2025 IEEE 19th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG) | 979-8-3315-1517-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/CPE-POWERENG63314.2025.11027278

Long Short-Term Memory-Based Anomaly Detection for State Estimation in Smart Grid Considering Anomalous Database

Kazem Haghdar
Department of Electrical Engineering
Sharif University of Technology
Tehran, Iran
haghdar@ee.sharif.edu

Ali Abbaspour Tehrani-Fard
Department of Electrical Engineering
Sharif University of Technology
Tehran, Iran
abbaspour@sharif.edu

Fei Wang
Department of Electrical Engineering
North China Electric Power University
Baoding, China
feiwang@ncepu.edu.cn

.

Mahmud Fotuhi-Firuzabad
Department of Electrical Engineering
Sharif University of Technology
Tehran, Iran
fotuhi@sharif.edu

***Abstract*— To achieve real-time state estimation in smart grids, Phasor Measurement Units (PMUs) have been extensively utilized. However, statistics show that PMU data can also contain various types of anomalies. In response to these challenges, anomaly detection has become a key area of research. Many existing studies on anomaly detection rely on databases, but they often assume that the historical data used are free of anomalies. In practice, however, data may be contaminated due to factors such as system detection errors, cyberattacks. Using such contaminated data in studies can significantly increase the vulnerability of the smart grid. Manual inspection of data on a case-by-case basis is both time-consuming and costly. This paper proposes a method for detection of anomalies against state estimation system considering anomalous database. A subset of the database is selected and a hybrid approach combining model-based and data-driven anomaly detection methods is implemented through residual-based analysis and subsequence correlation in both temporal and spatial dimensions of the data series. The data is assessed in terms of its adaptability to the smart grid and its pattern correlation in terms of anomalies identification. After identifying the anomalies, a powerful long short-term memory (LSTM) network is trained. Once the trained model is implemented in remaining parts of database, anomalies are also detected. The proposed method is applied to the IEEE 39-bus network. The results presented demonstrate its ability to successfully detect anomalies in comparison with support vector machine and decision tree algorithms.**

Keywords—anomaly, phasor measurement unit, anomalous data detection, LSTM, nearness, spatial-temporal correlation, smart grid.

I. Introduction

Today, with the development of smart grid, many complications have been added to its control and operation. Acceptable states estimation of grid can be very important and effective in the proper control and operation of the smart grid. The input data for estimating network states are different measurements from different points of the power grid. Traditionally, the Supervisory Control And Data Acquisition (SCADA) systems have been used to measure the voltage magnitude of the buses, the current magnitude of the lines, the power flow of the lines and the injected power to the buses. However, the low accuracy and speed of this system have caused the PMU-based system to be considered. Unlike SCADA, PMU can obtain dynamic measurements. According to IEEE Std C37.118.2-2011, PMU is a device to estimate synchronized phasor, frequency and rate of change of frequency from voltage and/or current signals and a time synchronizing signal with Global Positioning System (GPS) timestamps [1]. This approach leads to a linear relationship between the measurements and other variables of the network state, contrasting with the non-linear relationship in SCADA. As a result, the state estimation speed is significantly higher when utilizing PMUs. Additionally, PMUs offer superior accuracy. The real-time measurements provided by PMUs enable swift anomaly detection and prompt remedial actions in the system. However, data anomalies, whether unintentional, such as sensor errors and their improper functioning, or intentional, like cyberattacks, are inevitable. According to statistics provided by the California independent system operator, the detected anomalies in North American power grid range from approximately 10% to 17% [2] while, this number ranges from 10% to 30% in China [3]. These statistics are significant and emphasize the growing need for the development of robust data anomaly detection systems. Given its importance, this topic has attracted considerable attention from researchers, and two primary approaches have been proposed for designing anomaly detection systems: the first is based on network topology, and the second is data-driven.

In the first category, various methods, such as the Chi-square test and the largest normalized residual test, have been introduced [4-5]. However, these methods are vulnerable under certain conditions. Data-driven approaches, on the other hand, have garnered extensive attention from researchers. In [6], the Kullback-Leibler divergence is used to assess the similarity between two probability distributions, historical measurements and suspicious measurements. Also, in [7], the consistency between predicted and received measurements is evaluated using statistical testing methods. In [8], the Jensen-Shannon divergence is employed to determine the similarity of residual distributions. The majority of research efforts in identifying attacks involving erroneous data have utilized machine learning techniques. In [9], the authors investigated the use of support vector machines and statistical anomaly detection approaches to identify data injection attacks.

Anomaly detection using deep learning is investigated in [10-11]. Reference [10] proposes a detection framework based on conditional deep belief networks. Convolution neural network-based anomaly detection is presented in [13]. In [14], Dynamic Bayesian Networks and learning algorithms based on Boltzmann machines were used to detect undetectable attacks. A method based on Recurrent Neural Networks was introduced in [15]. A deep learning-based framework that combines convolutional neural networks with long short-term memory networks for detecting new FDI attacks was proposed in [16]. Also, [17] proposes a deep space clustering-based detector. Graph recurrent neural network-based approach for detecting attacks in smart grid wireless communication systems is investigated in [18]. Reference [19] proposes a few machine learning methods to detect anomalous data. An autoencoder generative adversarial network to investigate power market data of real-time locational marginal prices is investigated in [20]. Also, reference [21] proposes a combination of long short-term memory neural network and a Fourier transform to detect attacks in automatic generation control.

All of these methods require the use of database (historical data) to enable effective design. These studies assume that database is both completely clean and readily available. However, database may become contaminated due to various factors, such as malfunctioning of the existing detection systems or cyberattacks. In such cases, designing an anomaly detection system based on anomalous data can have a high risk. The database, in addition to designing an online anomaly detection system for smart grid, is also used in many other studies. Therefore, determining the quality of the database is of great importance. If a smart grid is equipped with a powerful anomaly detection system, it can effectively identify anomalies in the database. However, problems arise when the anomaly detection system has deficiencies and fails to accurately detect anomalies. The main objective of this paper is to detect anomalies in the database alongside designing an online anomaly detection system for the smart grid. This paper combines model-based and data-based methods to identify anomalies in the database. In this regard, due to anomalous database, just a portion of the data is initially selected. A combination of the traditional anomaly detection method based on residuals is used along with identifying time-series behavioral patterns at different locations and times. Through correlation, an index called "Nearness" is calculated to provide a better picture of the anomalies in the database. Subsequently, after identifying the anomalies in the selected section, the LSTM network is trained. The trained model is then applied to other remaining parts of database to identify anomalies across the entire database.

The remainder of this paper is structured as follows. Section II describes the structure of anomaly detection. Then, simulation and results are provided in section III. Finally, Section IV concludes the paper.

II. Anomaly Detection

With the expansion of smart grids, data anomalies have been increasing day by day. These anomalies pose a significant threat to the security of smart grids. Consequently, addressing these anomalies has drawn the attention of researchers. By studying various types of anomalies, researchers can investigate methods for their detection and mitigation. In the previous studies, it has been usually assumed that the database is clean. However, in reality, the database may become contaminated for various reasons. Therefore, this paper assumes that the database is contaminated with various types of anomalies. Databases are typically one of the main sources used in various studies. Consequently, their contamination can be catastrophic, as studies based on contaminated data may produce unrealistic results, making the smart grid highly vulnerable. Manual inspection of data on a case-by-case basis is both time-consuming and costly Thus, the primary objective of this paper is to propose a method for detecting anomalies in the smart grid, considering and identifying database contamination. This paper assumes that the studied network, which utilizes PMUs, still employs one of the traditional anomaly detection methods, such as the residual-based method, which needs improvement. While traditional methods have advantages, they also have limitations that necessitate efforts to enhance their performance. In this paper, to design the anomaly detection system, a three-stage method is introduced (Fig. 1). In the first stage, a portion of the database is selected, and the traditional residual-based method is used to identify possible anomalies related to complete replacement of authentic measurements with simulated network values. This evaluates the data in terms of adaptability to the studied network. Next, for data deemed non-anomalous by the residual-based method, their spatiotemporal correlation patterns are identified through

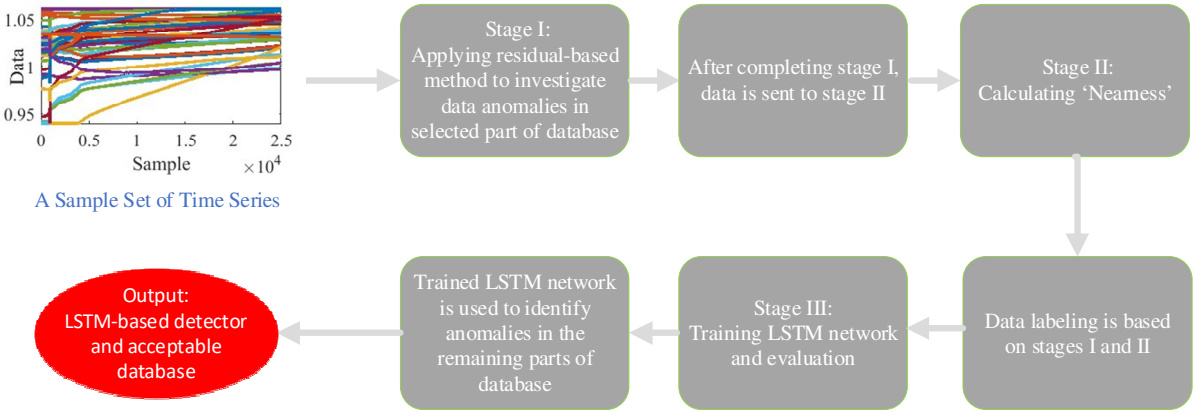

Fig. 1. Flowchart of LSTM-based detector design considering anomalous database

time-series subsequence correlation, which detects anomalies. In the third stage, classified data from the previous stages are used to train an LSTM model. Finally, the trained network can detect possible anomalies in rest of the database. A detailed explanation of these three stages is provided in the following sections.

A. Stage I: Residual-Based Detector

In this paper, it is assumed that this historical data pertains to a network currently in operation and includes a bad data detection system based on a residual threshold. Various measurements from different points across the smart grid are provided to the state estimation system to estimate the network variables. State estimation as core of energy management of smart grid is responsible for obtaining accurate estimates of the states. PMUs directly measure state variables, leading to the linearization of equations and significantly enhancing both the speed and accuracy of state estimation.

$$\boldsymbol{M} - \boldsymbol{H} \times \boldsymbol{V} = ב \tag{1}$$

$\boldsymbol{M}$ and $\boldsymbol{V}$ are measurement data vector and state variables, respectively. ב is measurement noise and random uncertainty and also, H is matrix to relate state variables to measurement. Also, weighted least squares estimator is used to estimate the states of the smart grid:

$$\min_{V} R(V) = [\boldsymbol{M} - \boldsymbol{H} \times \boldsymbol{V}]^t W^{-1} [\boldsymbol{M} - \boldsymbol{H} \times \boldsymbol{V}]$$
$$W = diag\{\sigma_1^2, \sigma_2^2, \dots, \sigma_{N_d}^2\}, \tag{2}$$

Where, R, σ_i and N_d are the measurement residual, standard deviation of i^{th} measurement and number of measurements, respectively. The estimated states V are assumed to be safe as long as (3) is valid [22]

$$\|M - HV\| < r \tag{3}$$

Where, r is a certain threshold. Inequality (3) is capable of detecting many types of anomalies. But in very rare cases with very low probability, it can be bypassed. Assume that by a bad data vector a added to measurement vector $\boldsymbol{M}$, state vector will be $\boldsymbol{V} + \boldsymbol{c}$

$$\|M_{bad} - H\hat{V}_{bad}\| = \|M + a - H(\hat{V} + \mathrm{c})\|$$
$$= \| M - H\hat{V} + (a - Hc) \| \tag{4}$$

As seen if $a = Hc$, anomalous data bypasses the inequality (3). Therefore, malicious adversaries by accessing to H and PMU(s) can bypass (3).

On one hand, access to PMUs, and on the other hand, access to all parameters and the topology of the smart grid, if not impossible, is highly unlikely. Therefore, inequality (3) remains highly effective. This paper proposes residual-based method to evaluate scenarios where the majority of grid data could be replaced by entirely fabricated network data, situations where data-driven approaches may prove inadequate.

B. Stage II: Saptial-Temporal Correlation-Based Detector

When N time series are obtained from PMUs, to identify anomaly patterns in different sections of each time series, it is necessary to compare different sections of each time series with other sections within that series and across other series. Assume that DS is a set of time series represented by $DS = \{TS_1, TS_2, \dots, TS_n, \dots, TS_N\}$ where $TS_i = [x_{i,1}, x_{i,2}, \dots, x_{i,j}, \dots, x_{i,T}]$, $TS_i \in DS$, $x_{i,j}$ is the j^{th} measurement from time series i and T is the total number of measurements in each time series. To analyze the behavior of each section of time series i, we define multiple subsequences of length l from it.

$$Sub_{k_{TS_i}} = [x_{i,k}, x_{i,k+1}, \dots, x_{i,k+l-1}],$$
$$k = 1, 2, \dots, T - l + 1 \tag{5}$$

For each subsequence $Sub_{k_{TS_i}}$, its length normalized Euclidean distance $D\left(\frac{Sub_{k_{TS_i}} - \mu_{Sub_{k_{TS_i}}}}{\sigma_{Sub_{k_{TS_i}}}}, \frac{Sub_{s_{TS_n}} - \mu_{Sub_{s_{TS_n}}}}{\sigma_{Sub_{s_{TS_n}}}}\right)$ with all subsequences of DS is calculated (Fig. 2). The smallest distance, after excluding the self-distance, is defined as the Nearness of $Sub_{k_{TS_i}}$ to the entire DS [23]. For simplicity $P = Sub_{k_{TS_i}}$ and $Q = Sub_{s_{TS_n}}$

$$D = \sqrt{\frac{1}{l}\sum\left(\left(\frac{P - \mu_P}{\sigma_P}\right)^2 + \left(\frac{Q - \mu_Q}{\sigma_Q}\right)^2 - 2\left(\frac{P - \mu_P}{\sigma_P}\right)\left(\frac{Q - \mu_Q}{\sigma_Q}\right)\right)}$$
$$= \sqrt{2\left(1 - \frac{P \odot Q - l\mu_P\mu_Q}{l\sigma_P\sigma_Q}\right)}, \tag{6}$$

Where, μ and σ are mean and standard deviation of each subsequence, respectively. $\odot$ stands for dot product of two subsequence. Nearness will be

$$\text{Nearness}_{k_{TS_i}} = min_{s_{TS_n}}(D) \tag{7}$$

Algorithm 1 Nearness

Input: DS and l
Output: Nearness
1: for n ←1 to N {*every time series in DS*}
2: for k ←1 to $T - l + 1$ {*every subsequence in* TS_i}
3: for n ←1 to N
4: for s ←1 to $T - l + 1$
{*computing distance of* $Sub_{k_{TS_i}}$ *to other subsequences in DS*}
5: $D\left(Sub_{k_{TS_i}}, Sub_{s_{TS_n}}\right)$
6: end
7: end
{*Removing self distance and computing mininum distances of* $Sub_{k_{TS_i}}$ *to other subsequences in DS*}
8: $\text{Nearness}_{k_{TS_i}}$
9: end
10: end

Fig. 2. Simple Presentation of Nearness algorithm

By injecting data anomalies into DS, this Nearness changes. Through experimental analysis and by calculating the Nearness based on the stage I results that has passed through (3), a threshold for anomaly detection can be determined.

C. Stage III: Long Short-Term Memory

Recurrent Neural Networks (RNNs) are presented to setup recurrent connections within hidden layers that allow them to get memory mechanism to remember information from the previous inputs. This mechanism enables RNNs to effectively process and generate sequential data using temporal context. However, RNNs have two major problems during training that are the vanishing and the exploding gradient problems. The vanishing gradient problem occurs when gradients decrease exponentially during backpropagation that make it difficult for the network to obtain long-term dependencies and make correct correlations between distant inputs. In contrast to vanishing gradient, the exploding gradient problem arises when gradients become excessively large that lead to unstable parameter updates and potential model divergence. To tackle these challenges, Long Short-Term Memory (LSTM) networks were introduced [24]. LSTMs consist of memory blocks, each containing a memory cell regulated by three gates: the input gate, forget gate, and output gate. These gates control the flow of information, determining which data should be retained, discarded, or updated. This structure enables LSTMs to obtain relevant information over sequences while discarding irrelevant information [25]. By addressing the vanishing and exploding gradient problems, LSTMs provide a robust solution for learning long-term dependencies, making them highly effective in applications such as time-series classifications. Let $\{X_1, X_2, \dots, X_t, \dots, X_k\}$ denote an input for an LSTM network, where $X_t \in \mathbb{R}^n$ at the t^{th} time step. Forget gate with forgetting threshold f_t determines which information to discard from the previous cell state C_{t-1}. Also, input gate using input threshold i_t decides which information will be added to the cell state. Output gate produces output information using O_t as a threshold

$$f_t = \sigma\left(W_f.[h_{t-1}, X_t] + b_f\right) \quad (8)$$
$$i_t = \sigma(W_i.[h_{t-1}, X_t] + b_i) \quad (9)$$
$$C_t = f_t\,.C_{t-1} + i_t.\widetilde{C_t} \quad (10)$$
$$\widetilde{C_t} = tanh(W_c.[h_{t-1}, X_t] + b_c) \quad (11)$$
$$O_t = \sigma(W_0.[h_{t-1}, X_t] + b_0) \quad (12)$$
$$h_t = O_t.tanh(C_t) \quad (13)$$

Where, W_f, W_i, W_c and W_0 are weights for gates and b_f, b_i, b_c and b_0 are bias terms for gates. Also, h_t and σ are hidden state at time t and sigmoid function, respectively.
In this paper, after labeling the data in the first two stages, the labeled data is used to train the LSTM network. Then, the trained model is applied to the test data for evaluation.

III. Simulation and Results

In this section to validate our proposed method, simulation results are presented. The IEEE 39-bus system shown in Fig. 3 is chosen for this study and 102482 data samples (observations) using MATPOWER are generated in different operating conditions from which 20% of the data is

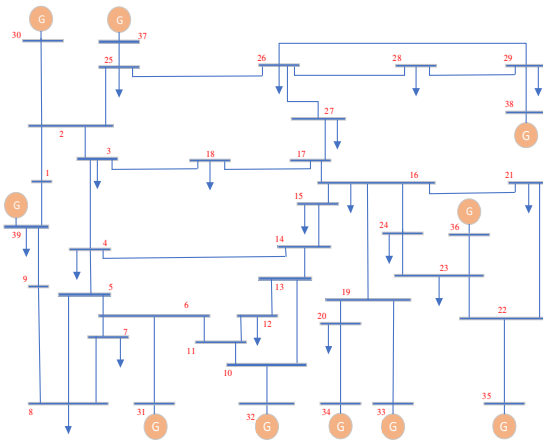

Fig. 3. IEEE 39-bus network

TABLE I
Proposed LSTM Network

Input Layer
2 LSTM Layers (64,64)
2 Dropout Layers (0.2,0.2)
1 Fully Connected Layer
Output Activation Function: Sigmoid
Loss Function: Binary Cross Entropy
Optimizer: Adam
Epochs (20)
Batch size (10)

contaminated by various types of anomalies. To get more realistic condition, four types of anomalies considered in this paper are as follows:

- **Random**: Some data points are randomly increased or decreased.
- **Ramp**: Data is altered with a positive or negative slope.
- **Scale**: Data is scaled using a positive or negative scaling factor.

PMU is installed at buses 3, 8, 11, 16, 20, 23, 25, and 29 in order to get full observability. In this paper, it is assumed that the maximum error of PMUs is 0.7 % for magnitude and 0.7 centiradian for phase angle [26]. For stage I, 40447 observations are chosen from database. At this stage, the data are examined for compliance with the equations of power grid that is under study to identify anomalies at the level of network. Stage I investigations confirm the absence of large-scale data pollution that would result from complete replacement of authentic measurements with simulated fake network values. Then in the second stage, the correlation of time-series subsequences is used to identify all anomalies. Based on the type and severity of anomalies, the majority were successfully identified by the end of the second phase, achieving a misclassification rate of 1.31%. Once the anomalies are identified and labeled, the process moves to the final stage, which involves designing an LSTM network. For this stage, approximately 80% of the data is allocated for training the network and 20% for test. The interesting point

is all misclassified samples have been placed in the training dataset. Also, the details of the designed LSTM network are presented in Table I. Fig. 4 shows the accuracy at each epoch during the training phase. An epoch refers to one complete pass of the entire training data through the neural network. Similarly, the loss function value at each epoch is shown in Fig. 5. As seen, binary cross entropy loss fuction value is against accuracy. As evidenced by the results, the LSTM requires ten epochs to successfully identify dominant patterns in the data features while achieving classification accuracy of 98.8% at final epoch. Fig. 6 illustrates the confusion matrix for the test data. A comparison of the performance of the proposed LSTM with the Support Vector Machine (SVM) and Decision Tree methods is presented in Table II, with the parameters used as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$F1_{Score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{17}$$

Where:
TP (True Positive): Correctly predicted positive instance.
TN (True Negative): Correctly predicted negative instance.
FP (False Positive): Incorrectly predicted positive instance.
FN (False Negatives): Incorrectly predicted negative instance.

As observed, the proposed method has performed significantly better across all four criteria All detectors have been tested multiple times under different contamination conditions, and in all cases, the results closely matched the values presented in Table II. The results demonstrate slightly better performance on the test data where accuracy is 99.78% compared to the training data, because all misclassified instances from the nearness stage were placed in the training dataset. Due to their limited quantity, these misclassified instances failed to significantly impact the LSTM's performance, allowing the test data to reveal the model's true predictive capability.

After ensuring the effectiveness of the proposed detector, it is used to identify anomalies in the remaining dataset. Fig. 7 shows the confusion matrix, illustrating the detector's performance on this portion of remaining dataset. As shown in this figure, the performance of the detector is highly remarkable. A more detailed analysis revealed that out of the

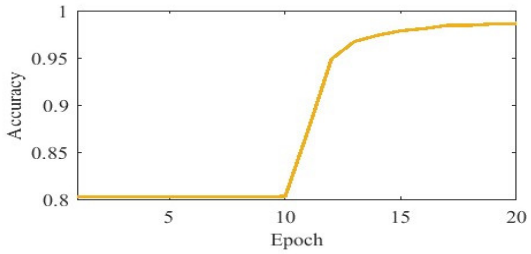

Fig. 4. Accuracy at each epoch during the training phase of the LSTM network.

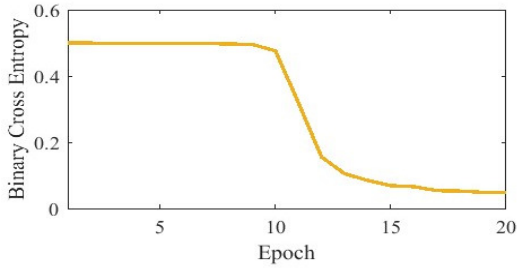

Fig. 5. Loss at each epoch during the training phase of the LSTM network.

67 samples incorrectly classified as normal, the anomalies were minimal and had no significant impact on the operation of the smart grid. Table III further demonstrates the performance of the proposed detector based on four evaluation metrics. As seen, the results indicate the high accuracy and pricision of the proposed method.

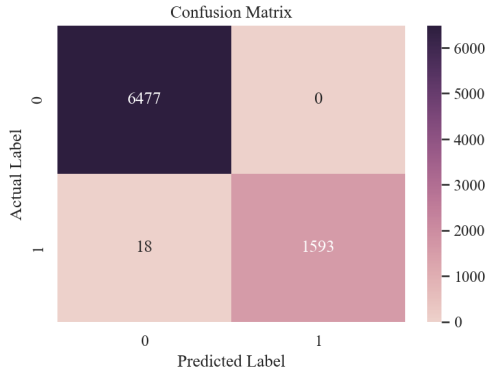

Fig. 6. Confusion matrix with heatmap visualization for test data (0: normal, 1: anomaly)

TABLE II
PROPOSED DETECTOR PERFORMANCE IN COMPARISON WITH OTHER DETECTORS ON TEST DATA

	Accuracy	Recall	F1_s.	Precision
SVM	0.8884	0.4358	0.6070	1
Decision Tree	0.5911	1	0.4966	0.3303
Proposed	0.9978	0.9888	0.9944	1

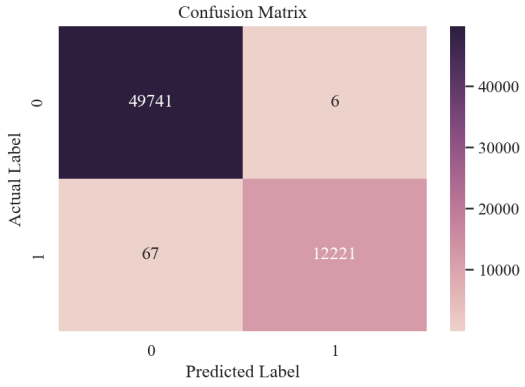

Fig. 7. Confusion matrix with heatmap visualization for remaining parts of database (0: normal, 1: anomaly)

TABLE III
LSTM-BASED DETECTOR PERFORMANCE ON REMINING PARTS OF DATABASE

Accuracy	0.9988
Precision	0.9995
Recall	0.9945
F1-Measure	0.9970

IV. CONCLUSION

This paper presents a hybrid approach that combines data-driven and model-based methods to detect anomalies considering anomalous database. In the proposed method, only a portion of the database is selected, and anomalies are investigated using a residual-based method combined with spatiotemporal correlation analysis of time-series subsequences. By integrating both approaches, we first verified the data's compliance with the studied power network equations, subsequently, through spatiotemporal correlation analysis, temporal and spatial features are extracted for anomaly identification Then, these data are used to train powerful LSTM networks. The trained network can then detect anomalies in other remaining parts of the database as well. Ultimately, this trained network can also serve as an online anomaly detector for smart grids. The proposed method was implemented on the IEEE 39-bus network, and the results demonstrated its significant effectiveness. Furthermore, a comparison with the SVM and decision tree methods highlighted the superiority of the proposed detector.

ACKNOWLEDGEMENT

Financial support provided by the Iran National Science Foundation (INSF) is acknowledged.

REFERENCES

[1] N. B. Bhatt, "Role of Synchrophasor Technology in the development of a smarter transmission grid," IEEE PES General Meeting, Minneapolis, MN, USA, 2010, pp. 1-4.

[2] California ISO, “Five year synchrophasor plan,” California ISO, Fol som, USA, Technical Report, Nov. 2011.

[3] W. Qi, “Comparison of differences between SCADA and WAMS real time data in dispatch center,” Proceedings of the 12th International Workshop on Electric Power Control Centers, Bedford Springs, USA, Jun. 2013, pp. 2-5.

[4] Y. Liu, P. Ning, M. K. Reiter, “False data injection attacks against state estimation in electric power grids”, ACM Transactions on Information and System Security, vol. 14, no. 1, 2011.

[5] M. A. Rahman and H. Mohsenian-Rad, “False data injection attacks with incomplete information against smart power grids”, IEEE Global Communications Conference (GLOBECOM), 2012, pp. 3153-3158.

[6] S. K. Singh, K. Khanna, R. Bose, B. K. Panigrahi, and A. Joshi, “Joint transformation- based detection of false data injection attacks in smart grid,” IEEE Trans. Ind. Informat., vol. 14, no. 1, pp. 89-97, Jan. 2018.

[7] J. Zhao, G. Zhang, M. La Scala, Z. Y. Dong, C. Chen, and J. Wang, “Short-term state forecasting-aided method for detection of smart grid general false data injection attacks,” IEEE Transactions on Smart Grid, vol. 8, no. 4, pp. 1580-1590, Jul. 2017.

[8] G. Cheng, Y. Lin, J. Zhao and J. Yan, “A Highly Discriminative Detector Against False Data Injection Attacks in AC State Estimation,” IEEE Transactions on Smart Grid, vol. 13, no. 3, pp. 2318-2330, May 2022.

[9] Mohammad Esmalifalak, Lanchao Liu, Nam Nguyen, Rong Zheng, and Zhu Han. Detecting stealthy false data injection using machine learning in smart grid. IEEE Systems Journal, vol. 11, no. 3, pp. 1644–1652, 2014.

[10] A. Khaleghi, M. S. Ghazizadeh and M. R. Aghamohammadi, "A Deep Learning-Based Attack Detection Mechanism Against Potential Cascading Failure Induced by Load Redistribution Attacks," IEEE Transactions on Smart Grid, vol. 14, no. 6, pp. 4772-4783, Nov. 2023.

[11] G. Avelino Sampedro, S. Ojo, M. Krichen, M. A. Alamro, A. Mihoub and V. Karovic, "Defending AI Models Against Adversarial Attacks in Smart Grids Using Deep Learning”, IEEE Access, vol. 12, pp. 157408-157417, 2024.

[12] Youbiao He, Gihan J Mendis, and Jin Wei, “Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism”, IEEE Transactions on Smart Grid, vol. 8, no. 5, pp. 2505–2516, 2017.

[13] J. Ji, Y. Liu, J. Chen, Z. Yao, M. Zhang and Y. Gong, "False Data Injection Attack Detection Method Based on Deep Learning With Multi-Scale Feature Fusion," IEEE Access, vol. 12, pp. 89262-89274, 2024.

[14] Hadis Karimipour, Ali Dehghantanha, Reza M Parizi, Kim-Kwang Raymond Choo, and Henry Leung, “A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids”, IEEE Access, vol. 7, pp. 80778–80788, 2019.

[15] Abdelrahman Ayad, Hany EZ Farag, Amr Youssef, and Ehab F El-Saadany “Detection of false data injection attacks in smart grids using recurrent neural networks” IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), pp. 1–5, 2018.

[16] Xiangyu Niu, Jiangnan Li, Jinyuan Sun, and Kevin Tomsovic, “Dynamic detection of false data injection attack in smart grid using deep learning” IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), pp. 1–6, 2019.

[17] A. Bhattacharjee, A. K. Mondal, A. Verma, S. Mishra and T. K. Saha, "Deep Latent Space Clustering for Detection of Stealthy False Data Injection Attacks Against AC State Estimation in Power Systems," IEEE Transactions on Smart Grid, vol. 14, no. 3, pp. 2338-2351, May 2023.

[18] W. Jiang, J. Wang, K. -L. Hsiung and H. -Y. Chen, “GRNN-Based Detection of Eavesdropping Attacks in SWIPT-Enabled Smart Grid Wireless Sensor Networks,” IEEE Internet of Things Journal, vol. 11, no. 22, pp. 37381-37393, 15 Nov.15, 2024.

[19] H. Goyel and K. S. Swarup, “Data Integrity Attack Detection Using Ensemble-Based Learning for Cyber–Physical Power Systems,” IEEE Transactions on Smart Grid, vol. 14, no. 2, pp. 1198-1209, March 2023.

[20] Z. Zhang, S. Bu, Y. Zhang and Z. Han, “Market-Level Integrated Detection Against Cyber Attacks in Real-Time Market Operations by Self-Supervised Learning,” IEEE Transactions on Smart Grid, vol. 15, no. 4, pp. 4128-4142, July 2024.

[21] F. Zhang, Y. Dubasi, W. Bao and Q. Li, "Detection and Localization of Data Forgery Attacks in Automatic Generation Control," IEEE Access, vol. 11, pp. 95999-96013, 2023.

[22] J. Liang, L. Sankar, and O. Kosut, “Vulnerability analysis and consequences of false data injection attack on power system state estimation,” *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3864–3872, Sep. 2016.

[23] J. Zakaria, A. Mueen and E. Keogh, "Clustering Time Series Using Unsupervised-Shapelets," 2012 IEEE 12th International Conference on Data Mining, Brussels, Belgium, 2012, pp. 785-794.

[24] S. Hochreiter, J. Schmidhuber, “Long Short-Term Memory”, Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[25] A. Pulver and S. Lyu, "LSTM with working memory," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 2017, pp. 845-851.

[26] A. Salehi, M. Fotuhi-Firuzabad, S. Fattaheian-Dehkordi, M. Gholami and M. Lehtonen, “Developing an Optimal Framework for PMU Placement Based on Active Distribution System State Estimation Considering Cost-Worth Analysis,” *IEEE Access*, vol. 11, pp. 12088-12099, 2023.