

Deep Learning-Based Models for Predicting Poorly Damped Low-Frequency Modes of Oscillations

Abdullahi Oboh Muhammed, Younes J. Isbeih, *Member, IEEE*,

Mohamed Shawky El Moursi, *Senior Member, IEEE*, and Khalifa Hassan Al Hosani, *Senior Member, IEEE*

Abstract— This work proposes a real-time deep learning-based model for predicting the small-signal stability of an electrical network. The trained models equip power system operators with an accurate and fast monitoring tool which can be used during online operation. To achieve this objective, three different model architectures are employed in this research; stacked long short-term memory (LSTM), convolutional neural network (CNN)-LSTM and Convolutional LSTM (Conv-LSTM). These models are trained using datasets which contain the oscillatory parameters (frequency and damping ratio) of both local and inter-area modes of oscillations. In addition, the voltage phasors at different buses are taken as the model input where the output comprises the real-time oscillatory patterns of the modes. Furthermore, the overall performance of proposed models is shown for the New-England 10-machine, 39-bus, IEEE 16-machine, 68-bus, 5-area, and IEEE 50-machine, 145-bus benchmark test cases. The main findings show that training CNN-LSTM and Conv-LSTM models provide better performance compared with the stacked-LSTM model. The former models have less number of parameters and thus shorter training time. In addition, CNN-LSTM and Conv-LSTM models are less prone to overfitting problems in the network and have a better ability in capturing spatial and temporal features inherent in input data.

Index Terms—Convolutional-LSTM (Con-LSTM), Convolutional neural network LSTM (CNN-LSTM), Power system, Small-signal stability predictions, Stacked-LSTM, time domain simulations.

I. INTRODUCTION

A. Motivation

THE development of deep learning techniques in the last decades has enabled power system planners, operators, technical experts and researchers to look for new paradigms in power network monitoring, protection and control. This coincides with the latest advancements in inverter-based interface renewable energy technologies and their growing penetration levels in power generation. Therefore, the dynamic characteristics of entire power system have been witnessing a good deal of change as a result of the stochastic nature of renewable power generation, replacement of synchronous generators and increased variability of load profiles [1]. Consequently, power systems are constantly being pushed towards their stability limits and thus accurate assessment and monitoring algorithms are needed to provide fast corrective actions to avoid the occurrence of cascaded events and subsequent network collapse. This work presents novel models for forecasting the small-signal stability of an electrical network using deep learning techniques. They can capture the inherent dynamic behavior of an electrical

network and thus provide accurate predictions about the stability status using collected data from a wide area measurement system (WAMS).

B. Literature Review

The insufficient damping torque has been practically found to largely contribute to the small-signal stability problem and presence of poorly damped low-frequency modes of oscillations (LFMOs). The amplitude of these modes can grow as time evolves when small disturbances are experienced and hence power transfer capacity of tie lines may be jeopardized. This phenomenon is usually studied through a modal analysis where the differential and algebraic equations (DAE) model of a power grid is linearized about an equilibrium point [2]. The real components of the eigenvalues which correspond to conjugate complex pairs determine the amount of damping ratio of low frequency modes. This analysis is usually carried out during offline studies as it relies on developing the DAE model of all system components and their control algorithms. Furthermore, different strategies have been followed to mitigate the risks these oscillations may impose on system's security and reliability. For example, novel optimal power flow formulations are proposed to find the power system operating point which also enhances its small-signal stability [3], [4]. In addition, proper design of power system stabilizers (PSS) [5] and power oscillator damper (POD) [6] has successfully contributed to improve the overall damping of LFMOs. Nevertheless, these studies are performed offline without considering the real-time operation which may be prone to the possibility of cascaded failures and blackouts [7].

Real-time assessment and monitoring of the overall damping of low-frequency modes have been the main focus of several studies in the literature. Initially, prony analysis is applied to single-channel measurements like generator active powers to estimate the oscillatory parameters (amplitude, frequency and damping ratio) of low-frequency modes [8]. However, this method is found very sensitive to the noise contained in the measured signals. Other estimation techniques are applied such as the Tufts-Kumaresan [9], Empirical Mode Decomposition [10], Kalman Filter-based estimation [11], wavelet transform (WT) [12], dynamic mode decomposition (DMD) [8], and extended subspace identification (ESI) [13]. Although accurate estimates are provided by the employment of these methods, they still suffer several drawbacks like an inaccurate estimation of the mode amplitudes, incorrect identification of near frequency oscillations and being computationally expensive, especially in the presence of multiple oscillations.

The recent developments in artificial intelligence and machine learning techniques have given researchers the opportunity to benefit from the ability to capture inherent patterns and features

This work was supported by Khalifa University under Award No. [CIRA-2018-37]. Abdullahi Oboh Muhammed, Younes J. Isbeih, Mohamed Shawky El Moursi and Khalifa Hassan Al Hosani are with the Advanced Power, and Energy Center, EECS Department, Khalifa University, Abu Dhabi 127788, UAE (emails: 100059847@ku.ac.ae, younes.isbeih@ku.ac.ae, mohamed.elmoursi@ku.ac.ae, khalifa.halhosani@ku.ac.ae). Mohamed Shawky El Moursi is also with the Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt.

contained in the data to enhance a power system's small-signal stability [14], [15]. For example, a time series model is trained using recurrent neural networks (RNN) for predicting the small-signal stability status [15]. The oscillatory modes are estimated in real-time using ELPROS which relies on propriety modal identification algorithm and is applied to collected active powers from PMUs. However, the reported results show that overfitting occurs at epoch 24 when the training is stopped. A support vector machine (SVM) with a kernel function is employed to explicitly derive small-signal stability constraints which are incorporated into the reported data-driven small-signal stability constrained optimal power flow (SSSC-OPF). Furthermore, the dataset generation strategy for predicting small-signal stability constraints comprises generator voltage magnitudes and minimum damping ratio as model inputs and output, respectively [16]. However, this approach can be employed in a real-time manner as the reported total CPU time is found to be 51.03 sec for the IEEE 118-bus system. As a result, the required computational time is expected to be longer for large-scale interconnected electrical networks. In [17], the authors have investigated several neural networks (NN) architectures such as a multi-layer perceptron, a fully-convolutional neural network, an inception network, a time convolutional NN, and a multi-channel deep convolutional NN for building a time series-based classifier for classifying small-signal stability. After linearizing the grid model around an equilibrium point, the damping ratio of dominant eigenvalue is computed to determine the corresponding label (stable/unstable) for a specific contingency. However, the proposed approach does not consider changes in the dominant modes across time. In addition, the sample learning method is employed to train a mapping model based on data obtained about steady-state operation [18]. However, multi-layer perceptron (MLP) network may be limited when estimating evolving low-frequency oscillations (time-series task) due to its lack of mechanism for learning the temporal dependencies between observations. In addition, it exhibits inability to learn time series tasks time-step by time-step. MLP network may becloud the interpretability of the factors that are driving the time series. An assessment and correction control model for small signal stability is proposed using the extreme gradient boosting (XGBoost) algorithm. Nevertheless, the maximum error of test set is found to be 19.14%, which may be harmful in some cases where wrong estimates are obtained. The t-distributed stochastic neighbor embedding approach is used to design a convolutional neural network (CNN) model that is deployed to the electrical distance of power grids for different conditions [14]. However, the suggested method ignores how oscillatory mode frequency and damping ratios change over time. To overcome the aforementioned challenges, the authors in [19] have proposed a unified online deep learning model which comprises of a CNN and an LSTM networks for the forecast of transient and small-signal stability. The LSTM network provides online predictions of damping ratios and frequency of local and inter-area oscillation modes as dynamic trajectories vary over time. Although good predictions are obtained for the actual oscillatory patterns, some detailed shortfalls are observed due to inability to capture abrupt low-frequency oscillatory patterns of power system, which result in overshooting of the values of the reconstructed damping rate with high prediction errors.

C. Models Features and Contributions

The conducted literature review reveals that a room for improvement does exist in the area of developing accurate online prediction models for assessing and enhancing small-signal stability of the power grid. In this context, the main contributions of this paper are summarized as follows:

- 1) Stacked-LSTM, CNN-LSTM and Conv-LSTM networks are trained using real-time data to provide accurate and online tracking of the frequency and damping ratio of local and inter-area modes of oscillations. The models exhibit very low generalization loss regardless of the system strength. Also, accurate estimations of low-frequency oscillations and reconstruction of damping rates are obtained for the system with renewable generations.
- 2) The proposed real-time low-frequency oscillation modes estimation networks mainly focus on achieving good performance on and explanation for, especially abnormal changes and abruptions of the modes of oscillation in the original time series data. In addition, the presented model architectures provide forecasts of small-signal stability that correspond to the real-time patterns of the original dataset.
- 3) The main findings show that training CNN-LSTM and Conv-LSTM models provide better performance compared with the stacked-LSTM model but at the expense of slightly larger hyper-parameter tuning. Conv-LSTM models achieves similar performance to the other networks with less number of layers while CNN-LSTM model with same depth as stacked-LSTM network possess less trainable parameters, thus shorter training time. In addition, CNN-LSTM and Conv-LSTM models are less prone to overfitting problems in the network and have the ability to capture spatial and temporal features inherent in input data.

D. Paper Organization

The rest of the paper is organized as follows: Section II discusses small-signal stability problems (SSSP). Section III presents the proposed prediction models for SSSP. Section IV puts forward the implementations of these models. Section V verifies the proposed models with test cases. Section VI concludes this paper.

II. THE SMALL-SIGNAL STABILITY PROBLEM

Small-signal stability of a power system is associated with its ability to remain in synchronism when a small-disturbance such as load change occurs. The disturbance size is small in magnitude such that the little deviation from equilibrium is governed by the linearized model of differential and algebraic equations (DAE) model. In addition, individual synchronous machines along with their control circuits mainly contribute to system ability to return to a state of equilibrium after being subjected to disturbance. They also determine the existing synchronizing and damping torques which significantly shape the system dynamic response and assist in bringing the deviation in equilibrium point to zero. As a result, the change in electrical torque should instantaneously exhibit sufficient damping and synchronizing torques to ensure small-signal stability. Moreover, the lack of sufficient torque components is experienced by the rotor angle of a synchronous generator in the form of a periodic drift and oscillatory response, respectively. In practical power systems, the small-signal stability problem is largely associated with the presence of poorly damped LFMOs.

The real-time operation of power systems reveals that the small-signal stability problem is largely contributed by insufficient dampening of oscillations which lie in a frequency range of 0.3 – 2 Hz. The higher side of this range (1-2 Hz) corresponds to the frequency of local modes of oscillations

which are triggered when synchronous machines swing against each other in one location. The other type of oscillations includes multiple synchronous generators that swing against each other at different locations within a frequency of 0.3-1 Hz. The presence of these modes of oscillations can be very harmful for system operation as it limits power transfer especially through weak tie-lines where large amount of power is to be transmitted. The dynamic response of a power system is described by a set of nonlinear differential and algebraic equations which can be expressed in a compact form as:

$$\dot{x} = \mathcal{F}(x, y) \quad (1)$$

$$0 = \mathcal{H}(x, y) \quad (2)$$

where x and y denote the state and algebraic variables. Furthermore, \mathcal{F} and \mathcal{H} represent the vectors of differential and algebraic equations. Eqns. (1-2) describe the time evolution of system states and algebraic variables when a disturbance is experienced. In addition, modal analysis can be applied to the DAE model after linearizing it about the equilibrium point (x_0, y_0) . Therefore, the small-signal model expressed in (3)–(4) is utilized to investigate the small-signal of an electrical network around an equilibrium point when subjected to a minor disturbance. This goal is achieved by calculating the roots of the characteristic equation (eigenvalues) as follows [20]:

$$\Delta \dot{x} = A \Delta x + B \Delta y \quad (3)$$

$$0 = C \Delta x + D \Delta y \quad (4)$$

$$A = \frac{\partial \mathcal{F}}{\partial x}, B = \frac{\partial \mathcal{F}}{\partial y}, C = \frac{\partial \mathcal{H}}{\partial x}, D = \frac{\partial \mathcal{H}}{\partial y} \quad (5)$$

$$\det(A_{sys} - \lambda I) = 0 \quad (6)$$

where $A_{sys} = A - B(D^{-1})C$ and $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_n]$. The estimated eigenvalues can be either real or complex ones which result in non-oscillatory or oscillatory responses. Further, complex eigenvalues exist in conjugate pairs, each of which indicates an oscillatory mode [20]. The i -th complex pair of eigenvalues can be expressed as $\lambda_i = \alpha_i \pm j\beta_i$ where α_i and β_i correspond to the real and imaginary part. As a result, the damping ratio ζ_i and frequency of oscillation f_i in Hz are expressed as

$$\zeta_i = \frac{-\alpha_i}{\sqrt{\alpha_i^2 + \beta_i^2}}, f_i = \frac{\beta_i}{2\pi} \quad \forall i \in \mathcal{M} \quad (7)$$

where \mathcal{M} denotes the set of low-frequency modes of oscillations that include both local and inter-area ones and have a frequency $f_i \leq 2\text{Hz}$ and a damping ratio $\zeta_i \leq 0.1$.

III. HYBRID CNN AND LSTM DEEP LEARNING MODELS FOR PREDICTING SMALL-SIGNAL STABILITY OF A POWER SYSTEM

In the previous section, the small-signal stability problem is described where modal analysis is employed to compute the parameters of low-frequency modes of oscillations. It can be seen from (3)–(4) that they are computed using the state matrices about an equilibrium point. Therefore, the traditional method for studying small-signal stability lacks the ability to provide useful information concerning changes of oscillatory parameters (damping ratio and frequency) during real-time operation of a power system. In other words, modal analysis fails to provide online tracking for the evolution of these modes as time elapses and changes in operating conditions are experienced in real time.

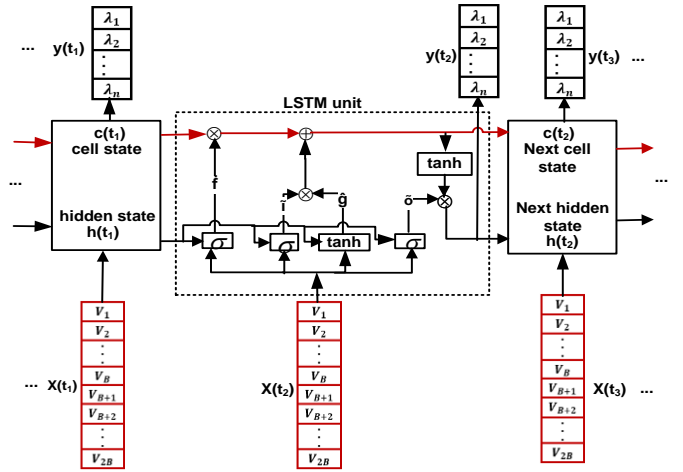


Fig. 1. LSTM Model for predicting small-signal stability.

In addition, obtaining a state matrix for large-scale power systems can be complicated and computationally expensive as it requires detailed mathematical models for all components and their control circuits [21]. On the other hand, power system operators can benefit from the availability of various measurements that are provided by WAMS across the entire network [22]. The collected data carries useful and non-linearized information concerning the real-time operation of a system and hence, dynamic signatures which can be deployed to construct an accurate mapping using machine learning techniques. The input to this mapping represents the collected measurements such as voltage magnitudes, angles and active power. The mapping will result in the real-time estimates of the oscillatory parameters for all modes of oscillations. Therefore, deep learning techniques provide the ability to capture inherent dynamic characteristics that exist in the dataset. As a result, the employment of machine learning and deep learning techniques assists in constructing fast and accurate models for predicting the oscillatory parameters for low-frequency modes of oscillations. In this section, theoretical background is presented for constructing accurate small-signal prediction models using deep learning techniques.

A. Convolutional Neural Network (CNN) Architecture

CNN is a type of feedforward artificial neural network (ANN) that is renowned for image visualization and classification. CNN can learn spatial-dependent patterns from data efficiently. Like an ANN, CNN consists of hidden layers that include neurons with learnable parameters. The neurons when fed with inputs, execute dot product operation to generate feature map and then non-linearity is applied to the feature map by an activation function for fast convergence [23]. However, CNN is different from multilayer (MLP) network because it uses convolutional layers, pooling, and nonlinearities such as ReLU. The convolutional layer comprises of a kernel and is utilized to locally slide through the width and height of an input image or row and column of a vector based input data, and compute dot product of the input's region and the weight learning parameters (8).

$$\begin{aligned} \hat{h}_i^{(\theta)} &= (w \otimes x)_i = \sum_{j=1}^n x_{i+j} w_j + b_j \\ i &= \{1, 2, \dots, m\} \end{aligned} \quad (8)$$

where w is filter weight; x is input vector and \otimes is convolution operation. Note that convolution acts over all channels concurrently for multi-channeled signals, merging the dynamics of several signals to generate one new signal with valuable spatial features. CNN has sparse connections at each layer and shares weight across its layers and these decrease trainable parameters and thus, help to speed computation [24]. It should be noted that despite the sparse connectivity of CNN at each layer, its receptive field expands with depth [25], which means signals that appear later may be identified earlier. The pooling layers down-sample the rectified feature map into a salient feature representation, allowing for the modeling of a small local invariance [25]. Following the convolutional layers, one or more fully connected (FC) layers are added to identify and interpret the extracted spatial pattern from previous layers. The entire CNN expresses the mapping between raw image pixels or raw time-series in 1-D vector values and their target values. Conventionally, the softmax function or linear function is used at the last layer of the network for classification or regression tasks. Nevertheless, in this study, CNN last layer's is coupled to LSTM layer(s) and then follow by dense layer (s).

B. LSTM Networks for SSP

A Long short term memory (LSTM) is increasingly deployed in research areas about sequential data such as video, text, and audio sequences and has offered a new perspective for analyzing power system time-series event-based operations. The gate functions in LSTM cell structure gives it the capability for handling and recalling long-term temporal correlations in the past sequences during training. The gating functions enable LSTMs to achieve more exciting results compared to recurrent neural networks (RNNs), which have gradient exploding and/ or gradient vanishing constraint during weight updates for more prolonged sequence. The main reason for this behavior is the presence of sigma or tanh cells in RNNs [26]. An LSTM network is made of a stack of cells. The long-term component of an LSTM is termed the memory cell (c). The network learns by encoding sequence data in stepwise hidden state and memory cell that are copied from time step to another. It is to be noted that integration of memory modules in the LSTM cell enables it to overcome the problem of gradient vanishing and explosion which is inherent in the original RNN. New sequence information is passed into the memory by summing updates; in consequent, the gradient expressions do not accumulate multiplicatively over time like RNNs. LSTM networks comprises three gating units: input gate (\tilde{i}), forget gate (\tilde{f}), and output gate (\tilde{o}), as well as a cell state ($c(t)$) and a hidden state ($h^{(\theta)}$). The gating units are interacting FC-ANNs that utilize elementwise multiplication of an appropriate temporal information to control the flow of sequential data points in the LSTM. The forget gate (\tilde{f}) control the memory cell element to carry forward to next time step or reset to zero. The input gate (\tilde{i}) updates additively the element of the memory cell with new information from the input vector at a specific time step. The output gate (\tilde{o}) regulates the element in the memory cell to transfer to the short-term memory ($h^{(\theta)}$), which plays a similar role to the hidden state in basic RNNs. The mathematical model which describes LSTM architecture can be written as:

$$\tilde{f} = \sigma(W^{(\tilde{f})}[h^{(\theta)}(t_1), x(t_2)] + b^{(\tilde{f})}) \quad (9)$$

$$\tilde{i} = \sigma(W^{(\tilde{i})}[h^{(\theta)}(t_1), x(t_2)] + b^{(\tilde{i})}) \quad (10)$$

$$\tilde{g} = \tanh(W^{(\tilde{g})}[h^{(\theta)}(t_1), x(t_2)] + b^{(\tilde{g})}) \quad (11)$$

$$\tilde{o} = \sigma(W^{(\tilde{o})}[h^{(\theta)}(t_1), x(t_2)] + b^{(\tilde{o})}) \quad (12)$$

$$c(t) = \tilde{f} * c(t-1) + \tilde{i} * \tilde{g} \quad (13)$$

$$h^{(\theta)}(t) = \tilde{o} * \tanh(c(t)) \quad (14)$$

where $*$ is the element-by-element product; $W^{(*)}$ is the weight parameter which represents the symbol for one of the four ANNs of the LSTM units (\tilde{f} , \tilde{i} , \tilde{o} and \tilde{g} , the candidate cell). Fig. 1 shows the overall architecture an LSTM unit. LSTM networks are well-suited to translating sequential data in real-time, which makes them perfect for SSP. LSTM networks may also learn to translate patterns in observed time-domain signals into the oscillatory modes of the system at every time instance; technically, $\hat{\lambda}(t) = H_{\phi}(x(t_n))$ by utilizing data from various operational situations. Since previous forecasts have an impact on future predictions, the LSTM network must minimize the mean squared error (MSE) between $\hat{\lambda}(t_n)_i$, and actual values $\lambda(t_n)_i$ post fault clearing time to the end of simulation, T , and across all i contingencies:

$$Loss_{MSE} = \frac{1}{|\mathcal{M}|} \sum_{i=1}^{|\mathcal{M}|} \sum_{t=1}^T (\lambda(t)_i - \hat{\lambda}(t)_i)^2 \quad (15)$$

The LSTM network's weight parameters, ϕ , include the weights of all fully connected (FC) networks as well as the LSTM units. Furthermore, LSTM networks may be made more sophisticated by stacking LSTM units such that one LSTM unit's hidden state combines with the input of another LSTM unit. FC layers are frequently added to the output of LSTM units for interpretation purposes.

C. CNN-LSTM Networks for SSP

A CNN-LSTM approach is proposed using the architecture depicted in Fig. 2 to extract the rich information from the spatial-temporal dynamic measurements across various time periods. CNN and LSTM layers make up the deep learning model. Convolution layers are used in the proposed architecture in place of CNN's fully connected layer to learn generic features since they may encode spatial information, which is essential for the subsequent LSTM. Conventionally, the design should demonstrate how the temporal tensors are chosen as inputs, and how features in these inputs are initially extracted by Convolution layers, followed by pooling layers to lower the spatial size of the representation learned by CNNs. Then, the output of the pooling layer is transformed into a 1-D array by a flatten layer and used as the input for the LSTM layer. Finally, LSTMs extract temporal features, which are interpreted by fully connected (FC) layers for regression analysis [27]. However, this study uses 1-D CNN, casual padding, but no pooling and no flatten layer—since the data at each time step are 1-dimensional vectors.

D. Conv-LSTM Networks for SSP

In this study, a convolutional LSTM (Conv-LSTM) network is presented for small signal predictions. The small signal is formulated as a spatiotemporal sequence forecasting problem that may be resolved using the generic framework for sequence-to-sequence learning. However, in this study, the sequence-to-sequence are of different lengths. Conv-LSTM which incorporates convolutional structures in both the input- to-state and state-to-state transitions is an extension of the FC-LSTM

concept that is effectively used to simulate the spatiotemporal interactions. An end-to-end trainable model is created for small-signal stability (SS) prediction by piling up Conv-LSTM layers and creating an encoding-forecasting structure. The use of full connections in input-to-state and state-to-state transitions when no spatial information is recorded is FC-LSTM's main limitation when processing spatiotemporal data. To solve this problem, all the inputs $\mathcal{X}(1), \dots, \mathcal{X}(t)$, cell output $\mathcal{C}(1), \dots, \mathcal{C}(t)$, hidden states $\mathcal{H}(1), \dots, \mathcal{H}(t)$, and gates \tilde{i} , \tilde{f} , and \tilde{o} of the Conv-LSTM are designed as 3-D tensors, where the last two dimensions represent spatial dimensions (rows and columns). To better understand the input and states, they can be foreseen as vectors positioned on a grid in space. Conv-LSTM uses the inputs and previous states to determine the future state of a specific cell in the grid. Using a convolution operator in the state-to-state and input-to-state transitions makes this simple to achieve (see Fig. 3) [28]. The main equations of Conv-LSTM are shown in (16)–(21),

$$\tilde{f} = \sigma(W^{(xf)} \odot \mathcal{X}(t_2) + W^{(hf)} \odot \mathcal{H}(t_1) + W^{(cf)} \otimes \mathcal{C}(t_1) + b^f) \quad (16)$$

$$\tilde{i} = \sigma(W^{(xi)} \odot \mathcal{X}(t_2) + W^{(hi)} \odot \mathcal{H}(t_1) + W^{(ci)} \otimes \mathcal{C}(t_1) + b^i) \quad (17)$$

$$\tilde{g} = \tanh(W^{(xg)} \odot \mathcal{X}(t_2) + W^{(hg)} \odot \mathcal{H}(t_1) + b^g) \quad (18)$$

$$\mathcal{C}(t_2) = \tilde{f} \otimes \mathcal{C}(t_1) + \tilde{i} \otimes \tilde{g} \quad (19)$$

$$\tilde{o} = \sigma(W^{(xo)} \odot \mathcal{X}(t_2) + W^{(ho)} \odot \mathcal{H}(t_1) + W^{(co)} \otimes \mathcal{C}(t_2) + b^o) \quad (20)$$

$$\tilde{\mathcal{H}} = \tilde{o} \otimes \tanh(\mathcal{C}(t_2)) \quad (21)$$

where \odot denotes the convolution operator and \otimes , as before, denotes the element-by-element multiplication. A Conv-LSTM with a bigger transitional kernel should be able to detect faster kernel if the states are depicted as the hidden representations of moving objects.

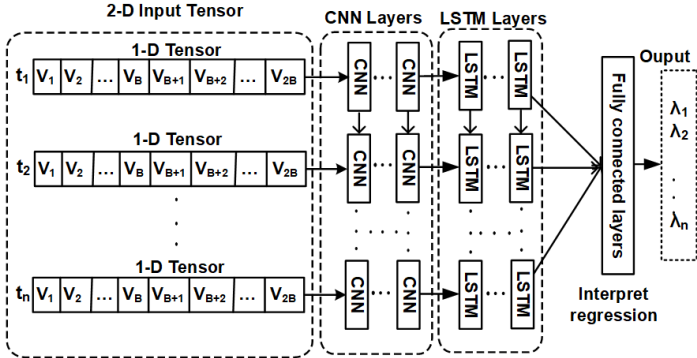


Fig. 2. CNN-LSTM Topology.

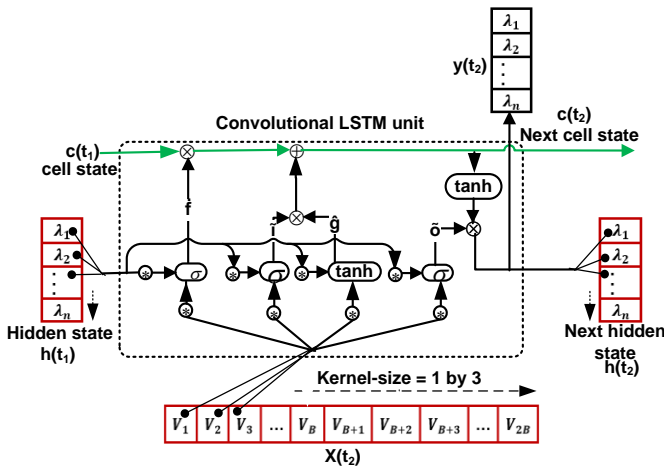


Fig. 3. Conv-LSTM architecture.

Additionally, in accordance with the viewpoint presented in [29], the inputs, cell outputs, and hidden states of the conventional FC-LSTM described by (9)–(14) can alternatively be thought of as 3-D tensors with the last two dimensions being equal to 1. FC-LSTM is basically a special case of Conv-LSTM in this regard, all features standing on a single cell. are all combined. The states must have the same number of rows and columns as the inputs in order to be valid. Furthermore, padding is required before using the convolution process, [28]. It is observed that the Conv-LSTM model used in this study significantly outperformed the FC-LSTM when trained and tested on the same training and testing datasets.

IV. IMPLEMENTATION OF LSTM, CNN-LSTM AND CONV-LSTM NETWORKS FOR SMALL-SIGNAL STABILITY PREDICTION

A. LSTM SSSP Model: Data Preparation and Implementation

Extensive time-domain simulations are conducted to generate the datasets for training and testing the small-signal stability prediction model. To achieve this objective, a library of possible fault types, locations, clearing times and loading levels is designed to cover a wide range of operating conditions and contingencies. For each contingency, the test system is simulated for 10 seconds until a steady state is attained. The contingency is thereafter applied for a specific time after which the fault is cleared where the system is allowed to enter the post-fault period. In this research, the dataset is constructed through stacking up the post-fault dynamic response of various system states and variables especially, the latest 850-time steps (equaling 8.5 s) for each contingency. In addition, bus voltage magnitudes and angles are designated as inputs to the prediction model whereas the oscillatory modes λ are used as target variables. A 1-dimensional row vector of input variables and output responses is formulated at each time step. This process is repeated for the subsequent time steps and the 1-D vectors are horizontally concatenated to form, at the end, a 2-D vector of supervised data. Equations (22)–(25) are mathematical descriptions of this process.

$$V(t) = \{|V_n(t)|, \angle V_n(t) \mid n = 1, \dots, B\} \quad (22)$$

$$\tilde{\lambda}(t) = \{\lambda_l(t) \mid l = 1, \dots, 2L\} \quad (23)$$

$$X = [V(t)^T \mid t = t_c, \dots, T] \quad (24)$$

$$\Lambda = [\tilde{\lambda}(t)^T \mid t = t_c, \dots, T] \quad (25)$$

$|V_n(t)|$ and $\angle V_n(t)$ are the voltage magnitude and angle for system node n at time t , respectively. Since the oscillatory mode λ is complex conjugate pair which does not occupy Euclidean space, (15) would not fit the output $\hat{\lambda}(t)$ well. We therefore opted for disjointed split of λ into real and imaginary parts:

$$\tilde{\lambda} = \{(\lambda_1), R(\lambda_2), \dots, R(\lambda_L), J(\lambda_1), J(\lambda_2), \dots, J(\lambda_L)\} \quad (26)$$

This approach is found effective since there is no complex analogue of the MSE loss function currently available in the literature. Similarly, this method is followed to evaluate the performance of deep learning models used in this work.

To model the complicated highly nonlinear dynamics of a power system, a stacked LSTM of two layers and L units is designed as shown TABLE I. The first LSTM unit collects a 3-

dimensional input signal from raw data and feeds it into the second LSTM unit together with its hidden state and so on. In addition, the training is carried out for different LSTM networks with 64, and 256 number of hidden units as in [19]. However, the model check-point technique is deployed with early stopping for a patience of 100 epochs to save the best model during long training processes. The early stopping criteria helps to stop the training once the loss function value does not improve after a patient of 100 epochs. Further, no dropout nor weight decay are employed since the adopted training technique results in an appropriately fit model. To achieve stability in LSTM training, Adam optimizer is used [30], with default settings ($\beta_1 = 0.9, \beta_2 = 0.999$).

TABLE I: THE PROPOSED LSTM ARCHITECTURES

Layer	Model parameter size		
	IEEE39 Bus	IEEE 68 Bus	IEEE145 Bus
Input step	$2B \times 1$	$2B \times 1$	$2B \times 1$
LSTM 1	64	256	256
LSTM 2	64	256	256
Dense 1	32	-	128
Dense 2	-	-	64
Output	18	10	12

B. CNN-LSTM SSSP Model: Data Preparation and Implementation

The hidden layers of a CNN network are used in the CNN-LSTM architecture to extract features from input data which are utilized by the LSTM network to provide sequence prediction. Subsequences of the main sequence will be read into the CNN-LSTM model as blocks, and each block's features are extracted before being sent to the LSTM for interpretation. One way to put this model into practice is to divide each window of n -time steps into subsequences that the CNN model can analyze. For instance, four subsequences of three-time steps may be created from each window of 12-time steps. Then, a CNN model may be created that reads sequences of 3-time steps in length and m -features. The same CNN model can be read in each of the four subsequences in the window by wrapping the complete CNN model in a time distributed layer. The extracted features are then flattened and provided to the LSTM model to extract its own features prior to making a final mapping to the set of predicted oscillatory modes (λ 's). This is carried out through hidden dense layer(s) to output dense layer. The basic structure utilized in the CNN-LSTM model is typically two consecutive CNN layers, followed by dropout and a max-pooling layer [31]. This model is similar to the LSTM model which is previously described except that two 1-dimensional CNNs in end-to-end concatenation with two-layered LSTMs are implemented. The final extracted features are thereafter passed to the dense layer to generate the output. However, pooling, dropout, and flatten layers are not used in this model. The architecture of CNN-LSTM is tabulated in TABLE II.

C. Conv-LSTM SSSP Model: Data Preparation and Implementation

This architecture represents an expansion of the CNN-LSTM SSSP model. For this model, convolutions of the CNN networks are carried out as part of the LSTM at each time step. The Convolutional LSTM (Conv-LSTM) is a combination similar to the CNN-LSTM as it is utilized for spatiotemporal data. In contrast to LSTM and CNN-LSTM models, the Conv-LSTM directly utilizes convolutions as part of feeding the inputs into LSTM units. It can be set up to forecast 1D multivariate time series. By default, the Conv-LSTM2D class anticipates that input data in (22) will be in the following shape: [samples, time steps, rows, columns, channels]. The definition of each data at any time step is an image of (rows \times columns)

TABLE II: THE PROPOSED CNN-LSTM ARCHITECTURES

Layer	Model parameter size		
	IEEE39 Bus	IEEE 68 Bus	IEEE145 Bus
Input	$2B \times 1$	$2B \times 1$	$2B \times 1$
CNN 1	32 kernels; kernel (3); stride (1); padding (causal)	64 kernels; kernel (1); stride (1); padding (causal)	64 kernels; kernel (1); stride (1); padding (causal)
CNN 2	64 kernels; kernel size (3); stride (1); padding (causal)	128 kernels; kernel (1); stride (1); padding (causal)	128 kernels; kernel (1); stride (1); padding (causal)
LSTM 1	64	256	256
LSTM 2	64	256	256
Dense 1	32	-	128
Dense 2	-	-	64
Output	18	10	12

data points. In this study, the samples are the total number of observations in the training or the testing datasets; the total number of time steps, rows, and columns are each set to 1 to make the network suitable for online prediction and the channels correspond to the total number of input variables in the training and testing data, in this case, ($2B$). Both the CNN and the LSTM must be configured for the Conv-LSTM2D class as shown in Fig. 3. Included are the number of filters, the size of the two-dimensional kernel—say 1 row and 3 columns of subsequence time steps, and the rectified linear unit (ReLU) activation function, in this case. The output should first be flattened into a single lengthy vector, much like with a CNN-LSTM model, in order to interpret it.

To explain the model's behavior, the Conv-LSTM network is first compared with the FC-LSTM one using the same supervised dataset. The model is run with the same number of layers and units as FC-LSTM and a fixed kernel size of (1×3) as shown in TABLE II. In addition, the same training approach is adopted along with the default Adam optimizer values. Two hidden layers are concatenated of Conv-LSTM with one dense hidden layer to compare the model performance with that of the FC-LSTM algorithm based on MSE, mean arctangent absolute percentage error (MAAPE) evaluation metrics for small-signal stability predictions as utilized in [19]. Moreover, the effectiveness of proposed model is demonstrated for the more difficult low-frequency oscillations and damping ratios forecasting problem. The simulation results indicate that Conv-LSTM is superior to FC-LSTM in handling spatiotemporal correlations. This is probably due to the size of the convolutional kernel is bigger than 1, which is critical for tracing the spatiotemporal patterns and deeper models can result in better results with fewer parameters [28].

D. Evaluating LSTM, Conv-LSTM and CNN-LSTM

The root mean squared error (RMSE) and mean arctangent absolute percentage error (MAAPE) metrics are used to evaluate the performance of LSTM-based, Conv-LSTM-based and CNN-LSTM-based SSSP models. Although the mean absolute percentage error (MAPE) is mostly used in the literature to quantify prediction errors of LSTM network, the MAAPE is found better metric while preserving the same semantic value due to the near-zero values of $\Re(\lambda)$. The damping ratio ξ , which is reconstructed from λ , is subjected to these measures. The average errors for all cases in the test set

TABLE III: THE PROPOSED CONV-LSTM ARCHITECTURES

Layer	Model Model parameter size		
	IEEE39 Bus	IEEE 68 Bus	IEEE145 Bus
Input	$2B \times 1 \times 1 \times 1$	$2B \times 1 \times 1 \times 1$	$2B \times 1 \times 1 \times 1$
Conv-LSTM 1	64 filters, filter size (1 × 3), ReLU activation, padding (same)	256 filters, filter size (1 × 10), ReLU activation, padding (same)	256 filters, filter size (1 × 10), ReLU activation, padding (same)
Conv-LSTM 2	64 filters, filter size (1 × 3), ReLU, padding (same)	256 filters, filter size (1 × 10), ReLU activation, padding (same)	256 filters, filter size (1 × 10), ReLU activation, padding (same)
Flatten			
Dense 1	32 units	128	128
Output	18 unit linear	10	12

are presented. The reconstruction and evaluation of ξ is based on (27)-(29).

$$\xi_j(t) = \frac{-\Re(\lambda_j(t))}{\sqrt{\Re(\lambda_j(t))^2 + \Im(\lambda_j(t))^2}} \quad (27)$$

$$\text{RMSE}(\xi_j) = \sqrt{\frac{1}{U} \sum_t^U (\hat{\xi}_j(t) - \xi_j(t))^2} \quad (28)$$

$$\text{MAAPE}(\xi_j) = \frac{100\%}{U} \sum_t^U \arctan\left(\left|\frac{\hat{\xi}_j(t) - \xi_j(t)}{\xi_j(t)}\right|\right) \quad (29)$$

where U is the time span of the testing window, which is similar to the training window (8.5 seconds) and corresponds to the damping ratio ξ reconstructed from $\hat{\lambda}$. The SSSP Models, if properly trained, should have RMSE and MAAPE values that are close to zero.

V. CASE STUDIES

The New England 39 bus system, five areas 68 bus system [32], and 50 machines, 145 bus system are used in this work to evaluate the performance of the proposed models. The power system analysis toolbox (PSAT) is used to implement time-domain simulations for these systems [33]. The generator's rotor angles, voltage magnitudes, and angles are collected from each bus in addition to the modes of oscillations. The dataset is generated using extensive time domain simulations, where the characteristics of the system are sampled at 100 Hz frequency (i.e., 0.01 second to obtain each data sequence). For each fault scenario, the simulation is performed in Matlab/Simulink using PSAT toolbox for a period of 10 seconds which results in 1,000 samples. These samples are distilled to 850 post-contingency samples by removing 150 samples that account for pre-fault and during-fault samples and as well as some delay time before (and after) fault is applied (and cleared). Figs. 4-5 show the dynamic response of voltage angles and magnitudes at all buses of the IEEE 39-bus system. A three-phase to ground fault is applied at bus 1 at time $t = 1\text{sec}$ and cleared after 100ms. The overall loading level of the system represents 60% of the base case. Furthermore, Figs. 6-7 show the temporal evolution of real and imaginary components of low-frequency modes of oscillations. The voltage magnitudes and angles are used in the training, validation, and test stages. The training, validation and test instances are obtained from the stable contingencies which exhibit maximum rotor angles below a 360° threshold. As a result, the dataset comprises dynamic trajectories for the oscillatory modes across a wide range of contingencies and operating conditions. Moreover, the dataset is split into 80%, 10%, and 10% for training, evaluation, and testing of the prediction model, respectively. The proposed deep learning networks are developed on Keras on top of Tensorflow and trained, validated

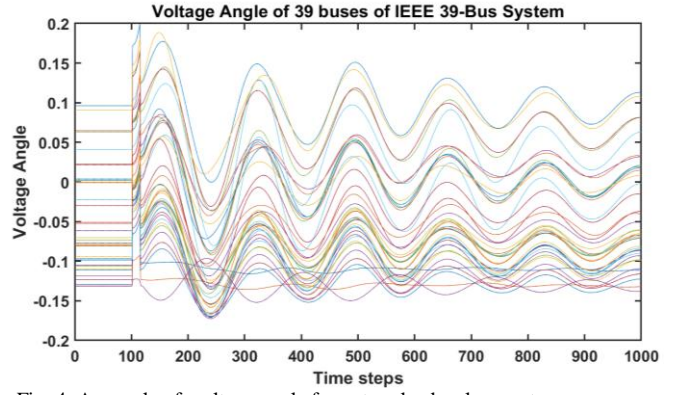


Fig. 4. A sample of voltage angle for networks development.

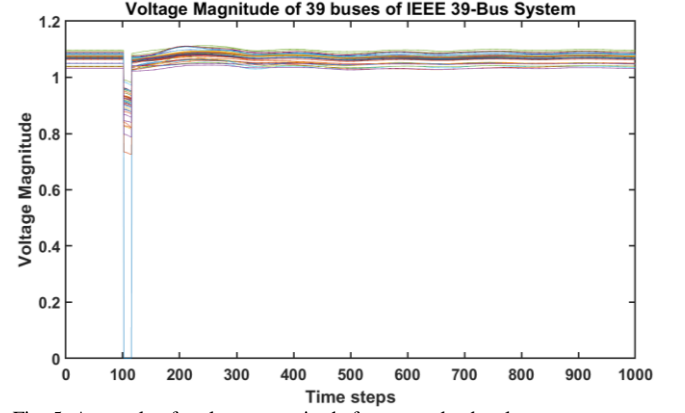


Fig. 5. A sample of voltage magnitude for networks development.

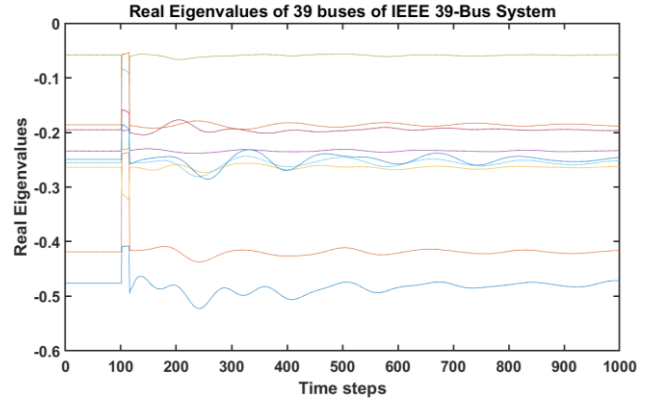


Fig. 6. A sample of real low-oscillation modes.

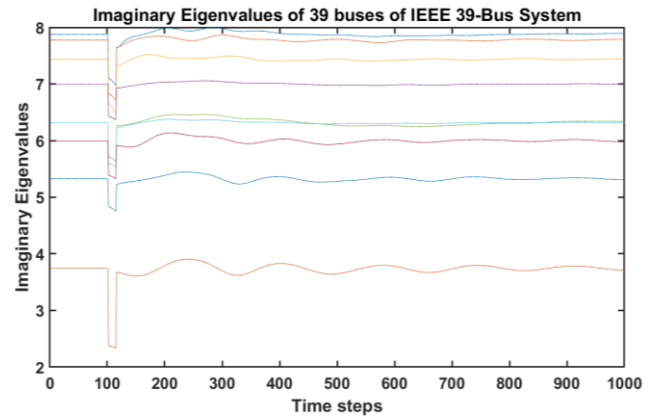


Fig. 7. A sample of imaginary low-oscillation modes.

and tested using Tesla K80 GPU on Google Colab. Using the best hyperparameters, each of the three proposed networks was

trained for average of 6 hours (i.e., 18 hours training time for the whole networks).

A. The New England 39 Bus Test System

The dynamic model of generator 1 is described using 3rd order model. The rest of synchronous generators is described by 4th order model. In addition, all generators have been equipped with type II exciters except generator 10 for which manual stimulation is used [19]. The generation of training, validation and test datasets involves extensive time-domain simulations of the power system dynamics. Three-phase to ground faults are applied at different locations because they are more severe compared to other ones. The system's loading level is adjusted in 5 percent increments from 60 to 100 percent. The base case loading is equivalent to 122.19 percent of the maximum loading condition. In addition, three-phase faults are applied at different buses and transmission lines at various locations (20%, 40%, 60% and 80%) of each line. The line is switched off when the fault is cleared after a duration of 0.1, 0.2, 0.3 and 0.4 seconds, respectively. Also, for the other fault cases, the fault clearance time was varied between 100 and 400 milliseconds [19], [34]. Consequently, 8,316 ($4 \times 9 (39 + 4 \times 48)$) TSA contingencies are produced.

TABLE IV: RECONSTRUCTED ξ RMSE FOR DIFFERENT Models.

	RMSE			
	IEEE 39-Bus			
	SSLSTM-128 [19]	Stacked LSTM	CNN-LSTM	CONV-LSTM
ξ_1	0.00270	0.00065	0.00053	0.00059
ξ_2	0.00340	0.00059	0.00042	0.00057
ξ_3	0.00360	0.00046	0.00045	0.00048
ξ_4	0.00200	0.00077	0.00052	0.00055
ξ_5	0.00510	0.00180	0.00230	0.00220
ξ_6	0.00570	0.00200	0.00240	0.00220
ξ_7	0.00310	0.00150	0.00130	0.00150
ξ_8	0.01700	0.00220	0.00230	0.00170
ξ_9	0.06000	0.00290	0.00240	0.00270
Average error	0.01140	0.00143	0.001402	0.001388

Model Evaluation: Three small-signal stability prediction models are trained for this test system; Two-stacked LSTM, CNN-LSTM and Conv-LSTM, respectively. In addition, each of these models is trained using 406,640 instances, 50,830 validation instances and tested using 50,830 ones. Moreover, the training and testing results are compared with those obtained by training SSLSTM-128 as reported in [19]. TABLE IV shows the computed RMSE value for all models and the observed modes of oscillations. It can be seen that the trained models which are proposed in this work achieve very high accurate reconstructions of the system modes ($\zeta_1, \zeta_2, \dots, \zeta_9$) in real time. In addition, the Conv-LSTM and CNN-LSTM models achieve slightly better accuracy compared to the two-stacked LSTM. Moreover, the proposed LSTM, CNN-LSTM, and Conv-LSTM achieve a prediction performance that is about 87.4% lower than that of SSLSTM-128. The MAAPE metrics shown in TABLE V indicate that the performance of each of the proposed models is about 75% better than the SSLSTM-128. The real-time predictions of the oscillatory modes for this test system are presented for LSTM models. The model predictions for the IEEE bus-39 test case with nine oscillation modes–eigenvalues (output signals): λ_{1-7} (local

modes) and λ_{8-9} (inter-area modes) are depicted in Figs. 8 and 9, respectively. It can be observed that the real and imaginary predictions $R(\lambda_j)$ and $J(\lambda_j)$ are captured and aligned very closely to the original oscillation patterns in the ground truth data. Further, to visualize the numerical accuracy of estimated eigenvalues (because the proposed models are regression models), reconstructed damping rates ξ_j are mathematically evaluated from estimated eigenvalues using (27). The reconstructed damping rate is then graphically matched (Fig. 10) against the damping rate measured from the ground-truth data. The ξ_j are favorably estimated and indicates that the

TABLE V: RECONSTRUCTED ξ MAAPE FOR DIFFERENT Models.

	MAAPE			
	IEEE 39-Bus			
	SSLSTM-128 [19]	Stacked LSTM	CNN-LSTM	CONV-LSTM
ξ_1	4.36	0.90	0.63	0.67
ξ_2	5.01	0.82	0.49	0.89
ξ_3	8.37	0.80	0.60	0.72
ξ_4	4.62	1.25	0.92	0.84
ξ_5	17.5	4.35	7.81	8.19
ξ_6	17.2	2.79	3.25	3.53
ξ_7	4.90	2.82	2.36	2.90
ξ_8	5.81	2.70	3.22	2.31
ξ_9	40.6	5.28	6.03	6.37
Average error	12.041	2.412	2.812	2.936

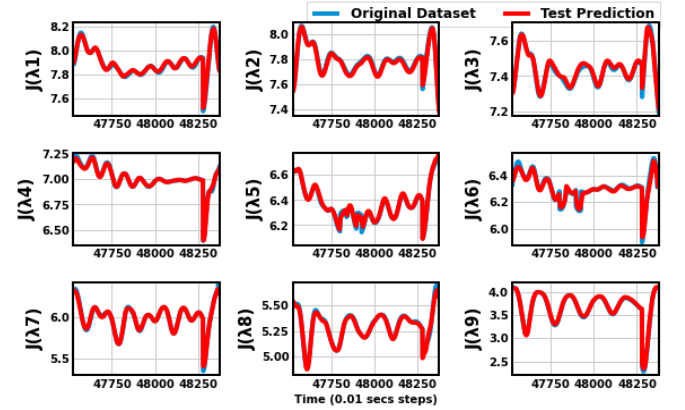


Fig. 8. Actual and proposed LSTM-64 predictions of $J(\lambda_j)$ for an arbitrary test case in the New England 39-Bus System.

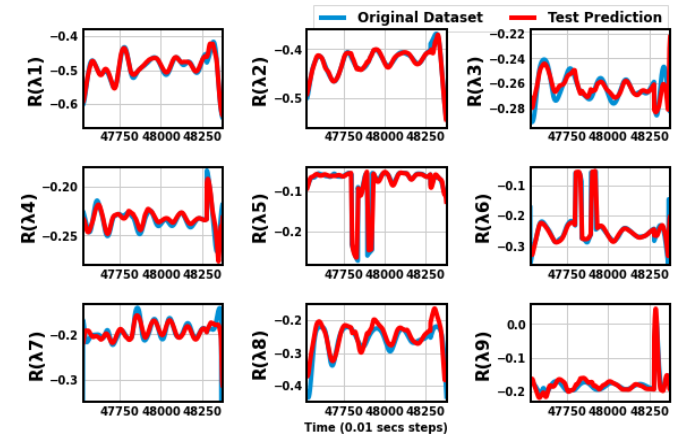


Fig. 9. Actual and LSTM-64 predictions of $R(\lambda_j)$ for an arbitrary test case in the New England 39-Bus System.

proposed models are capable of enhancing the system damping characteristics. Since there are no cases of underestimation and/or overestimation, the model possesses reasonable skills for accurate detection of the dynamics that spread across the 850 tested samples in the actual simulation. The model can capture the sudden dips and the hard edges of the troughs and crests of the evolving modes. However, some of the model predictions could not attain some of the peaks in the actual signal but in all, the predictions do not deviate from the original signal course and hence, do not suffer from prediction outliers. This performant model can be utilized to generate a reliable reference signal for an adaptive controller as well as a suitable candidate for analyzing the severity of fault on the system's small-signal stability.

B. The IEEE 16-Machine, 68-Bus Test System

The dynamic performance of each generator in the system is described using the sub-transient model with four equivalent rotor coils. In addition, generators G1–8 and G10–12 are equipped with IEEE standard DC exciter (DC4B); G9 uses fast static excitation (ST1A), and the remaining generators G13–16 have manual excitation. Additionally, two lead-lag compensation and washout filter blocks make up the standard power system stabilizers that are installed at the DC and fast static excitation systems. This system's test data is directly derived from [32]. Extensive time-domain simulations are performed for various loading conditions, fault locations, and clearing times in order to produce the training, evaluation and test dataset. All loads' active and reactive power have been independently and randomly varied between 80% and 120% of the basic loading level. Additionally, three-phase to ground faults are applied at either system buses and transmission lines with random distance in order to emulate practical contingencies. These faults are cleared after a random time within 0.06 to 0.5 seconds.

Model Evaluation: Three small-signal stability prediction models are trained for this test case; Two-stacked LSTM, CNN-LSTM and Conv-LSTM, respectively. In addition, each of these models is trained using 403,200 training observations, 50,400 validation samples, and 50,400 test samples. Further, the training and test results are compared with those obtained by training SSLSTM-256 as reported in [19]. TABLE VI and TABLE VII present the estimated RMSE and MAAPE metric values for all models and observed oscillation modes. It can be seen that the trained models which are proposed in this work achieve high accurate reconstructions of the

TABLE VI: RECONSTRUCTED ξ RMSE FOR DIFFERENT Models

	RMSE			
	IEEE 68-Bus			
	SSLSTM-256 [19]	Stacked LSTM	CNN-LSTM	CONV-LSTM
ξ_1	0.00410	0.00335	0.00294	0.00348
ξ_2	0.00230	0.00285	0.00254	0.00259
ξ_3	0.00450	0.00415	0.00373	0.00435
ξ_4	0.00050	0.00073	0.00044	0.00179
ξ_5	0.00200	0.00220	0.00190	0.00227
Average error	0.00268	0.00266	0.00231	0.00290
IEEE 145-Bus				
ξ_1	0.00040	0.00079	0.00088	0.00075
ξ_2	0.00020	0.00008	0.00009	0.00012
ξ_3	0.00040	0.00010	0.00014	0.00008
ξ_4	0.00140	0.00019	0.00100	0.00081
ξ_5	0.00140	0.00025	0.00079	0.00079
ξ_6	0.00010	0.00010	0.00011	0.00018
Average error	0.00065	0.00025	0.00050	0.00046

system modes ($\zeta_1, \zeta_2, \dots, \zeta_5$) in real time. Furthermore, the proposed CNN-LSTM and LSTM models achieve slightly better accuracy than Conv-LSTM. The proposed CNN-LSTM achieves 13.8% RMSE and 23.8% MAAPE prediction performance better than that of SSLSTM-256. Also, the propose LSTM achieves 0.7% RMSE and 6.3% MAAPE lower. However, Conv-LSTM corresponding error values are slightly higher than those of SSLSTM-256. This test system has five oscillation modes (output signals): eigenvalues λ_{1-3} and λ_{4-5} (local and inter-area modes, respectively).

TABLE VII. RECONSTRUCTED ξ MAAPE FOR DIFFERENT Models.

	MAAPE			
	IEEE 68-Bus			
	SSLSTM-256 [19]	Stacked LSTM	CNN-LSTM	CONV-LSTM
ξ_1	2.500	1.954	1.618	2.223
ξ_2	1.850	1.676	1.425	1.754
ξ_3	4.570	5.745	4.342	7.591
ξ_4	0.910	0.624	0.507	0.779
ξ_5	4.430	3.360	2.969	4.243
Average error	2.852	2.672	2.172	3.318
IEEE 145-Bus				
ξ_1	0.310	0.256	0.218	0.257
ξ_2	0.190	0.119	0.093	0.202
ξ_3	0.620	0.120	0.207	0.103
ξ_4	1.610	0.221	0.861	0.771
ξ_5	1.640	0.156	0.554	0.604
ξ_6	0.100	0.066	0.077	0.174
Average error	0.745	0.156	0.335	0.352

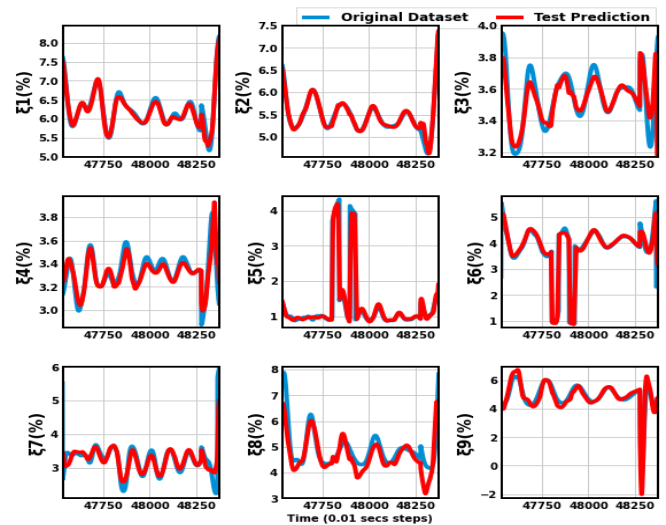


Fig. 10. Actual and proposed LSTM-64 reconstructed values of ξ_j for an arbitrary test case in the New England 39-Bus System.

It can be seen from Figs. 11 and 12 that predictions provided by the proposed model do not suffer overestimations. However, it observed the reconstructed damping rate ξ_2 and ξ_3 in Fig. 13 exhibit slight overestimations after the 750 samples and also, there are some obvious underestimations of ξ_4 . There are also irregular random spikes of varying amplitudes in λ_1 which appear to all the models as outliers and thus, affect the models' learning and generalization and well as the reconstruction of ξ_1 . Despite the above constraint, the trained model appears to

capture the dynamics stronger. Overall, our well-trained model has a better grasp on the mode oscillations and this makes the model more suitable for small signal stability forecasts.

C. The IEEE 50-Machine, 145-Bus Test System

The dynamic performance of synchronous generators G_{1-6} , and G_{23} are described using the sub-transient model. These generators come with two lead-lag PSS and fast static exciters (ST1A). Additionally, the dynamic performance of remaining generators is described using a classical model [19]. The training and test examples for this system are generated using two different types of faults. The first type is a three-phase to-ground fault that occurs at any bus and is naturally cleared. The second kind consists of three-phase-to-ground faults that can occur anywhere along a line at randomly chosen points. These faults are resolved by tripping the line from both ends. Additionally, a random duration between 0.06 and 0.5 seconds is chosen for fault clearance.

Model Evaluation: Here, the three small-signal stability prediction models are trained for this test case. Besides, each of the models is trained using 327,080 training observations, 40,885 validation samples, and 40,885 test samples. Further, the training and test results are compared with those obtained by training SSLSTM-256 as reported in [19]. TABLE VI and TABLE VII present the estimated RMSE and MAAPE metric values for all models and observed modes of oscillations. It can be seen that the trained models which are proposed in this work achieve high accurate reconstructions of the system modes ($\zeta_1, \zeta_2, \dots, \zeta_6$) in real-time.

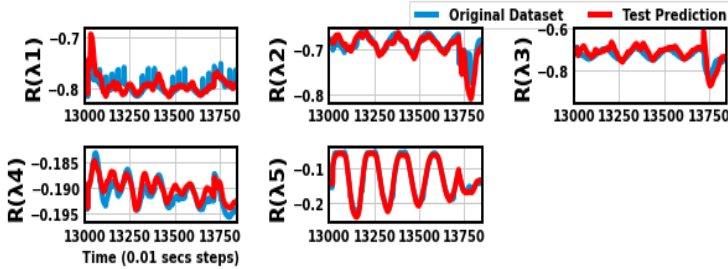


Fig. 11. Actual and proposed LSTM-256 predictions of $R(\lambda_j)$ for an arbitrary test case in the IEEE 68-Bus System.

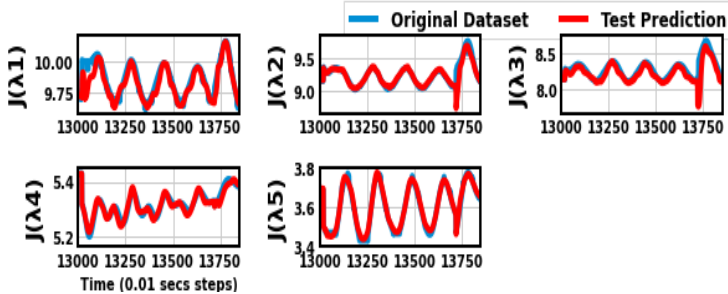


Fig. 12. Actual and proposed LSTM-256 predictions of $I(\lambda_j)$ for an arbitrary test case in the IEEE 68-Bus System.

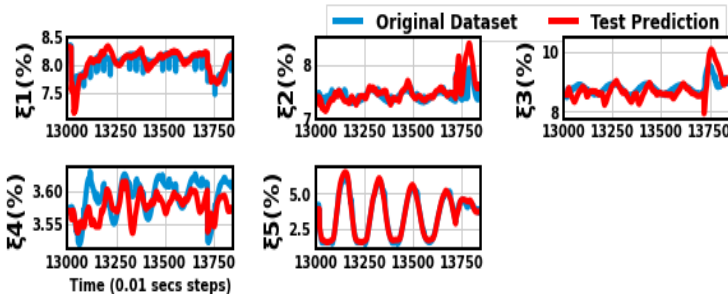


Fig. 13. Actual and proposed LSTM-256 reconstructed values of ζ_j for an arbitrary test case in the IEEE 68-Bus System.

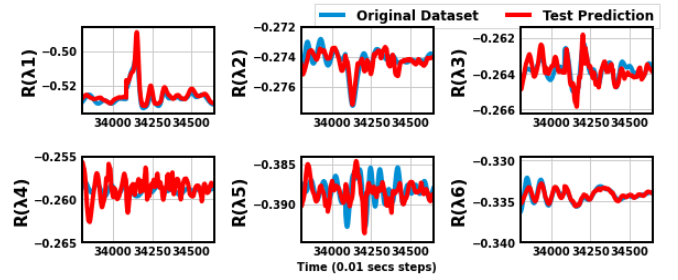


Fig. 14. Actual and proposed LSTM-256 predictions of $R(\lambda_j)$ for an arbitrary test case in the IEEE 145-Bus System.

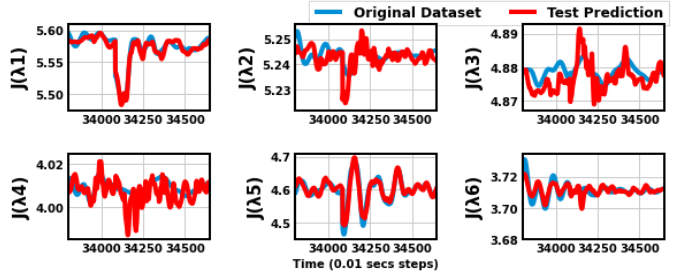


Fig. 15. Actual and proposed LSTM-256 predictions of $I(\lambda_j)$ for an arbitrary test case in the IEEE 145-Bus System.

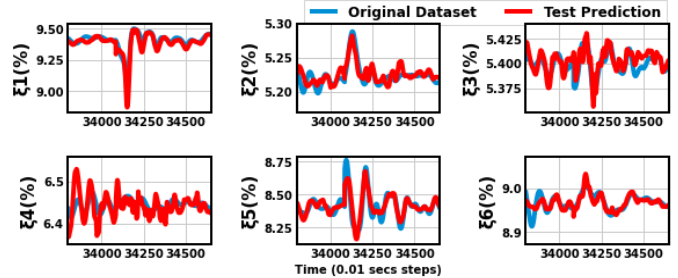


Fig. 16. Actual and proposed LSTM-256 reconstructed values of ζ_j for an arbitrary test case in the IEEE 145-Bus System.

Further, the proposed LSTM, CNN-LSTM and Conv-LSTM models achieve 61.5%, 23.1% and 29.2% RMSE values lower than SSLSTM-256. Similarly, the models gain 79.1%, 55.0% and 52.8% MAAPE values lower than the SSLSTM-256. The LSTM model predictions of the six oscillation modes (output signals): eigenvalues λ_{1-6} (inter-area oscillation modes) are displayed in Fig. 14 and 15. It is observed that the reconstructed damping rate ξ_{1-6} (Fig. 16) has a significantly lower variance. Nonetheless, there are some visible peaks in some of the real components (Fig. 14). In addition, the model could not draw a more smooth curve over the hard edges of the actual $J(\lambda_3)$ and $J(\lambda_4)$ imaginary components (Fig. 15). In general, some contingencies in the train, validation and test data of $J(\lambda_3)$ and $J(\lambda_4)$ are smooth with little or no oscillations and making the model see the intermittent oscillations as outliers leading to lower gradient improvements during training. However, the model exhibits a better fitting for λ signals with no noisy oscillations.

D. The IEEE 39-Bus Test System with 400 MW Wind Plants

To assess the effectiveness of the proposed approach in the presence of intermittent renewable power generation sources, the 39-bus test system is revised such that two 400MW wind power plants are installed at buses 16 and 30, respectively. The wind power plants run variable-speed wind turbines with doubly fed induction generators (DFIG) [1]. The wind speed for both plants is described by means of the Weibull's distributions.

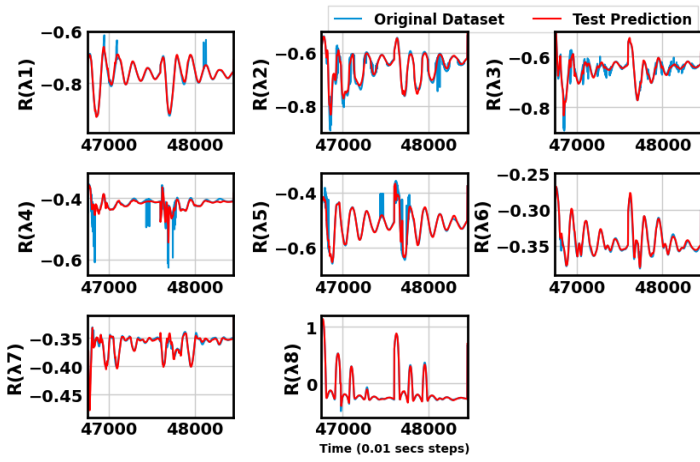


Fig. 17. Actual and proposed LSTM-128 predictions of $R(\lambda_j)$ for an arbitrary test case in the IEEE 39-Bus System with wind plants.

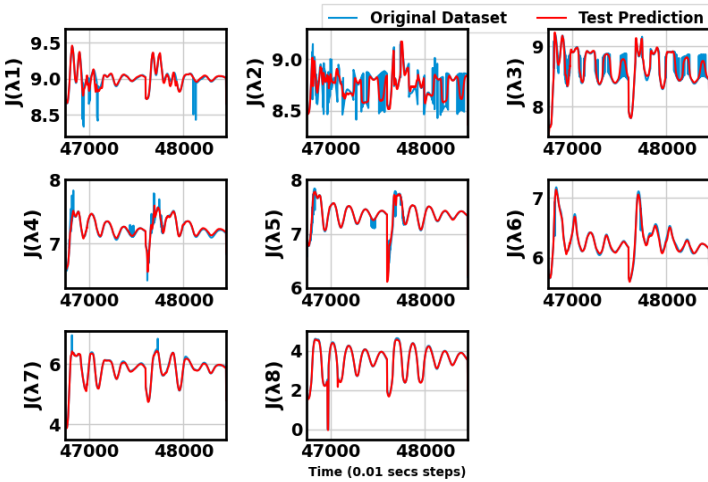


Fig. 18. Actual and proposed LSTM-128 predictions of $I(\lambda_j)$ for an arbitrary test case in the IEEE 39-Bus System with wind plants.

The stator and rotor flux dynamics of the DFIG generators are neglected as they are faster in comparison with grid dynamics. The converter is modeled as an ideal current source, where i_{qr} and i_{dr} are state variables which are used for the rotor speed and voltage control, respectively. A new dataset is generated through extensive time domain simulations that are carried out in PSAT for the revised 39-bus test system. The wind speed pattern is generated using Weibull's distribution for each contingency. In addition, three-phase-to-ground faults are applied at different locations with and without line tripping. For every disturbance, the simulation was run for 10 seconds. The low-frequency oscillations consisting of 1000 time steps corresponding to stable rotor angle are selected for each fault instance, while the unstable ones are discarded. After accounting for clearing time and delays, 850 samples (8.5 seconds) are considered for each fault scenario. A total of 816000 samples (960 fault cases) are aggregated for the proposed model training. The data samples are segmented into 80% (652800) train, 10% (81600) validation and 10% (81600) test sets. It's discovered that a four-layered LSTM architecture comprising the same neurons is most suitable for capturing the stochastic impact of wind generations in this study. Three sets of experiments for the architecture each of 64, 128, and 256 cell units are run for 250 epochs. The architecture including 128 neurons in each hidden layer explains quite well the low-frequency

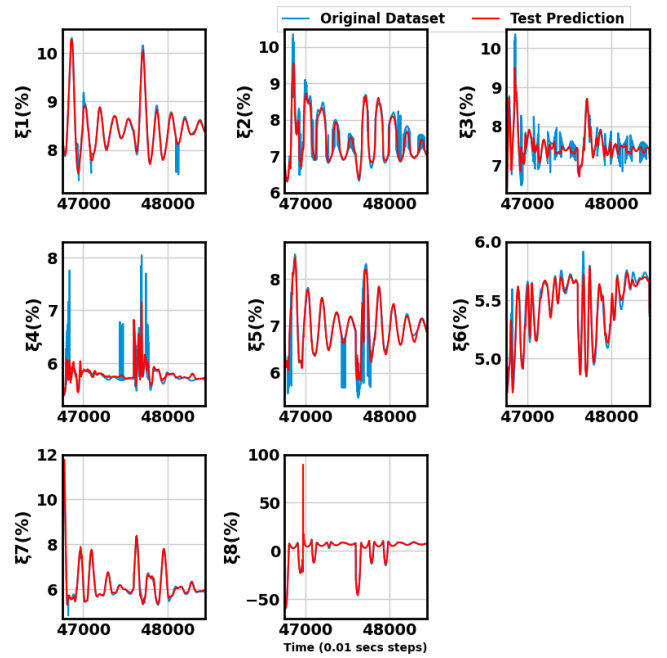


Fig. 19. Actual and proposed LSTM-128 reconstructed values of ξ_j for an arbitrary test case in the IEEE 39-Bus System with wind plants.

oscillations in grid-connected renewable energy systems with 0.067 RMSE of the test sample. While 64- and 256-based architecture display 0.070 and 0.072, respectively. Figs. 17, 18 and 19 depict the performance of the proposed model on power system with wind generation plants. It can be seen the model approximates and fits the evolving low frequency oscillations quite well. However, the model could not significantly explain the inconsistent oscillations patterns, which might be due to unusual fluctuations of renewable generations.

E. A Comparison of actual and estimated eigenvalues of few cases

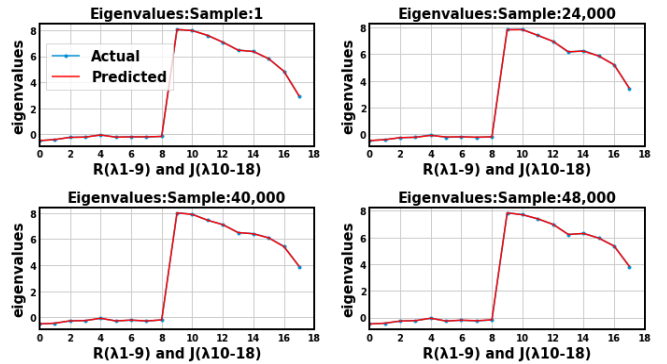


Fig. 20. Actual and estimated eigenvalues at time steps: $t_{(1)}$, $t_{(24,000)}$, $t_{(40,000)}$, and $t_{(48,000)}$

Fig. 20 depicts some arbitrary samples of estimated and actual eigenvalues of the test set (50830 samples) of the 39-bus system. Due to space limitation, the plot compares only the actual and estimated eigenvalues at time steps: $t_{(1)}$, $t_{(24,000)}$, $t_{(40,000)}$, and $t_{(48,000)}$. It can be inferred from Fig. 20 that the trained model can highly approximate the actual eigenvalues across the test samples.

VI. CONCLUSION

Real-time monitoring of poorly damped low-frequency modes of oscillations is crucial for safe and reliable operation of concurrent power systems. To achieve this objective, this paper benefits from the recent developments in the field of artificial intelligence and machine learning techniques. Online small-signal stability prediction models are trained using stacked long short-term memory (LSTM), convolutional neural network (CNN)-LSTM and Convolutional LSTM (Conv-LSTM) architectures. The main objective is to improve the overall forecasting performance in comparison with the described SSLSTM-128 and SSLSTM-256 in [19] and to show the effectiveness of the proposed model with renewable integrated power system. In this context, the average RMSE and MAAPE error values are computed after the reconstruction of damping ratios of each of the proposed models. It is found that these metrics are about 87.4% and 75% lower than SSLSTM-128 for the New England 39-Bus system. Furthermore, the proposed CNN-LSTM achieves 13.8% RMSE and 23.8% MAAPE, LSTM achieves 0.7% RMSE and 6.3% MAAPE prediction performance better than SSLSTM-256 for the IEEE-68 test case. However, the Conv-LSTM corresponding error values are slightly higher. Similarly, the proposed LSTM, CNN-LSTM and Conv-LSTM models achieve 61.5%, 23.1% and 29.2% RMSE values and 79.1%, 55.0% and 52.8% MAAPE prediction performance better than SSLSTM-256 for IEEE-145 test system. Overall, the well-trained proposed models can be utilized to effectively learn the operations of renewable coupled power system.

REFERENCES

- [1] A. O. Muhammed and M. Rawa, "A Systematic PVQV-Curves approach for investigating the impact of solar photovoltaic-generator in power system using powerworld simulator," *Energies*, vol. 13, no. 10, p. 2662, 2020.
- [2] Y. J. Isbeih, M. S. El Moursi, W. Xiao, and E. El-Saadany, "mixed-sensitivity robust control design for damping low-frequency oscillations with DFIG wind power generation," *IET Gener. Transm. Distrib.*, vol. 13, no. 19, pp. 4274–4286, 2019, doi: 10.1049/iet-gtd.2018.6433.
- [3] J. Condren and T. W. Gedra, "Expected-Security-Cost Optimal Power Flow With Small-Signal Stability Constraints," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1736–1743, Nov. 2006, doi: 10.1109/TPWRS.2006.882453.
- [4] R. Zarate-Minano, F. Milano, and A. J. Conejo, "An OPF Methodology to Ensure Small-Signal Stability," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1050–1061, Aug. 2011, doi: 10.1109/TPWRS.2010.2076838.
- [5] T. Surinkaew and I. Ngamroo, "Hierarchical Co-Ordinated Wide Area and Local Controls of DFIG Wind Turbine and PSS for Robust Power Oscillation Damping," *IEEE Trans. Sustain. Energy*, vol. 7, no. 3, pp. 943–955, Jul. 2016, doi: 10.1109/TSTE.2015.2508558.
- [6] L. Fan, H. Yin, and Z. Miao, "On Active/Reactive Power Modulation of DFIG-Based Wind Generation for Interarea Oscillation Damping," *IEEE Trans. Energy Convers.*, vol. 26, no. 2, pp. 513–521, Jun. 2011, doi: 10.1109/TEC.2010.2089985.
- [7] H. Nguyen-Duc, H. Cao-Duc, C. Nguyen-Dinh, and V. Nguyen-Xuan-Hoang, "Simulation of a power grid blackout event in Vietnam," in *2015 IEEE Power & Energy Society General Meeting*, Jul. 2015, pp. 1–5. doi: 10.1109/PESGM.2015.7286511.
- [8] M. I. Hossain, M. Shafiullah, M. Abido, and M. A. Emran, "Online Monitoring of Inter-Area Oscillations and damping of Power systems employing Prony Analysis," in *2018 10th International Conference on Electrical and Computer Engineering (ICECE)*, Dec. 2018, pp. 269–272. doi: 10.1109/ICECE.2018.8636737.
- [9] "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise | IEEE Journals & Magazine | IEEE Xplore." <https://ieeexplore.ieee.org/document/1163974> (accessed Sep. 12, 2022).
- [10] M. Barbosh, P. Singh, and A. Sadhu, "Empirical mode decomposition and its variants: a review with applications in structural health monitoring," *Smart Mater. Struct.*, vol. 29, no. 9, p. 093001, Aug. 2020, doi: 10.1088/1361-665X/aba539.
- [11] Y. Wang, Y. Sun, Z. Wei, and G. Sun, "Parameters Estimation of Electromechanical Oscillation With Incomplete Measurement Information," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5016–5028, Sep. 2018, doi: 10.1109/TPWRS.2018.2817214.
- [12] J. L. Rueda, C. A. Juarez, and I. Erlich, "Wavelet-Based Analysis of Power System Low-Frequency Electromechanical Oscillations," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1733–1743, Aug. 2011, doi: 10.1109/TPWRS.2010.2104164.
- [13] P. Sharma, V. Ajarapu, and U. Vaidya, "Data-Driven Identification of Nonlinear Power System Dynamics Using Output-Only Measurements," *IEEE Trans. Power Syst.*, vol. 37, no. 5, pp. 3458–3468, Sep. 2022, doi: 10.1109/TPWRS.2021.3131639.
- [14] D. Shi *et al.*, "Study on quick judgement of small signal stability using CNN," *J. Eng.*, vol. 2019, no. 16, pp. 826–829, 2019, doi: 10.1049/joe.2018.8839.
- [15] J. Cepeda, I. Gómez, F. Calero, and A. Vaca, "Big Data Platform for Real-Time Oscillatory Stability Predictive Assessment Using Recurrent Neural Networks and WAProtector's Records," in *2022 International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA)*, May 2022, pp. 1–6. doi: 10.1109/SGSMA51733.2022.9806014.
- [16] J. Liu, Z. Yang, J. Zhao, J. Yu, B. Tan, and W. Li, "Explicit Data-Driven Small-Signal Stability Constrained Optimal Power Flow," *IEEE Trans. Power Syst.*, vol. 37, no. 5, pp. 3726–3737, Sep. 2022, doi: 10.1109/TPWRS.2021.3135657.
- [17] S. A. Dorado-Rojas, T. Bogodorova, and L. Vanfretti, "Time Series-Based Small-Signal Stability Assessment using Deep Learning," *2021 North Am. Power Symp. NAPS*, Nov. 2021, doi: 10.1109/naps52732.2021.9654643.
- [18] Y. Yang, X. Zhang, and L. Yang, "Data-driven power system small-signal stability assessment and correction control model based on XGBoost," *Energy Rep.*, vol. 8, pp. 710–717, Aug. 2022, doi: 10.1016/j.egyr.2022.02.249.
- [19] S. K. Azman, Y. J. Isbeih, M. S. E. Moursi, and K. Elbassioni, "A Unified Online Deep Learning Prediction Model for Small Signal and Transient Stability," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4585–4598, Nov. 2020, doi: 10.1109/TPWRS.2020.2999102.
- [20] P. Kundur, N. J. Balu, and M. G. Lauby, *Power system stability and control*, vol. 7. McGraw-hill New York, 1994.
- [21] L. Liu, Y. Li, Y. Cao, F. Liu, W. Wang, and J. Zuo, "Transient Rotor Angle Stability Prediction Based on Deep Belief Network and Long Short-term Memory Network," *IFAC-Pap.*, vol. 52, no. 4, pp. 176–181, Jan. 2019, doi: 10.1016/j.ifacol.2019.08.175.
- [22] C. B. Saner, M. Kesici, M. Mahdi, Y. Yaslan, and V. M. I. Genc, "Wide Area Measurement-based Transient Stability Prediction using Long Short-Term Memory Networks," in *2019 7th International Istanbul Smart Grids and Cities Congress and Fair (ICSG)*, Apr. 2019, pp. 159–163. doi: 10.1109/ISGCF.2019.8782391.
- [23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, 2012, vol. 25. Accessed: Jul. 11, 2022. [Online]. Available: <https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>
- [24] C. M. Bishop, "Neural Networks for Pattern Recognition," p. 498.
- [25] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. Cambridge, MA, USA: MIT press, 2016.
- [26] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [27] J. Xie and W. Sun, "A Transfer and Deep Learning-Based Method for Online Frequency Stability Assessment and Control," *IEEE Access*, vol. 9, pp. 75712–75721, 2021, doi: 10.1109/ACCESS.2021.3082001.
- [28] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *Adv. Neural Inf. Process. Syst.*, vol. 28, 2015.
- [29] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation." arXiv, Mar. 08, 2015. doi: 10.48550/arXiv.1411.4038.
- [30] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization." arXiv, Jan. 29, 2017. Accessed: Jul. 11, 2022. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [31] J. Donahue *et al.*, "Long-term Recurrent Convolutional Networks for Visual Recognition and Description." arXiv, May 31, 2016. doi: 10.48550/arXiv.1411.4389.
- [32] A. K. Singh and B. C. Pal, "Report on the 68-bus, 16-machine, 5-area system," *IEEE PES Task Force on Benchmark Systems for Stability Controls*, Ver. Dec. 03, 2013. http://eicoc.pnnl.gov/benchmark/ieees/NETS68/New_England_New_York_68_Bus_System_study_report.pdf (accessed Jul. 13, 2022).
- [33] F. Milano, "An Open Source Power System Analysis Toolbox," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1199–1206, Aug. 2005, doi: 10.1109/TPWRS.2005.851911.
- [34] J. J. Q. Yu, D. J. Hill, A. Y. S. Lam, J. Gu, and V. O. K. Li, "Intelligent Time-Adaptive Transient Stability Assessment System," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1049–1058, Jan. 2018, doi: 10.1109/TPWRS.2017.2707501.



Abdullahi Oboh Muhammed received the B.Eng degree in electronics and electrical engineering and the M.Eng degree in electronics and Telecommunication engineering from the University of Benin, Nigeria, in 2013 and 2018, respectively. He also, received the M.Sc degree in electrical and computer engineering (Power and electrical machines) from King Abdulaziz University, KSA, in 2020. He is currently pursuing the PhD degree in electrical and computer engineering at Khalifa University of Science and Technology, UAE. His

research interests include applications of Artificial intelligence, Machine learning, deep learning, Power systems and renewable energy integration.



Younes J. Isbeih received a B.Sc. degree from the German Jordanian University, Jordan in 2011 and an M.Sc. and Ph.D. from Khalifa University of Science and Technology - Masdar Campus, UAE in 2015 and 2019, respectively. He is currently working as a Post-Doctoral Fellow Researcher at the Electrical Engineering and Computer Science department at Khalifa University of Science and Technology. His research interests are diverse, including power system dynamics, stability, robust control design, and integration of renewable energy resources like solar and wind.



Mohamed Shawky El Moursi (M'12, SM15) received the B.Sc. and M.Sc. degrees from Mansoura University, Mansoura, Egypt, in 1997 and 2002, respectively, and the Ph.D. degree from the University of New Brunswick (UNB), Fredericton, NB, Canada, in 2005, all in electrical engineering.

He was a Research and Teaching Assistant in the Department of Electrical and Computer Engineering, UNB, from 2002 to 2005. He joined McGill University as a Postdoctoral Fellow with the Power Electronics Group. He worked at Vestas Wind

Systems, Arhus, Denmark, in the Technology R&D with the Wind Power Plant Group. He was with TRANSCO, UAE, as a Senior Study and Planning Engineer. He is currently a Professor in the Electrical Engineering and Computer Science Department and the Acting Director of Advanced Power and Energy Center (APEC) at Khalifa University and seconded to a distinguished Professor position in the Faculty of Engineering, Mansoura University, Egypt and currently on leave. He was a Visiting Professor at Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. Dr. Shawky is currently an Editor of IEEE Transactions on Power Delivery, IEEE Transactions on Power Systems, Associate Editor of IEEE Transactions on Power Electronics, Associate Editor of IEEE Transactions on Smart Grid, Guest Editor of IEEE Transactions on Energy Conversion, Guest Editor-in-Chief for special section between TPWRD and TPWRS, Editor for IEEE Power Engineering Letters, Regional Editor for IET Renewable Power Generation and Associate Editor for IET Power Electronics Journals. His research interests include renewable energy systems (Wind and PV) integration, power system, FACTS technologies, VSC-HVDC systems, Microgrid operation and control and AI applications in Power Systems.



Khalifa Al Hosani (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Notre Dame, Notre Dame, IN, USA, in 2005 and 2007, respectively, and the Ph.D. degree in electrical and computer engineering from The Ohio State University, Columbus, OH, USA, in 2011. He is currently an Associate Professor with the Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi, United Arab Emirates. He is also an Engineering & Technology affiliate Member with the Mohammed bin Rashed Academy

of Scientists. He is the Co-Founder of the Power Electronics and Advanced Sustainable Energy Center Laboratory, ADNOC Research and Innovation Center (now Power Electronics and Sustainable Energy (PEASE) Research Lab at Khalifa University) Abu Dhabi, United Arab Emirates. His research interests include a wide range of topics including nonlinear control, sliding mode control, control of power electronics, power systems stability and control, renewable energy systems modeling and control, smart grid, microgrid and distributed generation, and application of control theory to oil and gas applications.