A Novel Transformer Approach for the Recompensed Measurement Generation and Accurate Topology Identification

Seshu Moturu*, Srinivasa Rao Gummadi, Madhu Valavala, Veera Vasantha Rao Battula, Sravanthi Kantamaneni

***Abstract*—Accurate topology identification (TI) is essential for various applications, including fault location, load flow analysis, state estimation and system planning in the distribution networks. However, TI is vulnerable to missing measurements that may arise due to meter malfunctions and communication failure due to denial of service (DoS) attacks. Thus, a robust and novel methodology for accurate TI is proposed in this study. The proposed methodology is divided into two steps: 1) generating the recompensed measurement data by attention mechanism-based transformer model and 2) topology identification from the recompensed measurement data. For evaluating the efficacy of the proposed methodology, a comprehensive study is conducted to assess the influence of renewable energy sources (RES) on the prediction performance. This investigation aims to quantify the degree to which the integration of RES influenced the proposed approach's efficacy and robustness. The proposed approach has been tested and evaluated with varying percentages of missing data such as 10%, 30% and 50%, for the modified IEEE 37 and 69 node system. In addition, a comparative study with various reference models for the missing measurements forecasting is also conducted. The case study results indicate that the proposed approach outperforms other approaches in missing measurement forecasting and TI, while also demonstrating resilience in the context of missing measurements. The proposed approach improved the performance of the system by more than 35% for the 10%, 30%, and 50% missing percentages in the test systems.**

***Index Terms*—topology, deep learning, missing, measurements, forecasting**

I. Introduction

CONTROL and monitoring were not important in conventional distribution system because of the radial connections, unidirectional power flow, and regular load patterns. The distribution system is becoming more market-ready due to the rapid growth of RES, mainly solar and wind energy. Due to frequent switching and control, distributed energy resources can frequently affect the network topology [1]. The objective of identifying distribution network topology is to determine the current operational topology based on the provided distribution system model by utilising measurements obtained in real-time. Hence, the accurate real-time distribution network topology identification is challenging task.

The majority of traditional TI methods are dependent on the state of switches and circuit breakers across the feeders to obtain the adjacent matrix. The utilisation of switch statuses as a priori topology information by the authors has been employed in the estimator to evaluate and rectify the topology [2]. In a similar manner, the authors in [3] employed a method of detecting changes in topology through the assumption of an initial known topology. However, the timely availability or completeness of updated topology information may be compromised. In metropolitan areas and industrial zones, feeders often comprise substantial underground networks, making the supervision of switches and breakers a formidable task. At the same time, the technology in the distribution system is advancing and generating more measuring data for analysis. Control operators have more measurement data because of advanced technology including phasor measuring units (PMUs), improved meters, line current sensors, and SCADA. Different parameters such as standard deviation of voltage drop [4], voltage covariance matrix [5], and the voltage phasors [6] are used to determine whether two nodes of the system are connected. The advancements in AI methodologies and measurement technologies have facilitated the processing of larger datasets for the purpose of training AI models to achieve optimal performance in the TI. In [7], the authors developed a TI model by formulating it as a multi-label classification problem. And in [8], the authors proposed a deep network model for the online TI. These models increase TI performance but are not resilient to missing data. DoS attacks interrupt the measuring infrastructure with malicious requests or disrupt communication channels, making them unavailable or unreliable. Topology identification fails due to inaccurate and missing measurements, affecting control centres severely.

Missing measurement compensation approaches range from the zero-order hold approach [9] to deep learning algorithms. The power system's non-linearity makes standard imputation methods ineffective. Deep neural network models are well-suited for power system non-linearity. It is noteworthy that the estimation of missing measurements can be considered a time series prediction problem in wider terms

Manuscript received June 10, 2023; revised December 28, 2023

Seshu Moturu is an Assistant professor in the Department of Electrical and Electronics Engineering, Prasad V Potluri Siddhartha Institute of Technology, Kanuru, Vijayawada, Andhra Pradesh, India (corresponding author, email: moturuseshu@gmail.com).

Srinivasa Rao Gummadi is an Associate Professor in the Department of Electrical and Electronics Engineering, V.R. Siddhartha Engineering College, Kanuru, Vijayawada, Andhra Pradesh, India (email: vasu1in@vrsiddhartha.ac.in).

Madhu Valavala is an Associate Professor in the Department of Electrical and Electronics Engineering, Swarnandhra College of Engineering & Technology, Naraspur, Andhra Pradesh, India (email: vmadhu.eee@swarnandhra.ac.in).

Veera Vasantha Rao Battula is an Assistant Professor in the Department of Electrical and Electronics Engineering, RVR&JC College of Engineering, Guntur, Andhra Pradesh, India (email: Veeravasantharao@gmail.com).

Sravanthi Kantamaneni is an Assistant Professor in the Department of Electronics and Communication Engineering, RVR & JC College of Engineering, Guntur, Andhra Pradesh, India (email: sravanthivasanth@gmail.com).

[10]. Hence, from the literature on time series prediction models, they are divided into two broad categories such as statistical models [11], and AI-based models. The working of statistical models involves the analysis of past data patterns and attributes to generate forecasts on future values. The statistical models include autoregressive integrated moving average (ARIMA) [12], [13], seasonal ARIMA (SARIMA), exponential smoothing methods, vector autoregression (VAR), and Bayesian structural time series (BSTS) [14]. However, these models only work with stationary data, small-scale, univariate systems and need extensive pre-processing. AI-based forecasting models include machine learning [12], [15] and deep learning models [16]. Support vector machine (SVM) is the popular ML model. However, ML models may not work for non-linear and large datasets of the power system. Deep learning models can interpret and generalise non-linear data, making them effective time series prediction models. Recurrent neural networks (RNNs) are designed for sequential data applications [17]. Long short-term memory (LSTM) [18], elman neural network (ENN) [17], gated recurrent unit (GRU) [19], and bidirectional LSTM (Bi-LSTM) [20] models are RNN architecture variants. For example, in [17], the authors proposed a ENN based model for the wind speed forecasting application showcasing its improved performance in comparison with other models. RNNs can capture sequential dependencies and handle input sequences of different lengths, making them suitable for time series applications. 1-dimensional convolutional neural network model (CNN) is also employed for time series applications because of its spatial feature extraction property [21], [22]. The authors used the hybrid version of the CNN and LSTM in [23], [24].

A. Demerits of the forecasting models

Most of the authors use LSTM and the variant versions of the LSTM model for the forecasting application. However, these forecasting models fail to forecast the missing measurement effectively because of the complex and huge size of the distribution system. These RNN models such as LSTM particularly possess certain limitations, including: a) computationally heavy b) require a larger dataset for training c) restricted receptive field d) more memory and time complexity. Thereby limiting the LSTM's ability to deliver optimal performance when processing lengthy sequential measurement data to forecast efficiently. The presence of missing samples highly impacts the performance of the TI models. The accurate identification of topology is crucial, thus efficient imputation of missing data is necessary.

B. Contributions

This study introduces a novel methodology based on the transformer architecture aimed at resolving the above-mentioned problems, pertaining to the forecasting and compensation of missing measurements and the accurate TI. The primary contributions of this study are as follows: To solve the shortcomings of the LSTM model, the transformer model is proposed for forecasting the missing measurement data. The proposed transformer model not only exhibits superior precision in missing measurements forecasting but also possesses parallel processing capability, a self-attention mechanism that enables a larger receptive field, the ability to interpret more asynchronous relationships between dynamic and complex measurement data and efficient memory usage. This study involves a comparison of the performance of the proposed transformer model with other reference models in the context of generating the recompensed measurement data for varying percentages of missing data, specifically 10%, 30%, and 50%. In evaluating the efficacy of the proposed methodology, a comprehensive study is conducted to assess the influence of RES on performance. The proposed approach resulted in the dominating performance in both the modified IEEE 37 node system with and without RES integration. The scalability test of the proposed approach is conducted using the modified IEEE 69 node system. And the results indicate more than 35% improvements over the second-best measurement forecasting approach when applied to the modified IEEE 37 and 69 node system for all three levels of missing in both cases.

C. Organization

The subsequent sections of the manuscript are structured in the following manner: In Section II, the TI formulation is provided. Section III outlines the methodology employed in the proposed approach. Experimental results are present in Section IV and the manuscript is concluded in Section V.

II. Problem Formulation

The topology of the distribution network changes more frequently than the mesh-connected transmission network. The dynamic nature of topology in comparison to the transmission system poses significant difficulties for conventional techniques in identifying real-time topology in distribution networks. The use of tie-switches, capacitor banks, and sectionalizing switches further guarantees the reliability, security, and simplicity of the distribution network. In the radial networks, it is postulated that the loads are subject to uniform scaling and that significant occurrences such as capacitor bank switching or topological modifications are infrequent. In this particular instance, the statistical properties of the voltages can be categorised into three primary characteristics: the uniformity of voltage reduction, the uniformity of voltage magnitude, and the point at which voltage measurement changes. The utilisation of voltage vectors can be employed for determining the connectivity or topology of the system's nodes. The efficacy of conventional models may be inadequate in addressing this issue. Despite the potential benefits of smart meters in collecting vast amounts of data, data-driven methodologies face challenges in the field of TI due to issues related to data quality, particularly accuracy. The determination of the system's topology is facilitated by the examination of the connectivity status of the switches between the nodes. The binary status of the nodes indicates their connectivity, with a value of either 0 or 1. Robust architectures should be developed for addressing non-linear problems due to the intricate relationship between node voltages and network structure. The occurrence of measurement loss results in topological misidentification. Thus, it is essential to achieve accurate topology even with the missing data.

III. Methodology

The proposed methodology consists of two steps. The first step involves the utilisation of a transformer to forecast the missing measurements, thereby producing the recompensed measurement data. In the second step of the study, the recompensed measurement data is provided to the topology identifier to accurately classify the topology of the system. Before the implementation of the two-step process, the proposed transformer model undergoes training to forecast measurements using historical measurements as the input training data. The efficacy of the proposed transformer model is exhibited through an assessment of its performance across different levels of missing data during this offline phase. The received network measurements are analysed in real-time for any missing data. If there is any missing data, the transformer model is used to forecast the missing measurements and then generate the recompensed measurement data. This recompensed data is sent into the topology classifier for the TI.

A. Transformer model for the generation of recompensed measurement data

The proposed approach forecast missing measurements and the missing positions in the input measurement data are replaced with forecasted values, resulting in the generation of recompensed measurement data. The transformer model receives historical measurement data as input and produces recompensed measurement data as output. The transformer model is a type of recurrent neural network model variation. The transformer model is comprised of various mechanisms, including self-attention and mutual attention mechanisms [25]. The neural network utilising an attention mechanism exhibits similarities to human retrieval processes. The utilisation of attention mechanisms enables the retrieval of relevant information from distant tokens. This is achieved through the assignment of attention weights based on the relevance of the information, which is determined by the previous state and simultaneously process all the tokens and calculate the corresponding attention weights. The proposed transformer model architecture is shown in Fig. 1. Transformer networks provide parallel processing, sequence segment linkages, novel embeddings, and long-term dependency minimization [26]. The transformer network has an encoder-decoder structure. Each input layer determines the importance of input measurement data, which is used for subsequent encoding levels. An extended sequence's relevance to newly created tokens is measured using a self-attention score. Highly relevant input data has a high self-attention score, whereas less relevant material has a lower score. Positional inputs help the transformer network handle temporal dependencies, aiding the self-attention process. The transformer network relies on multiheaded attention as its central mechanism. Specifically, the decoder block's multihead attention takes the encoding block's output as input. The decoder and encoder blocks are iterated multiple times across several layers.

To make the transformer model aware of the order positions, the utilisation of position embedding has been identified as a viable solution for managing the sequential order of measurement data points concerning time. The complete series is transmitted concurrently to the network, thereby eliminating the issues of gradient vanishing or exploding. To effectively handle large quantities of measurement data, it is necessary to assign varying weights to the most significant characteristics of the input. Self-attention mechanisms have been observed to facilitate the selective filtering of non-relevant input while prioritising the more important input components. The utilisation of self-attention is a means of addressing the difficulty posed by the parallel processing of all hidden state time stamps. This results in a current token encoding that is significantly influenced by the relevance of past tokens. Three parameters that hold significant importance for the transformer model in the missing measurement forecasting are Query (P), Key (Q), and Value (R). During the training process, three weight matrices are acquired for each token. These matrices are responsible for generating the query, key, and value vectors for the token.

B. Mathematical analysis and modelling of the proposed transformer network

The input embedding of the transformer is computed as follows:

$$i_M = EE_M + PE_M \tag{1}$$

Adding and normalizing the vectors: The equation for the addition:

$$\psi = \ I \ + \ O \tag{2}$$

Normalization,

$$\eta = \frac{1}{lM}\sum_{a=1}^{l}\sum_{b=1}^{M}\psi_{ab} \tag{3}$$

$$\sigma^2 = \frac{1}{lM}\sum_{a=1}^{l}\sum_{b=1}^{M}\left(\psi_{ab} - \eta\right)^2 \tag{4}$$

$$Z_{ab} = \frac{\psi_{ab} - \eta}{\sqrt{\sigma^2 + \varepsilon}} \tag{5}$$

$$\begin{aligned} I &= [i_1, i_2, \ldots, i_M] \\ O &= [o_1, o_2, \ldots, o_M] \\ P &= [p_1, p_2, \ldots, p_M] \\ Q &= [q_1, q_2, \ldots, q_M] \\ R &= [r_1, r_2, \ldots, r_M] \end{aligned} \tag{6}$$

The processing of self-attention mechanism:

$$[P = SW_P I, Q = SW_Q I, R = SW_R I] \tag{7}$$

$$B = \frac{Q^T P}{\sqrt{l}} \tag{8}$$

The equation for creating the new embedding using weighted average: $O = QSW$ The result of the multi-head attention: the concatenated output vector is represented as

$$O = [o_1, o_2, \ldots, o_H]\, O = SW^0 \widetilde{O} \tag{9}$$

The computation of the masked self-attention is as follows:

$$\begin{aligned} P &= SW_P I \\ Q &= SW_Q I \\ R &= SW_R I \\ Y &= \frac{K^T P}{\sqrt{l}} \end{aligned} \tag{10}$$

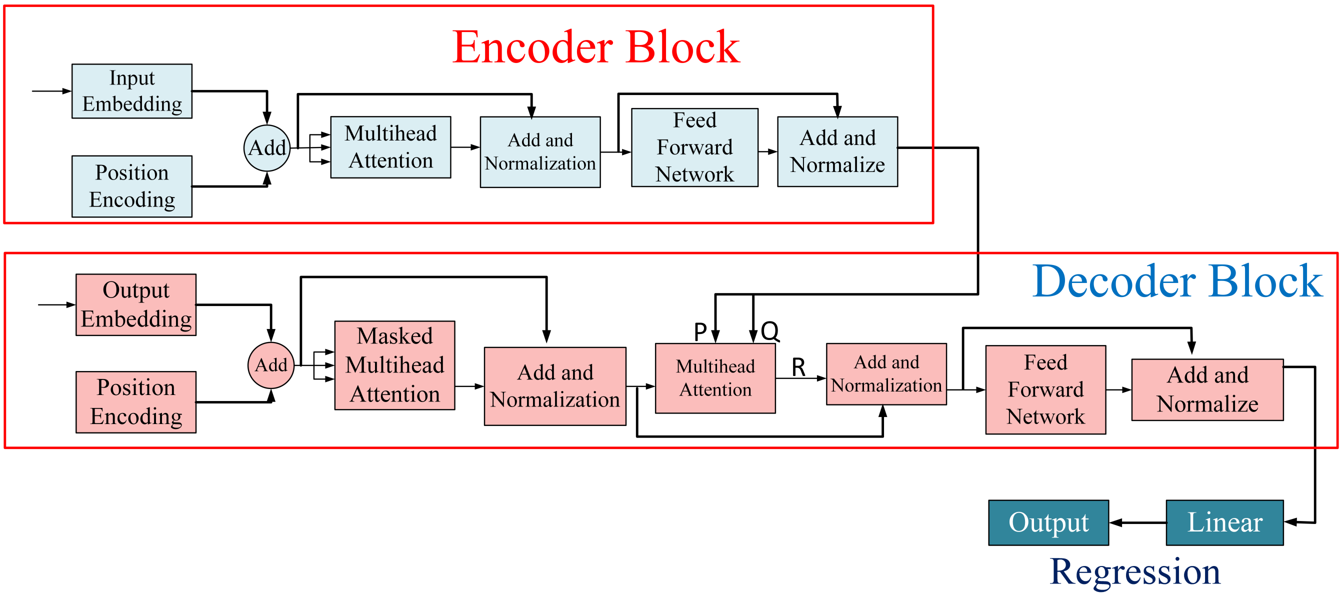

Fig. 1. Architecture of the transformer model

The computation of weighted average new embedding $O = RSW$

The computational details of the encoder and decoder self-attention mechanism are represented below:

$$\begin{aligned} I_d &= [i_1^d, i_2^d, \ldots, i_{Md}^d] \\ I_e &= [i_1^e, i_2^e, \ldots, i_{Me}^e] \\ P &= SW_P I_d \\ Q &= SW_Q I_e \\ R &= SW_R I_e \\ Weights, Z &= \frac{Q^T P}{\sqrt{L}} \\ SW &= soft\max(Z) \end{aligned} \tag{11}$$

Weighted average based new embedding $O = RSW$

The following provides the details of the symbols used in the above mathematical model, sm - SoftMax function, I- input of encoder, O- output of the decoder, I_e encoder output; I_d- decoder's input; M-total inputs, H- heads, l- length of embedding,Q^T- Q's transpose, EE_u- encoder embedding, PE_u- positional embedding, η-mean, $\sigma^2 - variance$, Y_{ab} normalized output, SW^0 - reduced dimensional matrix, ϵ- numerical stability constant, SW- Synaptic weight.

The utilisation of transformer normalisation has been observed to facilitate expedited training processes and mitigate the occurrence of covariate shifts. The inclusion of aid serves to enhance the maintenance of positional data and reinforce the strength of gradients. The current observations provide a resemblance to prior instances in which missing measurement forecasting has relied on a multi-layered approach to achieve accurate predictions. The attention mechanism is utilised to connect an encoder through which a portion of the sequence that is relevant to the present forecasting is identified. The proposed approach involved the implementation of transformer-based fusion, which effectively reduced the training time due to the parallelization technique employed. The encoding process involves the utilisation of the sine and cosine functions. The utilisation of position offset and masking techniques can ensure accurate forecasting by enabling the analysis of future data samples.

C. Proposed Transformer model for forecasting missing measurements and generation of recompensed measurement data

The workflow of the proposed approach for the generation of the recompensed measurement data from the forecasted result of the transformer model and the TI is shown in the Fig. 2 and also provided below:

Workflow of the Proposed approach:

1) Define the original measurement data (X) collected from the nodes of the network
2) Split the X into X_{train} and X_{test} and train the transformer model with X_{train} data.
3) Now define the missing percentage and the attack matrix which is a random matrix of the same size of M with the elements 0 and 1. The measurement is present if the value is 1, and the measurement is absent if the value is 0. Thus, first define and generate the attack matrix as:
   M=random (X, β) Where β is the missing percentage.
4) Using the below equation, generate the attack matrix X_m. $X_m = M. * X$
5) Use the transformer model with the attack matrix X_m to produce the X_{pred}.
6) Generate the recompensed measurement data X_{rec} from the X_{pred} by substituting the missing in the missing positions together with the non-missing measurements. $X_{rec} = X_m + X_{pred}. * (1 - M)$
7) Use the recompensed measurement matrix X_{rec} for the topology identification.

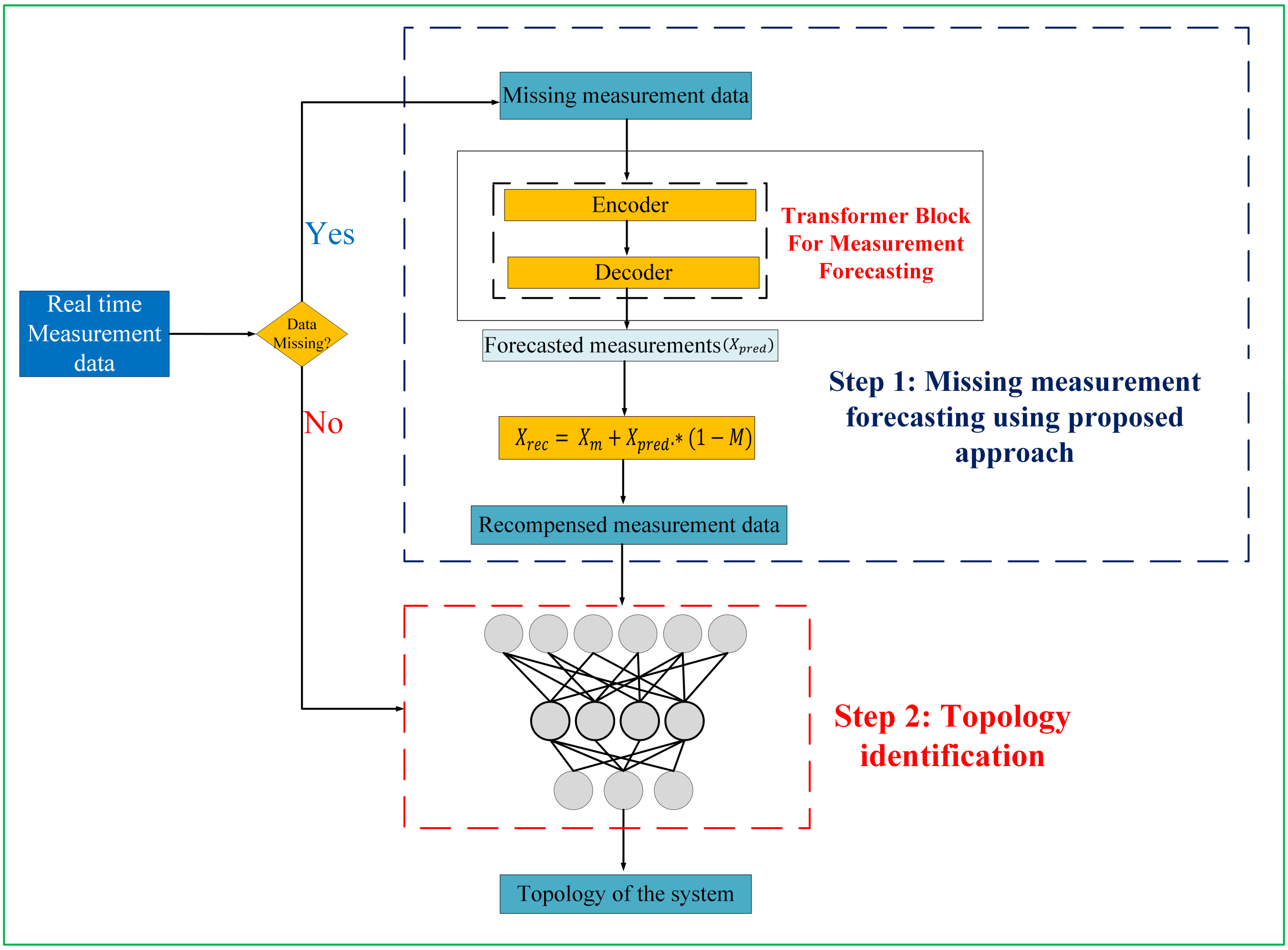

Fig. 2. The complete flowchart of the proposed methodology

8) Topology identification is completed, and the system topology and compensated measurement data are then provided for further analysis.

D. Topology classification using the recompensed measurement data

The missing measurements are filled with the proposed transformer forecasted measurements generating the recompensed measurement data from the previous step. The compensated measurement data is utilised in this stage to execute the TI. In this step, the output of the transformer is utilised to build a neural network model for classification. The proposed methodology involves a classifier where the input layer comprises the recompensed measurement data obtained from the transformer, and the output layer represents the topology of the system. During the online phase, the network operator will generate recompensed measurement data from the transformer in the event of any missing measurements. This recompensed data will be utilised as input for the topology classifier and hence the accurate topology can be identified.

IV. Experimental results and analysis

The experimental results of the proposed methodology for generating the recompensed measurement data and the topology classification are presented in this section.

A. IEEE 37 node system

The experimentation in this study focuses on the IEEE 37-node distribution system. However, certain modifications have been implemented, such as the treatment of distributed loads as lumped loads and the exclusion of transformers and regulators from consideration. The switching lines used in this study are 729-736, 708-733, 731-740, 725-731, and 713-704 as shown in Fig. 3. The formation of different topological structures is achieved by defining switching lines that adhere to specific conditions. These conditions include the preservation of radial network characteristics while avoiding the formation of loops, as well as ensuring that no node in the network is isolated. This study examines a total of 25 distinct topological configurations for the modified IEEE 37-node distribution system.

B. Dataset description

In this study, the hourly load data of 60 days from the National Renewable Energy Laboratory (NREL) database is acquired [27]. Rescaling the gathered load data to align with the IEEE 37 node system is necessary, given the different ranges present in the two datasets. This load data is passed to the distribution system power flow algorithm, resulting in the generation of voltage measurement data for a total of 1440 hours. This study considers 25 different possible topological structures and the total measurements are 108 in the modified IEEE 37 node system. Consequently, the dataset for voltage measurement is established, with a size of (36000x108). Hence, to further evaluate the proposed model, the investigation of the impact of RES on the topology classification is performed in this study. For this, the photovoltaic (PV) are considered at nodes 724, 725, 738, and 729 and wind turbine (WT) systems are considered to be present at nodes 703 and 728. For the topology classification, each topological structure is represented in the equivalent binary form. This results in the identification of five distinct binary classes, which in turn facilitates the classification process as a multi-label classification.

C. Evaluation indices

To evaluate the proposed approach and the comparative approaches for the missing measurement forecasting, root mean square error (RMSE) is employed in this study. It helps in evaluating the forecasting model's effectiveness in capturing the variability of the data. It is a metric that applies a greater penalty to larger forecast errors compared to other error metrics.

The mathematical formula is provided below:

$$RMSE = \sqrt{\frac{1}{M}\sum_{n=1}^{M}(t(n)-f(n))^2} \quad (12)$$

Where, n denotes the sample, M denotes the total no. of samples, t(n) and f(n) denote the nth sample of the true values and the forecasted values.

Similarly, in the TI, the accuracy metric is considered as shown in the following equation:

$$\text{Accuracy in } \% = \frac{(TP+TN)}{(TP+FP+TN+FN)} * 100 \quad (13)$$

TP: true positive; TN: true negative; FP: false positive; FN: false negative;

$$\text{Two-Norm error} = \sqrt{\sum_{n=1}^{M}\sum_{k=1}^{3}\left|v_{n,k}^{estimated} - v_{n,k}^{actual}\right|^2} \quad (14)$$

Where,
M: no. of nodes; $v_{n,k}^{estimated}$: $n^{th}node\ k^{th}$ phase estimated voltage magnitude; $v_{n,k}^{act}$: n^{th} node k^{th} phase true voltage.

D. Evaluating the performance of the proposed approach for forecasting missing measurements

1) Performance of proposed approach for the modified IEEE 37 node system without considering RES integration: This study assesses the efficiency of the proposed approach in forecasting missing measurements, utilising the modified IEEE 37 node system. Different percentages of the missing data that include 10%, 30%, and 50% are considered for the evaluation of the proposed approach. Various deep learning models, including Bi-LSTM, TCN, LSTM, CNN, and KNN are utilised as comparative models to evaluate the effectiveness of the proposed approach. To ensure homogeneity, the deep learning models are trained to utilise the same parameters, including the number of hidden layers, optimizer, learning rate and loss function. To mitigate the effects of randomness in the training of the deep learning model, a series of experiments are conducted and repeated 10 times. The results are then averaged and presented in Table I. From Table I, it is very clear that the proposed approach performance is dominating over the comparative models for the different missing percentages. Considering the 10% missing percentage, the best RMSE is achieved by the proposed approach and the large RMSE is achieved by the KNN model. With this RMSE of more than 4, it becomes evident that KNN faces limitations in accurately imputing missing measurements. This is consistent with the inherent characteristics of KNN: its reliance on proximity-based imputation which can falter when confronted with complex patterns or substantial data gaps. At the same time, the RMSE of the proposed approach is 1.3143 and this dominance of the proposed approach is mainly due to its self-attention mechanism that enables a larger receptive field, and the ability to interpret more asynchronous relationships between dynamic and complex measurement data. Whereas from the comparative models, the second best is achieved by the TCN model because of its spatial as well as temporal feature extraction capability. However, even with this advantageous characteristic, the TCN model still encountered challenges in effectively adapting to the dataset's altered dynamics. As a result, while its performance is good, it couldn't outperform the proposed approach in terms of RMSE.

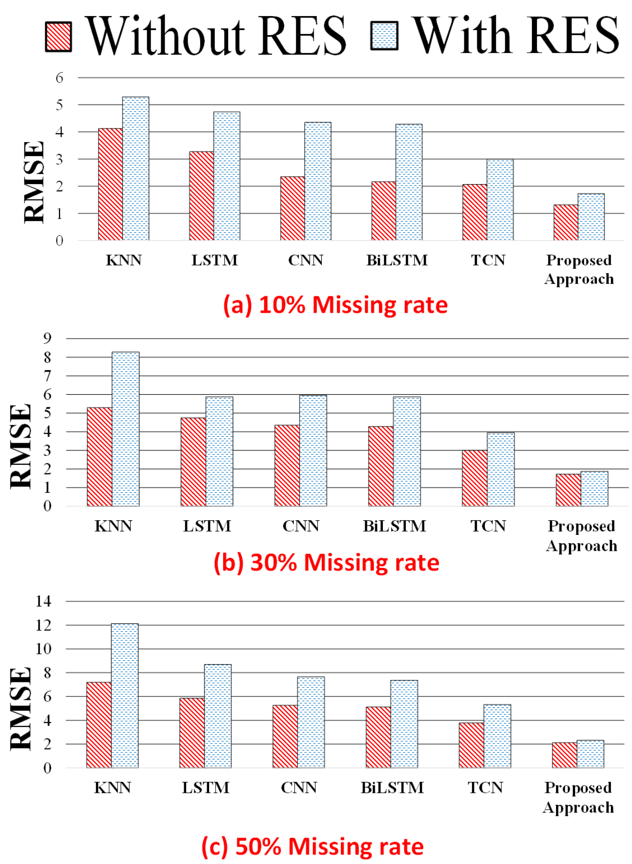

Fig. 4. Forecasting error of the proposed approach and the comparative models for 10%, 30% and 50% missing data in the modified IEEE 37 node system with and without RES integration

The third best is achieved by the Bi-LSTM model with its ability to leverage both forward and backward sequences

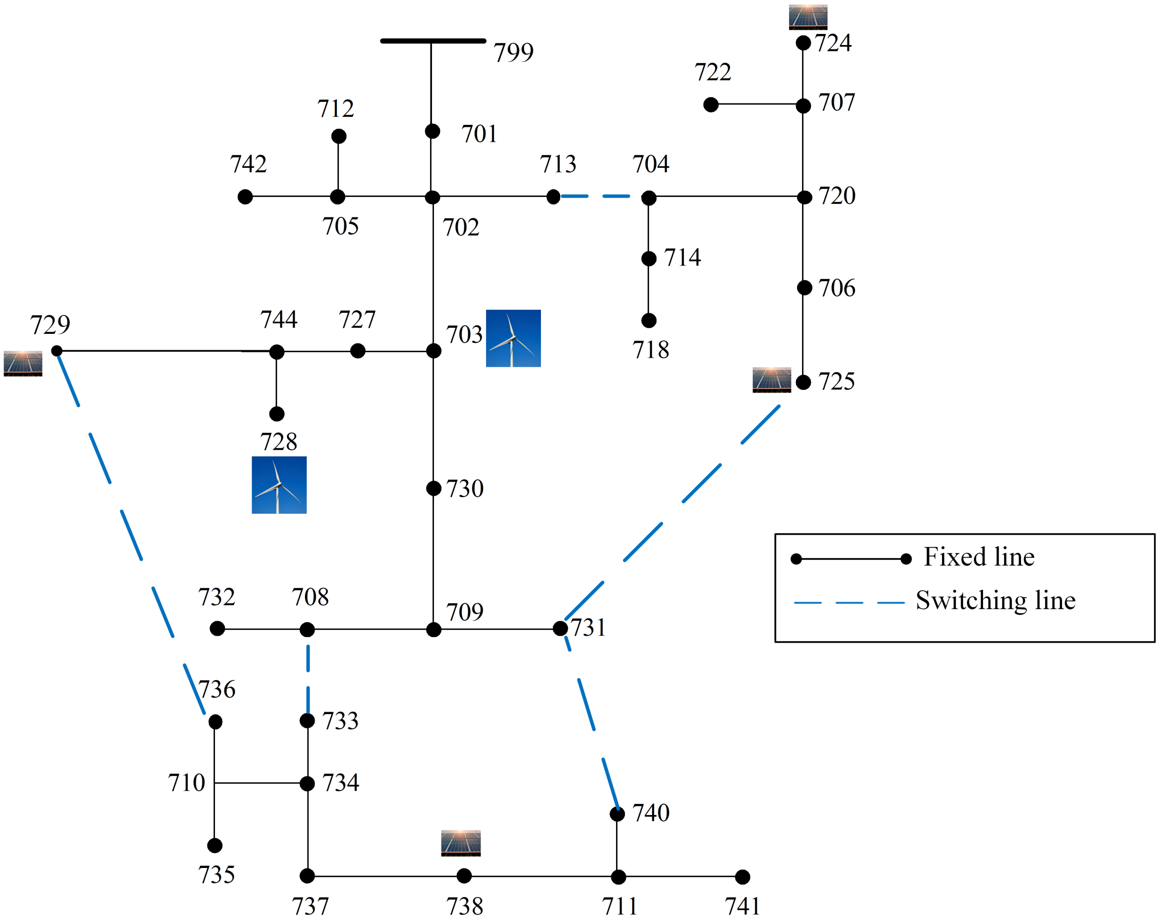

Fig. 3. Modified IEEE 37 node system single line diagram

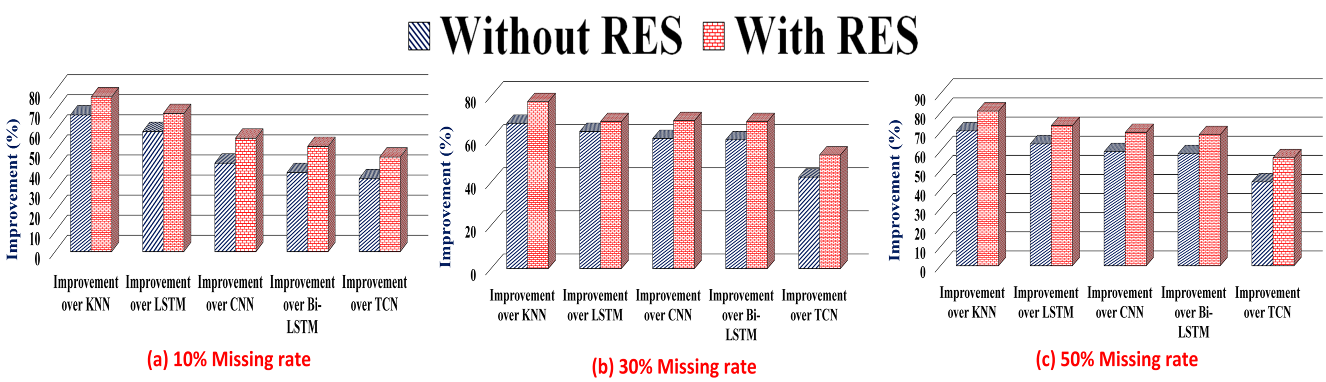

Fig. 5. Improvement of proposed approach over the comparative models for 10%, 30% and 50% missing data modified IEEE 37 node system with and without RES integration

TABLE I

RESULTS OF THE PROPOSED AND COMPARATIVE MODELS FOR THE MISSING MEASUREMENT IMPUTATION FOR THE MODIFIED IEEE 37 NODE SYSTEM

Model	**RMSE for different missing percentages**					
	Without RES			With RES		
	10%	**30%**	**50%**	**10%**	**30%**	**50%**
KNN	4.1242	5.2912	7.1982	5.9312	8.2721	12.1349
LSTM	3.2702	4.7377	5.8495	4.3523	5.8723	8.6924
CNN	2.3508	4.3488	5.2569	3.1233	5.9521	7.6521
BiLSTM	2.1672	4.2804	5.1229	2.8521	5.8683	7.3652
TCN	2.0629	2.9948	3.7817	2.5679	3.9459	5.3247
Proposed Approach	1.3143	1.7235	2.1245	1.3576	1.8654	2.2211

holds the potential for capturing intricate relationships within the data. Considering the proposed and the TCN model, the proposed approach achieved an improvement of 36% over the second-best model and 39% over the third-best model for the 10% missing data. Similarly considering the 30% missing data, the proposed approach achieved an improvement of 49% over the TCN model and 62% over the Bi-LSTM model. In both the 10% and 30% missing rates, the CNN and LSTM models occupied the fourth and fifth positions, respectively. The graphical representation of the error metrics is presented through a bar diagram in Fig. 4. For instance, the proposed approach forecasted measurements vs the true measurements of the three phases of node 6 for the random 150 hours from the test data is shown in Fig. 6. The improvement of the proposed approach over the other models is presented in Table II and through the bar chart representation in Fig. 5. The results indicate that the proposed approach outperformed the other models in the presence of missing rates of 10%, 30%, and 50%. One notable observation is that the proposed approach exhibits a significant increase in improvement percentage over the second-best model, ranging from 10% to 50% in the presence of missing data. This indicates that the proposed approach is well-suited for larger systems. The missing positions in the measurement data are compensated by utilising highly accurate forecasted measurements. The data is utilised by the topology classifier to make predictions regarding the system's topology.

2) Performance of proposed approach for the modified IEEE 37 node system considering RES integration: With the penetration of RES, the voltage amplitudes change more which impacts the performance of the predictive models. This RES penetration has led to alterations in the three statistical attributes of voltage amplitudes, notably impacting the coherence of voltage drops. Thus, the forecasting of the measurements accurately becomes more intricate. Hence, to evaluate the proposed approach's effectiveness, RES like WT, and PV are integrated at some random nodes in the modified IEEE 37 node system. Comparing forecasting metrics in Table I, it is evident that the proposed approach consistently outperformed other methods even with RES integration. The superior performance of the proposed transformer-based approach can be attributed to its inherent capabilities in effectively handling the challenges introduced by the integration of RES in the modified IEEE 37 node system. The RMSE values of reference models significantly increased when RES are integrated. By integrating RES into the modified IEEE 37 node system, the dynamics of voltage data distribution change and could lead to variations in the relationships between different variables, potentially impacting the efficacy of model's performance. For example, when missing rates of 10%, 30%, and 50% are introduced, the KNN model's imputation becomes less accurate due to the evolving data landscape shaped by the integration of RES. This results in higher RMSE values. The substantial RMSE increases of 44%, 56%, and 69% for the respective missing rates underscore the intricate interplay between KNN's imputation approach and the changing dynamics introduced by RES. Similarly, the second-best model, TCN, saw RMSE increases of 24%, 32% and 41% for the three missing rates. The reason for the drop in performance is due to the introduction of additional complexity and potential irregularities in the data by the RES, which can impact the model's ability to learn consistent patterns. Without a mechanism to address this, the TCN model failed to effectively adapt to the fluctuations introduced by RES, leading to less accurate forecasts. These performance drops are attributed to voltage characteristic changes, especially in voltage drop consistency due to RES integration. In contrast, the proposed approach remained robust and displayed consistent performance across all missing rates. Specifically, considering the proposed approach, the RMSE values registered only marginal increases: 3%, 8%, and 9% for the respective missing rates. Comparing the improvement indices, the proposed approach achieved a high improvement of 77%, 78% and 80% over the KNN model, and an improvement of 47%, 53% and 56% over the second-best TCN model, for the three missing rates respectively. The comparison of the predicted and the actual measurement values at node 724 with the PV integration is shown in Fig. 7, and at node 703 with the WT integration is shown in Fig. 8. The pivotal strength of the proposed approach lies in the incorporation of a self-attention mechanism that endows the proposed model with the ability to dynamically interpret temporal relationships. This mechanism facilitates the adaptive concentration on evolving voltage patterns, profoundly influenced by the integration of RES. By effectively capturing these intricate relationships, the proposed approach surpasses traditional models like KNN and even advanced counterparts like TCN in effectively forecasting the missing values. This imputed data is utilized by the topology classification algorithm for estimating the topology of the system.

E. Topology classification from the recompensed measurement data

1) Evaluation for the modified IEEE 37 node system without RES integration for TI: The recompensed measurement data obtained from the previous step is utilised as input for the topology classifier, which is responsible for identifying the network topology. The classifier is trained to utilise k-fold validation methodology to mitigate potential biassing issues. Table III represents the topology classification results derived from the recompensed measurement data of the proposed approach and comparative approaches for system with RES integration and without RES integration.

The accuracy of the ideal scenario, where no data is missing, is also included in the Table III. Based on the data presented in Table III for the absence of RES case, it is apparent the recompensed measurement data obtained by the proposed approach has resulted in the highest accuracy in the TI. The performance of the classifier using KNN-imputed data consistently falls below 85% across all missing rates. This decline in accuracy can be attributed to the inherent limitations of KNN when generating forecasts. The imprecision in KNN's forecasting results in deviations in the imputed data, adversely affecting the overall TI performance. TCN imputed measurement data yields improved accuracy compared to other models. Notably, the accuracy percentages for the three missing rates are significant: 94.05%, 92.40%, and 90.77%. This enhanced accuracy can be attributed to TCN's proficiency in capturing temporal dependencies and patterns present in the data. This, in turn, augments the

TABLE II
IMPROVEMENT OF THE PROPOSED APPROACH OVER THE COMPARATIVE MODELS FOR 10, 30, AND 50% MISSING RATES FOR THE MODIFIED IEEE 37 NODE SYSTEM

	Without RES			With RES		
Missing rates	**10%**	**30%**	**50%**	**10%**	**30%**	**50%**
Improvement over KNN (%)	68.13	67.43	70.49	77.11	77.45	81.70
Improvement over LSTM (%)	59.81	63.62	63.68	68.81	68.23	74.45
Improvement over CNN (%)	44.09	60.37	59.59	56.53	68.66	70.97
Improvement over Bi-LSTM (%)	39.35	59.74	58.53	52.40	68.21	69.84
Improvement over TCN (%)	36.29	42.45	43.82	47.13	52.73	58.29

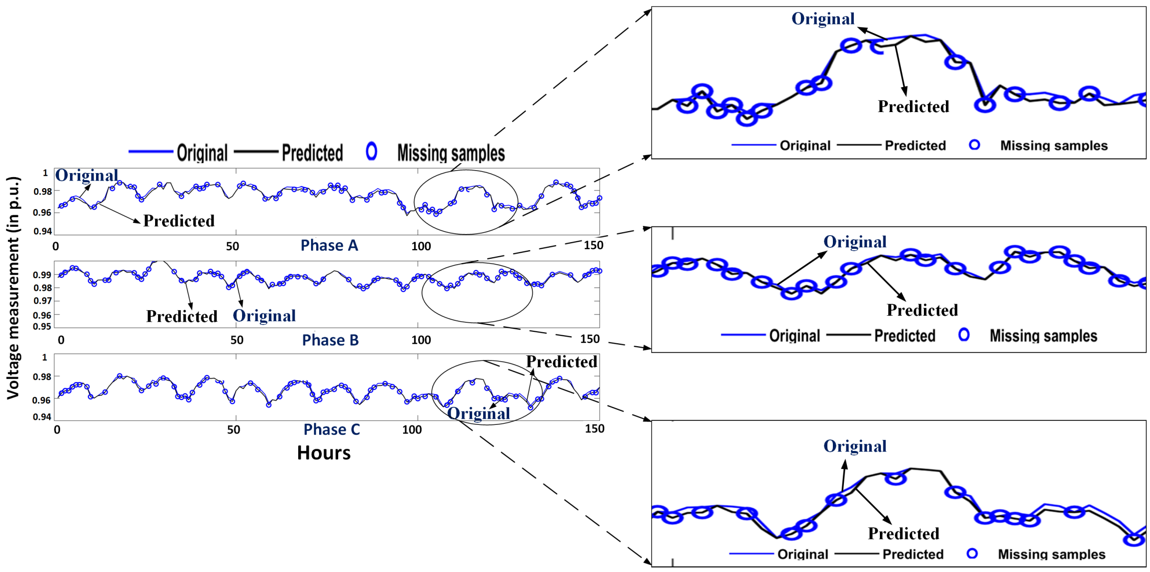

Fig. 6. Comparison of true measurements vs the forecasted measurements by the proposed approach for the three phases of node 6 of modified IEEE 37 node system

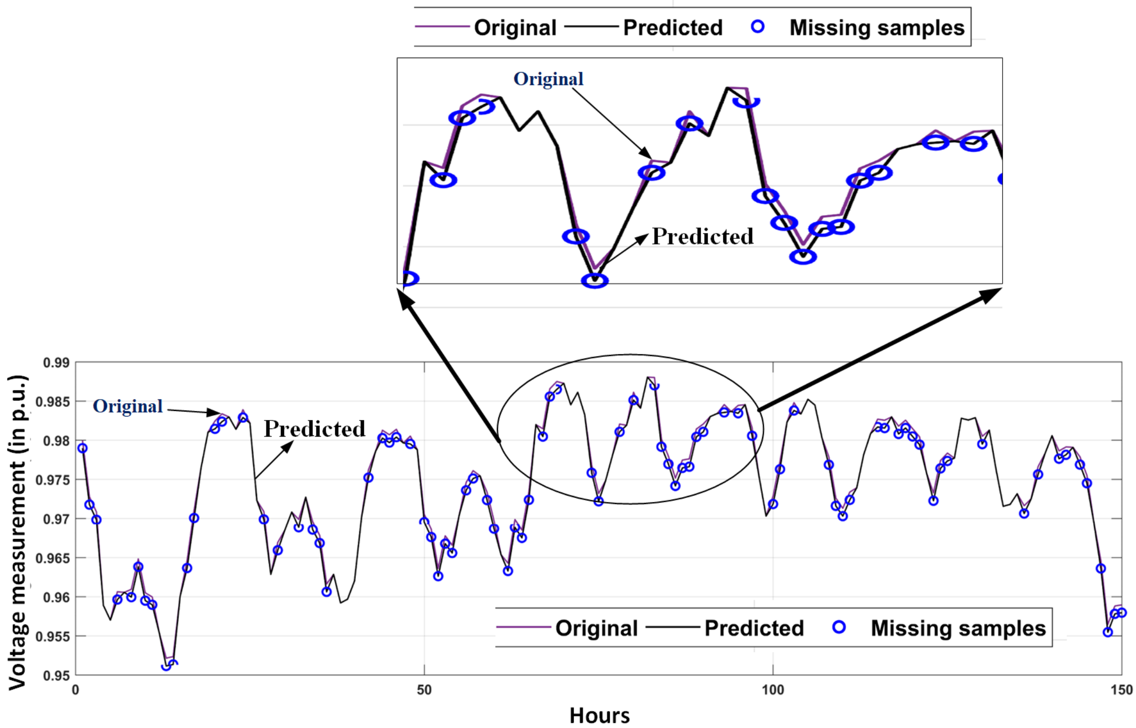

Fig. 7. Comparison of true measurements vs the forecasted measurements by the proposed approach at the node 724 with PV integration in the modified IEEE 37 node system

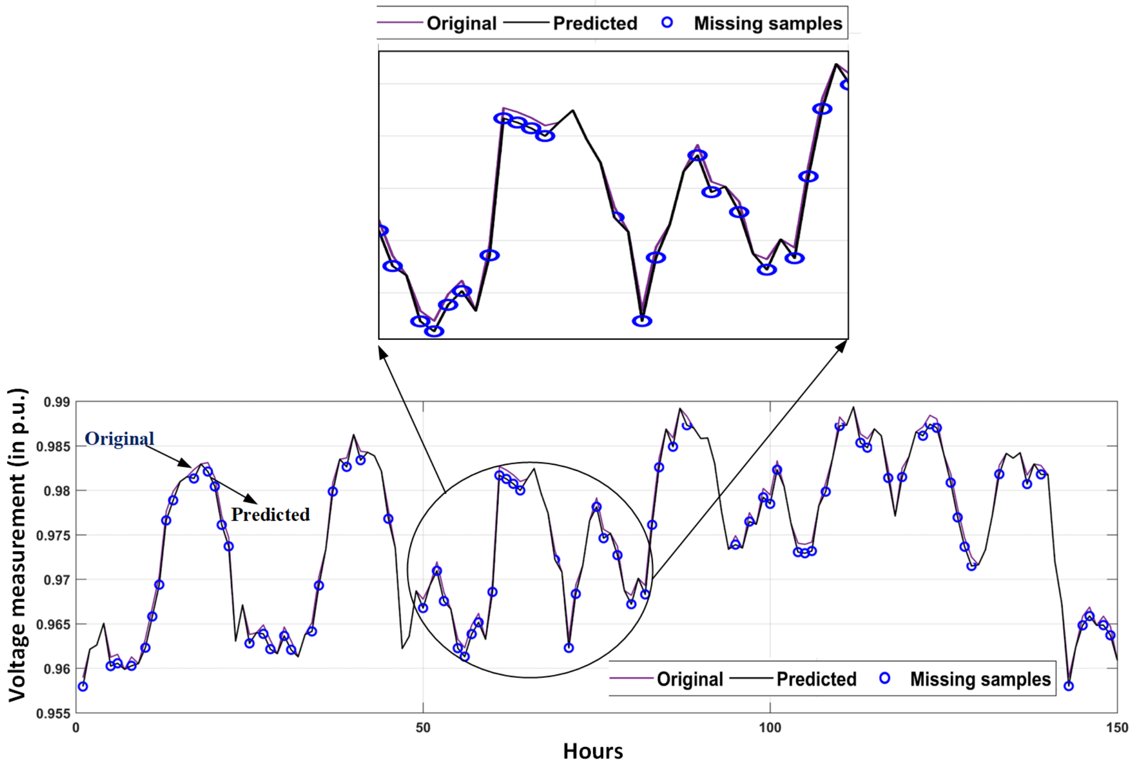

Fig. 8. Comparison of true measurements vs the forecasted measurements by the proposed approach at the node 703 with WT integration in the modified IEEE 37 node system

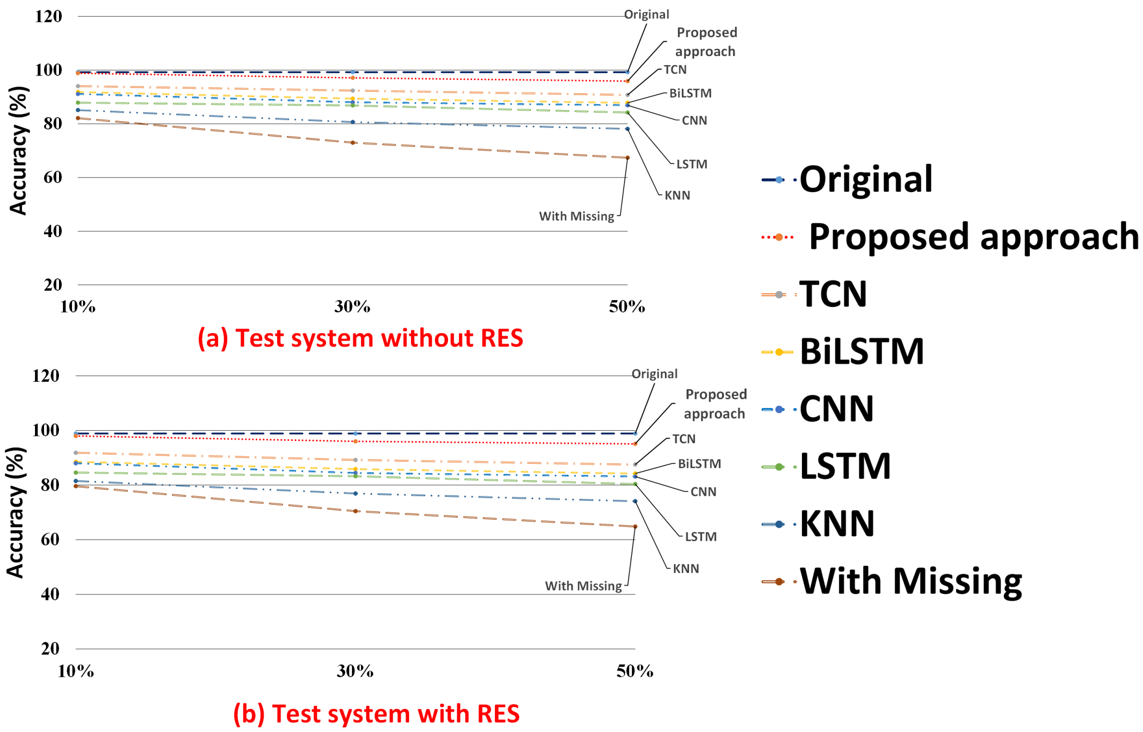

Fig. 9. Variation of the accuracy of the TI with the proposed approach and the other models recompensed measurement data for the modified IEEE 37 node system

TABLE III
COMPARISON OF ACCURACY OF THE TI WITH THE RECOMPENSED MEASUREMENT DATA FROM THE PROPOSED APPROACH AND THE OTHER MODELS FOR THE MODIFIED IEEE 37 NODE SYSTEM

	Without RES			With RES		
Models	**10%**	**30%**	**50%**	**10%**	**30%**	**50%**
Original	99.25	99.25	99.25	98.89	98.89	98.89
Proposed approach	98.87	97.14	95.92	97.98	96.02	95.11
TCN	94.05	92.40	90.77	91.52	89.87	88.24
Bi-LSTM	91.86	89.41	87.81	89.33	86.88	85.28
CNN	91.12	88.00	86.90	88.59	85.47	84.37
LSTM	87.89	86.85	84.23	85.36	84.32	81.70
KNN	85.13	80.67	78.12	82.60	78.14	75.59
With missing	82.12	72.98	67.36	79.60	70.46	64.83

quality of imputed data, leading to a more accurate classifier. However, with the recompensed measurement data from the proposed approach, the accuracy is higher than the other comparative models. For instance, the accuracies are 98.87%, 97.14%, and 95.02% for three missing rates. There is not much variation in the original accuracy without any missing samples. This remarkable performance of the proposed approach can be attributed to the effective imputation of measurement data, which stems from the inherent self-attention mechanism embedded within the proposed model.

2) Evaluation for the modified IEEE 37 node system with RES integration for TI: From Table III, the accuracies decreased significantly for the modified IEEE 37 node system because of the RES integration. In comparison to KNN-imputed measurement data, the accuracy is notably lower across all missing rates. This drop in accuracy is particularly striking when considering the difference between the modified IEEE 37-node system without RES and with RES. This decrease is considerably more significant than what's observed in other models. Imputed measurement data from various comparative models, like LSTM, CNN, Bi-LSTM, and TCN, exhibit relatively better accuracy compared to the KNN model. Among these models, TCN's imputed data yields the best accuracy for the TI. Importantly, the accuracy reduction between the modified IEEE 37 node system without RES and with RES remains comparatively modest for the TCN model. This indicates that TCN effectively handles RES-related variations, leading to consistent accuracy. Yet, the promising performance comes from the proposed approach's imputed measurement data. Here, the accuracy decline is minimal, outperforming other comparative models. This showcases the robustness of the proposed approach. Despite the influence of RES and missing rates, the imputed data from the proposed approach consistently resulted in higher accuracy levels.

Fig. 9 illustrates the accuracy variation across the three missing rates for both cases. Notably, the accuracy level with the proposed approach imputed data closely approaches the original accuracy when no measurements are missing, underscoring the precision of the proposed methodology. Comparatively, the accuracy of the comparative models decreases when evaluated with imputed measurement data. This decline is particularly notable for missing rates of 30% and 50%, signifying the challenges these models face in accurately compensating for missing data. In contrast, the proposed model achieved the best performance with its robust capability consistently outperforming the other models. In summary, the proposed model demonstrates exceptional resilience, showcasing its superior performance across modified IEEE 37 node systems, both with or without RES.

TABLE IV
RESULTS OF THE PROPOSED AND COMPARATIVE MODELS FOR THE MISSING MEASUREMENT IMPUTATION FOR IEEE 69 NODE SYSTEM

Models	**10%**	**30%**	**50%**
KNN	5.5530	9.7251	11.9756
LSTM	4.9426	8.8086	7.7223
CNN	6.7460	7.5671	8.6087
BiLSTM	5.6507	6.7302	8.1099
TCN	5.1551	7.9678	7.9617
Proposed approach	**3.2768**	**4.1356**	**5.1829**

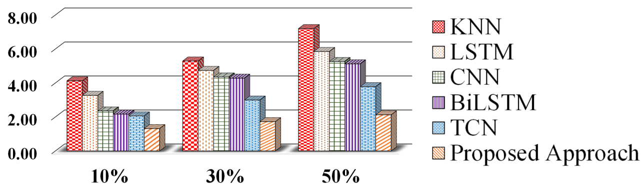

Fig. 10. Forecasting error of the proposed approach and the comparative models for 10%, 30% and 50% missing data for the IEEE 69 node system

F. Impact of the incorrect TI on the state estimation

To underscore the significance of accurate TI, a comprehensive investigation using state estimation is conducted to illustrate the consequences of topology misidentification. This examination involved inducing a fault in branch 702-713, leading to the isolation of certain nodes within the system. Consequently, line 725-731 is switched on to establish connections among all nodes, thereby altering the system's topology. This transformation underscores the criticality of precisely predicting topology changes, a task effectively

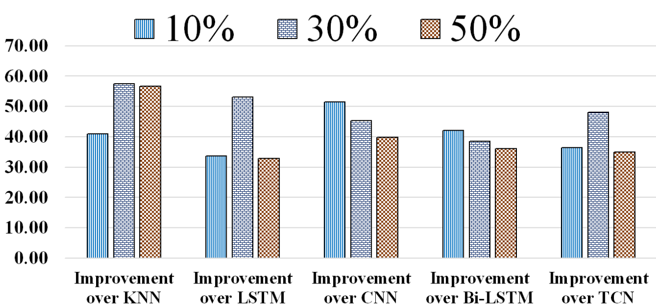

Fig. 11. Improvement of proposed approach over the comparative models for 10%, 30% and 50% missing data for the IEEE 69 node system

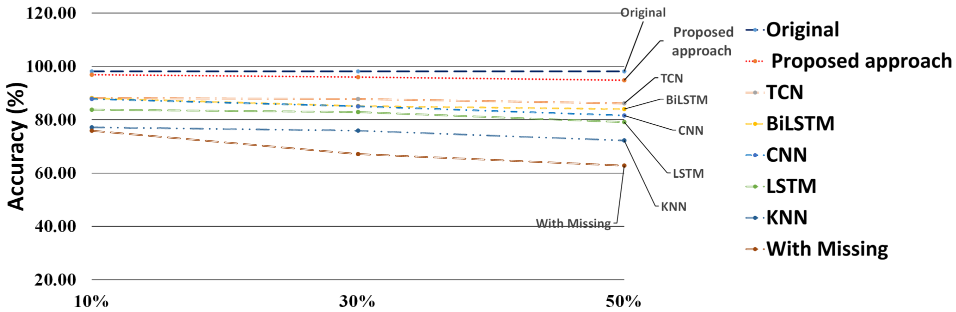

Fig. 12. Variation of the accuracy of the TI with the proposed approach and the other models recompensed measurement data for the IEEE 69 node system

TABLE V
IMPROVEMENT OF THE PROPOSED APPROACH OVER THE COMPARATIVE MODELS FOR 10, 30, AND 50% MISSING RATES FOR IEEE 69 NODE SYSTEM

Missing Percentages	10%	30%	50%
Improvement over KNN	40.99	57.47	56.72
Improvement over LSTM	33.70	53.05	32.88
Improvement over CNN	51.43	45.35	39.79
Improvement over Bi-LSTM	42.01	38.55	36.09
Improvement over TCN	36.44	48.10	34.90

TABLE VI
COMPARISON OF ACCURACY OF THE TI WITH THE RECOMPENSED MEASUREMENT DATA FROM THE PROPOSED APPROACH AND THE OTHER MODELS FOR IEEE 69 NODE SYSTEM

Models	10%	30%	50%
Original	98.12	98.12	98.12
Proposed approach	96.89	95.99	94.81
TCN	88.12	87.77	86.12
Bi-LSTM	88.04	85.12	83.98
CNN	87.78	85.00	81.58
LSTM	83.76	82.87	79.11
KNN	77.13	75.89	72.21
With missing	75.87	67.11	62.77

addressed by the proposed TI model. With this new topology, the state estimator should get the correct topology information in real-time scenarios, ensuring effective state estimation. The results of the state estimation in terms of two-norm error (Eq. 14), considering both the incorrect topology and the accurate topology obtained from the preceding step, are presented in Table IV.

TABLE VII
ERROR COMPARISON BETWEEN CORRECT AND INCORRECT TI FOR THE MODIFIED IEEE 37 NODE SYSTEM

Type	Magnitude	Phase
With correct TI	6.1×10^{-2}	0.52×10^{-2}
With incorrect TI	20.3×10^{-2}	4.23×10^{-2}

Based on the data presented in Table IV, a notable observation emerges: when the topology is accurately predicted using the proposed TI, the corresponding error is substantially lower compared to the error associated with an incorrect topology assumption. This results underscores the critical importance of accurate TI in achieving accurate state estimation results.

G. Scalability test of the proposed approach for the larger scale system

The modified IEEE 69 node system is considered in this study to demonstrate the performance of the proposed approach for larger scale power system. The results of the experiments are presented through three key tables: Firstly, Table IV delineates the metrics employed for missing value imputation, accompanied by graphical representation in Fig. 10. Secondly, Table V vividly portrays the significant improvement percentages achieved by the proposed approach when benchmarked against reference models, further elucidated in Fig. 11. Lastly, Table VII offers valuable insights into the accuracy of TI leveraging recompensed measurement data from different models, with accompanying visualizations of accuracy variations in Fig. 12. This comprehensive assessment sheds light on the efficacy of the approach in the context of larger and more complex power systems. According to Table IV, the proposed approach outperforms all comparative models, demonstrating lower error metrics. Notably, the RMSE values for various missing percentages are 3.2768, 4.1356, and 5.1829. This represents a substantial improvement, with an average improvement of 40% compared to the second-best TCN model, as detailed in Table V. Furthermore, the accuracy achieved with the recompensed measurement data from the proposed approach consistently exceeds 94.8% for all cases. From Figure 12, it is evident that the accuracy variation closely approximates the accuracy obtained with the original data. In contrast, the accuracy of all other models deviates significantly from the original data, showing a pronounced decrease. This robust performance underscores the suitability of the proposed approach for larger-scale power systems.

V. CONCLUSIONS

This study introduces a novel approach with the attention mechanism-based transformer model to mitigate the impact of DoS attacks on the accuracy of TI by forecasting missing measurements. The first step of the proposed methodology involves forecasting missing measurements and generating the recompensed measurement data. The subsequent step involves providing the topology identifier with the recompensed data to get the accurate topology. The experimental methodology involved assessing the efficiency of the proposed approach under different conditions of missing data, specifically at rates of 10%, 30%, and 50%, and utilizing various reference models for the modified IEEE 37 node system with and without RES integration. The proposed approach consistently exhibited dominant performance in both scenarios. Furthermore, proposed approach's effectiveness is also evaluated on a higher-order modified IEEE 69 node system. The experiments yielded compelling results, showcasing that the proposed approach led to a substantial improvement of over 35% across all three missing rates in the initial phase, for both with and without RES integration and two distinct systems. Additionally, the recompensed measurement data obtained from this approach resulted in heightened accuracy for TI, closely approaching the ideal scenario where no measurements are missing. These results underscore the capability of the proposed approach to deliver desirable TI accuracy while displaying resilience to missing measurements, even in the context of higher-order power systems.

REFERENCES

[1] V. Telukunta, J. Pradhan, A. Agrawal, M. Singh, and S. G. Srivani, "Protection challenges under bulk penetration of renewable energy resources in power systems: A review," *CSEE journal of power and energy systems*, vol. 3, no. 4, pp. 365–379, 2017.

[2] V. Freitas and A. S. Costa, "Integrated state & topology estimation based on a priori topology information," in *2015 IEEE Eindhoven PowerTech*. IEEE, 2015.

[3] S. S. Alam, B. Natarajan, and A. Pahwa, "Distribution grid state estimation from compressed measurements," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1631–1642, 2014.

[4] D. Deka, S. Backhaus, and M. Chertkov, "Structure learning in power distribution networks," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1061–1074, 2017.

[5] G. Cavraro, V. Kekatos, and S. Veeramachaneni, "Voltage analytics for power distribution network topology verification," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 1058–1067, 2017.

[6] G. Cavraro *et al.*, "Distribution network topology detection with time-series measurements," in *2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2015.

[7] N. Duan and E. M. Stewart, "Deeplearningbased power distribution network switch action identification leveraging dynamic features of distributed energy resources," *IET Generation, Transmission & Distribution*, vol. 13, no. 14, pp. 3139–3147, 2019.

[8] D. Gotti, H. Amaris, and P. L. Larrea, "A deep neural network approach for online topology identification in state estimation," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5824–5833, 2021.

[9] S. Sun, "Optimal linear filters for discrete-time systems with randomly delayed and lost measurements with/without time stamps," *IEEE Transactions on Automatic Control*, vol. 58, no. 6, pp. 1551–1556, 2012.

[10] X. Xu *et al.*, "A novel hybrid cnn-lstm compensation model against dos attacks in power system state estimation," *Neural Processing Letters*, pp. 1–25, 2022.

[11] L. Zhao *et al.*, "Robust pca-deep belief network surrogate model for distribution system topology identification with ders," *International Journal of Electrical Power & Energy Systems*, vol. 125, p. 106441, 2021.

[12] P. M. Broersen, "Modified durbin method for accurate estimation of moving-average models," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 5, pp. 1361–1369, 2009.

[13] J. Che and H. Zhai, "Wt-arima combination modelling for short-term load forecasting," *IAENG Int. J. Comput. Sci*, vol. 49, pp. 542–548, 2022.

[14] M. T. Ismail and R. S. Al-Gounmeein, "Overview of long memory for economic and financial time series dataset and related time series models: A review study." *IAENG International Journal of Applied Mathematics*, vol. 52, no. 2, pp. 261-269, 2022.

[15] Z. Wang and Y. Lou, "Hydrological time series forecast model based on wavelet de-noising and arima-lstm," in *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. IEEE, 2019.

[16] Huang Sen, "Time Series Prediction based on Improved Deep Learning," *IAENG International Journal of Computer Science*, vol. 49, no. 4, pp. 1133-1138, 2022.

[17] L. Demidova and M. Ivkina, "Development and research of the forecasting models based on the time series using the random forest algorithm," in *2020 2nd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA)*. IEEE, 2020.

[18] A. Tokgöz and G. Ünal, "A rnn based time series approach for forecasting turkish electricity load," in *2018 26th signal processing and communications applications conference (SIU)*. IEEE, 2018.

[19] X. Li *et al.*, "Time-series production forecasting method based on the integration of bidirectional gated recurrent unit (bi-gru) network and sparrow search algorithm (ssa)," *Journal of Petroleum Science and Engineering*, vol. 208, p. 109309, 2022.

[20] T. Liang *et al.*, "Wind speed prediction based on multi-variable capsnet-bilstm-mohho for wpccc," *Energy*, vol. 250, p. 123761, 2022.

[21] W. N. Al-Sharu, A. M. Alqudah, S. Qazan, and A. Alqudah, "Detection of valvular heart diseases using fourier transform and simple cnn model," *IAENG International Journal of Computer Science*, vol. 49, no. 4, pp. 985–993, 2022.

[22] M. Turuk, R. Sreemathy, S. Kadiyala, S. Kotecha, and V. Kulkarni, "Cnn based deep learning approach for automatic malaria parasite detection," *IAENG Int. J. Comput. Sci*, vol. 49, pp. 745–753, 2022.

[23] S. Mehtab and J. Sen, "Analysis and forecasting of financial time series using cnn and lstm-based deep learning models," in *Advances in Distributed Computing and Machine Learning: Proceedings of ICADCML 2021*. Springer, 2022.

[24] Venkatesh, Siddhanth U Hegde, Satish B Basapur, and Nagaraju Y, "DistilBERT-CNN-LSTM Model with GloVe for Sentiment Analysis on Football Specific Tweets" *IAENG International Journal of Computer Science*, vol. 49, no. 2, pp. 420-432, 2022.

[25] H. Zhang, W. Cui, T. Shi, Y. Tao, and J. Zhang, "ATMLP: Attention and Time Series MLP for Fall Detection," *IAENG International Journal of Applied Mathematics*, vol. 53, no. 1, pp. 58-65, 2023.

[26] D. M. Madhiarasan, P. Roy, and P. Pratim, "Hybrid transformer network for different horizons-based enriched wind speed forecasting," *arXiv preprint arXiv:2204.09019*, 2022.

[27] "Nrel data," https://www.nrel.gov/, accessed: June 9, 2023.