Graph Neural Network-Assisted Methods for Resolving State Estimator Divergence

Tuna Yildiz and Ali Abur
Department of Electrical and Computer Engineering
Northeastern University
Boston, USA
(yildiz.tu, a.abur)@northeastern.edu

S. Tripathy, V. Singhvi, M. Patel, J. Dondeti
Transmission Operations and Planning
Electric Power Research Institute
Palo Alto, CA
(stripathy, vsinghvi, mpatel, jdondeti)@epri.com

***Abstract*—Accurate and continuous monitoring of the power network is crucial for ensuring reliable and secure system operations. State Estimation (SE) plays a central role in this process. Conventional SE approaches rely on the assumption of a known network model. However, in practice, there may be unreported topology changes that can cause divergence, resulting in loss of system visibility, or worse, yield incorrect estimates that may lead to inappropriate operator actions. Similarly, challenges may arise when substation measurements become inaccessible with communication failure, forcing reliance on stale measurements as pseudo-measurements that may bias SE results.**

This work builds upon a previously established framework for non-divergent SE to isolate erroneous measurements causing divergence, while maintaining reliable estimates for the rest of the system. Expanding on that foundation, this paper introduces Graph Neural Network (GNN) based methods to address the root causes of divergence, specifically focusing on issues due to unreported topology changes and unchanged measurements.

***Index Terms*—State Estimation, Weighted Least Square, Divergence, Lagrange Multiplier, Topology Error Detection, Graph Neural Network, Missing Data Imputation.**

I. Introduction

State estimation (SE) remains central to modern power system monitoring, enabling operators to evaluate real-time system conditions and make informed operational decisions. Its accuracy, however, relies heavily on correct topology information - a condition often compromised by unreported configuration changes or inaccurate network parameters. Such discrepancies can lead to divergence or erroneous results, preventing operators from reliably assessing system conditions and responding effectively. This can result in severe consequences, including inaccurate power dispatch decisions, financial inefficiencies in the energy market, and a heightened risk of unexpected outages. Additionally, state estimators may occasionally fail to converge due to measurement losses caused by events such as communication failures or malfunctioning measurement devices. In such situations, using pseudo measurements to force convergence can lead to biased SE results [1]–[4].

In [4], a non-divergent state estimator was developed by integrating "Resilient" and "Recursive" SE techniques, successfully isolating problematic areas that caused the divergence of SE and enabling convergence for the majority of the system. This paper explores ways to address the underlying causes of divergence in such problematic areas, specifically focusing on issues arising from unreported switch status changes that lead to split-bus errors or from static measurements that fail to capture system dynamics.

To address issues due to unreported switch status changes, a switch status detection method is developed based on the well-known Lagrange Multiplier technique [5]. However, since the isolated section of the system from [4] cannot converge, the standard Lagrange Multiplier approach cannot be directly applied. To overcome this limitation, two alternative methods are considered to establish an anchor point for identifying topology changes responsible for divergence:

1) **Brute Force Approach:** Starting with the initial circuit breaker (CB) statuses, this approach iteratively tests all possible CB configurations until the state estimator converges. Once convergence is achieved, the CB statuses are corrected accordingly. If the size of the removed area from [4] is small, this method has low computational cost and, being offline, prioritizes accuracy over speed. However, for large-scale or online applications, the approach may become computationally intensive and unsuitable. For this reason, it is discussed conceptually but not evaluated in this work.
2) **GNN-Assisted Topology Error Correction:** Building on prior studies that applied Graph Neural Networks (GNNs) for topology identification [6]–[8], this work proposes a hybrid approach as the primary solution. The GNN predicts likely CB statuses from available measurement data, providing an initial starting point for the Lagrange Multiplier method. This eliminates the need for exhaustive brute force searches, significantly reducing computational overhead while maintaining accuracy.

To address issues caused by unchanged measurements, a GNN-based imputation framework is also introduced. While various neural network models, such as Generative Adversarial Imputation Networks (GAINs), Autoencoders (AEs), and Graph Autoencoders (GAEs), have been applied for data imputation in the past [9]–[11], most rely on temporal correlations to reconstruct missing values [12], [13]. In contrast, the scenario considered here involves measurements that may

979-8-3315-8138-1

remain unchanged for extended periods - potentially several hours - making temporal data unsuitable. The proposed approach uses GAEs to estimate missing measurements directly from available system snapshots, enabling accurate imputation without requiring time-series data.

The details of the proposed methods are presented in the following sections.

II. Unreported Switch Status Changes

A. Brute Force Approach

If the Weighted Least Squares (WLS) method with Lagrange Multipliers diverges, a converged anchor point must first be established [5]. To obtain this, all possible switch status combinations are initialized totaling $2^N - 1$, with N being the number of switches. For example, with $N = 5$, 31 combinations are tested employing a brute-force search, excluding the original diverged configuration. Each combination is evaluated iteratively until a single converged case is found, after which the WLS method with Lagrange Multipliers is applied to determine the correct switch statuses and identify the actual topology.

Although this process is computationally intensive, the "Recursive" SE in [4] may isolate a problematic area, which is much smaller than the overall system, keeping the number of combinations manageable. Since the method is executed offline, accuracy takes priority over speed, making the brute-force approach a practical solution for correcting topology errors. However, its exhaustive nature may make it impractical for situations where the number of possible configurations grows exponentially with larger isolated areas or for online applications.

B. Graph Attention Network (GAT) for Switch Status Identification

While the brute-force method is effective for offline correction of topology errors in smaller isolated areas, further efficiency can be achieved by reducing the number of combinations that need to be tested. To this end, a GNN-based CB status identification method is proposed. The trained model leverages available measurement data to predict likely unreported CB statuses, providing an initial starting point for the Lagrange Multiplier method and thereby minimizing the extent of brute-force exploration.

While prior studies have applied GNNs for topology identification [14]–[16], this work adopts a hybrid approach. Rather than relying solely on GNN predictions, the proposed method uses GNNs as a supplementary tool to guide the Lagrange Multiplier method. Specifically, the GNN provides an initial estimate of topology errors replacing the brute-force search after which the Lagrange Multiplier method refines and finalizes the CB statuses in the problematic area.

For switch status identification, a Graph Attention Network (GAT) is employed, as it naturally incorporates edge features into the model, allowing effective use of both node and edge information [17]. In a GAT, attention coefficients are computed for each connected node pair, incorporating both node and edge features.

Each node i has a feature vector $h_i \in \mathbb{R}^F$, and each edge (i, j) has a feature vector $e_{ij} \in \mathbb{R}^E$. Node features are linearly transformed as:

$$h'_i = W h_i, \tag{1}$$

where $W \in \mathbb{R}^{F' \times F}$ is a trainable weight matrix. Attention coefficients incorporate node and edge attributes:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}(a^\top [h'_i \parallel h'_j \parallel e_{ij}])\right)}{\sum_{k \in N(i)} \exp\left(\text{LeakyReLU}(a^\top [h'_i \parallel h'_k \parallel e_{ik}])\right)}, \tag{2}$$

where $a \in \mathbb{R}^{2F'+E}$ is a learnable attention vector and $\parallel$ denotes concatenation. Node updates are computed as:

$$h''_i = \sigma \sum_{j \in N(i)} \alpha_{ij} h'_j, \tag{3}$$

where σ is a non-linear activation (e.g., ReLU). This formulation enables GAT to jointly leverage node and edge features for richer graph representations.

After GAT processing, node embeddings (containing both structural and attribute information) are concatenated for each edge, producing a 2×hidden_channels vector. This is passed through a Multi-Layer Perceptron (MLP) for edge classification.

Thus, the MLP predicts edge properties, such as whether a CB is open or closed, using features from both connected nodes. By combining GAT-based feature extraction with MLP-based classification, the proposed hybrid model:

- Reduces computational overhead compared to brute force approach.
- Leverages both graph structure and deep learning.
- Improves robustness by integrating GNN predictions with physical system constraints.

The architecture used for CB status identification is shown in Fig. 1.

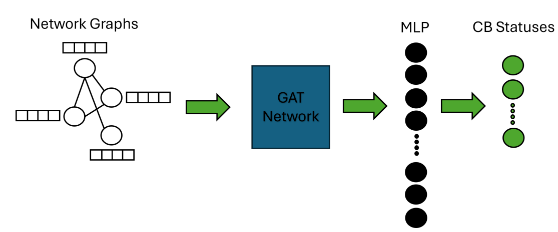


Figure 1: Proposed model for the CB status identification.

III. Missing Data Imputation

To address the issue of unchanged measurements, this work presents a GNN-aided missing data imputation method that uses observable voltage states to estimate those in unobservable regions, improving SE accuracy and reliability. Previously, Denoising Autoencoders (DAEs) and Denoising Graph Autoencoders (DGAEs) have been widely applied for reconstructing missing measurements [18]–[20]. Following this approach, a DGAE is employed here to effectively handle the issue of unchanged measurements in SE.

979-8-3315-8138-1

A. Proposed DGAE Model

The DGAE is designed to impute missing node feature values in power systems through *graph-based feature learning*. It employs Graph Attention Networks (GATv2) [21] as the encoder to capture graph-structured dependencies and an MLP decoder to reconstruct missing values.

The GAT Encoder is a multi-layer Graph Attention Network (GATv2) [21] that processes node features and their relationships to learn latent representations. It consists of four GATv2Conv layers, each refining node embeddings by aggregating information from neighboring nodes and incorporating edge features (admittance values) into the attention mechanism. Layer Normalization (LayerNorm) is applied after each attention layer to stabilize training, followed by ReLU activations to introduce non-linearity. The encoding process is:

$$h^{(1)} = \text{ReLU}(\text{LN}(\text{GATv2Conv}_1(X, A, E))) \tag{4}$$

$$h^{(2)} = \text{ReLU}(\text{LN}(\text{GATv2Conv}_2(h^{(1)}, A, E))) \tag{5}$$

$$h^{(3)} = \text{ReLU}(\text{LN}(\text{GATv2Conv}_3(h^{(2)}, A, E))) \tag{6}$$

$$Z = \text{LN}(\text{GATv2Conv}_4(h^{(3)}, A, E)) \tag{7}$$

where X is the input node feature matrix, A is the adjacency matrix, E denotes edge features, GATv2Conv_i is the i^{th} attention layer, LN is Layer Normalization, and Z is the final encoded representation.

The MLP decoder reconstructs missing node features from Z using four fully connected layers with ReLU activations between layers. The DGAE integrates the encoder and decoder to reconstruct missing features while preserving structural consistency. A masking mechanism ensures that known values remain unchanged while missing values are replaced by predictions.

The proposed model uses voltage magnitudes ($|V|$) and voltage angles (θ) from the observable part of the system to estimate states in unobservable areas affected by unchanged measurements, leveraging correlations between interconnected nodes. The overall network architecture is shown in Fig. 2.

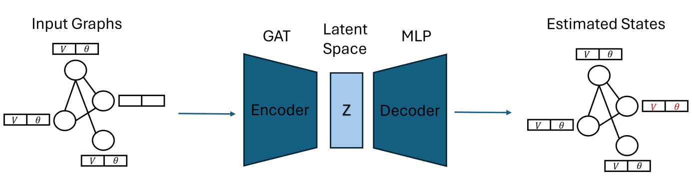


Figure 2: Proposed DGAE model for missing data imputation.

IV. Simulation Results

A. Unreported Switch Status Changes

While the brute-force method is discussed for context, all experimental results presented here are based on the GNN-based approach described in Section II-B.

To evaluate the proposed algorithm for topology error detection in a diverged state estimator, the IEEE 30-bus system (Fig. 3) was used, representing a portion of the network isolated using the non-divergent SE scheme from [4]. In this setup, if some switches were modeled as closed while the actual measurements indicated they were open, multiple split-bus errors would occur. In such cases, the conventional state estimator would fail to converge for the isolated area, and the Lagrange Multiplier method would be unable to provide a solution.

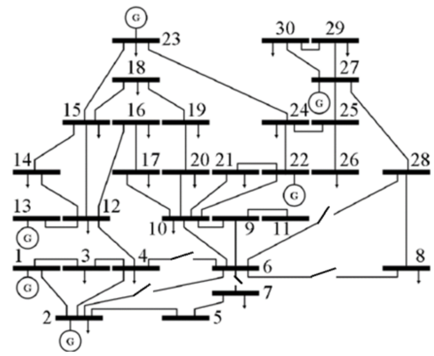


Figure 3: IEEE 30-bus system network.

To generate the dataset for GNN training, the IEEE 30-bus system was modified using the Pandapower library in Python. Branches at selected switch locations (2-6, 4-6, 6-8, 6-7, 6-28) were removed and replaced with low-impedance lines to represent CBs. Multiple breaker configurations were then applied. Based on the defined switch statuses, branches were either kept in service for closed switches or removed for open switches. The load at each bus was scaled using load factors derived from a representative daily load curve, with values updated at five-minute intervals. After updating the load values, power flow analyses were performed to obtain voltage magnitudes, voltage angles, real and reactive power injections, and real and reactive power flows for all lines and transformers. To prevent trivial detection of breaker statuses, power flows on breaker branches were set to zero.

The resulting measurement set was converted into GNN input features as shown in Table I. Node features included voltage magnitude ($|V|$), real power injection (P_{inj}), and reactive power injection (Q_{inj}). Edge features included real and reactive power flow (P_{flow}, Q_{flow}) and a binary indicator for breaker identification, where 1 denoted a breaker and 0 denoted a regular branch. Thus, the dataset contained $n \times 3$ node features and $(m - \#CB) \times 3$ edge features, where n was the number of buses and m was the number of branches.

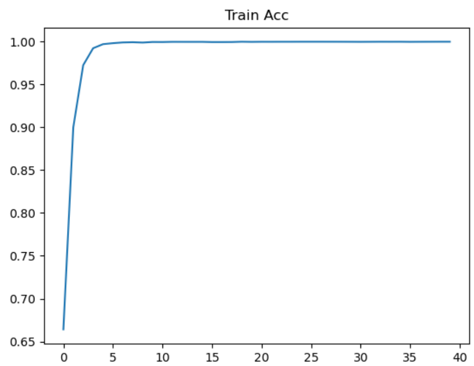
Table I: Node and Edge Features

Node Features	Edge Features
$\lvert V \rvert$	P Flow
P Injection	Q Flow
Q Injection	1 for Breakers 0 for Lines

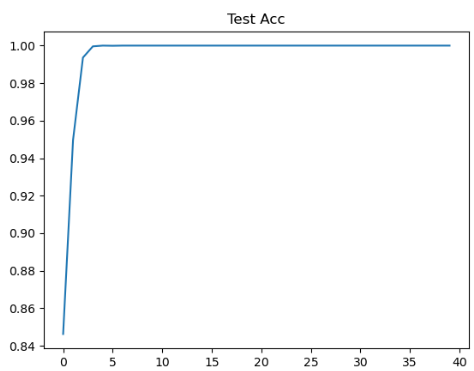
The model was then trained on 80% of the prepared dataset and evaluated on the remaining 20% (unseen data). The

979-8-3315-8138-1

training and test accuracies (vs epoch) are shown in Fig. 4. As observed, the model achieved 100% accuracy in detecting breaker statuses for both the training and test datasets using node and edge features.



(a) Train accuracy



(b) Test accuracy

Figure 4: Training and test results of GAT Network for switch status detection.

To further evaluate the model performance, an unseen set of breaker configurations was considered. Power flow simulations were then performed, covering all breaker status combinations, for a load scenario. The corresponding voltage magnitudes, power injections, and power flows were extracted as node and edge features. The trained model achieved 100% accuracy on this unseen dataset, successfully identifying all breaker statuses.

Finally, the well-known Lagrange Multiplier method was integrated into the proposed approach. The predicted breaker statuses were then used as inputs to the Lagrange Multiplier method, removing the need for brute-force initialization. When predictions were fully correct, convergence occurred in a single iteration; otherwise, the method corrected any misclassifications, ensuring an accurate final solution.

B. Missing Data Imputation

For this task, the IEEE 30-bus system was kept in its original configuration without introducing switches, representing a portion of the network isolated using the non-divergent SE scheme from [4]. A 30-day load profile was created, providing load multipliers at 5-minute intervals. These multipliers were applied to the original load values during power flow simulations to generate synthetic measurements that mimic real-world SCADA data under varying loading conditions.

To replicate the SE process, SE was first performed on the full system using these synthetic measurements to obtain the ground truth. For each loading scenario, voltage magnitudes, angles, branch data, and the admittance matrix were extracted for GNN input. Voltage magnitude and angle values served as node features, with 30% of them replaced by `NaN` to simulate missing data. An additional node feature indicating missingness (1 for missing, 0 for available) was included. The dataset was split 80/20 for training and testing, and the model was trained for 200 epochs. The trained model reconstructed voltage magnitudes and angles for nodes with missing data, replacing the `NaN` entries with predicted values. The loss curves in Fig. 5 show the Mean Square Error (MSE) between predicted states and the ground truth, quantifying imputation accuracy during training and testing.

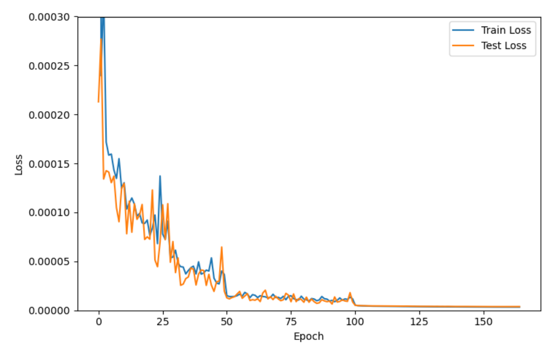


Figure 5: Loss curves for training and testing phases over 200 epochs.

To further evaluate the proposed data imputation scheme, a new measurement set was generated by running power flow based on a different load scenario. SE was first solved using these synthetic measurements for the full system to obtain the ground truth. Next, measurements at selected buses were removed to create single unobservable points at buses 8, 9, 11, 14, 15, 18, 19, 28, 29, and 30. SE was then executed on the remaining 20-bus observable system, where voltage magnitudes ($|V|$) and angles (θ) for unobservable states were replaced with `NaN` values. Observable states from the partial SE solution were used to generate input features for the GNN model.

The trained GNN model predicted the missing states, replacing the `NaN` entries in the SE results. The reconstructed states were then compared against the ground truth to calculate the MSE metric. Robustness was further assessed under two contingency scenarios while keeping the loading level unchanged: (i) removing the line between Bus 4 and Bus 6, and (ii) removing both lines between Bus 4–Bus 6 and Bus 9–Bus 10.

979-8-3315-8138-1

Table II: MSE results for reconstructed system states under different test scenarios

Test Scenario	Voltage Mag. MSE	Voltage Ang. MSE
Base case (33% states missing)	8.4104×10^{-6}	5.1261×10^{-6}
Line 4–6 removed (33% states missing)	6.7678×10^{-6}	1.0644×10^{-5}
Lines 4–6 and 9–10 removed (33% states missing)	3.9191×10^{-5}	1.0657×10^{-5}

Table III: MSE comparison: GNN vs. pseudo measurements

Approach	Frozen Measurement	Voltage Magnitude MSE	Voltage Angle MSE	J(x)
GNN-based approach	–	6.7261×10^{-6}	2.6719×10^{-5}	13.2571
Pseudo measurement	1% load change since recorded	5.3402×10^{-5}	8.9431×10^{-5}	275.9749
Pseudo measurement	10% load change since recorded	N/A	N/A	Diverged

The MSE results for all scenarios are summarized in Table II.

The proposed scheme was also tested on the larger IEEE 300-bus system using the same procedure as for the IEEE 30-bus case. After training the model, test scenarios were created by introducing states that conventional SE could not estimate. The GNN-based imputation algorithm accurately reconstructed these missing states, achieving MSE values on the order of 10^{-6} for both voltage magnitudes and angles, demonstrating its scalability and effectiveness for larger systems.

Lastly, for comparison, a pseudo-measurement approach was tested on the IEEE 300-bus system, where historical measurements were frozen for states that could not be estimated. While this sometimes reduced MSE, it significantly increased the estimator's cost function, increasing the risk of false bad-data detection and, in extreme cases, causing divergence when the interval between frozen and updated measurements grew, as shown in Tables III. In contrast, the GNN-based method consistently converged by using only the observable portion of the system, though its performance may decline when the unobservable region is large or weakly connected.

V. Conclusion

This work investigated and addressed the root causes of divergence in problematic areas identified by the "Resilient" and "Recursive" state estimators, with a focus on topology errors and unchanged measurements. A framework to address topology error was developed by integrating a Graph Attention Network (GAT) for switch status detection with the Lagrange Multiplier method, eliminating the need for exhaustive brute-force searches. The proposed approach was validated through case studies, demonstrating its effectiveness and efficiency.

For unchanged measurements, a missing data imputation technique based on Graph Autoencoders (GAEs) was implemented to reconstruct unavailable states. Comparative analysis against a pseudo-measurement-based method confirmed the improved performance of the proposed approach. Overall, these schemes enhanced state estimator robustness, improving reliability and observability under challenging operating conditions.

References

[1] U.-C. P. S. O. T. Force, "Final report on the august 14, 2003 blackout in the united states and canada," *Dept. Energy*, Apr. 2004 [Online] https://www.energy.gov/sites/default/files/oeprod/DocumentsandMedia/BlackoutFinal-Web.pdf.

[2] FERC and NERCH, "Arizona-southern california outages on september 8, 2011: Causes and recommendations," Apr. 2012.

[3] N. A. E. R. Corporation., "SCADA failure resulting in loss of monitoring function," [Online] http://www.nerc.com/files/LL20130203SCADAFailureResultinginLossofMonitoringFunction.pdf.

[4] T. Yildiz, A. Abur, S. Tripathy, V. Singhvi, and M. Patel, "Implementation of a state estimator ensuring visibility for the largest possible set of buses," in *2024 IEEE Power Energy Society General Meeting (PESGM)*, pp. 1–5, 2024.

[5] K. Clements and A. Costa, "Topology error identification using normalized lagrange multipliers," *IEEE Transactions on Power Systems*, vol. 13, no. 2, pp. 347–353, 1998.

[6] C. Zhao, X. Li, and Y. Cai, "A power grid topology detection method based on edge graph attention neural network," *Electric Power Systems Research*, vol. 239, p. 111219, 2025.

[7] C. Wang, J. An, and G. Mu, "Power system network topology identification based on knowledge graph and graph neural network," *Frontiers in Energy Research*, vol. 8, p. 613331, 2021.

[8] A. Ebtia, M. Ghafouri, M. Debbabi, M. Kassouf, and A. Mohammadi, "Power distribution network topology detection using dual-graph structure graph neural network model," *IEEE Transactions on Smart Grid*, 2024.

[9] Y. Zhang, R. Zhang, and B. Zhao, "A systematic review of generative adversarial imputation network in missing data imputation," *Neural Computing and Applications*, vol. 35, no. 27, pp. 19685–19705, 2023.

[10] W. H. L. Pinaya, S. Vieira, R. Garcia-Dias, and A. Mechelli, "Autoencoders," in *Machine learning*, pp. 193–208, Elsevier, 2020.

[11] A. Salehi and H. Davulcu, "Graph attention auto-encoders," *arXiv preprint arXiv:1905.10715*, 2019.

[12] Y. Raghuvamsi and K. Teeparthi, "Distribution system state estimation with transformer-bi-lstm-based imputation model for missing measurements," *Neural Computing and Applications*, vol. 36, no. 3, pp. 1295–1312, 2024.

[13] W. Khan, N. Zaki, A. Ahmad, M. M. Masud, L. Ali, N. Ali, and L. A. Ahmed, "Mixed data imputation using generative adversarial networks," *IEEE Access*, vol. 10, pp. 124475–124490, 2022.

[14] C. Zhao, X. Li, and Y. Cai, "A power grid topology detection method based on edge graph attention neural network," *Electric Power Systems Research*, vol. 239, p. 111219, 2025.

[15] C. Wang, J. An, and G. Mu, "Power system network topology identification based on knowledge graph and graph neural network," *Frontiers in Energy Research*, vol. 8, p. 613331, 2021.

[16] A. Ebtia, M. Ghafouri, M. Debbabi, M. Kassouf, and A. Mohammadi, "Power distribution network topology detection using dual-graph structure graph neural network model," *IEEE Transactions on Smart Grid*, 2024.

[17] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[18] S. Ryu, M. Kim, and H. Kim, "Denoising autoencoder-based missing value imputation for smart meters," *IEEE Access*, vol. 8, pp. 40656–40666, 2020.

[19] J. Zhang and P. Yin, "Multivariate time series missing data imputation using recurrent denoising autoencoder," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 760–764, IEEE, 2019.

[20] Y. Zhang, "Missing data imputation for health care big data using denoising autoencoder with generative adversarial network," *Scalable Computing: Practice and Experience*, vol. 25, no. 5, pp. 3850–3857, 2024.

[21] M. Fey and J. E. Lenssen, "Fast graph representation learning with pytorch geometric," *arXiv preprint arXiv:1903.02428*, 2019.

979-8-3315-8138-1