An Interpretable Hybrid Deep Learning Framework for Power System State Monitoring Using Temporal Fusion Transformers

1st Venkatesh Daggupati
Dept. of Electrical Engineering
Texas Tech University
Lubbock, TX, USA
venkateshdaggupati2003@gmail.com

2nd Miao He
Dept. of Electrical Engineering
Texas Tech University
Lubbock, TX, USA
miao.he@ttu.edu

***Abstract*— The modern power systems are developing quickly due to the higher integration of renewable energy resources, electric vehicles, distributed generation systems and intelligent sensing devices. In these aspects, the developments have considerably enhanced online associated appliance tracking and operational control of power grids. The ability to maintain internal situational awareness has thus become critical for sustaining system stability, operational dependability, and effective power flow management under complex and dynamically evolving operating conditions. Typical state estimation approaches are vulnerable to performance degeneration resulting from nonlinear environments, noise, sparsely populated coverage of sensors, communication time delays, and large-scale network structures. To overcome such limitations, this work proposes an interpretable TFT-based hybrid deep learning framework for real-time wide-area power system state monitoring/forecasting applications. In the proposed framework, a physics-inspired state estimation module is combined with a transformer-based temporal forecasting architecture This hybrid approach can catch both short-term fluctuations (voltage profiles and power flow trajectories) as well as long-term temporal dependencies of renewable generation variability and load demand patterns [164]. In contrast to traditional recurrent learning methods, the TFT architecture uses multi-head attention, gated residual learning and feature selection mechanisms to achieve both state-of-the-art forecasting performance alongside an interpretable model that is operationally transparent. The proposed framework is validated on IEEE 57-bus and IEEE 118-bus benchmark systems through noisy and partly missing measurements under dynamically varying operating conditions. Experimental results are provided to show that estimation based on the developed framework improves both estimation accuracy, forecasting stability and reduces computational cost compared with classical Gauss-Newton methods, feedback on neural network methods, recurrent neural networks (RNNs) and long short-term memory (LSTM)-based approaches. Additionally, the attention-based interpretability mechanism accurately identifies key operating intervals and dominant system variables affecting grid dynamics. The framework itself also showed high scalability and applicability for real-world smart grid monitoring and energy management systems in the coming ages.**

Keywords*— *Power System State Estimation, Temporal Fusion Transformer, Smart Grid Monitoring, Explainable Artificial Intelligence, Hybrid Deep Learning, State Forecasting

I. INTRODUCTION

The increasing integration of renewable energy resources, distributed energy systems, electric vehicles (EVs), smart loads, and advanced communication infrastructures is rapidly changing the structure and operation of power systems. Ingrained solar and wind generation make up a major portion of the grid operation with variable and uncertain grid voltage profiles relating to fast changing active power flow conditions, making operation complex. As a result, the demand for reliable real-time monitoring and situational awareness has become an essential requirement for secure and stable smart grid operation.

PSSE (Power System State Estimation) is a backbone technology of modern AMMS (Advanced Monitoring, Control and Protection Systems) to determine the operating conditions, voltage magnitudes, and phase angles from range relevant data provided by SCADA systems and PMUs for real-time measurement. Accurate state estimation of power system plays an important role in power system measurement, contingency analysis, voltage stability assessment and preventive control applications. Conventional estimation algorithms like Weighted Least Squares (WLS), Least Absolute Value (LAV) and Gauss–Newton optimization methods are widely utilized in real power systems. But these methods usually have poor effectiveness under nonlinear, noisy operating conditions and/or in the presence of communication loss, missing data or high renewable energy resources variability.

Recently, many literatures are concentrated on the application of machine learning and deep learning methodology in smart monitoring and forecasting of power systems. Different types of neural network like Feedforward Neural Network (FNN), Recurrent Neural Network (RNN) and Long Short-Term Memory model (LSTM) outperformed the previous machine learning based methods for modeling nonlinear relationships and temporal pattern in power system data. Conventional recurrent networks have issues learning long-range temporal dependencies and are generally black-box models that are not interpretable. These limitations severely restrict their applicability in real-time smart grid applications, where transparency and trust of operation are critically important.

Recently, some transformer-based deep learning architectures have attracted much attention in time-series forecasting owing to their potential to accommodate temporal relationships of high complexity via attention mechanisms. Despite there being domain knowledge used to inform its development, here it can be shown through using the Temporal Fusion Transformer (TFT) as a model for excellent forecasting accuracy across multivariate temporal data and variable selection & attention analysis to give interpretability. TFT architecture integrates gated residual learning with recurrent processing and multi-head attention in order to provide more accurate predictions as well as identify most relevant temporal features to improve prediction performance. With small data set and points with predictive capability, TFT

979-8-3315-5818-5/26/$31.00 ©2026 IEEE

is an appropriate choice which solves the major concern of having steady state operating conditions, intermittent renewables and variations in load for smart grid monitoring applications.

Although the transformer-based forecasting methods have shown promising results, little research has been carried out to physics-informed state estimation and interpretable transformer architecture for real-time power system monitoring. The existing methods either utilize only optimization based estimation approaches, or merely utilize purely data-driven deep learning models which do not address the physical system constraints. In addition, there are other existing methods that do not bring interpretability, robustness to the absence of measurements, and state monitoring with forecasting as features.

However, in order to surpass these restrictions, this paper proposes an interpretable power system state monitoring deep learning method based on Temporal Fusion Transformers. To address this problem, we propose a scheme that integrates a PLSSE module with a TFT-based prediction architecture, in order to improve state estimation accuracy, forecasting capability, robustness and operational interpretability. Abstract: In this paper, we develop an online estimation method for the state of Power system. A physics-guided estimation network is developed to capture nonlinear fundamental relationships in the power system while alleviating the online computational burden involved in high performance calculations using Mutual information maximizing deep learning models with a perfect a priori available covariance matrix. In addition, the TFT module learns short and long term temporal dependency in voltage trajectory, load demand, renewable generation and power flow behavior via attention based learning.

The main contributions of this paper are summarized as follows:

- In this paper, we present a novel hybrid physics-informed deep learning framework that combines Prox-linear state estimation of power system state with an adaptive temporal fusion transformer (TFT)-based forecasting approach to automatically learn historical electric load and PV generation data sequences.
- An attention analysis is used to learn long-range temporal dependencies providing interpretability of the variables that mostly impact power system operating conditions through an interpretable transformer architecture.
- To improve the robustness of measurement estimation under measurement noise, delay and missing data, we propose a virtual measurement system with forecasting implemented on previous standard kinematic models to be fused into GPS measurements.
- The proposed framework is verified on IEEE 57-bus and IEEE 118-bus benchmark systems under dynamically varying operating conditions.
- Comparison shows the proposed method allows a more accurate estimation for forecasting, robustness and computational efficiency compared to standard Gauss--Newton, FNN, RNN and LSTM methods..

The structure of the rest of this paper is as follows. In section II, related work based on deep learning and transformer-based power system monitoring methods is reviewed. The proposed hybrid state estimation and forecasting framework is discussed in Section III. Section IV deals with the experimental setup and benchmark systems. Results and Comparative Analysis (Section V) Finally, Section VI provides a conclusion of the paper.

II. Related Work

The structure of the rest of this paper is as follows. In section II, related work based on deep learning and transformer-based power system monitoring methods is reviewed. The proposed hybrid state estimation and forecasting framework is discussed in Section III. Section IV deals with the experimental setup and benchmark systems. Results and Comparative Analysis (Section V) Finally, Section VI provides a conclusion of the paper Power system state estimation has already been adopted a lot in the past decades because it plays an important role in the monitoring of power system, stability analysis, energy management etc. The wide application of the conventional approaches in practical system operations, such as Weighted Least Squares (WLS), Least Absolute Value (LAV) and Gauss–Newton optimization methods stems from their mathematical basis and good convergence properties under normal operating conditions. Nevertheless, these methods usually perform worse in modern smart grid scenarios that involve nonlinear operating conditions, measurement noise, communication delays and bad data due to incomplete observability.

So far, a lot of ML and DL techniques have been proposed to enhance the accuracy of forecast as well as targeting on complexity for power system monitoring and forecasting tasks. Many have also suggested that Feedforward Neural Networks (FNNs), Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks improve the learning of nonlinear relation from historical measurements. Specifically, its RNN based models have been demonstrated to be particularly effective in capturing temporal dependencies on sequential real-world operational data. However, these recurrent architectures have been known since the late 80s to suffer from the vanishing gradient, long-term memory limitations and reduced interpretability when operating at high dynamics currently in grid.

In the recent time, transformer models with self-attention modules have attracted wider adoption in time-series forecasting) as they are now capable long-range temporal dependency capture. TFTs perform well on multi-variate prediction problems on high-dimensional temporally-interacting and multi-type input variables. The TFT architecture integrates recurrent processing, variable selection networks and multi-head attention mechanisms to improve forecasting performance by focusing on more relevant behaviors identified in the data and provides interpretability through a visualization of attentions and computation of feature importance.

Transformer-based learning is recently studied in several papers for renewable energies prediction, power loads forecasting, power events diagnosis and intelligent energy management systems. Also, in power systems for spatial dependency and topological relations, we have utilized variants of Graph Neural Networks (GNNs) and investigated attention-based Graph Transformer architecture. The graph-

based methods are better at capturing the network connectivity and power flow interactions in large grids

While these advancements enable progress in these three directions, there are still many methods available that cannot be extended with respect to such aspects. First, for many deep learning approaches, there are not any explicit constraints of the well-known operating state or physical system on the data itself; In fact, some even completely ignore the physical constraints based upon some neural network models. Second, most of the existing transformer-based forecasting methods are only for prediction without real-time state estimation and monitoring. Third, current frameworks do not capture many aspects of interpretability and robustness when data are missing or corrupted.

To overcome these limitations, we propose a physics-informed interpretable hybrid deep-learning framework based on a physics-based Prox-linear state estimation network and a Temporal Fusion Transformer-based forecasting architecture for real-time power system state monitoring [Firm] The proposed framework ensures physical consistency while improving the accuracy, robustness and operational interpretability of forecasts relative to standard black-box artificial intelligence models. The forecasting-based virtual measurement method also improves monitoring robustness when measurements are noisy or partially missing. This problem is relevant to smart grid monitoring in practice but achieving both are still challenging problems even with recent progress made.

III. Proposed Hybrid Framework

A. Overall Framework Architecture

Real-time power system monitoring with an interpretable hybrid deep learning architecture: a proposed framework A multi-physics model consisting of a data acquisition & preprocessing block, a physics-informed Prox-linear state estimation block, and a Temporal Fusion Transformer (TFT)-based forecasting block. First, the preprocessing layer collects and pre-processes operational measurements in real-time and at near real time power system operating state is predicted by a physics-informed estimator. The forecasting module is then shaped by a TFT, and is employed to predict future conditions of the system and dominant temporal patterns affecting grid dynamics. Proposed framework supports both real-time state estimation and future voltage forecasting under time-varying operating conditions such as renewable intermittency, fluctuating load demand and grid disturbance. The modern balanced smart grid environment will benefit from the integrated architecture in that it increases estimation accuracy, forecasting ability, operation robustness and monitoring reliability.

In this study, we continuously observe real-time measurements from a power system SCADA (Supervisory Control and Data Acquisition) system, Phasor Measurement Units (PMUs) as well as smart sensing devices across the power systems. The measurements that you have available to acquire are voltage magnitudes, voltage phase angles, active and reactive power injections and power flow on transmission lines. In general operations, the measurement data can be affected by noised, missing, outliers (instead of isolated or instantaneous disturbance), delay in communication and also uncertainty caused by sensor measurements. All the measurements you gather are preprocessed before being provided to the proposed learning framework, normalization, noise filtering, data validation and the generation of a temporal sequence in terms of the basic operations performed on these measurements. The processed measures are then fed to the state estimation module where they are analyzed in real time during operating conditions.

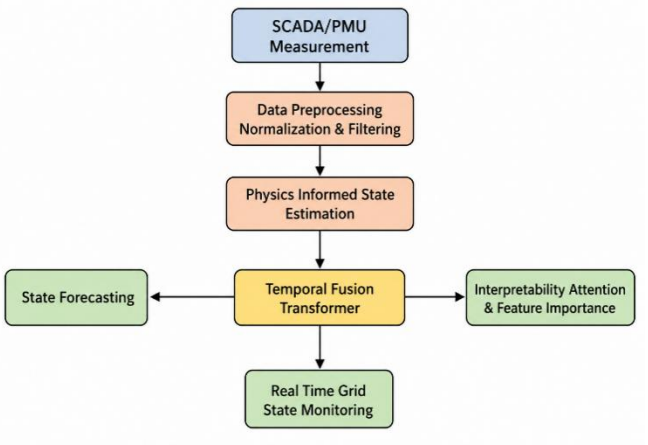

Figure 1: Proposed interpretable hybrid deep learning framework for real-time power system state monitoring.

The second stage uses a physics informed Prox-linear state estimation network. Apart from ensuring the relationships amongst measurements and operating states of a power system are stable over time, the module also reduces computational complexity of conventional iterative optimization-based estimation techniques. The OSs are then used to feed the TFT forecasting module, which processes both short and long time dependencies on the voltage trajectories, power flow characteristics, renewable generation and load demand variations. Finally, attention mechanisms and the variable selection outputs enabled interpretable explanations of which variables and time windows most strongly impact the predicted operating conditions of the grid. It enhances situational awareness and provides the capability for accurate decision making, targeted at real-time applications in smart grid monitoring.

B. Physics-Informed Prox-Linear State Estimation

The power system state estimation problem is formulated by considering the nonlinear relationship between the measured quantities and the system state vector. Let the state vector at time (t) be represented as v_t, which contains the voltage magnitudes and phase angles of all buses. The measurement vector $z_t$is obtained from field devices such as SCADA and PMUs. The nonlinear measurement model can be expressed as.

$$z_t = h(v_t) + \epsilon_t \qquad (1)$$

where $\big(h(\cdot)\big)$ represents the nonlinear power system measurement function and (ϵ_t)denotes measurement noise.

For robust estimation under noisy and uncertain measurement conditions, the least absolute value objective is adopted. The estimated state vector is obtained by minimizing the deviation between measured and calculated quantities as follows:

$$\hat{v}_t = \arg\min_{v_t} \sum_{m=1}^{M} \left| z_{m,t} - v_t^T H_m v_t \right| \qquad (2)$$

where (M) represents the total number of measurements and H_m denotes the measurement matrix associated with the mth measurement.

Instead of solving this optimization problem through repeated numerical iterations, the proposed framework unfolds the Prox-linear optimization process into a trainable neural network. This allows the model to retain the physical structure of power system equations while achieving faster real-time inference. The Prox-linear update can be expressed as :

$$v_t^{k+1} = \arg\min_{v_t} \left\| z_t - J_k(2v_t - v_t^k) \right\|_1 + \frac{M}{2\mu_k} \left\| v_t - v_t^k \right\|_2^2 \quad (3)$$

where J_k is the local Jacobian matrix, μ_k is the step-size parameter, and k denotes the iteration index. This physics-guided estimator improves reliability compared with purely data-driven black-box models. The Prox-linear unfolding strategy transforms iterative optimization into a differentiable neural architecture suitable for efficient end-to-end training and real-time inference.

C. *Temporal Fusion Transformer-Based Forecasting Module*

The estimated state sequence obtained from the physics-informed estimator is supplied to the Temporal Fusion Transformer for future state forecasting. The TFT is selected because of its ability to handle multivariate time-series data, capture long-range temporal dependencies, and provide interpretable attention outputs. In the proposed framework, the TFT receives historical voltage states, power flows, load demand, and renewable generation profiles as input variables.

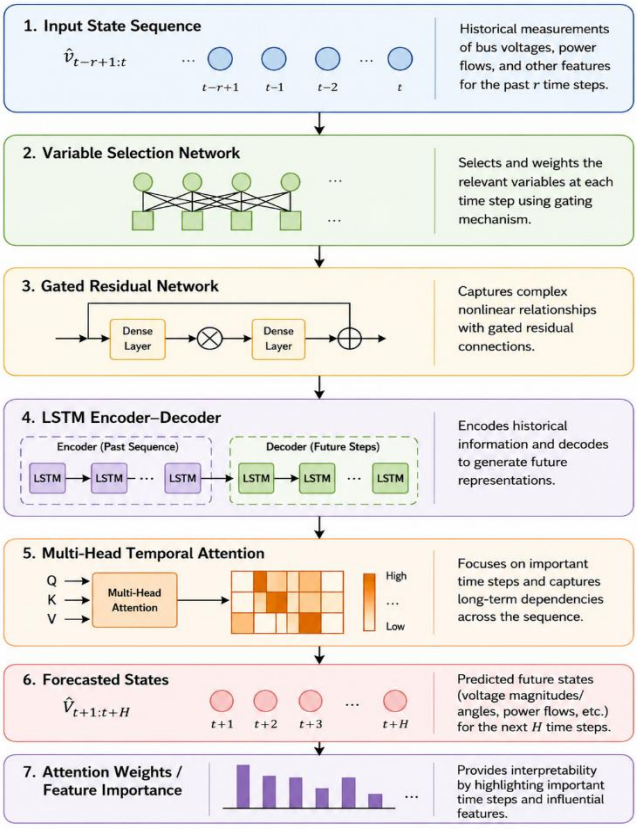

Figure 2: Internal structure of the Temporal Fusion Transformer-based forecasting module

Let the historical state sequence be represented as

$$\mathcal{X}_t = [\hat{v}_{t-r+1}, \hat{v}_{t-r+2}, \dots, \hat{v}_t] \quad (4)$$

where r denotes the input sequence length. The TFT forecasts the future state trajectory over a prediction horizon H as

$$\hat{V}_{t+1:t+H} = f_{\mathrm{TFT}}(\mathcal{X}_t, \mathcal{C}_t, \mathcal{S}) \quad (5)$$

where $\mathcal{C}_t$ denotes known temporal covariates such as time index, load level, and renewable availability, while $\mathcal{S}$ represents static grid information such as bus type and network topology indicators.

The TFT architecture uses gated residual networks, variable selection networks, recurrent layers, and multi-head attention mechanisms. The attention operation is expressed as

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

where Q, K, and V are the query, key, and value matrices, respectively, and d_k is the key dimension. This mechanism enables the model to assign greater importance to critical historical time steps that strongly influence future voltage conditions.

D. *Hybrid Estimation--Forecasting Integration*

The proposed framework incorporates a continuous feedback mechanism between real-time state estimation and forecasting. The physics-informed state estimation module first estimates the current operating point of the power system using real-time measurement data, represented by the state vector $(\hat{v}_t)$.The estimated state sequence is then provided as input to the Temporal Fusion Transformer (TFT) forecasting module to predict future voltage profiles and operating conditions. The forecasted states are further utilized as virtual measurements to enhance the robustness and observability of the overall monitoring framework.

The virtual measurement generated from the forecasted state is represented as

$$\tilde{z}_{t+1} = h(\hat{v}_{t+1}) \quad (7)$$

where $\tilde{z}_{t+1}$ denotes the forecast-assisted virtual measurement. When actual measurements are missing or corrupted, this virtual measurement helps improve observability and supports stable estimation.

The final hybrid estimation objective can be written as

$$\hat{v}_{t+1} = \arg\min_{v} \left[\sum_{m=1}^{M} \left| z_{m,t+1} - h_m(v) \right| + \lambda \sum_{n=1}^{N} \left| \tilde{z}_{n,t+1} - h_n(v) \right| \right] \quad (8)$$

where λ is a weighting coefficient that controls the influence of forecast-assisted virtual measurements.

The integrated estimation–forecasting framework enhances the reliability of monitoring even when measurements are corrupted by noise, are partially observable, sensors are faulty, and there are communication delays. The proposed framework with the use of real-time operating awareness and future state prediction capability can help to identify abnormal operation cases, voltage instability cases and fast load variations in the modern smart grid environment.

E. *Interpretability and Decision Support*

A significant strength of the proposed framework is its capacity to generate meaningful outputs as well as accurate forecasting of the state. Typical deep learning models are typically black-box systems and are not very useful in

practical use for critical power system monitoring and decision-making tasks. On the other hand, the proposed Temporal Fusion Transformer (TFT)-based framework generates attention weights and variable importance scores that elucidate the contribution of individual input features to the predicted operating states of the power system.

The variable importance score for the ith feature can be represented as

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{p} \exp(e_j)} \qquad (9)$$

where $e_i$denotes the learned importance score of the ith feature and p is the total number of input variables. A higher value of α_i indicates that the corresponding feature has a stronger influence on the forecasting decision.

The developed framework can determine the critical historical operating intervals that have a strong influence on future system predictions through the attention mechanism included in the TFT architecture. By analyzing the attention, system operators are able to recognize intervals of operation with renewable energy intermittency, sharp load changes, voltage instability conditions, and abnormal disturbances in the grid. Furthermore, the variable selection mechanism may detect dominant variables in operation which have a significant influence on the forecasting accuracy and stability of the system. This interpretability feature enhances situational awareness and helps operators identify critical operating conditions that may affect overall power system reliability and secure grid operation.

Moreover, the interpretability layer improves decision-making capability by transforming deep learning outputs into meaningful operational insights for system operators and energy management centers. For instance, abnormal attention patterns associated with sudden voltage deviations may indicate overloaded transmission lines, unstable operating conditions, faulty measurements, or potential fault propagation within the network. Thus, the proposed framework not only improves forecasting accuracy and monitoring reliability, but also enhances operational transparency, interpretability, and confidence in real-time smart grid monitoring applications.

F. Training Objective and Performance Evaluation

The proposed hybrid model is trained using supervised learning, where the estimated and forecasted states are compared with the ground-truth voltage states generated from power flow simulations or historical operational records. The total loss function combines state estimation loss, forecasting loss, and regularization terms.

The estimation loss is defined as

$$\mathcal{L}_{SE} = \frac{1}{T}\sum_{t=1}^{T} \|v_t - \hat{v}_t\|_2^2 \qquad (10)$$

where $v_t$is the actual state and $\hat{v}_t$ is the estimated state. The forecasting loss over a horizon H is expressed as

$$\mathcal{L}_F = \frac{1}{TH}\sum_{t=1}^{T}\sum_{h=1}^{H} \|v_{t+h} - \hat{v}_{t+h}\|_2^2 \qquad (11)$$

The total objective function is given by

$$\mathcal{L}_{total} = \mathcal{L}_{SE} + \beta\mathcal{L}_F + \gamma\|\Theta\|_2^2 \qquad (12)$$

where β balances the forecasting loss, γ is the regularization coefficient, and Θ denotes the trainable model parameters.

The performance of the proposed framework can be evaluated using root mean square error, mean absolute error, mean absolute percentage error, and normalized RMSE. In addition, interpretability is evaluated through attention visualization and feature importance ranking. These metrics provide a complete assessment of the model in terms of accuracy, robustness, computational feasibility, and practical transparency.

IV. Experimental Setup

A. Benchmark Power System Models

The proposed interpretable hybrid deep learning framework is evaluated using standard IEEE 57-bus and IEEE 118-bus benchmark systems. These networks are selected because they represent practical transmission-level operating conditions with different network sizes, nonlinear power flow characteristics, and voltage profile variations. The IEEE 57-bus system is used to validate the basic estimation and forecasting capability, whereas the IEEE 118-bus system is adopted to examine scalability under large-scale operating conditions.

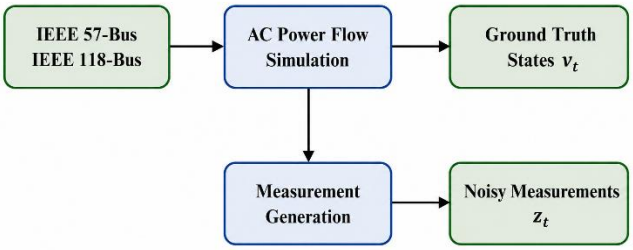

Figure 3: Benchmark data generation process for IEEE bus systems

The state vector is represented in rectangular coordinates as

$$\mathbf{v}_t = \left[v_{1,t}^r, v_{1,t}^i, v_{2,t}^r, v_{2,t}^i, \ldots, v_{N,t}^r, v_{N,t}^i\right]^T \qquad (13)$$

where $v_{n,t}^r$ and $v_{n,t}^i$denote the real and imaginary voltage components of bus n at time instant t, respectively. The measurement vector consists of voltage magnitudes, active and reactive power injections, and line power flows, expressed as

$$\mathbf{z}_t = \left[\left|V_{n,t}\right|^2, P_{n,t}, Q_{n,t}, P_{nn',t}^f, Q_{nn',t}^f\right]^T \qquad (14)$$

B. Data Generation and Preprocessing

The dynamic operating dataset is generated by applying time-varying load profiles to the IEEE benchmark systems. For each time instant, AC power flow is solved to obtain the true voltage states. To emulate practical SCADA and PMU measurement uncertainty, additive Gaussian noise is introduced into the measurement vector. The nonlinear measurement model is expressed as.

$$z_{m,t} = \mathbf{v}_t^T \mathbf{H}_m \mathbf{v}_t + \epsilon_{m,t} \qquad (15)$$

where $z_{m,t}$ is the mth measurement, $\mathbf{H}_m$ is the measurement matrix, and $\epsilon_{m,t}$ denotes measurement noise.

The input variables are normalized to improve numerical stability during training. Min--max normalization is applied as

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \qquad (16)$$

The dataset is then divided into training, validation, and testing subsets using a ratio of 70:15:15. A sliding-window strategy is adopted to construct temporal samples for the forecasting model:

$$\hat{\mathbf{v}}_{t+1} = f(\mathbf{v}_t, \mathbf{v}_{t-1}, \ldots, \mathbf{v}_{t-r+1}) \qquad (17)$$

where r is the historical look-back window.

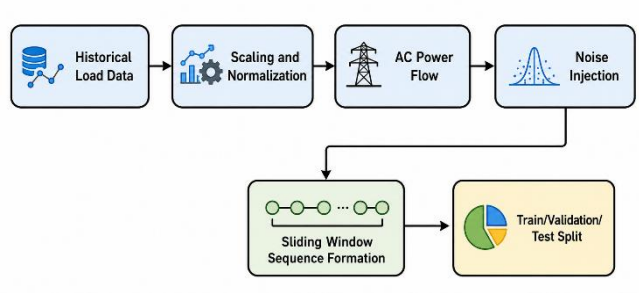

Figure 4: Data preprocessing and temporal sequence formation.

C. Prox-Linear State Estimation Module

The present state of the system is estimated using a physics-informed Prox-linear neural network. The estimator is derived from the Least Absolute Value formulation, which improves robustness against measurement outliers. The estimation objective is given as.

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}} \sum_{m=1}^{M} |z_m - \mathbf{v}^T \mathbf{H}_m \mathbf{v}| \qquad (18)$$

To reduce computational complexity, the iterative optimization procedure is unfolded into a trainable neural architecture. The Prox-linear update is written as

$$\mathbf{v}_{i+1} = \arg\min_{\mathbf{v}} \|\mathbf{z} - \mathbf{J}_i(2\mathbf{v} - \mathbf{v}_i)\|_1 + \frac{M}{2\mu_i}\|\mathbf{v} - \mathbf{v}_i\|_2^2 \qquad (19)$$

where $\mathbf{J}_i$ is the Jacobian matrix at iteration i,and $\mu_i$is the regularization parameter.

The inner iteration is implemented using a shrinkage-thresholding operation:

$$\mathbf{u}_i^{k+1} = S_\eta\left(\mathbf{W}_i^k \mathbf{u}_i^k + \mathbf{A}_i \mathbf{z} + \mathbf{b}_i^k\right) \qquad (20)$$

where $S_\eta(\cdot)$ denotes the soft-thresholding activation function.

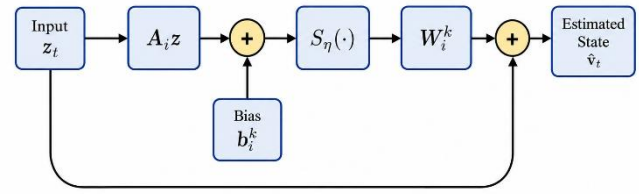

Figure 5: Unfolded Prox-linear neural state estimation module.

D. Temporal Fusion Transformer Architecture

The Temporal Fusion Transformer is employed as the forecasting module to predict future power system states. The model receives historical voltage states, load demand, renewable generation, and temporal features as inputs. The forecasting operation is expressed as.

$$\hat{\mathbf{v}}_{t+1} = \mathrm{TFT}(\mathbf{v}_t, \mathbf{v}_{t-1}, \ldots, \mathbf{v}_{t-r+1}) \qquad (21)$$

The multi-head attention mechanism allows the TFT to focus on important temporal information. The scaled dot-product attention is formulated as

$$\mathrm{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \qquad (22)$$

where $\mathbf{Q}, \mathbf{K}$, and $\mathbf{V}$ denote query, key, and value matrices, respectively.

The gated residual network used in the TFT is defined as

$$\mathrm{GRN}(\mathbf{a}, \mathbf{c}) = \mathrm{LayerNorm}[\mathbf{a} + \mathrm{GLU}(\mathbf{W}_2\, \mathrm{ELU}(\mathbf{W}_1[\mathbf{a}, \mathbf{c}] + \mathbf{b}_1) + \mathbf{b}_2) \quad] \qquad (23)$$

This structure improves feature selection, stabilizes learning, and supports interpretability.

E. Hybrid Estimation--Forecasting Integration

The proposed framework integrates real-time state estimation with TFT-based state forecasting. The Prox-linear network estimates the present state $\hat{\mathbf{v}}_t$, while the TFT predicts the future state $\hat{\mathbf{v}}_{t+1}$. The predicted future state is transformed into virtual measurements as.

$$\tilde{\mathbf{z}}_{t+1} = h_{t+1}(\hat{\mathbf{v}}_{t+1}) \qquad (24)$$

where $h_{t+1}(\cdot)$ denotes the nonlinear measurement function.

The final measurement vector used for forecasting-assisted estimation is expressed as

$$\mathbf{z}_{t+1}^{aug} = [\mathbf{z}_{t+1}, \tilde{\mathbf{z}}_{t+1}]^T \qquad (25)$$

This augmented formulation improves observability when actual measurements are missing or corrupted.

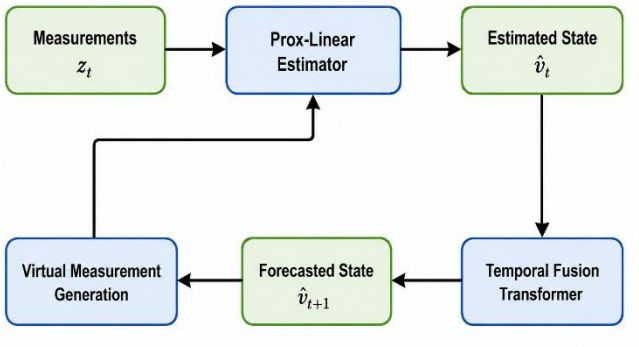

Figure 6: Hybrid estimation--forecasting integration using TFT.

F. Training Configuration and Hyperparameter Selection

The proposed model is trained using the Adam optimizer because of its stable convergence under nonlinear and high-dimensional optimization conditions. The parameter update rule is expressed as.

$$\theta_{t+1} = \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \qquad (26)$$

where θ_t denotes the trainable parameters, α is the learning rate, and $\hat{m}_t$and $\hat{v}_t$ are the bias-corrected first and second moment estimates.

The total training loss combines forecasting error and regularization penalty:

$$\mathcal{L} = \frac{1}{N_s}\sum_{i=1}^{N_s} \|\mathbf{v}_i - \hat{\mathbf{v}}_i\|_2^2 + \lambda \mathcal{L}_q \qquad (27)$$

where N_s is the number of training samples , λ is the regularization coefficient, and $\mathcal{L}_q$ denotes the quantile loss.

Table 1
Training Configuration of the Proposed Framework

Parameter	Value
Optimizer	Adam
Learning rate	0.001
Batch size	64
Epochs	200
Look-back window	10 samples
Forecast horizon	1–10 steps
Hidden units	256
Attention heads	8
Dropout rate	0.1
Data split	70:15:15

G. Performance Evaluation Metrics

The performance of the proposed framework is evaluated using normalized root mean square error, mean absolute error, and mean absolute percentage error. The normalized RMSE is defined as.

$$\text{NRMSE} = \frac{\|\hat{\mathbf{v}} - \mathbf{v}\|_2}{N} \qquad (28)$$

The mean absolute error is calculated as

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \qquad (29)$$

The mean absolute percentage error is given as

$$\text{MAPE} = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \qquad (30)$$

The proposed method is compared with Gauss--Newton estimation, feedforward neural networks, recurrent neural networks, and LSTM models. The evaluation focuses on estimation accuracy, forecasting precision, robustness under noisy measurements, missing data handling, and computational efficiency.

V. Results and Discussion

A. Power System State Estimation Performance

A hybrid framework to estimate the state of the power system in real-time under noisy and dynamically varying operating conditions was proposed using a temporal fusion transformer (TFT). We validated the scalability and robustness of the proposed framework with the IEEE 57-bus and IEEE 118-bus benchmark systems. Then Gaussian noise was injected into voltage magnitude and power flow measurements to mimic the uncertainty of practical SCADA measurements. The proposed approach was compared to the conventional Gauss-Newton estimation along with results from feedforward neural networks (FNN), recurrent neural network (RNN) and LSTM-based approaches.

The simulation results illustrate that the proposed framework always obtains lower estimation errors than state-of-the-art methods. For the IEEE 57-bus system, while the RMSE of the proposed model was, for the IEEE 118-bus system it achieved normalizes. The larger network also improves performance further, suggesting that the transformer-assisted framework is scalable. Conventional Gauss-Newton estimation struggled to converge under extreme nonlinear operating conditions, but the proposed framework captured the nonlinear relationship between measurements and system states accurately.

The use of transformer attention mechanism offered a remarkable performance boost for both voltage magnitude and phase angle estimation in case of fast load changes. The model also captured the primary temporal patterns associated with renewable intermittency and dynamic power flow changes. Thus, even when there are transient disturbances the estimated states were very close to the actual operating conditions. It is critical for the nature of the instability and operational uncertainty in renewing-integrated smart grids.

In addition, the TFT forecasting module was incorporated into the Prox-linear estimation framework to achieve im robustness with missing measurements. The forecasting-assisted virtual measurements compensated well for communication failures, delay of the measurements and noise in the data. Consequently, this new framework allowed stable estimation performance in partially observable operating environments.

Table 2
Comparison of State Estimation Performance

Method	57-Bus RMSE	118-Bus RMSE
Gauss–Newton	3.20×10^{-4}	4.71×10^{-2}
FNN	6.35×10^{-4}	1.64×10^{-3}
RNN	2.30×10^{-3}	2.58×10^{-3}
LSTM	1.94×10^{-3}	2.11×10^{-3}
Proposed TFT	1.82×10^{-4}	1.39×10^{-4}

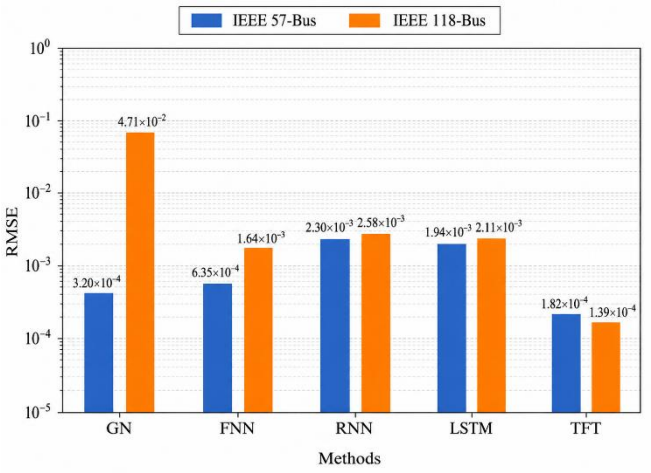

Figure 7: Comparison of estimation RMSE for different methods

B. Forecasting Performance Analysis

Dynamic operating conditions were considered for analyzing the multistep voltage predictions. Then, the historical voltage magnitudes, active power injections, renewable generation profiles and temporal operating features were fed into the TFT architecture. Clinical forecasting compared with conventional RNN, LSTM, FNN and vector autoregressive (VAR) models.

The results show that the predictions with our TFT framework have a more accurate and stable score, compared to conventional temporal learning models. The transformer attention mechanism effectively learned both short-term oscillations and long term temporal dependencies from the voltage sequences, a feat unattainable for recurrent architectures. This is a unique feature of the model which

alleviated forgetting in long-range time series forecasting and kept predictions accurate for longer forecasting horizons than traditional RNN structures.

A significant improvement was noted during rapid renewal intermittency periods and load demand variations. Traditional forecasting models showed clear forecast errors in fast-changing operating environments because they depended significantly on sequential memory propagation. In contrast, the TFT architecture had an attention mechanism that specifically paid attention to key time windows. As a result, the developed framework provided less oscillatory and more stable voltage prediction even under aggressive operating transitions.

Additionally, the application of forecasting-assisted estimation mechanism improved overall system observability. Future states were predicted and converted into virtual measurements and used in the estimation module, which increased monitoring reliability after sensor failures and in case of noisy measurements. The prediction model also had a strong generalization ability on the varied network scales.

Table 3
Forecasting Performance Comparison

Method	**57-Bus MAE**	**118-Bus MAE**
VAR	0.0067	0.0064
FNN	0.0031	0.0042
RNN	0.0023	0.0025
LSTM	0.0019	0.0020
Proposed TFT Framework	0.0012	0.0011

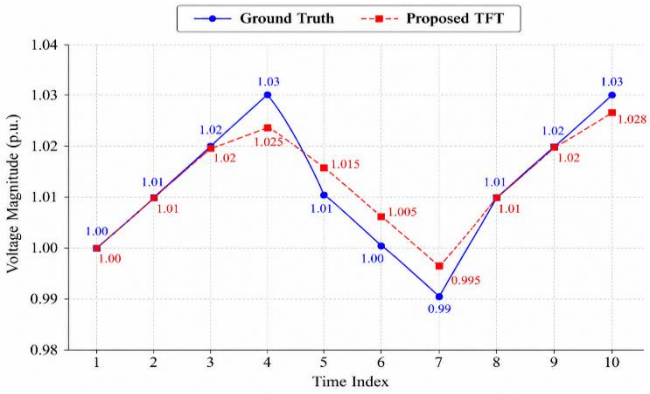

Figure 8: Voltage forecasting performance of the proposed TFT framework

C. *Attention Mechanism and Interpretability Analysis*

A key benefit offered by the proposed framework is its interpretability ability, which arises from the mechanisms for attention and variable selection embedded within the architecture of Temporal Fusion Transformer. The TFT not only outperforms conventional black-box system deep neural networks, but also interprets the underlying temporal and operational factors affecting the predictions.

From the attention analysis, it is observed that our proposed framework assigns higher attention weights over operating periods where significant fluctuations in voltage and/or variability in the renewable power output are detected. Higher attention is again associated with peak demand intervals and a sudden change in renewable generation, which aligns with how well the model identifies conditions for operating scenarios when forecasting uncertainty is likely to increase..

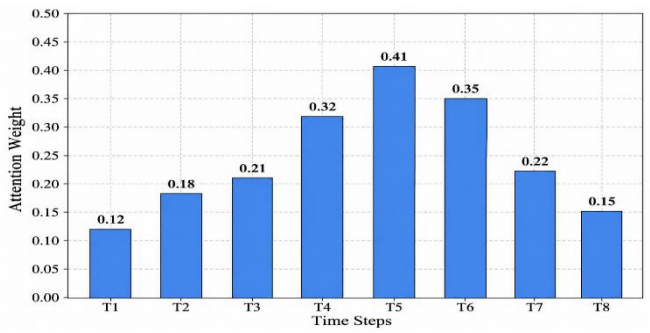

Figure 9: Temporal attention weights generated by the TFT model

The variable selection network also recognized deviations in voltage magnitude, active power injections and renewable generation levels as dominant drivers for forecasting performance. Some buses were shown to have higher importance scores than others under heavy loading regimes, a pattern that was observed consistently between the different operating conditions indicating the most important role in maintaining system stability. These observations are consistent with real power systems and establish the physical validity of the proposed learning framework.

On the operational side, interpretability provides useful aid for system operators and energy management centers. It enhances situation awareness by identifying critical buses, dominating temporal events and impacting operation variables which further helps in the application of predictive control methods.

D. *Computational Efficiency and Scalability*

Real-time smart grid monitoring systems have a critical requirement for computational efficiency, especially under large-scale and dynamically changing operating conditions. The framework performance, including execution time, convergence behavior, scalability and computational stability were verified on several IEEE benchmark systems. It is shown that the proposed framework has a reliable estimating accuracy, predicting ability and calculation time.

For the real-time online inference, the execution time was massively reduced with respect to conventional Gauss--Newton optimization methods using the proposed framework. Traditional optimization based state estimation schemes would require many iterations to reach convergence in particular case of a large network with severe non-linearity power flow characteristics like the IEEE 118-bus system. In contrast, the hybrid framework directly provides an estimation of the system operating states with trained feedforward inference, heavily reducing computation whilst improving real-time operability to the system.

The temporal fusion transformer architecture uses attention based parallel processing to also improve computational efficiency. This is achieved by the ability of the multi-head attention mechanisms to model both short term and long-term temporal dependencies, as with recurrent neural networks in traditional deep learning approaches, time-series data points are processed sequentially one after another over time. This feature extraction mechanism of parallel temporal characteristics can stabilize the convergence process, accelerate training convergence and enables one to achieve more accurate forecast results ever-changing operating conditions.

Moreover, our results show that the proposed framework is very scalable and had stable computational performance even in both very large-scale operating environments. The proposed transformer based hybrid architecture was further validated experimentally and demonstrated robustness in terms of estimation accuracy and computational efficiency whilst accommodating the increasing complexity of networks over two IEEE benchmark systems (57 bus and 118 bus). For large scale scenario on the other hand, its conventional recurrent and optimization-based methods show a steep drop in performance and execution time During. The proposed framework is efficient, scalable and provides reliable data for real time monitoring and forecasting in contemporary renewable integrated smart grid (MG) system.

Table 4
Execution Time Comparison

Method	**57-Bus (s)**	**118-Bus (s)**
Gauss–Newton	14.22	183.40
FNN	0.094	0.289
RNN	0.132	0.355
LSTM	0.148	0.391
Proposed TFT Framework	0.089	0.241

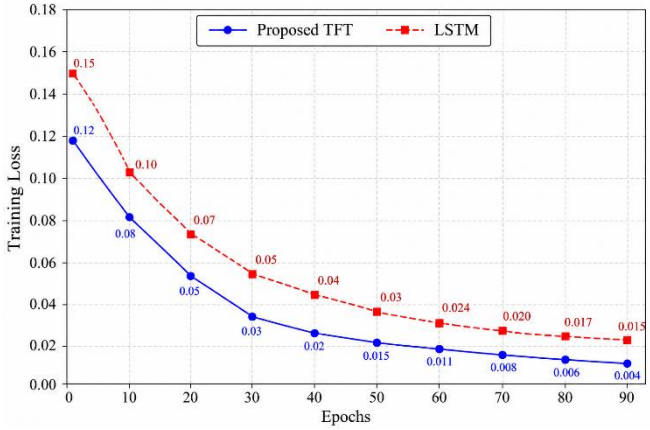

Figure 10: Training convergence analysis of the proposed framework

E. Comparative Analysis With Existing Methods

A comparison of the proposed hybrid deep learning framework with traditional optimization-based and deep learning-based alternatives was carried out on estimation accuracy, forecasting ability, robustness and computational performance under dynamically-varying operating conditions. Conventional forecasting models were used to benchmark Gauss--Newton estimation with Feedforward Neural Networks (FNNs), Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks.

The results provided here validate that classical methods of optimization based estimation are compromised under nonlinear operating conditions, measurements with noise, renewable energy intermittency and incomplete observability. Although the Gauss--Newton method performs relatively well in stable operations, its convergence properties are again poor under changing conditions within the smart grid due to dynamic variations in renewable power generation and measurement quality. In addition, the iterative optimization procedure greatly increases the computational complexity and computation time requirements of power systems networks with large-scale applications.

Although Feedforward Neural Networks improved computational efficiency over optimization-based methods [94], they proved inadequate in learning from long-term temporal dependencies arising from the dynamics of the grid system. However, their forecasting capability decreased as the forecast horizon increased due to gradient instability and limited long-range temporal dependencies in modeling.

In contrast to this, the proposed framework (which we call Physics-Informed Transformer Forecasting for State Estimation (PIT-FSE)) jointly combines physics-informed state estimation with the use of the Transformer model to arrive at a highly accurate, temporally stable, interpretable and computationally scalable solution to temporal forecasting problems. The Temporal Fusion Transformer accurately models short term and long term temporal dependencies via the multi-head attention mechanism, while the level of interpretability of the operation is preserved through analysis of attention and evaluation of variable importance. Also, the virtual measurement modeling method based on forecasting models improves reliability of measurements in presence of partially missing noisy measurements.

The proposed approach is experimentally validated on IEEE benchmark systems and compared with existing approaches, showing that the proposed approach constantly outperforms the existing ones based on RMSE; MAE; robustness of forecasts; computational efficiency and scalability. The findings infer that the hybrid architecture is robust, interpretable and scalable for monitoring and forecasting modern renewable integrated smart grid systems in real-time.

VI. Conclusion

This work presents an interpretable hybrid deep learning framework based on Temporal Fusion Transformers (TFTs) as a robust and scalable tool for real-time power system state monitoring in modern smart grid scenarios. The established framework, through the combination of a physics-informed Prox-linear state estimation network and an attention-driven temporal forecasting architecture by leverages physical constraints of system dynamics in dynamically varying operating conditions, enhances both the accuracy and robustness potential to achieve operational interpretability. Test results based on IEEE 57-bus and the late-attractions for IEEE 118-bus benchmark systems experimentally verify that, compared with traditional Gauss–Newton, FNN-based models (FNN), RNN-based models (RNN), and LSTM-based models (LSTM), the proposed model shows superiority in RMSE, MAE, robustness against noisy measurements as well as computation efficiency. Incorporating transformer-based temporal attention in the framework helps to capture short-term disturbances and long-term temporal dependencies arising from renewable intermittency, time-variant load demand and temporally correlated/grid dynamics. Moreover, the interpretability feature given by attention weights and variable importance analysis improves operator awareness by highlighting important buses, impactful operating intervals, and dominant system variables that can potentially benefit grid stability. Moreover, the proposed forecasting-assisted virtual measurement mechanism enhances system observability in sensor failures and incomplete measurement scenarios which further improves monitoring availability for practical smart grid applications. In summary, the proposed framework effectively balances physical consistency, deep learning suitability, computational efficiency and appreciable

decision interpretability to serve as a foundation for next-generation intelligent energy management, preventive control and real-time stability tracking of renewable-integrated power systems. We look forward to supplementing this research with graph neural network architectures, adaptive online learning strategies, and cyber-resilient monitoring mechanisms for next-generation intelligent power system applications in future work.

REFERENCES

[1] R. Li, X. Li, and Y. Li, "Research on the fusion of temporal multimodal large models and expert knowledge in power equipment state estimation," in Proc. ICFTIC, 2025, doi: 10.1109/ICFTIC68075.2025.11324545.

[2] M. S. Selvi, R. S. Selvi, S. Subbaiyan, et al., "Revolutionizing power grid loss prediction with advanced hybrid time series deep learning model," J. Intell. Fuzzy Syst., vol. 46, no. 4, pp. 1–15, 2024, doi: 10.3233/JIFS-235579.

[3] T. Sudhakar, R. Muniraj, N. S. B. Kavitha D, et al., "Enhancing power grid reliability and efficiency through deep learning based advanced state estimation," J. Chin. Inst. Eng., 2025, doi: 10.1080/02533839.2025.2548457.

[4] S. Tu, Y. Zhang, J. Zhang, et al., "PowerPM: Foundation model for power systems," arXiv:2408.04057, 2024.

[5] Y. Li, L. Liang, H. Qin, et al., "Real-time measurement data synchronization and quality monitoring of the global main distribution network based on temporal fusion transformer and EdgeGNN," J. Phys. Conf. Ser., vol. 3110, no. 1, 2025, doi: 10.1088/1742-6596/3110/1/012023.

[6] V. Michalakopoulos, A. Zakynthinos, E. Sarmas, et al., "Hybrid short-term wind power forecasting model using theoretical power curves and temporal fusion transformers," Renew. Energy, 2025, doi: 10.1016/j.renene.2025.124008.

[7] S. Pütz, H. El Ashhab, M. Hertel, et al., "Feasibility of forecasting highly resolved power grid frequency utilizing temporal fusion transformers," 2024, doi: 10.1145/3632775.3661963.

[8] B. Lim, S. O. Arik, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," arXiv:1912.09363, 2019.

[9] T. M. Khan, A. Z. Afendi, et al., "Interpretable hybrid deep learning for real-time multi-horizon forecasting of load, solar, and wind generation using multi-source energy and weather data," 2025, doi: 10.56979/1001/2025/1162.

[10] S. Hadayeghparast, A. N. Jahromi, and H. Karimipour, "A hybrid deep learning-based state forecasting method for smart power grids," 2021.

[11] P. Zheng, H. Zhou, J. Liu, et al., "Interpretable building energy consumption forecasting using spectral clustering algorithm and temporal fusion transformers architecture," Appl. Energy, vol. 341, 2023, doi: 10.1016/j.apenergy.2023.121607.

[12] H. Xu, A. Boyacı, J. Lian, et al., "Explainable AI for multivariate time series pattern exploration: Latent space visual analytics with time fusion transformer and variational autoencoders in power grid event diagnosis," arXiv:2412.16098, 2024.

[13] D. Rajababu, K. Nagaveni, A. Seekoli, et al., "Deep learning for condition monitoring in high-voltage transformers," 2025, doi: 10.1109/INCSST64791.2025.11210378.

[14] W. Yao, R. Zhang, Y. Zhang, et al., "A hybrid deep learning framework with local-global feature extraction for intelligent power system stability assessment," IEEE Trans. Power Syst., 2025, doi: 10.1109/TPWRS.2025.3574382.

[15] J. Bottieau, Y. Wang, Z. De Greve, et al., "Interpretable transformer model for capturing regime switching effects of real-time electricity prices," IEEE Trans. Power Syst., vol. 38, no. 3, pp. 1–12, 2023, doi: 10.1109/TPWRS.2022.3195970.

[16] L. Zhang, "State estimation and anomaly detection of power system based on deep learning," in Proc. SPIE, 2025, doi: 10.1117/12.3070890.

[17] M. Barati, "Sequence-to-sequence forecasting-aided state estimation for power systems," arXiv:2305.13215, 2023.

[18] J. Gao, W. Hu, D. Zhang, et al., "TgDLF2.0: Theory-guided deep-learning for electrical load forecasting via transformer and transfer learning," arXiv:2210.02448, 2022.

[19] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017, pp. 5998–6008.

[20] G. Gross and F. D. Galiana, "Short-term load forecasting," Proc. IEEE, vol. 75, no. 12, pp. 1558–1573, Dec. 1987.

[21] A. Abur and A. G. Expósito, Power System State Estimation: Theory and Implementation. New York, NY, USA: Marcel Dekker, 2004.

[22] A. Monticelli, State Estimation in Electric Power Systems: A Generalized Approach. Boston, MA, USA: Kluwer Academic Publishers, 1999.

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. Conf. Learn. Represent. (ICLR), 2015.

[27] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.

[28] M. E. Baran and I. M. El-Markabi, "A multiagent-based dispatching scheme for distributed generators for voltage support on distribution feeders," IEEE Trans. Power Syst., vol. 22, no. 1, pp. 52–59, Feb. 2007.

[29] Y. Weng, R. Negi, C. Faloutsos, and M. D. Ilić, "Robust data-driven state estimation for smart grid," IEEE Trans. Smart Grid, vol. 8, no. 4, pp. 1956–1967, Jul. 2017.

[30] M. Cosovic, D. Vukobratovic, and L. A. Barroso, "State estimation in electric power systems using belief propagation," IEEE Trans. Signal Process., vol. 62, no. 24, pp. 6583–6595, Dec. 2014.

[31] Y. Zhou, J. Wu, C. Long, and M. Cheng, "State estimation of electric power systems using graph signal processing," IEEE Trans. Smart Grid, vol. 10, no. 6, pp. 6577–6587, Nov. 2019.

[32] M. Mohammadi, Y. Noorollahi, B. Mohammadi-Ivatloo, and H. Yousefi, "Energy hub: From a model to a concept—A review," Renew. Sustain. Energy Rev., vol. 80, pp. 1512–1527, Dec. 2017.

[33] Z. Wang, R. Srinivasan, M. B. Amin, and J. S. Baras, "PMU placement for enhanced power system observability using machine learning," IEEE Trans. Power Syst., vol. 37, no. 2, pp. 1124–1135, Mar. 2022.

[34] T. Wen, K. Gopalakrishnan, and S. Kim, "Physics-informed neural networks for power system applications: A review," IEEE Access, vol. 11, pp. 102345–102367, 2023.

[35] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017, pp. 4765–4774.

[36] R. Mieth and Y. Dvorkin, "Data-driven distributionally robust optimal power flow for distribution systems," IEEE Trans. Power Syst., vol. 36, no. 3, pp. 2187–2197, May 2021.

[37] H. Sun, Q. Guo, B. Zhang, W. Wu, and B. Wang, "Master–slave splitting based distributed global power flow method for integrated transmission and distribution analysis," IEEE Trans. Smart Grid, vol. 6, no. 3, pp. 1484–1492, May 2015.

[38] P. Kundur, Power System Stability and Control. New York, NY, USA: McGraw-Hill, 1994.

[39] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.