

# Robust Wind Speed Forecasting: A Deep Spatio-Temporal Approach

Mohsen Saffari  
Department of Computer Science  
University of Tulsa  
Tulsa, OK, USA  
mos0968@utulsa.edu

Michael Williams  
Department of Computer Science  
University of Tulsa  
Tulsa, OK, USA  
mtw6402@utulsa.edu

Mahdi Khodayar  
Department of Computer Science  
University of Tulsa  
Tulsa, OK, USA  
mahdi-khodayar@utulsa.edu

Miadreza Shafie-khah  
School of Technology and Innovations  
University of Vaasa  
Vaasa, Finland  
mshafiek@univaasa.fi

João P. S. Catalão  
INESCTEC  
Faculty of Engineering of University of Porto  
Porto, Portugal  
catalao@fe.up.pt

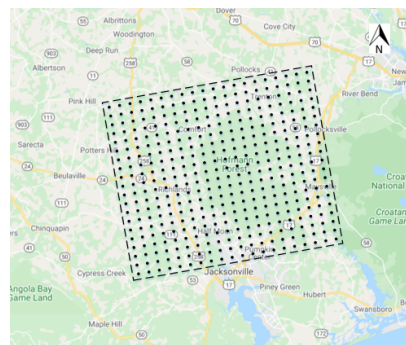
**Abstract**—With the significant increase in wind speed usage as a clean source of energy, an accurate wind speed forecasting system is a must for more effective utilization of this energy. Failure to consider the inherent spatio-temporal features of wind speed time series leads to the lack of generalization capacity for current wind speed forecasting approaches. This paper proposes an end-to-end deep neural network framework, i.e., convolutional rough long short-term memory (ConvRLSTM), to extract spatio-temporal wind correlations and mitigate the inherent uncertainties in wind time series by incorporating the Rough set theory into a combination of convolution neural network (CNN) and LSTM units. Our proposed model receives the historical data of wind speed for a  $20 \times 20$  array of wind turbines in North Carolina, US. Several ConvRLSTM layers extract the most relevant features for the forecasting task, and finally, fully connected layers predict 400 wind speed values using the spatial features obtained by the CNN and temporal features computed by the LSTM. Through analyzing the numerical forecasting results, it can be inferred that the proposed approach outperforms the mainstream and recently published forecasting strategies in terms of the RMSE metric.

**Keywords**—Convolutional neural networks, Rough set theory, Long short term memory, Spatio-Temporal correlations, Wind speed forecasting

## I. INTRODUCTION

In recent years, among all sustainable and green energy sources such as wave and hydropower, there is a significant interest in modern societies' wind energy usage. Due to the correlation between the wind power generated by a wind turbine and wind speed, proposing an accurate and reliable wind speed forecasting method leads to the improvement of wind energy predictions [1]. During the past decade, several approaches have been considered for wind speed forecasting: 1) Numerical calculation methods, 2) Statistical methods, and 3) Machine Learning (ML) based methods. Numerical calculations including Weather Research and Forecasting (WRF) [2] and Mesoscale Model (MM5) [3] mainly exploit satellite

images and numerical simulations to predict wind speeds. Commonly used statistical methods for wind speed forecasting include the Gaussian process model [4], auto-regressive integrated moving average (ARIMA) [5], as well as the moving average (MA) [6]. Although statistical methods have shown a reliable performance for ultra-short-term and short-term wind speed forecasting, the general linear assumptions of these approaches cannot model existing stochastic, uncertain, and highly nonlinear nature of wind speed data.



**Fig. 1:** A  $20 \times 20$  wind turbine cluster in North Carolina. Each spot illustrates a wind turbine.

ML-based approaches are data-driven methods that have been exploited for a variety of regression [7] and classification [8] tasks, recently. Due to the shortcomings of primary methods, among all recent wind forecasting approaches, ML-based and hybrid methods have attracted the attention of a significant number of studies in this regard [9]. Liu et al. proposed a hybrid multi-module deep model for wind speed forecasting, which uses extreme learning machine (ELM), outlier robust ELM (ORELM), and deep belief neural network (DBN) for short-term forecasting tasks [10]. To reduce the nonstationarity of wind speed, preprocessing approaches are widely considered in wind speed sequence forecasting models. For

instance, Duan et al. [11] developed a forecasting model which extracts IMFs of raw wind speed based on Variational modal decomposition and forecasts the future values of wind speed time series by long-short term memory structure (LSTM). In this line of research, authors of [12] first extract the most representative information by Jaya optimization algorithm, then train a support vector machine (SVM) model for the prediction task.

The majority of studies merely capture the particular pattern in raw wind speed sequences in the time domain (temporal information); however, it is shown that there is a high correlation in neighboring wind turbine sites [13]. Hence, several recent studies have addressed the spatio-temporal forecasting methods. For example, in [14] authors developed a graph convolutional deep learning architecture (GCDLA) that includes LSTM neural networks (NNs) to capture the relevant spatio-temporal features and predict the future values of time series with higher accuracy. More recently, Lu et al. [15] proposed a short-term forecasting framework based on spatio-temporal analysis and multi-output support vector machine (MSVM). The authors used the Pearson correlation coefficient and partial autocorrelation function to investigate the spatio-temporal correlation between wind sites, and their optimized MSVM is employed to predict 15 sites' time series.

In this paper, we address the problem of robust wind speed forecasting for multiple wind sites simultaneously. We model the temporal information of 400 wind turbines farm in North Carolina state, USA, as a  $20 \times 20$  intensity map (Fig. 1). Recently, Convolutional NNs (CNN) and LSTM NNs have shown significant performance in the extraction of spatial and temporal features, respectively. Due to these models' superiority, in this paper, we proposed a novel end-to-end framework for wind speed forecasting. The proposed model, convolutional rough LSTM (ConvRLSTM), benefits from the CNN layer's advantages in the extraction of the spatial features and LSTM NN for capturing the temporal information of raw wind speed time series. Moreover, by making use of Rough set theory [16], [17], we use the interval convolution in the ConvRLSTM model to mitigate the uncertainties in wind speed time series. The main contribution of this study include:

- 1) This is the first study for robust spatio-temporal feature extraction and forecasting in the wind speed forecasting research. In contrast to previous works, this model studies how to incorporate interval knowledge into deep learning to develop a robust machine learning model for wind speed prediction tasks.
- 2) A novel combination of CNNs and LSTMs is proposed for the first time for accurate prediction of wind speed measurements at multiple neighboring sites. The CNN captures the spatial patterns while the LSTM learns complex temporal patterns in the highly varying wind datasets.
- 3) To the best of our knowledge, this work is the first attempt to develop a Rough neural network for the spatiotemporal prediction of wind speed time series. We incorporate the Rough set theory into the proposed

hybrid deep neural network to handle noise and uncertainties exist in wind data.

This paper is organized as follows: Section II discusses the problem formulation of wind speed forecasting for 400 wind sites. In Section III the details of the proposed ConvRLSTM model is explained. The experimental results and comparison of the proposed approach with state-of-the-art wind speed forecasting methodologies are made in Section IV. Finally, Section V provides the conclusion of this paper.

## II. PROBLEM FORMULATION

Our aim for wind speed forecasting is to predict the future values of the wind speed time series corresponding to multiple neighboring wind sites. Let us consider the location of each site by two dimensional coordinates  $(i, j)$ . The spatial information of the underlying sites at time  $t$  is stored in a  $I \times J$  array denoted by  $S^t$ :

$$S^t = \begin{pmatrix} s_{(1,1)}^t & s_{(1,2)}^t & \cdots & s_{(1,J)}^t \\ s_{(2,1)}^t & s_{(2,2)}^t & \cdots & s_{(2,J)}^t \\ \vdots & \vdots & \ddots & \vdots \\ s_{(I,1)}^t & s_{(I,2)}^t & \cdots & s_{(I,J)}^t \end{pmatrix} \quad (1)$$

where  $s_{(i,j)}^t$  describes the wind speed values for site  $(i, j)$  at time  $t$ . Our goal is to predict the next wind frame,  $\hat{S}_{t+h}$ , corresponding to the forecasting horizon,  $h$ , by the historical values of wind speeds  $\tilde{S}$ . More formally, the problem of wind speed prediction is formulated by:

$$\hat{S}_{t+h} = \underset{\theta}{\operatorname{argmax}} \mathcal{P} \left( S_{t+h} | \tilde{S}_{t-l+1}, \tilde{S}_{t-l+2}, \dots, \tilde{S}_t \right) \quad (2)$$

where  $\theta$  is the machine learning model parameters, and  $l \geq 1$  is the historical time lag. Since the wind speed information is highly nonlinear in nature, similar to [18], [19] to compute the nonlinear correlation between the time series values in different times, mutual information (MI) is employed. Hence, to form the spatial wind speed matrix  $S$  in (1) and to select the input variables for the forecasting framework the MI value between  $S_{t+1}$  and  $S_{t-l+1}$  corresponding  $l \geq 1$  is calculated. The wind speed data related to the time lags with higher MI values than  $\lambda$  are considered as an input set for our proposed algorithm. Note that,  $\lambda$  is a hyperparameter determined by empirical experiments.

## III. CONVOLUTIONAL ROUGH LONG SHORT-TERM MEOMORY

### A. Rough Set Theory

Rough set theory is a mathematical tool introduced by Pawlak [16] that has shown its superiority to handle the uncertainty and noise in time series in several recent studies [14], [19]. The theory defines an information system,  $S = \langle U, A, V, f \rangle$ , where the finite non-empty sets  $U$  and  $A$  are called the universe of primitive objects and set of attributes, respectively. Each attribute  $a \in A$  belongs to a specific domain set  $V_a$  and  $V$  consists of all attribute domains

in  $A$ ,  $V = \bigcup_{a \in A} V_a$ .  $S$  defines a total information function,  $f : U \times A \rightarrow V$ , that for every  $a \in A$  and  $x \in U$ ,  $f(x, a) \in V_a$ .

Consider  $Z$  as a subset of attributes in  $A$  ( $Z \subseteq A$ ), two objects  $v_1$  and  $v_2$  in information system  $S$  are indiscernible from each other if and only if  $\forall z \in Z : f(v_1, z) = f(v_2, z)$ . Using the entire knowledge in  $Z$  the Rough set theory defines two types of estimation for any concept set  $X \subseteq U$ :  $\overline{ZX}$  and  $\underline{ZX}$  which are called Z-Upper bound and Z-Lower bound estimation,

$$\begin{aligned}\overline{ZX} &= \cup\{O \in U | Z : O \cap X \neq \emptyset\} \\ \underline{ZX} &= \cup\{O \in U | Z : O \subseteq X\}\end{aligned}\quad (3)$$

And the  $Z$ -boundary region of  $X$  is obtained by,

$$BND_Z(X) = \overline{ZX} - \underline{ZX} \quad (4)$$

$BND_Z(X)$  describes the vagueness of  $X$ .  $\underline{ZX}$  denotes all objects in  $U$  that certainly belong to  $X$ ; however,  $\overline{ZX}$  shows a set of objects in  $U$  that can possibly be classified as a member of  $X$  using the knowledge of attributes in  $Z$ . If  $BND_Z(X) = \emptyset$ , then  $X$  is called a crisp set in respect of  $Z$  set, otherwise  $X$  is a rough set.

### B. Proposed Model

LSTM is a deep recurrent neural network (RNNs) that has proven its superiority in various recent sequence modeling applications [11], [14], [20]. The significant innovation of LSTM lies in its memory blocks,  $C_t$  that play a crucial role in the mitigation of the vanishing gradient issue by representing the long-term dependencies in time series. Although traditional LSTM structures have shown significant performance in modeling temporal correlation of time series, they cannot handle the spatial correlations.

Fig. 2 depicts the proposed robust spatiotemporal model. The mathematical formulations of a ConvRLSTM block are shown in (5)-(9):

$$\begin{aligned}i_t &= \text{ReLU}(\alpha_{S_i} \sigma(W_{S_i}^U \otimes S_t + b_{S_i}^U) + \beta_{S_i} \sigma(W_{S_i}^L \otimes S_t + b_{S_i}^L) \\ &+ \alpha_{H_i} \sigma(W_{H_i}^U \otimes H_{t-1} + b_{H_i}^U) + \beta_{H_i} \sigma(W_{H_i}^L \otimes H_{t-1} + b_{H_i}^L) \\ &+ \alpha_{C_i} \sigma(W_{C_i}^U \odot C_{t-1} + b_{C_i}^U) + \beta_{C_i} \sigma(W_{C_i}^L \odot C_{t-1} + b_{C_i}^L))\end{aligned}\quad (5)$$

$$\begin{aligned}f_t &= \text{ReLU}(\alpha_{S_f} \sigma(W_{S_f}^U \otimes S_t + b_{S_f}^U) + \beta_{S_f} \sigma(W_{S_f}^L \otimes S_t + b_{S_f}^L) \\ &+ \alpha_{H_f} \sigma(W_{H_f}^U \otimes H_{t-1} + b_{H_f}^U) + \beta_{H_f} \sigma(W_{H_f}^L \otimes H_{t-1} + b_{H_f}^L) \\ &+ \alpha_{C_f} \sigma(W_{C_f}^U \odot C_{t-1} + b_{C_f}^U) + \beta_{C_f} \sigma(W_{C_f}^L \odot C_{t-1} + b_{C_f}^L))\end{aligned}\quad (6)$$

$$\begin{aligned}C_t &= f_t \odot C_{t-1} + i_t \odot \tanh(\alpha_{S_c} \sigma(W_{S_c}^U \otimes S_t + b_{S_c}^U) \\ &+ \beta_{S_c} \sigma(W_{S_c}^L \otimes S_t + b_{S_c}^L) + \alpha_{H_c} \sigma(W_{H_c}^U \otimes H_{t-1} + b_{H_c}^U) \\ &+ \beta_{H_c} \sigma(W_{H_c}^L \otimes H_{t-1} + b_{H_c}^L))\end{aligned}\quad (7)$$

$$\begin{aligned}O_t &= \text{ReLU}(\alpha_{S_o} \sigma(W_{S_o}^U \otimes S_t + b_{S_o}^U) + \beta_{S_o} \sigma(W_{S_o}^L \otimes S_t + b_{S_o}^L) \\ &+ \alpha_{H_o} \sigma(W_{H_o}^U \otimes H_{t-1} + b_{H_o}^U) + \beta_{H_o} \sigma(W_{H_o}^L \otimes H_{t-1} + b_{H_o}^L) \\ &+ \alpha_{C_o} \sigma(W_{C_o}^U \odot C_t + b_{C_o}^U) + \beta_{C_o} \sigma(W_{C_o}^L \odot C_t + b_{C_o}^L))\end{aligned}\quad (8)$$

$$H_t = O_t \odot \tanh(C_t) \quad (9)$$

where  $\otimes$  and  $\odot$  denote convolution and element-wise product, respectively. The Sigmoid function is used as a latent activation function  $\sigma$  in ConvRLSM. As shown in Fig. 2, at time step  $t$ , the ConvRLSTM is fed by the spatial matrix,  $S_t$ , as well as the previous hidden state and memory cell,  $H_{t-1}$  and  $C_{t-1}$ , respectively. The model applies convolution on inputs  $S_t$  while the hidden state  $H_t$  is computed using the interval (upper- and lower- bounded) kernels and biases,  $\langle W^H, b^H \rangle$  and  $\langle W^L, b^L \rangle$ , respectively. Each time that a new input is inserted into a layer, the corresponding information is accumulated in the ConvRLSTM cell if the input cell  $i_t$  is fired. Also the past memory cell  $C_{t-1}$  must be forgotten if the forget gate  $f_t$  is activated. Moreover,  $O_t$  controls the propagation of the current memory cell  $C_t$  to the final hidden state  $h_t$ .

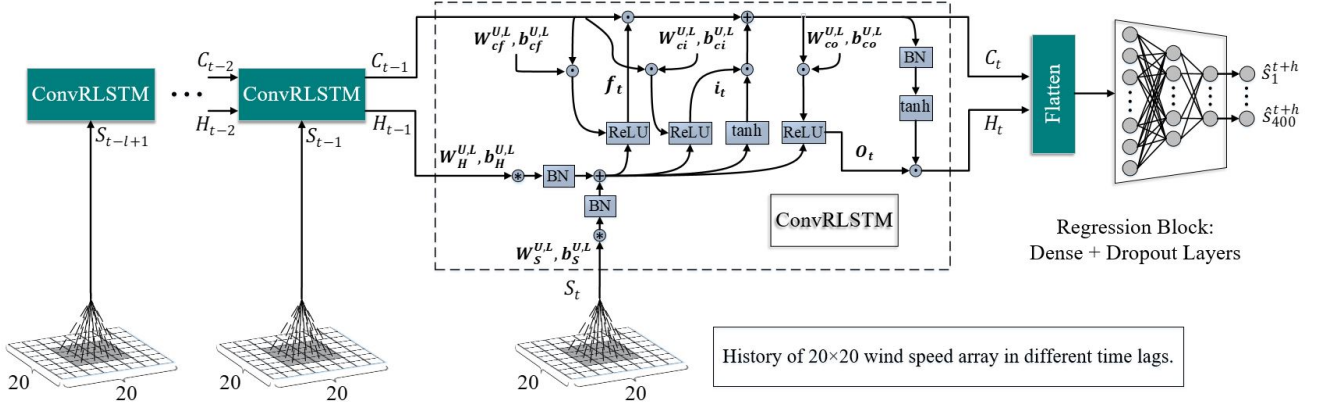
As shown in (5)-(9), the proposed ConvRLSTM has a convolution operator in both state-to-state and input-to-state transition. Also, in contrast to the classic LSTM, the input and output elements are 3D tensors rather than 1-dimensional vectors; hence, they are capable of providing more informative knowledge with larger parameter space. Making use of convolution operations on the spatial matrix and hidden state information leads to capturing complex spatial features corresponding to an array of the time series. As shown in Fig. 2, on top of the forecasting framework, multiple fully connected layers are used to yield predictions. Using gradient descent with root mean squared error between the predicted values and the actual values of the wind speed time series, the whole parameters of the model can be updated in an end-to-end manner.

As shown in Fig. 2, the proposed framework for the wind speed forecasting task has multiple stacked ConvRLSM cells. The number of ConvRLSTM cells (i.e., the number of input variables),  $l$ , significantly affects the model performance. This number helps us have a trade-off between the completeness of input information and the model's complexity. With a too small  $l$ , the historical information is not rich enough for the prediction task, whereas the large  $l$  results in an unnecessarily large hyperparameter space for the model. Similar to previous studies, to determine hyperparameter  $l$ , we used mutual information (MI) analysis. Considering  $s_t$  as the wind speed values at time  $t$ , the MI of the  $s_{t-l+1}$  and  $s_t$  is computed for  $l \in [1, 100]$  in the validation set. Fig. 3 shows the result of MI analysis corresponding to the  $l$  values. As shown in this plot, with the increase of our time lag,  $l$ , the correlation between the current and previous wind speed values decreases. In this work, by setting a threshold value  $\lambda$ , we choose the wind speed values corresponding to time-lags with MI more than  $\lambda = 0.25$ . This results in the incorporation of time-lags from  $l = 1$  to  $l = 24$ . Hence, in the proposed model we set  $\langle s(t-24), s(t-22), \dots, s(t) \rangle$  as the model's inputs.

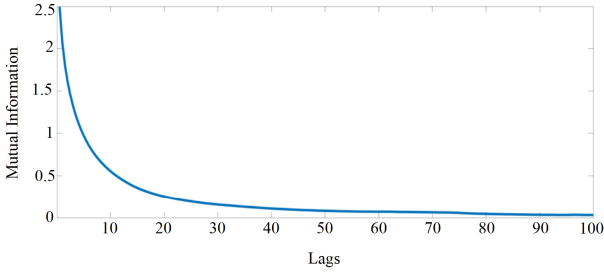
## IV. NUMERICAL RESULTS

### A. Evaluation Criteria

In this work, two famous criteria, the root mean square (RMSE) and mean absolute error (MAPE) are employed to



**Fig. 2:** The proposed ConvRLSTM for wind speed prediction. Symbols  $\otimes$  and  $\odot$  depict the interval convolution and the Hadamard product.



**Fig. 3:** Mutual Information of various time lags.

evaluate the prediction accuracy at each site  $(i, j)$  where  $1 \leq i, j \leq 20$ . The  $RMSE_{i,j}$  and  $MAE_{i,j}$  metrics are calculated as:

$$RMSE_{i,j} = \sqrt{\frac{1}{N} \sum_{n=1}^N (s_{i,j}(n) - \hat{s}_{i,j}(n))^2} \quad (10)$$

$$MAE_{i,j} = \frac{1}{N} \sum_{n=1}^N |s_{i,j}(n) - \hat{s}_{i,j}(n)|$$

where  $N$  is the number of samples, also the actual values and predicted values of time series are shown by  $s$  and  $\hat{s}$ , respectively. According to the defined criteria, the error indices for the whole array can be defined as,

$$RMSE_{Total} = \frac{1}{I \times J} \sum_{i=1}^I \sum_{j=1}^J RMSE_{i,j} \quad (11)$$

$$MAE_{Total} = \frac{1}{I \times J} \sum_{i=1}^I \sum_{j=1}^J MAE_{i,j}$$

where  $i$  and  $j$  determine the total number of sites and in this paper are equal to 20. The smaller the values of defined criteria are, the better the performance of the model is.

### B. Experimental Settings

We trained the proposed model on the wind speed measurements of 400 sites from the Wind Integration National Dataset,

a publicly available dataset by the National Renewable Energy Laboratory [21]. The entire dataset includes the wind speed information of more than 126,000 sites in the US for 2007 – 2013. We consider the 75% and 25% of the data from 2007 till 2012 as training and validation data, respectively. Moreover, we test our model on the prediction of wind speed for the year 2013. Using the MI analysis, the number of ConvRLSTM cells is set to 24, and we adopt empirically three hidden layers with dimension  $\{578, 800, 400\}$  in the MLP regression block. To mitigate the overfitting effect we employ determinantal point process dropout proposed in [22] in regression block. In this work, we consider  $3 \times 3$  and zero padding kernels with a stride coefficient of one for applying convolution to the input and hidden states. Moreover, all tunable parameters are trained with Adam learning algorithm with 0.001 and 100 the learning rate and the number of epochs, respectively. All the experiments are carried out using GPU-based Tensorflow [23] on Python 3. The simulations are processed in a system with a 10-core CPU having Intel Core-i7 Processors, an NVidia Quadro RTX 6000 GPU, and a 256-GB RAM.

### C. Performance Analysis

In this paper, we compare the proposed model with several shallow and deep structures, recurrent NNs, and state-of-the-art Spatio-temporal structures including, feed-forward neural network (FFNN), time-delay neural network (TDNN), nonlinear autoregressive neural network (NARNN), stacked autoencoder (SAE) [1], deep belief network (DBN) [1], gated recurrent unit (GRU), long short term memory (LSTM), spatio-temporal neural network (STNN) [24], and convolutional graph autoencoder (CGA) [18].

Tables I and II compare the  $RMSE_{Total}$  and the  $MAE_{Total}$  of the forecasting methods over 10-min to 3-hour horizon time, respectively. As shown by the tables, both criteria increase with the extension of the time horizons. As shown in the tables, FFNN has acceptable performance for ultra-short (i.e., 10-min) forecasting; however, it shows a poor performance with the increase in the horizon time steps. Generally speaking,

TDNN and NARNN outperform FFNN by 0.334 and 0.518 improvement on average  $RMSE_{Total}$  over all time horizons, respectively. This superiority originates from the TDNNs and NARNNs structure, whereby recurrent signals help the model to capture more relevant temporal features from the wind speed time series. SAE, which applies a set of auto-encoders to extract the nonlinear manifolds, works more accurately than NARNN. For example, considering the Table II, SAE reaches 0.942 and 1.982, while, NARNN reaches 0.976 and 2.007 in 30-min and 3-hour time horizon, respectively. The DBN model consists of multiple Restricted Boltzmann machines trained with Contrastive Divergence method [25]. In the tables, one can observe the decreasing in the averaged  $RMSE_{Total}$  and the  $MAE_{Total}$  of the DBN over the NARNN by 0.414 and 0.223, respectively. These improvements of SAE and DBN prove the deep structures' higher generalization capacity compared to the shallow neural networks due to having more nonlinear latent layers, which helps these models propose more accurate forecasting.

Compared to DBN and SAE, the recurrent deep structures, i.e., GRU and LSTM, show more reliable performance, especially on larger forecasting time horizons. As shown in the Table I, considering 2-hour and 3-hour time horizons, the GRU achieves 1.786 and 2.126, respectively, while for the DBN these values are 1.803 and 2.212. The results are further improved on both  $RMSE_{Total}$  and the  $MAE_{Total}$  when LSTM is applied. In comparison with DBN, the LSTM decreases the averaged  $RMSE_{Total}$  and the  $MAE_{Total}$  by 0.520 and 0.354, respectively. This improvement in the results shows the superiority of LSTM over previous methods in modeling the temporal correlations in the time series.

By considering the concept of spatio-temporal sequence forecasting, STNN and CGA methods result in more accurate results than LSTM. Based on the Table I, STNN which combines GRU with convolutional neural networks decreases the averaged  $RMSE_{Total}$  and the  $MAE_{Total}$  of LSTM by 0.138 and 0.148, respectively. The CGA that is originally proposed for probabilistic spatio-temporal solar irradiance forecasting reaches the lowest  $RMSE_{Total}$  and the  $MAE_{Total}$  among mentioned baselines. This model extracts the spatio-temporal features by applying a convolution graph to the input time series. The CGA model works slightly better than STNN, for example, in 10-min and 3-hour ahead forecasting, the averaged STNN's  $MAE_{Total}$  over all time horizons is 1.180 while this value for CGA is 1.136. Finally, the proposed ConvRLSTM model dominates all underlying baselines. In comparison with CGA as the best baseline, ConvRLSTM decreased the averaged  $RMSE_{Total}$  and the  $MAE_{Total}$  of CGA by 0.343 and 0.278. The better performance of ConvRLSTM is due to making use of the LSTM model and applying interval convolution, which mitigates the stochastic behavior of wind speed time series. Fig. (4(a)) and (4(b)) compare the relative performance of the proposed model with traditional forecasting methods and the state-of-the-art approaches for 250 test sample on July 26th in 3-hour horizon time, respectively. As shown in the plot, the proposed model follows the actual values

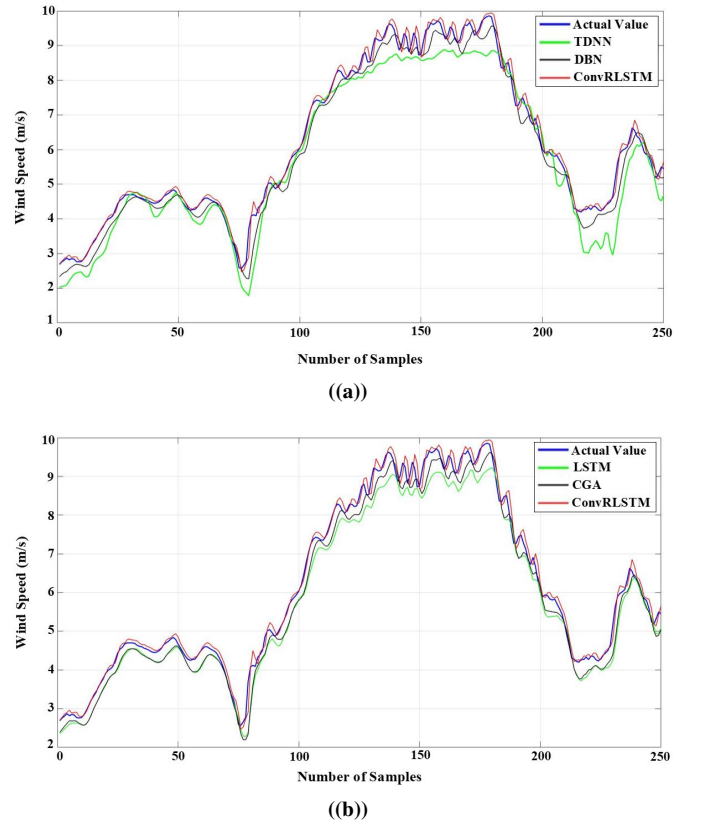
**TABLE I:** Obtained  $RMSE_{Total}$  for wind speed forecasting approach for 400 turbines array on several time horizons

Model	10-min	30-min	1-hr	2-hr	3-hr
FFNN	0.789	1.298	1.682	2.012	2.486
TDNN	0.753	1.221	1.621	1.981	2.357
NARNN	0.713	1.181	1.589	1.960	2.306
SAE	0.681	1.159	1.563	1.841	2.256
DBN	0.632	1.143	1.545	1.803	2.212
GRU	0.609	1.069	1.511	1.786	2.126
LSTM	0.562	1.002	1.492	1.754	2.005
STNN	0.533	0.987	1.472	1.701	1.984
CGA	0.510	0.962	1.376	1.672	1.923
ConvRLSTM	<b>0.492</b>	<b>0.931</b>	<b>1.211</b>	<b>1.578</b>	<b>1.888</b>

**TABLE II:** Obtained  $MAE_{Total}$  for wind speed forecasting approach for 400 turbines array on several time horizons

Model	10-min	30-min	1-hr	2-hr	3-hr
FFNN	0.591	1.029	1.496	1.911	2.266
TDNN	0.566	0.987	1.452	1.853	2.139
NARNN	0.504	0.976	1.403	1.736	2.007
SAE	0.476	0.942	1.374	1.701	1.982
DBN	0.451	0.931	1.366	1.689	1.966
GRU	0.420	0.903	1.310	1.632	1.923
LSTM	0.388	0.880	1.274	1.611	1.896
STNN	0.351	0.856	1.252	1.578	1.864
CGA	0.312	0.816	1.217	1.521	1.815
ConvRLSTM	<b>0.276</b>	<b>0.784</b>	<b>1.109</b>	<b>1.489</b>	<b>1.745</b>

of wind speed more precisely compared to recent deep learning benchmarks.



**Fig. 4:** Comparison of the 3-hour ahead prediction outputs of ConvRLSTM and the baselines for the test samples of 10th wind site on July 26th 2013.

To further analyze the effect of time horizon on prediction



accuracy, we expand the time horizon parameter from 1 to 24 hours. Fig. 5 illustrates the obtained results for this experiment. As shown in the plot, the increase in the time horizons yields consistent drops in forecasting methods' performance. For a small-time horizon (i.e., 1 hour), the methods show close performance metrics; however, a significant difference between the models is observed with the expanding in time horizons. As shown in Fig. 5, the simplest baseline, the FFNN, is generally dominated by other methods. The TDNN and NARNN improved the results by making use of recurrent signals in their models. Further improvement in the results is observed with LSTM and GRU models that consider temporal correlations in the time series. Finally, one can observe the superiority of spatio-temporal models. Among these techniques, the best results are obtained by the proposed framework due to high generalization power and effective robustness to uncertainties.

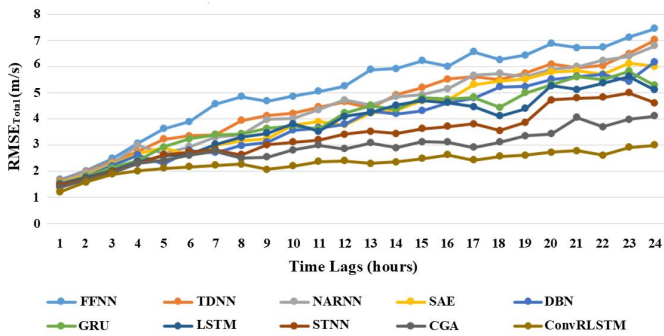


Fig. 5: RMSE<sub>Total</sub> comparison of the forecasting methods over different time horizons.

## V. CONCLUSION

This paper presents a novel deep structure to extract Spatio-temporal correlations for wind speed forecasting task. The integration of Rough set theory with convolutional-LSTM layer leads to a more robust forecasting approach. The model applies the interval convolution and LSTM operators on the historical data intensity map of  $20 \times 20$  wind speed turbines on different time lags; then, the captured spatio-temporal correlations are fed to several fully connected layers for forecasting task. The experimental results prove the superiority of the proposed structure over the state-of-the-art wind speed methodologies in 1-hr up to 24-hr ahead tasks.

## REFERENCES

- [1] M. Khodayar, J. Wang, and M. Manthouri, "Interval deep generative neural network for wind speed forecasting," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3974–3989, 2018.
- [2] T. E. Kibona, "Application of wrf mesoscale model for prediction of wind energy resources in tanzania," *Scientific African*, vol. 7, p. e00302, 2020.
- [3] Y. Kikuchi, M. Fukushima, and T. Ishihara, "Assessment of a coastal offshore wind climate by means of mesoscale model simulations considering high-resolution land use and sea surface temperature data sets," *Atmosphere*, vol. 11, no. 4, p. 379, 2020.
- [4] J. Hu, J. Wang, and L. Xiao, "A hybrid approach based on the gaussian process with t-observation model for short-term wind speed forecasts," *Renewable Energy*, vol. 114, pp. 670–685, 2017.
- [5] S. Singh, A. Mohapatra, et al., "Repeated wavelet transform based arima model for very short-term wind speed forecasting," *Renewable energy*, vol. 136, pp. 758–768, 2019.
- [6] L.-Y. Wei, C.-H. Cheng, and H.-H. Wu, "A hybrid anfis based on n-period moving average model to forecast taiex stock," *Applied Soft Computing*, vol. 19, pp. 86–92, 2014.
- [7] M. Khodayar, J. Wang, and Z. Wang, "Energy disaggregation via deep temporal dictionary learning," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 5, pp. 1696–1709, 2019.
- [8] M. Khodayar, G. Liu, J. Wang, and M. E. Khodayar, "Deep learning in power systems research: A review," *CSEE Journal of Power and Energy Systems*, 2020.
- [9] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management*, vol. 198, p. 111799, 2019.
- [10] H. Liu, R. Yang, T. Wang, and L. Zhang, "A hybrid neural network model for short-term wind speed forecasting based on decomposition, multi-learner ensemble, and adaptive multiple error corrections," *Renewable Energy*, vol. 165, pp. 573–594, 2021.
- [11] J. Duan, P. Wang, W. Ma, X. Tian, S. Fang, Y. Cheng, Y. Chang, and H. Liu, "Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and coreentropy long short-term memory neural network," *Energy*, vol. 214, p. 118980, 2021.
- [12] M. Liu, Z. Cao, J. Zhang, L. Wang, C. Huang, and X. Luo, "Short-term wind speed forecasting based on the jaya-svm model," *International Journal of Electrical Power & Energy Systems*, vol. 121, p. 106056, 2020.
- [13] L. Xie, Y. Gu, X. Zhu, and M. G. Genton, "Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 511–520, 2013.
- [14] M. Khodayar and J. Wang, "Spatio-temporal graph deep neural network for short-term wind speed forecasting," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 670–681, 2018.
- [15] P. Lu, L. Ye, W. Zhong, Y. Qu, B. Zhai, Y. Tang, and Y. Zhao, "A novel spatio-temporal wind power forecasting framework based on multi-output support vector machine and optimization strategy," *Journal of Cleaner Production*, vol. 254, p. 119993, 2020.
- [16] Z. Pawlak, *Rough sets: Theoretical aspects of reasoning about data*, vol. 9. Springer Science & Business Media, 2012.
- [17] H. Liao, S. Ding, M. Wang, and G. Ma, "An overview on rough neural networks," *Neural Computing and Applications*, vol. 27, no. 7, pp. 1805–1816, 2016.
- [18] M. Khodayar, S. Mohammadi, M. E. Khodayar, J. Wang, and G. Liu, "Convolutional graph autoencoder: A generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 571–583, 2019.
- [19] M. Khodayar, O. Kaynak, and M. E. Khodayar, "Rough deep neural architecture for short-term wind speed forecasting," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 2770–2779, 2017.
- [20] M. Cui, M. Khodayar, C. Chen, X. Wang, Y. Zhang, and M. E. Khodayar, "Deep learning-based time-varying parameter identification for system-wide load modeling," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6102–6114, 2019.
- [21] C. Draxl, A. Clifton, B.-M. Hodge, and J. McCaa, "The wind integration national dataset (wind) toolkit," *Applied Energy*, vol. 151, pp. 355–366, 2015.
- [22] M. Saffari, M. Khodayar, M. S. Ebrahimi Saadabadi, A. F. Sequeira, and J. S. Cardoso, "Maximum relevance minimum redundancy dropout with informative kernel determinantal point process," *Sensors*, vol. 21, no. 5, 2021.
- [23] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pp. 265–283, 2016.
- [24] Y. Liu, H. Qin, Z. Zhang, S. Pei, Z. Jiang, Z. Feng, and J. Zhou, "Probabilistic spatiotemporal wind speed forecasting based on a variational bayesian deep learning model," *Applied Energy*, vol. 260, p. 114259, 2020.
- [25] C.-Y. Zhang, C. P. Chen, M. Gan, and L. Chen, "Predictive deep boltzmann machine for multiperiod wind speed forecasting," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1416–1425, 2015.