A VAE-Based Bayesian Bidirectional LSTM for Renewable Energy Forecasting

Devinder Kaur, Shama Naz Islam, and Md. Apel Mahmud

Abstract—The advancement in distributed generation technologies in modern power systems has led to a widespread integration of renewable energy generation at customer side. However, the intermittent nature of renewable energy pose new challenges to the network operational planning with underlying uncertainties. This paper proposes a novel Bayesian probabilistic technique for forecasting renewable power generation by addressing data and model uncertainties by integrating bidirectional long short-term memory (BiLSTM) neural networks while compressing the weight parameters using variational autoencoder (VAE). Existing Bayesian deep learning methods suffer from high computational complexities as they require to draw a large number of samples from weight parameters expressed in the form of probability distributions. The proposed method can deal with uncertainty present in model and data in a more computationally efficient manner by reducing the dimensionality of model parameters. The proposed method is evaluated using pinball loss, reconstruction error, and other forecasting evaluation metrics. It is inferred from the numerical results that VAE-Bayesian BiLSTM outperforms other probabilistic deep learning methods in terms of forecasting accuracy and computational efficiency for different sizes of the dataset.

Index Terms—Bayesian deep learning, dimensionality reduction, renewable energy generation, Bidirectional long-short term memory, variational autoencoders

I. INTRODUCTION

The advancement in distributed generation (DG) technologies has led to a widespread integration of distributed renewable energy sources (RES) such as solar panels and wind turbines at the user end [1]. The renewable power generation systems often pose new challenges to distribution networks due to its intermittent nature [2], [3]. In this context, energy forecasting has a crucial role to play in planning and managing the operations of modern power systems often regarded as smart grid (SG) systems [4], [5]. Since the RES is an intermittent source of energy generation, its accurate forecasting ahead of time can be of key importance to increase the operational efficiency of the grid and make it more sustainable.

Recent research on energy forecasting highly focuses on deep learning-based methods such as recurrent neural networks (RNN) [6] and long short-term memory (LSTM) [7], [8] which utilize the available big data from SG infrastructure. Furthermore, bidirectional RNN combined with LSTM formulates the architecture for bidirectional LSTM (BiLSTM) and exhibits the advantage of long-term memory tackling vanishing gradients in addition to an extensive learning process. Bidirectional processing leads to proper exploitation of all data points and thus, improves the forecasting accuracy. In this regard, BiLSTM-based techniques have been implemented in [9], [10] to improve the learning process for the application of short-term energy forecasting in power markets. Deep learning techniques are often incorporated with optimisation techniques to improve the effectiveness of the training process [11].

Although the extensive learning often leads to increased computation time, a trade-off needs to be incorporated between the forecasting accuracy and time complexity of the model. In addition, standard deep learning forecasting methods such as RNN and its variants generate point forecasts and often struggle with challenges of overfitting and appropriate hyper-parameter tuning. Various regularization techniques have been proposed in the past to tackle the issue of overfitting, such as dropout and early stopping [12], [13].

Despite the recent advancements in deep learning techniques, the uncertainties in the energy data due to external factors such as changes in weather, as well as the uncertainty presented in the model parameters pose a significant challenge to accurate energy forecasting. In this context, probabilistic approaches have been widely incorporated in energy forecasting due to their ability to manage uncertainties. For example, the authors in [14] utilized probabilistic methods integrated with vector autoregression for solar power forecasting. In a similar manner, non-parametric Bayesian approach for renewable energy forecasting is implemented in [15]–[17]. An extreme learning machine-based probabilistic approach has been adopted in [18] for wind power forecasting. The authors in [19] adopted analog ensemble-based probabilistic forecasting for solar generation. Due to their ability to address uncertainty, probabilistic methods can be integrated with deep neural layers to exploit the advantages of both the probabilistic and deep learning techniques [20]. Kendall et. al. in [21] highlighted the problem of uncertainty and proposed Bayesian neural networks (BNN) based on the conditional probability to quantify the epistemic or model uncertainty, represented by a probability distribution on weight parameters of neural network layers. Though the epistemic uncertainty can be explained away by adding more data, uncertainty inherited from the input data in the form of noise or aleatoric uncertainty cannot be reduced with more data.

Very recently, Bayesian probabilistic approach integrated with deep learning methods have been proposed for demand forecasting [22]. Such methods give forecasting results in the form of prediction intervals to quantify the uncertainties associated with model parameters and energy consumption, contrary to the traditional deep neural networks that are deterministic in nature [2]. Bayesian neural networks have
been proposed in [23] for solar irradiation forecasting. The authors in [24] apply Bayesian model averaging to estimate the predictors of ensemble forecasting of solar power generation.

Furthermore, variational inference (VI) techniques have emerged as an effective method to obtain posterior distribution of weight parameters by minimizing the divergence between true and approximate posterior distributions in Bayesian probabilistic methods [25]. The authors in [26] implemented VI with convolutional gate recurrent unit (GRU) network for forecasting solar irradiation. VI is integrated with LSTM in [27] for solar irradiation forecasting while ensuring that training data is not shared among multiple solar generation sites. Though VI can offer a tractable solution to obtain posterior distributions, the use of probability distributions to represent weight parameters significantly expands the model parameter space, especially for weight matrices. Thus, the ability of Bayesian deep learning techniques to deal with model and stochastic uncertainties comes at the cost of computational complexity arising from sample size of weight parameters.

In this context, variational autoencoders (VAE), widely used in the existing deep learning literature for dimensionality reduction [28], can be implemented to improve the computational efficiency of Bayesian deep learning techniques. VAE represents the encoder with probability distributions which transforms the input data to adopt a lower dimensional structure [29]. VAE is implemented in [30] to forecast the solar generation data. Furthermore, reconstructed VAEs integrated with deep learning have been widely utilized for fault and anomaly detection in time series energy data using reconstruction scores from the decoding process [31]–[33].

A. Motivation and research contributions

Although, existing research has focused on the benefits of Bayesian deep learning methods to address the issue of uncertainty. However, the integration of advanced deep learning methods such as BiLSTM and VAE with Bayesian models remain an open problem for energy forecasting applications and representing such models with more tractable approximations of weight parameters. In addition, an effective solution to manage the increasing number of weight parameters for Bayesian deep learning methods have not been proposed yet.

To address the aforementioned challenges, we propose a new Bayesian BiLSTM forecasting technique integrated with VAE to tackle the uncertainties in the granular renewable energy generation data and high dimensionality in the model parameters of probabilistic methods. The specific contributions are as follows:

1) A Bayesian BiLSTM model for renewable energy generation forecasting is developed by implementing mean field VI to obtain a tractable approximation of the posterior distribution of weight parameters.

2) To improve the computational efficiency of the Bayesian BiLSTM forecasting, a VAE-based dimensionality reduction technique is proposed to compress the sample size of weight parameters using variational encoding layer. The information loss is monitored using decoding layer by generating reconstructed original data.

3) The proposed VAE-Bayesian BiLSTM model is evaluated with variable data sizes for solar generation and compared with a number of probabilistic deep learning methods. The proposed method demonstrates superior results in terms of pinball losses, prediction intervals, and root mean square error (RMSE), while achieving a much lower reconstruction loss for weight parameters.

B. Paper organization

The rest of the paper is organized in the following manner. Section II outlines the problem formulation regarding weight compression and optimized distribution of weight parameters for forecasting. Section III elaborates how the proposed VAE-BiLSTM model is implemented for forecasting renewable generation. Numerical evaluation and comparison results for the proposed scheme have been presented in Section IV. Finally, the paper is concluded in Section V.
Bayesian probabilistic model is represented using following equation [25]:

\[
p(w|D) = \frac{p(D|w)p(w)}{p(D)} \tag{1}
\]

where \(p(w|D)\) represents the posterior probability of weight parameters given observed data using Bayes rule of conditional probability and \(p(D|w)\) denotes the likelihood of observed data given weight parameters. Furthermore, \(p(D)\) represents the marginal probability for the observed parameters. Therefore, equation (1) is expressed as [25]:

\[
p(w|D) = \frac{p(D|w)p(w)}{\int p(D|w)p(w)dw} \tag{2}
\]

Due to the large number of data values presented in the integral of (2), it becomes intractable to analytically compute the exact posterior. So, instead of computing the true posterior, an approximated distribution \(q(\theta|\phi)\) parameterized over \(\phi\) is computed using mean field VI(MFVI).

In MFVI, a factorized distribution is used as the approximate posterior distribution. Thus, the approximate posterior probability distribution satisfies the following condition [25]:

\[
q_\phi(w|D_x) = \prod_{t=1}^{N} q_\phi(w_t|D_t) \tag{3}
\]

where \(q_\phi(w_t|D_t)\) represents the factorized posterior distribution for weight parameters \(w\) given observed data \(D\) over \(N\) samples. Then, the difference between real and approximated posterior is minimized using Kullback-Leibler (KL) divergence [21], given as:

\[
KL(q_\phi(w)||p(w)) = \log p(D) - \int q_\phi(w) \log \frac{p(D|w)}{q_\phi(w)}dw \tag{4}
\]

s.t. \(KL(q_\phi(w)||p(w)) \geq 0 \tag{5}\)

KL divergence is asymmetric in nature and also referred as information gain or relative entropy. Finally, the cost function for the proposed scheme is to minimize the KL divergence as in:

\[
\arg \min_{q_\phi(w)} KL(q_\phi(w)||p(w)) \tag{6}
\]

To minimize KL divergence, evidence lower bound (ELBO) is maximized. ELBO is defined as lower bound on the log marginal probability of observed data points [21]. The marginal probability of data points can be expressed as following:

\[
\log p(D) = \log \int p(D, w)dw \tag{7}
\]

\[
\log p(D) = \log \int \frac{p(D, w) * q_\phi(w)}{q_\phi(w)}dw \tag{8}
\]

\[
\log p(D) = \log \mathbb{E}_{q_\phi(w)} \left[ \frac{p(D, w)}{q_\phi(w)} \right] \tag{9}
\]

\[
\log p(D) \geq \mathbb{E}_{q_\phi(w)} \log \left[ \frac{p(D, w)}{q_\phi(w)} \right] \tag{10}
\]

Comparing (4) and (10), it is evident that

\[
\log p(D) = ELBO(\theta) + KL(q_\phi(w)||p(w)) \tag{11}
\]

Thus, maximizing variational ELBO leads to KL minimization and the optimisation problem in (6) is reformulated as:

\[
\arg \max_{q_\phi} ELBO \tag{12}
\]

Now, the approximated posterior in the aforementioned Bayesian BiLSTM method is a distribution \(q_\phi(w)\) rather than a single weight sample. This results into large dimensionality of weight parameters drawn from the optimal distribution. As a result, VAE is integrated to encode the large sample space during posterior approximation into lower dimensional space \(z\) representation using following equation:

\[
p(z|x) = \frac{p(z|x)p(z)}{p(x)} \tag{13}
\]

where \(p(z|x)\) represents the encoder on the input samples \(x\) and \(p(x|z)\) denotes the decoder to represent the likelihood of reconstructed data over weight parameters in a VAE process. \(P(x)\) symbolizes the marginalized probability of observed variables which can further be defined as:

\[
p(x|z) = \frac{p(x|z)p(z)}{\int (p(x,z)p(z)dz) \tag{14}
\]

VI can be used to approximate posterior distribution \(q_\phi(z|x)\) over parameters \(\phi\) as a representative of the true posterior distribution \(p(z|x)\). In this context, the cost function for VAE consists of reconstruction loss and Variational loss. The variational loss of VAE is computed using KL divergence as in (4). Thus, the aggregated loss for VAE is defined as:

\[
\tau_{loss} = ||(x - \hat{x})^2|| + KL(q_\phi(z)||p(z)) \tag{15}
\]

Here, \(\hat{x}\) represents the reconstructed data samples from the latent distribution. Thus, the optimised encoder can be obtained by solving:

\[
\arg \min_{q_\phi(z)} \tau_{loss} \tag{16}
\]
This section outlines the proposed framework for the integrated VAE-Bayesian BiLSTM probabilistic forecasting method. The methodology for the proposed algorithm is illustrated in Fig. 4 in the form of a detailed flow chart. In this paper, Bayesian BiLSTM method is employed to address data and model uncertainties integrated with VAE to deal with high dimensionality presented in weight parameters due to Bayesian probabilistic method implemented on BiLSTM layers. Bayesian probability incorporated with neural networks gives forecasting results in the form of prediction intervals to quantify uncertainties associated with model parameters and renewable energy generation.

Algorithm 1 VAE-Bayesian BiLSTM framework for probabilistic forecasting

**Input:** Actual observed data, \( D = (x_t, y_t) \)

**Output:** Prediction intervals (PIs), predictive mean \( \hat{y} \)

1: Obtain observed data \( D \);
2: Scale and normalize \( D \) in (-1,1);
3: Split training data \( D_i \), form \( D \);
4: Feed \( D \) to VAE-Bayesian BiLSTM;

**Encoding-decoding stage using VAE**
5: Prior probability on latent space \( p(z) \);
6: Approximate posterior \( q(z|D) \);
7: Encoder \( \implies \) posterior;
8: Decoder \( \implies \) likelihood;
9: Minimize \( \tau_{loss} \) using (15);

**Forecasting stage using Bayesian BiLSTM stage**
10: Sample from \( z \);
11: for \( i=1, i \leq n, i++ \) do
12: Formulate prior trainable on weight parameters \( p(w) \);
13: Use BiLSTM layer as hidden layer;
14: Use Dense Variational layer as output layer;
15: Approximate independent \( p(w|D) \) using (5);
16: Maximize ELBO using (11);
17: end for
18: Compute pinball loss using (15);
19: Compute \( \mu \) and \( \sigma \) and obtain PIs;
20: Compute predictive mean;
21: Compute RMSE, MAE, R score using (19), (20), and (21) respectively and evaluate forecasting performance;
22: Evaluate the model performance using time complexity and reduction in \( w \);

A. **Bayesian BiLSTM**

Epistemic uncertainty also known as model uncertainty is the uncertainty related to the model parameters such as weights and biases in a neural network. Furthermore, aleatoric uncertainty is defined as the uncertainty related to data while it is being sensed and also referred to as stochastic uncertainty. It captures the noise inherent in the observed data, either due to exogenous factors such as weather and human behavior patterns or from sensor devices during data acquisition. Although, it is not feasible yet to reduce the aleatoric uncertainty with the help of more data. However, it can be quantified using Bayesian deep learning approach by keeping probability distributions on the output layer. Furthermore, by optimizing weights and reducing weight uncertainty prediction accuracy is improved for the future forecasts.

Using proposed Bayesian approach, aleatoric uncertainty is quantified by placing a prior Normal distribution over the objective function. Whereas, epistemic uncertainty is handled by placing a prior over weight parameters of the BiLSTM network. Furthermore, it can be reduced with the help of hyperparameter tuning and optimizing the learning process.

To be specific, both the uncertainties are represented by placing prior on weights and objective function using a densovariaional layer incorporated with BiLSTM hidden layer. Then, the approximate posterior distribution is computed using MFVI based on evidence provided by the observed data by minimizing the KL loss using (4). In addition, BiLSTM works in forward and backward passes as illustrated in Fig. 1. The architecture of BiLSTM network is shown in Fig. 2.

B. **VAE for weight compression**

To reduce the sample space for weight parameters, VAE is integrated. VAEs are defined as the class of generative models. It is difficult for decoder to backpropagate to original data. This task is done by re-parameterization trick. Because, VAE tries to learn a distribution from the latent space. So, in the following equation is used to parametrize the VAE network with Normal distribution having \( \mu_z \) as mean and \( \sigma_z \) as the standard deviation for the latent distribution:

\[
    z = \mu_z + \sigma_z \odot \epsilon
\]  

(17)

The architecture of VAE is shown in Fig. 3. Finally, the trained model is used to obtain probabilistic forecasts of the renewable energy generation. The predictive mean and standard deviation of the predicted distribution is then computed.

Algorithm 1 represents the pseudo-code for the proposed scheme. After performing data pre-processing on input data, it is fed to the VAE stage to encode the higher time lags into lower dimensions. Data is encoded using variational autoencoder and decoded from latent space to monitor the information loss during the encoding process (line 5-9). After achieving the least value of \( \tau_{loss} \), the encoded representation is used for forecasting stage employed on Bayesian probability. The learning process is mainly carried out with the help of densovariaotional layer integrated with BiLSTM hidden layer (line 12-15). With the concept of MFVI, posterior distribution is obtained for the compressed weight parameters. The training in forecasting stage is optimized by maximizing the ELBO (line 12). After achieving the minimum loss, predictive mean is calculated from the sampled mean and variance and the proposed method is tested on the evaluation metrics (line 19-22).

IV. RESULTS AND DISCUSSIONS

This section presents the implementation details, evaluation criteria, and discussions on results for the proposed VAE-Bayesian BiLSTM method using solar generation time series data taken from Ausgrid [34]. The highly granular smart
meter data is acquired in kWh for year 2010-2013 from 300 individual smart homes having rooftop solar panels. For implementation purpose, first three months of solar generation data is taken from house no. 2076 and divided into training and testing set with split ratio of 80:20. The implementation results are obtained using an i7 processor, 16 GB RAM, Nvidia graphics processing unit (GPU), employed with tensorflow framework of python programming language.

A. Evaluation metrics

To evaluate the proposed VAE-Bayesian BiLSTM and comparative probabilistic deep learning-based forecasting methods, average pinball loss is computed focusing on the sharpness and consistency of the approximated distribution. In this regard, least value of pinball loss is more desirable. For actual data values \(y_t\) and predictions at \(t^{th}\) time-stamp \((\hat{y}_{t,q})\), pinball loss over percentile \(q \in [0,1]\) is formulated as following:

\[
\text{pinball}(y_t, \hat{y}_{t,q}, q) = \begin{cases} 
(y_t - \hat{y}_{t,q})q & y_t \leq \hat{y}_{t,q} \\
(\hat{y}_{t,q} - y_t)(1 - q) & y_t > \hat{y}_{t,q}
\end{cases}
\] (18)

Furthermore, forecasting methods are evaluated by computing root-mean square error (RMSE) and mean absolute error (MAE) of the differences between actual and predicted data values. The respective equations for RMSE and MAE are given as below:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (\hat{y}_t - y_t)^2}
\] (19)

\[
MAE = \frac{1}{n} \sum_{t=1}^{n} | \hat{y}_t - y_t |
\] (20)
where \( \hat{y}_t \) and \( y_t \) symbolize the predictive mean and actual values at time stamp \( t \) and \( n \) represents the total number of data samples from testing set. In addition, coefficient of determination symbolized as \( R \) is also obtained to represent the goodness of a fit by forecasting model. It is given as below:

\[
R = 1 - \frac{\sum (\hat{y}_t - y_t)^2}{\sum (y_t - \bar{y}_t)^2}
\]  

(21)

where \( \bar{y}_t \) represents the mean value for actual data points.

B. Framework Evaluation and Comparative Analysis

This section presents the implementation results for the proposed VAE-Bayesian BiLSTM framework. For training purpose, Adam optimizer [35] is utilized and the algorithm is optimized with respect to negative log likelihood and KL divergence over 100 epochs. Moreover, to avoid the false positive results validation data with 0.2 split is further used during training and thus, validation loss is also monitored. Since deep learning methods are highly exposed to the problem of overfitting, the technique of early stopping is employed with patience rate of 10. The input is fed in the form of 96 time lags representing two days of solar generation data and VAE model intends to encode the input to 24 time lags using latent distribution.

Additionally, a comparative analysis is conducted to evaluate the performance of the proposed method with state-of-the-art deep learning-based probabilistic methods as reflected in Table I. From the table, it can be observed that the number of weight parameters significantly decrease for Bayesian BiLSTM when integrated with VAE (M1,M4). Note that a trade-off is observed between the two methods for time complexity and forecasting error, as VAE contributes to some information loss during data encoding which leads to compromised values in accuracy. In this regard, it is constructive to outline that while employing VAE, reconstruction error has to be monitored and minimized with respect to the error threshold value. As reflected in the table, our proposed method outperforms VAE-Bayesian LSTM and VAE-Bayesian RNN in terms of forecasting error and pinball loss highlighting its best performance for probabilistic forecasting and uncertainty quantification. A comparison graph for RMSE and pinball loss is provided in Fig. 5 for methods (M1-M7). Additionally, higher values of R-score for BiLSTM-based methods justify the goodness of fit and learning performance claimed by bidirectional LSTM layers.

Furthermore, we affirm the efficacy of our proposed method by performing an extensive analysis with different data sizes in comparison to the Bayesian BiLSTM method. In this regard, Table II and III compares above two methods on the basis of evaluation metrics for six months and one year of solar generation data respectively. It can be observed from the results of these two data variations that with an increase in data size, there data encoding using VAE significantly improves the computational efficiency in terms of CPU time and weight dimensionality. Precisely, With one year of data in Table III execution time falls by one third with the proposed method (M1). Furthermore, with a considerable weight reduction from 76, 4224 to 7, 494 a slight increase in the forecasting error is noticed. However, pinball loss exhibits the competitive performance for probabilistic forecasting by M1 with M4. Note that the reported CPU execution time in the results refers to the learning process explicitly. Execution time for common computations between all methods is not included such as for data normalization, error computation, etc. Also, the best results are summarized in this section after appropriate hyperparameter tuning, for brevity.

Furthermore, Fig. 6(a) and Fig. 6(b) demonstrate two days and one week of forecasting horizons respectively, plotted with ground truth from test set and predictive mean given by proposed VAE-Bayesian BiLSTM method. Additionally, 90 % and 50 % PIs are plotted which reflect the future probabilities on different percentiles in a relation to the predictive mean and ground truth. It is inferred from the figures that the proposed method is capable to quantify uncertainties along with tighter bounds on future probabilities in form of PIs. Similarly, Fig. 7 exhibits the PIs generated by VAE-Bayesian LSTM method on two horizons. From the graphs, it is clear that VAE-Bayesian with LSTM has wider prediction coverage for future probabilities corresponding to less reliability and sharpness in comparison to the proposed method. Based on above results, it can be concluded that the proposed VAE-Bayesian BiLSTM method effectively quantifies the uncertainties with tighter PIs and lower computational cost. However, a trade-off needs to be maintained between the computational efficiency and forecasting performance, besides maintaining a lower reconstruction error.

V. CONCLUSION

In this paper, a new VAE-based Bayesian BiLSTM technique for renewable energy generation forecasting is presented to quantify the model and stochastic uncertainties, while optimising the number of weight parameters to improve the computational efficiency. The proposed technique outperforms benchmark point and probabilistic deep learning forecasting techniques in terms of evaluation metrics. Numerical results presented in the results section demonstrate the superior forecasting and computational performance of the proposed method in comparison to the other probabilistic deep learning methods. Furthermore, uncertainties are addressed by pro-
Table I: Comparative analysis

| Method no. | Method name            | RMSE (avg) | MAE  | R-score | pinball error | CPU Time (min) | weights |
|------------|------------------------|------------|------|---------|---------------|----------------|---------|
| M1         | Proposed VAE-Bayesian BiLSTM | 0.0985     | 0.0679 | 0.8853  | 0.0386        | 5.49           | 7,494   |
| M2         | VAE-Bayesian LSTM       | 0.1040     | 0.0705 | 0.8722  | 0.0491        | 2.5            | 3,846   |
| M3         | VAE-Bayesian RNN        | 0.1470     | 0.1172 | 0.7447  | 0.0548        | 2.6            | 1,974   |
| M4         | Bayesian BiLSTM         | 0.0898     | 0.0493 | 0.9046  | 0.0321        | -              | 16.32   |
| M5         | Bayesian ANN            | 0.2333     | 0.1846 | 0.7871  | 0.0957        | -              | 0.21    |
| M6         | VAE-BiLSTM              | 0.0953     | 0.0509 | 0.8943  | 0.1427        | 3.21           | 7,369   |
| M7         | VAE-LSTM                | 0.1108     | 0.0754 | 0.8548  | 0.1512        | 0.6941         | 1       |

Table II: Comparative analysis with six months of data for BiLSTM with VAE

| Method no. | Method name            | RMSE (avg) | MAE  | R-score | pinball error | CPU Time (min) | weights |
|------------|------------------------|------------|------|---------|---------------|----------------|---------|
| M1         | VAE-Bayesian BiLSTM    | 0.0998     | 0.0697 | 0.8476  | 0.0207        | 6.369          | 4.28    |
| M4         | Bayesian BiLSTM        | 0.0872     | 0.0532 | 0.8838  | 0.0413        | -              | 9.45    |

Table III: Comparative analysis with one year of data for BiLSTM with VAE

| Method no. | Method name            | RMSE (avg) | MAE  | R-score | pinball error | CPU Time (min) | weights |
|------------|------------------------|------------|------|---------|---------------|----------------|---------|
| M1         | VAE-Bayesian BiLSTM    | 0.0630     | 0.0503 | 0.8842  | 0.0421        | 7.317          | 9.7     |
| M4         | Bayesian BiLSTM        | 0.0582     | 0.0532 | 0.8838  | 0.0305        | -              | 29.47   |

Fig. 6. Forecasting results for proposed VAE-Bayesian BiLSTM method efficiently in the form of future prediction intervals.

REFERENCES

[1] W. El-Khattam and M. M. Salama, “Distributed generation technologies, definitions and benefits,” Electric power systems research, vol. 71, no. 2, pp. 119–128, 2004.

[2] M. Sun, T. Zhang, Y. Wang, G. Strbac, and C. Kang, “Using bayesian deep learning to capture uncertainty for residential net load forecasting,” IEEE Transactions on Power Systems, vol. 35, no. 1, pp. 188–201, 2019.

[3] O. Gandhi, D. S. Kumar, C. D. Rodriguez-Gallegos, and D. Srinivasan, “Review of power system impacts at high pv penetration part i: Factors limiting pv penetration,” Solar Energy, vol. 210, pp. 181–201, 2020.

[4] Y. Wang, Q. Chen, T. Hong, and C. Kang, “Review of smart meter data analytics: Applications, methodologies, and challenges,” IEEE Transactions on Smart Grid, vol. 10, no. 3, pp. 3125–3148, 2018.

[5] D. Kaur, S. N. Islam, M. Mahmud, Z. Dong et al., “Energy forecasting in smart grid systems: A review of the state-of-the-art techniques,” arXiv preprint arXiv:2011.12598, 2020.

[6] H. Shi, M. Xu, and R. Li, “Deep learning for household load forecasting: a novel pooling deep rnn,” IEEE Transactions on Smart Grid, vol. 9,
(a) Predictive mean and PIs for 2 days

(b) Predictive mean and PIs for one week

Fig. 7. Forecasting results for VAE-Bayesian LSTM

[7] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, “Short-term residential load forecasting based on lstm recurrent neural network,” IEEE Transactions on Smart Grid, vol. 10, no. 1, pp. 841–851, 2017.

[8] D. Kaur, R. Kumar, N. Kumar, and M. Guizani, “Smart grid energy management using mnn-lstm: A deep learning-based approach,” in 2019 IEEE Global Communications Conference (GLOBECOM). IEEE, 2019, pp. 1–6.

[9] H. Jahangir, H. Tayarani, S. S. Gougher, M. A. Golkar, A. Ahmadian, and A. Elkamel, “Deep learning-based forecasting approach in smart grids with micro-clustering and bi-directional lstm network,” IEEE Transactions on Industrial Electronics, 2020.

[10] J.-F. Toubeau, J. Bottieau, F. Vallée, and Z. De Greve, “Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets,” IEEE Transactions on Power Systems, vol. 34, no. 2, pp. 1203–1215, 2018.

[11] N. Amjady, F. Keynia, and H. Zareipour, “Short-term wind power forecasting using ridgelet neural network.” Electric Power Systems Research, vol. 81, no. 12, pp. 2099–2107, 2011.

[12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” The journal of machine learning research, vol. 15, no. 1, pp. 1929–1958, 2014.

[13] Y. Gal and Z. Ghahramani, “Dropout as a bayesian approximation: Representing model uncertainty in deep learning,” in International conference on machine learning. PMLR, 2016, pp. 1050–1059.

[14] R. J. Bessa, A. Trindade, C. S. Silva, and V. Miranda, “Probabilistic solar power forecasting in smart grids using distributed information,” International Journal of Electrical Power & Energy Systems, vol. 72, pp. 16–23, 2015.

[15] A. Bracale, P. Caramia, G. Carpinelli, A. R. Di Fazio, and P. Varilone, “A bayesian-based approach for a short-term steady-state forecast of a smart grid,” IEEE Transactions on Smart Grid, vol. 4, no. 4, pp. 1760–1771, 2013.

[16] W. Xie, P. Zhang, R. Chen, and Z. Zhou, “A nonparametric bayesian framework for short-term wind power probabilistic forecast,” IEEE Transactions on Power Systems, vol. 34, no. 1, pp. 371–379, 2018.

[17] C. Ning and F. You, “Data-driven adaptive robust unit commitment under wind power uncertainty: A bayesian nonparametric approach,” IEEE Transactions on Power Systems, vol. 34, no. 3, pp. 2409–2418, 2019.

[18] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, “Probabilistic forecasting of wind power generation using extreme learning machine,” IEEE Transactions on Power Systems, vol. 29, no. 3, pp. 1033–1044, 2014.

[19] T. Carriere, C. Vernay, S. Pitaval, and G. Kariniotakis, “A novel approach for seamless probabilistic photovoltaic power forecasting covering multiple time frames,” IEEE Transactions on Smart Grid, vol. 11, no. 3, pp. 2281–2292, 2020.

[20] Y. Gal, “Uncertainty in deep learning,” University of Cambridge, vol. 1, no. 3, p. 4, 2016.

[21] A. Kendall and Y. Gal, “What uncertainties do we need in bayesian deep learning for computer vision?” in Advances in neural information processing systems, 2017, pp. 5574–5584.

[22] Y. Yang, W. Li, T. A. Gulliver, and S. Li, “Bayesian deep learning-based probabilistic load forecasting in smart grids,” IEEE Transactions on Industrial Informatics, vol. 16, no. 7, pp. 4703–4713, 2019.

[23] R. Yacef, M. Benghanem, and A. Melliti, “Prediction of daily global solar irradiation data using bayesian neural network: A comparative study,” Renewable Energy, vol. 48, pp. 146–154, 2012.

[24] M. Q. Raza, N. Mithulananthan, and A. Sumnerfield, “Solar output power forecast using an ensemble framework with neural predictors and bayesian adaptive combination,” Solar Energy, vol. 166, pp. 226–241, 2018.

[25] C. Zhang, J. Büttepage, H. Kjellström, and S. Mandt, “Advances in variational inference,” IEEE transactions on pattern analysis and machine intelligence, vol. 41, no. 8, pp. 2008–2026, 2018.

[26] Y. Liu, H. Qin, Z. Zhang, S. Pei, C. Wang, X. Yu, Z. Jiang, and J. Zhou, “Ensemble spatiotemporal forecasting of solar irradiation using variational bayesian convolutional gated recurrent unit network,” Applied Energy, vol. 253, p. 113596, 2019.

[27] X. Zhang, F. Fang, and J. Wang, “Probabilistic solar irradiation forecasting based on variational bayesian inference with secure federated learning,” IEEE Transactions on Industrial Informatics, pp. 1–1, 2020.

[28] P. Bachhav, M. Todisco, and N. Evans, “Latent representation learning for artificial bandwidth extension using a conditional variational auto-encoder,” in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 7010–7014.

[29] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” arXiv preprint arXiv:1312.6114, 2013.

[30] A. Dairi, F. Harrou, Y. Sun, and S. Khadroura, “Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach,” Applied Sciences, vol. 10, no. 23, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/23/8400

[31] J. Pereira and M. Silveira, “Unsupervised anomaly detection in energy time series data using variational recurrent autoencoders with attention,” in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2018, pp. 1275–1282.

[32] X. Wang, P. Cui, Y. Du, and Y. Yang, “Variational autoencoder based fault detection and location method for power distribution network,” in 2020 8th International Conference on Condition Monitoring and Diagnosis (CMD). IEEE, 2020, pp. 282–285.

[33] S. Biswas, R. Meyur, and V. A. Centeno, “Devlearn: A deep visual reasoning framework for determining the location of temporary faults in power systems,” in 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm). IEEE, 2020, pp. 1–6.

[34] Ausgrid. [Online]. Available: https://www.ausgrid.com.au/Industry/ Our-Research/Data-to-share/Solar-home-electricity-data

[35] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.