2.45%. For horizons such as h+6, where the observed MAPE increased to 8.02%, operators should treat the single-model forecast with greater caution and consider additional validation or complementary forecasting models. These implications are conditional on the present dataset and should be validated through rolling-origin testing, multiple seasons, and operationally relevant error measures before deployment in dispatch or grid-control workflows.

Future research should evaluate BRNN using longer and more diverse time-series datasets, repeated or rolling-origin validation, and comparisons with contemporary forecasting approaches such as LSTM, GRU, and XGBoost. Additional sensitivity analysis of input variables would also be useful to further assess the robustness and generalizability of the forecasting model.

REFERENCES

- [1]. T. Hong and S. Fan, 2016, *Probabilistic Electric Load Forecasting: A Tutorial Review*, International Journal of Forecasting, vol. 32, no. 3, pp. 914–938, doi: 10.1016/j.ijforecast.2015.11.011.
- [2]. S. Akhtar, S. Shahzad, A. Zaheer, H. S. Ullah, H. Kilic, R. Gono, M. Jasiński, and Z. Leonowicz, 2023, *Short-Term Load Forecasting Models: A review of Challenges, Progress, And The Road Ahead*, Energies, vol. 16, no. 10, Art. no. 4060, doi: 10.3390/en16104060.
- [3]. F. Rodrigues, C. Cardeira, J. M. F. Calado, and R. Melicio, 2023, *Short-Term Load Forecasting Of Electricity Demand For The Residential Sector Based On Modelling Techniques: A Systematic Review*, Energies, vol. 16, no. 10, Art. no. 4098, doi: 10.3390/en16104098.
- [4]. Y. Eren and İ. B. Küçükdemiral, 2024, *A Comprehensive Review on Deep Learning Approaches for Short-Term Load Forecasting*, Renewable and Sustainable Energy Reviews, vol. 189, Art. no. 114031, doi: 10.1016/j.rser.2023.114031.
- [5]. A. Baliyan, K. Gaurav, and S. K. Mishra, 2015, *A Review of Short Term Load Forecasting Using Artificial Neural Network Models*, Procedia Computer Science, vol. 48, pp.121–125,doi:10.1016/j.procs.2015.04.160.
- [6]. A. Román-Portabales, M. López-Nores, and J. J. Pazos-Arias, 2021, *Systematic Review of Electricity Demand Forecast Using ANN-Based Machine Learning Algorithms*, Sensors, vol. 21, no. 13, Art. no. 4544, doi: 10.3390/s21134544.
- [7]. A. I. Arvanitidis, D. Bargiotas, A. Daskalopulu, V. M. Laitsos, and L. H. Tsoukalas, 2021, *Enhanced Short-Term Load Forecasting Using Artificial Neural Networks*, Energies, vol. 14, no. 22, Art. no. 7788, doi: 10.3390/en14227788.
- [8]. J. Sun, H. Dong, Y. Gao, Y. Fang, and Y. Kong, 2021, *The Short-Term Load Forecasting Using An Artificial Neural Network Approach With Periodic And Nonperiodic Factors: A Case Study of Tai'an, Shandong Province, China*, Computational Intelligence and Neuroscience, vol. 2021, Art. no. 1502932, doi: 10.1155/2021/1502932.
- [9]. C. Hora, F. C. Dan, G. Bendea, and C. Secui, 2022, *Residential Short-Term Load Forecasting During Atypical Consumption Behavior*, Energies, vol. 15, no. 1, Art. no. 291, doi: 10.3390/en15010291.
- [10]. L. Li, C. J. Meinrenken, V. Modi, and P. J. Culligan, 2021, *Short-Term Apartment-Level Load Forecasting Using a Modified Neural Network With Selected Auto-Regressive Features*, Applied Energy, vol. 287, Art. no. 116509, doi: 10.1016/j.apenergy.2021.116509.
- [11]. H. S. Hippert and J. W. Taylor, 2010, *An Evaluation of Bayesian Techniques for Controlling Model Complexity And Selecting Inputs in a Neural Network For Short-Term Load Forecasting*, Neural Networks, vol. 23, no. 3, pp. 386–395, doi: 10.1016/j.neunet.2009.11.016.
- [12]. P. Lauret, E. Fock, R. N. Randrianarivony, and J.-F. Manicom-Ramsamy, 2008, *Bayesian Neural Network Approach to Short Time Load Forecasting*, Energy Conversion and Management, vol. 49, no. 5, pp. 1156–1166, doi: 10.1016/j.enconman.2007.09.009.

- [13]. D. Niu, H.-F. Shi, and D. Wu, 2012, *Short-Term Load Forecasting Using Bayesian Neural Networks Learned by Hybrid Monte Carlo Algorithm*, Applied Soft Computing, vol. 12, no. 6, pp. 1822–1827, doi: 10.1016/j.asoc.2011.07.001.
- [14]. G. Tziolis, C. Spanias, M. Theodoride, S. Theocharides, J. Lopez-Lorente, A. Livera, G. Makrides, and G. E. Georghiou, 2023, *Short-Term Electric Net Load Forecasting For Solar-Integrated Distribution Systems Based On Bayesian Neural Networks And Statistical Post-Processing*, Energy, vol. 271, Art. no. 127018, doi: 10.1016/j.energy.2023.127018.
- [15]. A. Fauzi, Tarmizi, and R. Adriman, 2025, *Short-term electricity load forecasting using Bayesian regularization-based neural network: A case study in Langsa City*, CIRCUIT: Jurnal Ilmiah Pendidikan Teknik Elektro, vol. 9, no. 2, pp. 264–272, 2025, doi: 10.22373/crc.v9i2.25549.
- [16]. H. Eskandari, M. Imani, and M. P. Moghaddam, “Convolutional and recurrent neural network based model for short-term load forecasting,” Electric Power Systems Research, vol. 195, Art. no. 107173, 2021, doi: 10.1016/j.epsr.2021.107173.
- [17]. A. M. Tudose, I. I. Picioroaga, D. O. Sidea, C. Bulac, and V. A. Boicea, 2021, *Short-Term Load Forecasting Using Convolutional Neural Networks in COVID-19 Context: The Romanian Case Study*, Energies, vol. 14, no. 13, Art. no. 4046, 2021, doi: 10.3390/en14134046.
- [18]. A. Stratigakos, A. Bachoumis, V. Vita, and E. Zafiroopoulos, 2021, *Short-Term Net Load Forecasting With Singular Spectrum Analysis and LSTM Neural Networks*, Energies, vol. 14, no. 14, Art. no. 4107, doi: 10.3390/en14144107.
- [19]. M. Lopez-Martin, A. Sanchez-Esguevillas, L. Hernandez-Callejo, J. I. Arribas, and B. Carro, 2021, *Novel Data-Driven Models Applied To Short-Term Electric Load Forecasting*, Applied Sciences, vol. 11, no. 12, Art. no. 5708, doi: 10.3390/app11125708.
- [20]. F. U. M. Ullah, N. Khan, T. Hussain, M. Y. Lee, and S. W. Baik, 2021, *Diving Deep Into Short-Term Electricity Load Forecasting: Comparative Analysis And A Novel Framework*, Mathematics, vol. 9, no. 6, Art. no. 611, doi: 10.3390/math9060611.
- [21]. Q. Zhao, X. Wen, H. Huang, J. Wang, and J. Fang, 2022, *Optimised Extreme Gradient Boosting Model For Short-Term Electric Load Demand Forecasting of Regional Grid System*, Scientific Reports, vol. 12, Art. no. 19282, doi: 10.1038/s41598-022-22024-3
- [22]. C. Sankalpa, S. Kittipiyakul, and S. Laitrakun, 2022, *Forecasting Short-Term Electricity Load Using Validated Ensemble Learning*, Energies, vol. 15, no. 22, Art. no. 8567, doi: 10.3390/en15228567.
- [23]. D. Li, G. Sun, S. Miao, Y. Gu, Y. Zhang, and S. He, 2022, *A Short-Term Electric Load Forecast Method Based On Improved Sequence-To-Sequence Gru With Adaptive Temporal Dependence*, International Journal of Electrical Power & Energy Systems, vol. 137, Art. no. 107627, 2022, doi: 10.1016/j.ijepes.2021.107627.
- [24]. B. ul Islam and S. F. Ahmed, 2022, *Short-Term Electrical Load Demand Forecasting Based on LSTM and RNN Deep Neural Networks*, Mathematical Problems in Engineering, vol. 20, Art. no. 2316474, doi: 10.1155/2022/2316474.
- [25]. F. U. M. Ullah, N. Khan, T. Hussain, M. Y. Lee, and S. W. Baik, 2022, *Deep Learning-Assisted Short-Term Power Load Forecasting Using Deep Convolutional Lstm And Stacked GRU*, Complexity, vol. 2022, Art. no. 2993184, doi: 10.1155/2022/2993184.
- [26]. C. Cai, Y. Li, Z. Su, T. Zhu, and Y. He, 2022, *Short-Term Electrical Load Forecasting Based On VMD and GRU-TCN Hybrid Network*, Applied Sciences, vol. 12, no.13,Art.no.6647,doi: 10.3390/app1213667
- [27]. N. Wang, S. Xue, and Z. Li, 2023, *Short-Term Load Forecasting Based On Vmd And Combined Deep Learning Model*, IEEE Transactions on Electrical and Electronic Engineering, vol. 18, no. 7, pp. 1067–1075, doi: 10.1002/tee.23810.
- [28]. N. Houben et al., 2023, *Short-Term Electricity Load Forecasting—A Systematic Approach From System Level To Secondary Substations*, Applied Energy, vol. 332, Art. no. 120493, doi: 10.1016/j.apenergy.2022.120493.
- [29]. A. Jahani, K. Zare, and L. M. Khanli, 2023, *Short-Term Load Forecasting For Microgrid Energy Management System Using Hybrid SPM-LSTM*, Sustainable Cities and Society, vol. 98, Art. no. 104775, doi: 10.1016/j.scs.2023.104775.
- [30]. L.-L. Peng, X.-Y. Yang, X.-X. Hua, G.-F. Fan, Y.-J. Wang, A. J. Umbarkar, and W.-C. Hong, 2026, *Short-Term Electric Load Forecasting Based On Deep Learning And Multi-Modal Fusion*, Electric Power Systems Research, vol. 254, Art. no. 112593, doi: 10.1016/j.epsr.2025.112593.