Temporal Convolutional Network-based Locational Detection of False Data Injection Attacks in Power System State Estimation

V.S.N. Kalyani
Department of Electrical Engineering
NIT Andhra Pradesh
Tadepalligudem, India
kalyanivandanapu18@gmail.com

S. Naga Geetha
Department of Electrical Engineering
NIT Andhra Pradesh
Tadepalligudem, India
geethasnjr@gmail.com

Kiran Teeparthi
Department of Electrical Engineering
NIT Andhra Pradesh
Tadepalligudem, India
kiran.t39@nitandhra.ac.in

Y Raghuvamsi
Dept. of Electrical Engineering
NIT Andhra Pradesh
Tadepalligudem, India
raghuvamsi.sclr@nitandhra.ac.in

Y Srimannarayana
Department of Electrical Engineering
NIT Andhra Pradesh
Tadepalligudem, India
srimannarayanayekkirala@gmail.com

***Abstract*— In power system state estimation, it is important to detect and remove false data (FD) from measurements to ensure accurate estimation of the system state. The locational detection of false data injection attacks (FDIAs) is crucial for the timely and effective mitigation of the attack. It depends on the analysis of the measurement data at different nodes of the transmission system. The practical execution of deep learning (DL) models is used to decide the exact locations of FD in real time. In this paper, the FDIA detection will be addressed by employing the temporal convolution neural network (TCN) model in the classification mode and its performance is compared with other state-of-the-art DL models. A TCN has the ability to capture the spatial as well as temporal dynamics of the data in the transmission system and this feature can be used to identify the FDIAs. The TCN can be trained on historical data to understand the normal behavior of the system and then to detect any deviations from this normal behavior, which could indicate an FDIA. The FDIAs include structured and unstructured attacks which can be detected with the TCN model in association with the traditional bad data detection (BDD) algorithm. Further, the location of data intrusions can also be identified among the measurement data. The simulation studies are carried out on the IEEE 118-bus system by employing multilabel classification.**

Keywords*— *False Data Injection Attack, Deep Learning, Temporal Convolutional Network, Bad Data Detection

I. INTRODUCTION

Power system plays a crucial role in providing reliable, safe, and efficient electrical energy to support consumers. The arrival of information and communication technology in the conventional power system paves the way for real-time monitoring and control, which leads to improved reliability and efficiency but also poses new security challenges. To mitigate these security risks, it is important to implement robust cybersecurity measures such as encryption, firewalls, and intrusion detection systems. State estimation (SE) uses data from measurement devices to estimate the current state of a system. Compromised system state estimation can potentially disrupt power system operations. Liang et al. [1] and Deng et al. [2] conducted extensive reviews on the effects of cyberattacks on SE. These surveys covered various impacts such as line congestion, power outages, and communication blocks. One type of cyberattack that targets power system SE involves injecting FD into measurement devices. This attack is known as an FDIA, which aims to compromise the accuracy and reliability of the estimation process. FDIAs can be designed in a way that evades detection by conventional BDD in modern SCADA systems. This makes FDIAs one of the most difficult threats to address when it comes to state estimation. An instance of the economic impact of false data injection was demonstrated in [3] through the induction of transmission line congestion. Numerous studies have been conducted to explore potential methods for building FDIA, as outlined in [2]. As an example, the authors in [4] presented a covert assault that illustrates the ability of this type of falsified information to evade detection by the control center's BDD.

Simultaneously, considerable research has been dedicated to mitigating FDIA, which can be broadly categorized into two groups: physical-based defensive approaches [5]–[7] and data-dependent identification techniques [8]–[14]. Gai et al. [6] employed dynamic programming as a means to achieve the highest possible level of privacy protection for devices with limited resources, resulting in an optimal solution. Yang et al. [7] provide an instance of this by compromising the least number of sensors for data manipulation, and a defensive technique against FDIA is developed with an algorithm for optimal PMU placement. Different methods that rely on data have been suggested to explore the problem of detecting FDIA, such as the utilization of maximum-likelihood estimation [8], Gaussian mixture distribution methods [9], sparse optimization [10], Kalman filters [11], similarity matching [12], and network theory [13]. In [12], the authors have utilized generation schedules, load forecast data, and synchro phasor information to obtain a statistical representation of the differences between forecast-based predictions and SCADA-based state estimates in order to identify outliers. The efficacy of many existing approaches is heavily reliant on understanding the attack model and having access to power system structure with the assumption that power grids are nearly chordal sparse. In recent times, there have been proposals for data-driven detection techniques using deep learning. These methods differ from traditional

approaches by allowing the system to learn attack and power system models directly from training data, rather than relying on pre-defined algorithms. Although the current methods detect the existence of a malicious attack, it is also an inevitable and crucial aspect to identify the attack location, which is necessary for rapidly deploying effective countermeasures. Location identification captures inconsistency and co-occurrence dependency, which in turn, creates additional opportunities for improving the performance of presence detection.

This article aims to fill the gap by proposing a deep-learning mechanism for detecting the location of FDIA. Specifically, the problem is defined as a multilabel classification task and the solution to this problem involves a single architecture that combines the DL approach namely, temporal convolutional network (TCN) and a standard BDD algorithm. The BDD algorithm is utilized to eliminate bad data such as unstructured FDIAs (outliers), while the TCN functions as a classifier with multiple labels and it captures the co-occurrence dependency and inconsistency introduced by FDIA.

The paper structure is arranged as follows: Section II describes the basics of the power system SE and FDIAs. Section III presents the mechanism of the proposed method and TCN model architecture. Section IV demonstrates the simulations experiments of the detection of FDIA locations with different parameters of the TCN model. Finally, Section V provides the conclusion of the article.

II. Preliminary

A. Power System State Estimation (PSSE)

SE is an algorithm of determining the current state of the power system by processing the data available from measurement devices by excluding the noise present in the measurements. This paper focuses on PSSE in DC micro-grids rather than AC considering the advantages of the former like high reliability, easier control, and highly efficient when integrating with renewable energy sources (RES). The mathematical expression which relates the measurement data (z) and state estimates (x) is shown below:

$$z = Hx + e \quad (1)$$

Here 'H' is the mapping matrix connecting measurements/Jacobian matrix and 'e' is the noise vector.

The traditional BDD algorithm detects bad data based on the following condition:

$$\|z - Hx\|_2^2 \geq \tau \quad (2)$$

where τ is a predefined threshold value. Based on the above condition, the BDD algorithm indicates whether there exist any bad/compromised measurements and the presence of an attack.

B. False Data Injection Attack

The motivation behind an FDIA is to corrupt the measurements and manipulate the estimates of the state. Let the corrupted set of measurements $\hat{z} = z + a$ led to a false set of state estimates $\hat{x} = x + c'$ where $c' \neq 0$. Based on the BDD algorithm, the ℓ_2 norm is given by:

$$\|\hat{z} - H\hat{x}\| = \|(z + a) - H(x + c')\| = \|z - Hx\| \leq \tau \quad (3)$$

For the attack to avoid the BDD condition, attack vector (a) should be $a = Hc$. In practice, the unstructured 'a' vector gets easily detected by BDD. However, the structured FDIAs are formulated as $a = Hc$, which can bypass the BDD. Since the system information is highly secured and confidential, accessing the entire information of H is not possible which involves high cost and effort. The authors in [5] proved that a successful structured FDIA can be constructed using a minimum number of meters and limited information about system parameters and the corresponding attacker's cost of the resources can be effectively solved as a min-cut problem.

This article proposes a data-directed mechanism that can identify the location of false data injection. This type of problem is considered as a multi-label classification problem whose goal is to find out the compromised meters.

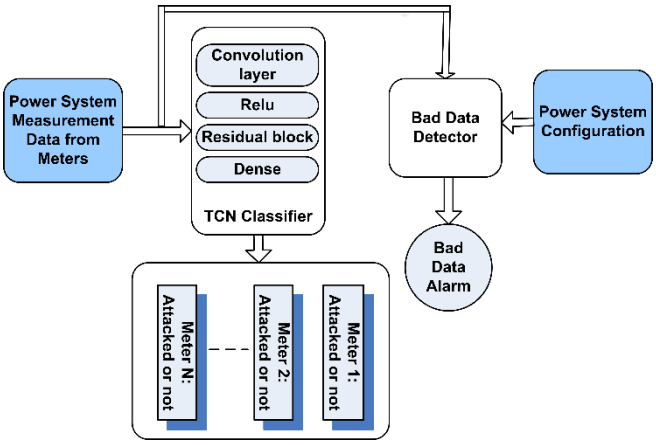

Fig. 1 FDIA locational detection mechanism

III. Proposed Methodology of Locational Detection

The mathematical task of detecting FDIA can be viewed as a classification approach where the measurement vector is classified into two categories - existence or non-existence of FDIA. In the case of a single output variable, the problem can be categorized as a simple classification problem with a single label. Since the PSSE involves a large number of measurements and identifying the attack location among these measurement data is carried out by the classification task on each measurement which is divided into two classes, thus finally a multilabel classification problem is formulated. Multilabel classification is a complex and widely applicable problem that has attracted much research interest. Even though the DL models have become highly successful in single-label classification, the complexity of classification for multiple classes remains challenging. These problems can be processed with a multitude of quality actions, which are often contradictory in nature. Moreover, the classes of multilabel classification approaches are typically very unbalanced, compared to the single-label classification approach and thus the problem is likely ineffective. To tackle this problem, an effective DL model is required and hence a temporal convolutional network (TCN) is developed as shown in Fig. 1, which can extract spatio-temporal features from the associated data information for obtaining satisfactory performance in the multilabel classification task.

Furthermore, the performance of the TCN model is evaluated with different layers and parameters.

A. *Methodology*

In the domain of transmission system state estimation, this paper proposes a method for identifying false data through the use of the TCN model. These networks are a form of deep learning architecture that excels at processing sequential data. One of the key strengths of TCNs is their ability to incorporate casual network convolutions, ensuring that past information does not influence future predictions. Additionally, TCNs are capable of handling variable input sizes, similar to RNNs, while also looking further into the past to make predictions, much like LSTMs. This unique combination of features makes TCNs highly effective, with auto-regressive properties and long-term memory. Additionally, TCNs are parallelizable due to their efficient memory usage and flexible input size.

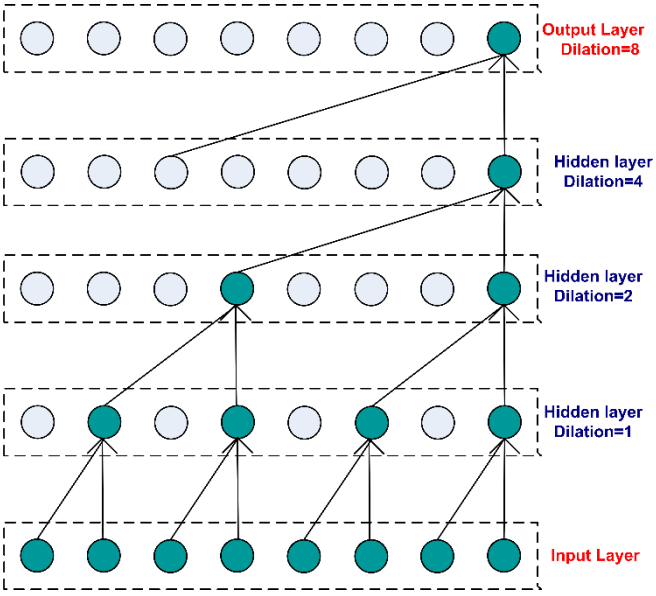

Fig. 2 Dilated causal convolution with dilation rates (1, 2, 4, 8)

To implement casual convolutions, TCN uses a 1D fully convolutional network (FCN) and dilated convolutions. Dilated convolutions, proposed by [15], allow a large receptive field [16], making it easier to apply casual convolutions in sequence tasks, particularly when observing back into the past. dilated convolutions are an operation that uses an input sequence and a filter to produce an output element. The flow of data through a dilated causal convolution with dilation rates of 1, 2, 4, and 8 is shown in Fig. 2. To ensure network stability, the TCN's receptive field depends on the parameters such as filter size, depth, and dilatation rate. Further, the TCN layers have a number of extraction filters which are fed to a residual module of convolution layers to achieve network stability. Fig. 3 depicts the residual block which consists of rectified linear unit (ReLU) activation, dilated convolutional layers, dropout layer applied on each dilated convolution, and weighted normalization applied to convolutional kernels. This approach ensures the network stability required to implement casual convolutions and achieve effective prediction.

The input to the dilated TCN is a sequence of measurements and it outputs a probability score for each measurement being false. A binary cross-entropy loss function is employed for each output and the model has been optimized using stochastic gradient descent. So, the dilated TCN was trained on a dataset of simulated power system measurements, which includes a variety of false data injection attack patterns at different locations. A baseline method had also been trained using a convolutional neural network (CNN) with a similar architecture. Then the performance of the dilated TCN and CNN was evaluated on a test set of simulated measurements with different types and levels of false data injection attacks.

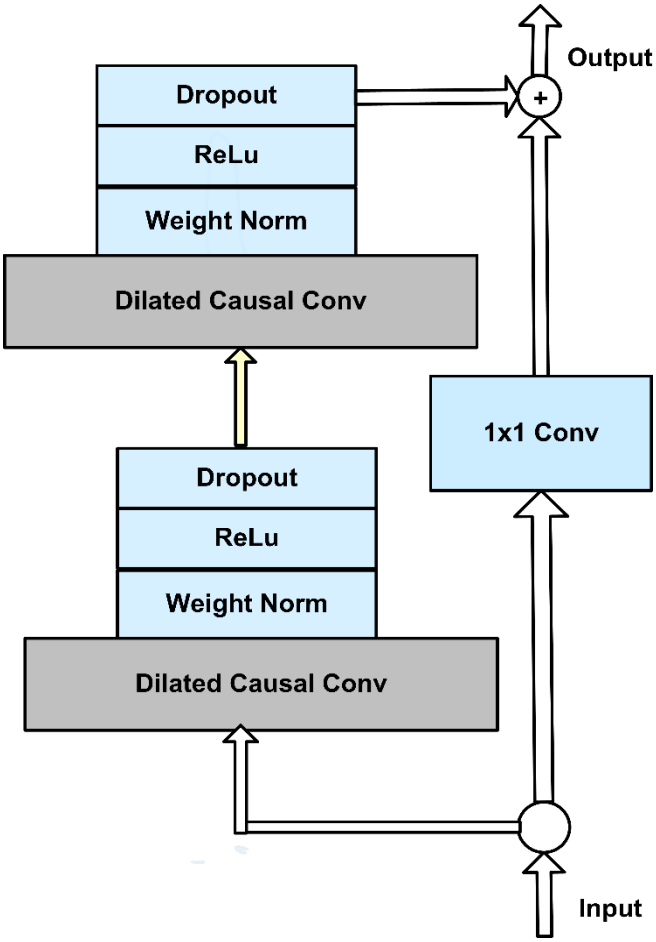

Fig. 3 TCN Residual block

The dilated causal convolution function over subsequent layers can be structured as given below:

$$x_h^t = f\left(\sum_{k=0}^{k-1} w_h^k * x_{(h-1)}^{(t-(k*d))} + b_h\right) \quad (4)$$

The output of a neuron in layer *h* at time *t*, denoted as x_h^t, is computed using a convolutional kernel of width *k* and a weight w_h^k corresponding to position *k*. The convolution operation is modified by a dilation factor *d*. Additionally, a bias term b_h is added to the output. The ReLU activation function is used, which means that the output of each neuron is equal to the maximum of '0' and the sum of its input weights and bias. The output of TCN is classified as

$$y_j^t = \begin{cases} 1, & \textit{the meter } j \textit{ at time } t \textit{ is attacked} \\ 0 & \textit{otherwise} \end{cases} \quad (5)$$

The network consists of a dense layer that is fully connected to *n* nodes in the output layer. The output layer is activated using the sigmoid function for the classification of each measurement. The output of multilabel classification is achieved as follows:

$$y_j^t = sigmoid(w_D \times x_t + b_D) \quad (6)$$

This is how a neural network is used to classify the type of measurements taken by a power meter. The output layer is a dense layer which is fully connected to *n* nodes, and the sigmoid function is used to squash the output of the dense layer between 0 and 1, allowing us to interpret it as the probability of each label.

B. Training

To implement the proposed scheme of detection of FDIA location using a dilated temporal convolutional neural network, an architecture is constructed consisting of TCN (Temporal Convolutional Network) layers followed by a dense layer with sigmoid activation. The TCN layers are designed with varying numbers of filters and dilation rates, and the ReLU activation function for the layers in this model is selected. The training procedure is as follows:

- The output from each TCN layer is passed as input to the next layer, resulting in a sequence of output tensors.
- Further, the output of the final TCN layer flattens into a one-dimensional vector followed by a dense layer with the specified number of dense units and sigmoid activation.
- The model is created with the input layer as its input and the dense layer as its output.
- The optimal learning parameter set is sought by introducing a loss function to get the difference between the target and the actual output within every mini-batch.
- As the proposed structure is extended to multilabel classification, the cross-entropy function is chosen as the loss function of the proposed dilated temporal convolutional neural network. The cross-entropy loss function over a mini-batch $\theta = \{t_1,\ldots,t_{200}\}$ is expressed as:

$$y(\theta) = \sum_{t\in\theta} \frac{-1}{n} \sum_{i=1}^{n} \{y_i^t \log(y_i^t) + ((1-\hat{y}_i^t)\log(1-y_i^t))\} \quad (7)$$

- With the loss function clearly defined, the Adam optimizer with an initial learning rate of 0.001.
- During the fitting process, the mini-batch gradient descent technique is utilized in the training process to avoid overfitting, and for getting better convergence. Each mini-batch consists of 100 instances of data with a random selection from the training data while calculating the gradient. The entire data is separated into a training set with 8/10 of the data and a validation set with 2/10 of the data for each batch.

In this way, the learning parameters need to be optimized before classification. This involves filter tuning, weights, and biases in each layer. Overall, it defines a neural network architecture that uses TCN layers with various hyperparameters to process input data and produce output for multilabel classification. The number of layers along with their shapes and parameters of the proposed TCN model are represented in Table I.

TABLE I. PARAMETERS OF PROPOSED TCN MODEL
(Total parameters: 6,183,860)

Layer	Output Shape	Parameters
Input_1	(180, 1)	0
TCN_1	(180, 128)	247168
TCN_2	(180, 256)	1115904
TCN_3	(180, 128)	377728
TCN_4	(180, 128)	295680
Flatten	(None, 23040)	0
Dense	(None, 180)	4147380

IV. SIMULATION EXPERIMENTS

This section utilized a temporal convolutional network to identify the co-occurrence dependency and inconsistency caused by FDIAs among the measurements in the IEEE 118-bus power system. The simulations have been carried out using MATLAB environment for obtaining the measurement data from the power flow algorithm. The machine learning (ML) and deep learning (DL) models have been implemented using the Keras framework.

A. Dataset Preparation

The system consists of 118 buses, 186 transmission lines, and a total of 180 measurements are considered. Out of these measurements, 70 are injection measurements which measure the power being injected into the power system at certain buses, and 110 are flow measurements, which measure the flow of power along the transmission lines. To create the dataset, meter measurements were organized based on the network topology using the proposed FDIA locational detection mechanism. Instead of relying on CNN, which analyzes the adjacent meter measurements to obtain features, temporal convolutional neural networks were employed. The line flow meters were indexed from *p=1* by first indexing the unindexed meters connecting bus *p* and incrementing *p*. The indexing process was ended once *p* exceeds 118; otherwise, the process turned back to 1. The line meters were then indexed, and the injection meters were labeled using the ascending order of the bus index. The following are the steps for the preparation of data:

i) Base Load: Initially, uncompromised data were gathered by extending real-time data and each bus load is generated artificially by randomly collecting samples from a normal distribution such that the baseload is chosen as the mean with standard variance selected as 1/6 of the baseload [9], [18]. Later, the compromised data were generated with partial system knowledge using the min-cut FDIA model developed in [17], considering only stealthy FDIAs since non-stealthy FDIAs can be easily detected by the conventional BDD algorithm, which is part of the proposed framework.

ii) Attack Implementation: The compromised data were obtained using the min-cut FDIA model such that the adversary is having limited system knowledge, which

does not involve much cost and effort for getting the information of transmission line parameters. A discrete uniform distribution with (2,10) is selected for choosing the number of target state variables. As discussed in [17], the information about a particular line impedance is obtained in a similar way and accordingly the cost was selected. The injection data's 2-norm is changed from 1 to 5 randomly for the experiments.

iii) Measurement Error: Additionally, random Gaussian noise was appended to the measurement values to account for dynamic errors encountered from the meters and the communication. The error's standard derivation for all the measurements is taken as 0.1 to 0.5.

iv) Preparation of training and testing datasets: A training set consists of 10,000 stealthy FDIA instances and 1,00,000 unattacked instances are generated for each level of attack. Whereas the testing dataset is generated for evaluating the model's performance by using 500 stealthy FDIA instances and 500 unattacked instances. Ten sets of testing data which are independent of each other are averaged to obtain each value presented in this section.

B. Locational Detection Performance

The performance of the proposed model is evaluated using the standard classification metrics which are given below:

$$Accuracy = \frac{TPR+TNR}{TPR+TNR+FPR+FNR} \quad (8)$$

$$Precision = \frac{TPR}{TPR+FPR} \quad (9)$$

$$Recall = \frac{TPR}{TPR+FNR} \quad (10)$$

$$F1-score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (11)$$

In the above formulae, TPR - True Positive Rate, FPR - False Positive Rate, FNR - False Negative Rate, and TNR – True Negative Rate. TPR is the probability of an FDIA location being correctly categorized as FDIA, FPR is the probability of a non-FDIA location being incorrectly labeled as an FDIA location, and FNR is the probability of an FDIA location being incorrectly labeled as non-FDIA. Additionally, the F1-score was also employed to balance the precision and recall, which is the geometric mean of both parameters.

V. RESULTS AND DISCUSSIONS

In this section, the performance of the TCN model is analyzed with a different number of layers and dilation rates. Once the optimized TCN model is achieved, its performance is compared with other ML/DL models.

From Table II, it can be observed that with the highest values of accuracy and precision, the four-layer TCN model outperforms the other models with an accuracy of 99.67% and a precision of 99.64%. However, it is clearly observed that as the number of layers increases (> 4), the model's performance decreases due to the overfitting problem. Therefore, it is crucial to select the appropriate number of layers in a TCN model for achieving the best performance while keeping the model complexity in check.

TABLE II. PERFORMANCE METRICS FOR A TCN MODEL WITH DIFFERENT NUMBERS OF LAYERS

Structure	Accuracy (%)	Precision (%)	Recall (%)	F1-score
TCN (1 layer)	93.36	92.56	81.48	86.66
TCN (2 layers)	95.17	94.82	86.52	90.48
TCN (3 layers)	99.31	96.48	89.44	92.83
TCN (4 layers)	**99.67**	**99.84**	**98.80**	**99.63**
TCN (5 layers)	99.47	99.19	95.78	97.45
TCN (6 layers)	99.54	99.14	95.76	97.42

As the number of layers in the TCN model increases, there is an overall improvement in the performance metrics. The accuracy metric increases from 93.36% with a single layer to a maximum of 99.67% with four layers, and then slightly decreases with additional layers. Similarly, the precision, recall, and F-score metrics also show an increasing trend with an increase in the number of layers, reaching maximum values for the four-layer TCN model.

The performance of the TCN model with 4 layers for different dilation rates is evaluated and corresponding metrics are mentioned in Table III.

TABLE III. PERFORMANCE METRICS AND TOTAL PARAMETERS OF TCN MODEL WITH 4 LAYERS FOR DIFFERENT DILATION RATES

TCN (Dilation rates)	Accuracy (%)	Precision (%)	Recall (%)	F1-score	Total parameters
TCN (1, 2)	98.87	92.02	84.85	88.29	5,494,452
TCN (1, 2, 4)	**99.67**	**99.84**	**98.8**	**99.63**	**6,183,860**
TCN (1, 2, 4, 8)	96.45	85.78	81.23	83.44	6,873,268

From Table III, the results show that increasing the dilation rate for all layers (TCN (1, 2, 4)) results in the best performance with an accuracy of 99.67%, a precision of 99.84%, recall of 98.8%, and F-score of 99.63%. When the dilation rates are 1 and 2 for the layers (TCN(1, 2)), there is a decrease in the performance metrics compared to the TCN (1, 2, 4) model, with an accuracy of 98.87%, precision of 92.02%, recall of 84.85%, and F1-score of 88.29%. However, the TCN (1, 2, 4) model has higher total parameters (6,183,860) than the TCN(1, 2) model indicating that increasing the dilation rate can improve the performance by increasing the model's complexity. When the TCN model is having dilation rates of 1, 2, 4, and 8 (TCN(1, 2, 4, 8)), there is a decrease in performance, with an accuracy of 96.45%, a precision of 85.78%, recall of 81.23%, and F1-score of 83.44%. This model also has the highest number of total parameters (6,873,268), indicating that the increase in parameters may not necessarily result in better performance. Thus the increase in parameters may lead to an overfitting problem which reduces the model's effectiveness. Overall, these results suggest that increasing the dilation rate for all TCN layers can improve the performance of the model. However, increasing the dilation rate beyond a certain point can lead to diminishing returns and

may not improve the model's performance due to overfitting issues. Hence, the number of parameters needs to be considered when selecting the dilation rates to balance model complexity and performance. Based on the above analysis, the proposed TCN model is selected with 4 layers and dilation rates of (1, 2, 4).

Table IV presents a comparison of the performance of the proposed TCN model with Multi-Layer Perceptron (MLP), CNN, k-Nearest Neighbors (kNN), and Decision Tree (DT) algorithms for the IEEE 118-bus system, using four metrics. The results show that TCN outperforms the other comparison models and the metrics indicating the effectiveness of the proposed mechanism. It is important to note that the superior accuracy achieved by the TCN structure is attributed to its ability to capture the co-occurrence dependency and inconsistency of adjacent measurements caused by FDIA. This highlights the significance of the proposed TCN structure in addressing the challenges posed by FDIA. The performance metrics represented in Table IV are depicted using a bar chart as shown in Fig. 4.

TABLE IV. COMPARISON OF PERFORMANCE METRICS BETWEEN DIFFERENT MODELS

Structure	Accuracy (%)	Precision (%)	Recall (%)	F1-score
DT	95.38	89.41	91.62	90.50
kNN	95.32	92.09	93.84	92.95
MLP	96.46	97.62	91.96	94.70
CNN	97.52	99.26	94.35	96.74
TCN (1, 2, 4)	**99.67**	**99.84**	**98.8**	**99.63**

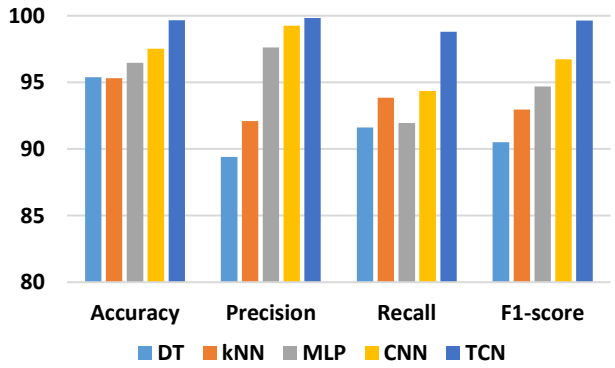

Fig. 4 Comparison of performance metrics of the proposed TCN model with other ML/DL models

V. CONCLUSION

In this paper, the locations of false data injection attacks in power system measurements are identified using a classification approach of multiple classes with BDD-TCN architecture. The TCN was able to seize the co-occurrence dependency and inconsistency introduced by FDIA, resulting in a model-free mechanism. The existing BDD system was utilized to build the architecture, requiring no alteration of the existing BDD algorithm, and the runtime of the detection process is very fast once the model is trained. Furthermore, extensive simulations were carried out on IEEE 118-bus system with different dilation rates and layers of TCN to obtain more efficient and effective performance. The dilated TCN outperformed a baseline method using convolutional neural networks in terms of locational detection accuracy. It was suggested that the dilated TCNs with a proper selection of layers and dilation rates have the potential to improve the accuracy and reliability of false data injection attack detection in power systems.

REFERENCES

[1] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A Review of False Data Injection Attacks Against Modern Power Systems," in *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630-1638, July 2017.

[2] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, "False data injection on state estimation in power systems—Attacks, impacts, and defense: A survey," IEEE Trans. Ind. Informat., vol. 13, no. 2, pp. 411–423, Apr. 2017.

[3] L. Xie, Y. Mo, and B. Sinopoli, "Integrity data attacks in power market operations," IEEE Trans. Smart Grid, vol. 2, no. 4, pp. 659–666, Dec. 2011.Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[4] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," ACM Trans. Inf. Syst. Security, vol. 14, no. 1, pp. 1–33, Jun. 2011.

[5] S. Bi and Y. J. Zhang, "Using covert topological information for defense against malicious attacks on DC state estimation," IEEE J. Sel. Areas Commun., vol. 32, no. 7, pp. 1471–1485, Jul. 2014.

[6] K. Gai, K.-K. R. Choo, M. Qiu, and L. Zhu, "Privacy-preserving content-oriented wireless communication in Internet-of-Things," IEEE Internet Things J., vol. 5, no. 4, pp. 3059–3067, Aug. 2018.

[7] Q. Yang, D. An, R. Min, W. Yu, X. Yang, and W. Zhao, "On optimal PMU placement-based defense against data integrity attacks in smart grid," IEEE Trans. Inf. Forensics Security, vol. 12, no. 7, pp. 1735–1750, Jul. 2017.

[8] R. Moslemi, A. Mesbahi, and J. M. Velni, "A fast, decentralized covariance selection-based approach to detect cyber attacks in smart grids," IEEE Trans. Smart Grid, vol. 9, no. 5, pp. 4930–4941, Sep. 2018.

[9] S. A. Foroutan and F. R. Salmasi, "Detection of false data injection attacks against state estimation in smart grids based on a mixture Gaussian distribution learning method," IET Cyber Phys. Syst. Theory Appl., vol. 2, no. 4, pp. 161–171, 2017.

[10] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on the power grid by sparse optimization," IEEE Trans. Smart Grid, vol. 5, no. 2, pp. 612–621, Mar. 2014.

[11] K. Manandhar, X. Cao, F. Hu, and Y. Liu, "Detection of faults and attacks including false data injection attack in smart grid using Kalman filter," IEEE Trans. Control Netw. Syst., vol. 1, no. 4, pp. 370–379, Dec. 2014.

[12] A. Ashok, M. Govindarasu, and V. Ajjarapu, "Online detection of stealthy false data injection attacks in power system state estimation," IEEE Trans. Smart Grid, vol. 9, no. 3, pp. 1636–1646, May 2018.

[13] Y. Guan and X. Ge, "Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks," IEEE Trans. Signal Inf. Process. Netw., vol. 4, no. 1, pp. 48–59, Mar. 2018.

[14] J. J. Q. Yu, Y. Hou, and V. O. K. Li, "Online false data injection attack detection with wavelet transform and deep neural networks," IEEE Trans. Ind. Informat., vol. 14, no. 7, pp. 3271–3280, Jul. 2018.

[15] A. v. d. Oord, S. Dieleman, et al., "Wavenet: A generative model for raw audio". arXiv preprint arXiv:1609.03499, 2016.

[16] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions". arXiv preprint arXiv:1511.07122, 2015

[17] S. Wang, S. Bi, and Y. -J. A. Zhang, "Locational Detection of the False Data Injection Attack in a Smart Grid: A Multilabel ClassificationApproach," in IEEE Internet of Things Journal, vol. 7, no. 9, pp. 8218-8227, Sept. 2020.

[18] H. Sedghi and E. Jonckheere, "Statistical structure learning to ensure data integrity in smart grid," IEEE Trans. Smart Grid, vol. 6, no. 4, pp. 1924–1933, Jul. 2015.