

Article

An Explainable Deep Learning Framework with Multi-Head Attention and SHAP for Power Stability Monitoring in IoE-Enabled Smart Cities

Hend Alshede

Self-Development Skills Department, Common First Year Deanship, King Saud University,
Riyadh 12211, Saudi Arabia; halsheddi@ksu.edu.sa

Abstract

The growing reliance of Internet of Energy (IoE)-enabled smart city infrastructures has significantly intensified the complexity of modern urban energy environments due to the integration of renewable resources, electric vehicles, and interconnected IoT devices. These highly dynamic environments introduce critical challenges related to power stability, operational reliability, and intelligent energy management. Therefore, developing accurate, adaptive, and explainable monitoring frameworks has become essential for ensuring resilient urban energy infrastructures. This paper proposes an Explainable Artificial Intelligence (XAI)-driven deep learning framework integrating Multi-Head Attention and SHapley Additive exPlanations (SHAP) for intelligent power stability monitoring in IoE-enabled smart cities. The proposed framework employs an attention-based deep learning architecture to classify stable and unstable operational conditions using multivariate operational power parameters. Furthermore, SHAP-based explainability analysis is incorporated to improve model transparency and identify influential operational factors affecting stability behavior. Using the Electrical Grid Stability Simulated Dataset, the proposed framework achieved 97.35% accuracy and 0.997 ROC-AUC, outperforming several traditional machine learning and deep learning baselines. The explainability results show that temporal response parameters have the strongest impact on stability decisions. This work offers not only high predictive performance but also valuable interpretability, which is essential for practical deployment in real-world smart city energy systems.

Keywords: internet of energy e; smart cities; power stability monitoring; explainable artificial intelligence; multi-head attention; deep learning; shapley additive explanations; intelligent energy management; smart energy systems; urban energy infrastructure

Academic Editor: Michael
Negnevitsky

Received: 6 July 2026
Revised: 29 July 2026
Accepted: 31 July 2026
Published: 5 August 2026

Copyright: © 2026 by the author.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The rapid advancement of smart cities has fundamentally transformed modern urban ecosystems through the extensive integration of Internet of Things (IoT) technologies, distributed renewable energy resources, intelligent communication infrastructures, cloud-enabled services, and Internet of Energy (IoE)-based systems [1]. Contemporary urban environments increasingly rely on highly interconnected power systems to support practical applications such as renewable energy integration, distributed energy resources, electric vehicle charging infrastructures, microgrids, and smart building energy

management systems. Maintaining stable operation across these applications is essential for ensuring reliable and resilient power delivery under dynamic operating conditions. Although these technologies significantly enhance operational efficiency, sustainability, and intelligent energy management capabilities, they simultaneously introduce substantial challenges related to power stability, operational resilience, dynamic load balancing, and secure energy orchestration under continuously fluctuating urban conditions [2].

Maintaining power stability is a fundamental operational requirement for ensuring the reliable and uninterrupted operation of modern IoE-enabled energy systems. In these environments, instability may arise from renewable energy intermittency, the integration of distributed energy resources, electric vehicle charging dynamics, communication latency, stochastic consumption patterns, and the heterogeneous behavior of large-scale interconnected devices. Such instability can adversely affect the operation of IoE-enabled energy applications, leading to degraded system performance, reduced energy efficiency, increased operational complexity, and interruptions in reliable power delivery [3].

Conventional power stability monitoring approaches commonly depend on mathematical modeling, statistical analysis, and rule-based monitoring techniques for evaluating operational conditions and detecting abnormal system behavior [4]. Despite their effectiveness in relatively static environments, these approaches frequently suffer from limited scalability and insufficient capability in handling highly nonlinear and dynamically evolving operational data generated within modern smart city ecosystems. Furthermore, traditional analytical methods often experience difficulties in accurately capturing hidden dependencies and complex feature interactions among multivariate operational parameters under rapidly changing urban energy conditions [5].

To address these limitations, intelligent data-driven approaches have recently attracted considerable attention in smart energy monitoring and operational stability analysis applications. Several studies have employed machine learning-based techniques to improve instability detection capability and enhance operational monitoring performance in complex energy environments. Nevertheless, many conventional intelligent monitoring approaches still exhibit limitations in learning sophisticated nonlinear feature interactions and temporal operational dependencies from large-scale multivariate energy data. In addition, a significant proportion of existing approaches primarily focus on predictive performance while providing limited interpretability regarding the operational factors contributing to instability conditions [6].

Recent advances in deep learning have enabled more sophisticated analysis of complex operational energy data through hierarchical feature representation learning and nonlinear dependency modeling. Attention-based deep learning architectures have demonstrated remarkable capability in selectively emphasizing influential feature interactions and improving learning effectiveness in high-dimensional environments. Multi-Head Attention mechanisms further enhance model capability by simultaneously learning multiple contextual feature representations, thereby improving the extraction of complex operational relationships associated with dynamic stability behavior in intelligent energy systems [5].

Despite the substantial performance improvements achieved by deep learning models, the black-box nature of these architectures remains a critical challenge for deployment within safety-critical smart city infrastructures [7]. In practical urban energy environments, monitoring systems must provide not only high predictive accuracy but also transparent and interpretable decision-making mechanisms to support reliable operational management and trustworthy intelligent control. Consequently, Explainable Artificial Intelligence (XAI) has emerged as an important research direction for improving transparency, interpretability, and operational trustworthiness in intelligent monitoring frameworks [8].

Among recent XAI approaches, SHapley Additive exPlanations (SHAP) has demonstrated strong effectiveness in interpreting complex deep learning decisions and quantifying feature importance within multivariate operational environments. SHAP-based explainability enables comprehensive analysis of both global and local contributions, thereby providing valuable insights into the operational parameters influencing power stability behavior in dynamic IoE-enabled smart city environments [9].

Motivated by these challenges, this paper proposes an XAI-driven deep learning framework integrating Multi-Head Attention and SHAP for intelligent power stability monitoring in IoE-enabled smart cities. The proposed framework employs an attention-based deep learning architecture to classify stable and unstable operational conditions using multivariate operational power parameters. Furthermore, SHAP-based explainability analysis is incorporated to improve model transparency and identify influential operational factors affecting stability behavior. The proposed framework is evaluated using the Electrical Grid Stability Simulated Dataset and comparatively analyzed against multiple intelligent classification models using several performance evaluation metrics.

The main contributions of this paper are summarized as follows:

- Developing an attention-based deep learning architecture capable of learning complex nonlinear feature interactions from multivariate operational energy data.
- Integrating SHAP-based explainability analysis to improve model transparency and identify influential operational parameters affecting stability behavior.
- Proposing an XAI-driven deep learning framework integrating Multi-Head Attention and SHAP for intelligent power stability monitoring in IoE-enabled smart cities.
- Performing comparative evaluation against multiple intelligent classification approaches using several performance metrics.
- Achieving high monitoring performance with an accuracy of 97.35% and a ROC-AUC score of 0.997.
- Providing an interpretable, scalable, and computationally efficient intelligent monitoring framework for reliable urban energy management in future smart city environments.

Compared with existing intelligent energy monitoring approaches that primarily focus on prediction-oriented or forecasting-based tasks, the proposed framework introduces a unified XAI-driven architecture that simultaneously integrates Multi-Head Attention-based contextual operational learning and SHAP-based explainability analysis for intelligent power stability monitoring in IoE-enabled smart city environments. The proposed framework not only improves stability classification performance under highly dynamic urban energy conditions but also enhances model transparency and operational interpretability through both global and local feature-level explanations. Furthermore, the integration of attention-driven contextual learning mechanisms enables more effective identification of influential operational dependencies associated with instability behavior in modern distributed smart city energy infrastructures.

The remainder of this paper is organized as follows. Section 2 reviews recent studies related to intelligent power stability monitoring and explainable deep learning approaches. Section 3 presents the proposed framework and methodology. Section 4 describes experimental setup and evaluation metrics. Section 5 discusses experimental results and explainability analysis. Finally, Section 6 concludes the paper and outlines future research directions.

2. Related Works

2.1. Deep Learning-Based Power Stability Monitoring

Deep learning techniques have recently emerged as promising solutions for intelligent power stability monitoring due to their capability in learning nonlinear operational relationships and extracting complex hidden patterns from large-scale multivariate energy data. The increasing complexity of IoE-enabled smart city environments, characterized by renewable energy intermittency, distributed energy integration, and highly dynamic consumption behaviors, has accelerated the adoption of deep learning architectures for intelligent stability analysis and operational monitoring applications.

Early research efforts primarily focused on utilizing standalone deep learning architectures for stability prediction and operational classification tasks. For instance, the study presented in [10] employed a Deep Neural Network (DNN)-based framework for electrical grid stability prediction and demonstrated improved predictive capability compared with conventional analytical approaches. Although the proposed model achieved promising classification performance, the study mainly emphasized prediction accuracy without addressing intelligent instability interpretation or adaptive monitoring requirements in dynamic urban energy environments. Similarly, the work in [11] investigated deep learning techniques for short-term voltage stability assessment in power systems and reported effective instability detection performance under varying operational conditions. Nevertheless, the proposed framework primarily focused on voltage-oriented analysis with limited consideration of comprehensive multivariate operational monitoring.

To improve temporal dependency learning capability, recurrent deep learning architectures have increasingly been adopted in intelligent energy monitoring applications. In particular, the multidirectional Long Short-Term Memory (LSTM) framework proposed in [12] demonstrated strong capability in learning sequential operational behaviors from multiple temporal directions, thereby improving stability prediction accuracy in dynamic energy environments. Furthermore, the neural network-based framework introduced in [13] addressed the challenge of incomplete operational measurements by developing a robust stability prediction model capable of handling missing input values. Although these approaches improved temporal learning capability and operational robustness, they still relied mainly on standalone recurrent architectures with limited capability in capturing complex feature interactions and contextual operational dependencies.

Recent studies have increasingly shifted toward hybrid deep learning architectures to enhance both feature extraction and temporal learning performance within complex energy environments. In this context, the hybrid Convolutional Neural Networks and bidirectional Long Short-Term Memory (CNN-BiLSTM) framework proposed in [14] integrated convolutional feature extraction with bidirectional temporal dependency learning, resulting in improved stability forecasting capability compared with conventional standalone models. Likewise, the MLP-ELM-based framework introduced in [15] demonstrated efficient classification performance and fast learning capability for intelligent stability prediction applications. Despite the improved learning effectiveness achieved by these hybrid approaches, most existing frameworks still lack advanced attention-based mechanisms capable of selectively emphasizing influential operational patterns associated with instability behavior.

With the growing complexity of modern IoE-enabled energy infrastructures, Transformer-driven architectures have recently attracted significant attention due to their capability in modeling long-range temporal dependencies and high-dimensional operational relationships. For example, the Transformer-based hybrid neural network presented in [16] demonstrated strong effectiveness in capturing multiscale temporal patterns and improving forecasting reliability under highly dynamic renewable energy conditions.

However, most Transformer-oriented studies primarily focused on forecasting and estimation tasks rather than comprehensive power stability monitoring and intelligent operational analysis within smart city environments.

In addition to stability prediction, several studies have explored deep learning applications for operational energy management and intelligent monitoring tasks. The work presented in [17] highlighted the importance of reactive power demand forecasting for maintaining stable operational conditions under fluctuating load environments. Similarly, the deep learning-based dynamic state estimation framework proposed in [18] demonstrated promising capability in monitoring frequency stability under renewable energy fluctuations. Nevertheless, the computational complexity and limited interpretability of several existing frameworks may restrict their applicability within large-scale real-time smart city energy systems.

Overall, existing studies demonstrate the effectiveness of deep learning architectures in intelligent power stability prediction and operational monitoring applications. However, several important limitations remain unresolved, including insufficient interpretability, limited transparency in decision-making processes, inadequate integration of attention-driven feature prioritization mechanisms, and restricted capability in identifying influential operational factors affecting stability behavior within highly dynamic IoE-enabled smart city environments. Table 1 summarizes recent deep learning-based approaches for intelligent power stability monitoring and prediction.

Table 1. Summary of recent deep learning-based approaches for intelligent power stability monitoring and prediction.

Study Year	Method	Strengths	Limitations
[10] 2024	DNN	Improved prediction accuracy.	Limited early monitoring capability.
[11] 2021	DNN	Effective voltage instability assessment.	Limited monitoring scope.
[12] 2020	Multidirectional LSTM	Strong temporal learning capability.	Limited local feature extraction.
[13] 2022	ANN	Robustness against missing data.	No advanced hybrid architecture.
[14] 2024	CNN + BiLSTM	Improved local and temporal feature learning.	No attention mechanism.
[15] 2023	MLP-ELM	Fast learning capability.	Weak long-term temporal learning.
[16] 2025	Transformer- Neural Network	Strong multiscale temporal learning.	Focused mainly on wind forecasting.
[17] 2024	Forecasting Network	Accurate reactive power forecasting.	Focused mainly on forecasting.
[18] 2023	Dynamic State Estimation	Effective under renewable energy fluctuations.	High computational complexity.

2.2. Attention-Based Energy Monitoring Approaches

Attention-based deep learning architectures have recently emerged as powerful solutions for intelligent energy monitoring applications due to their capability in selectively emphasizing influential operational patterns and improving contextual feature representation within complex multivariate environments. Unlike conventional recurrent architectures that process sequential information with relatively uniform importance, attention mechanisms dynamically prioritize critical operational dependencies associated with abnormal energy behavior, thereby enhancing monitoring accuracy and stability analysis performance in highly dynamic IoE-enabled systems.

Recent studies have demonstrated the effectiveness of attention-driven architectures in intelligent energy applications under fluctuating operational conditions. For instance, the attention-enhanced deep learning framework proposed in [19] improved voltage stability monitoring performance in electric vehicle-integrated energy environments by emphasizing influential temporal patterns associated with dynamic charging behaviors and rapidly varying energy demand. The study reported enhanced instability detection sensitivity compared with conventional deep learning approaches. However, the proposed framework primarily focused on voltage-oriented analysis without addressing

comprehensive intelligent power stability monitoring within large-scale smart city infrastructures.

Similarly, the hybrid attention-based deep learning framework introduced in [20] demonstrated strong capability in capturing complex temporal dependencies and improving renewable energy forecasting performance under highly dynamic wind power environments. The integration of attention mechanisms significantly enhanced temporal feature representation and improved learning effectiveness in fluctuating operational scenarios. Nevertheless, the study mainly concentrated on renewable energy forecasting tasks rather than intelligent operational stability monitoring and explainable decision analysis.

Transformer-oriented architecture has also highlighted the growing importance of attention mechanisms in modeling long-range temporal dependencies and contextual operational interactions within multivariate energy environments. In this context, the hybrid LSTM-Transformer framework presented in [21] improved temporal dependency analysis and operational stability prediction reliability by integrating recurrent learning capability with Transformer-based attention mechanisms. Likewise, the Transformer-enhanced hybrid framework proposed in [22] demonstrated effective multivariate energy consumption modeling with improved long-term contextual learning capability. Despite the promising performance achieved by these architectures, most existing Transformer-driven studies primarily focus on forecasting and energy estimation applications while providing limited transparency regarding the operational factors contributing to instability conditions.

Although recent attention-based approaches have demonstrated substantial improvements in intelligent energy monitoring performance, several important limitations remain unresolved. Existing studies often prioritize forecasting accuracy and temporal learning performance while providing limited emphasis on explainability, operational transparency, and interpretable instability analysis in safety-critical smart city environments. Furthermore, only limited research has investigated the integration of attention-based deep learning architectures with XAI mechanisms capable of identifying influential operational parameters affecting power stability behavior within dynamic IoE-enabled urban energy systems.

Consequently, integrating Multi-Head Attention mechanisms with SHAP-based explainability analysis represents a promising research direction for developing transparent, adaptive, and intelligent monitoring frameworks capable of supporting reliable power stability management in future smart city infrastructures. Table 2 summarizes recent attention-based and Transformer-enhanced approaches for intelligent energy monitoring and stability analysis.

Table 2. Summary of recent attention-based and Transformer-enhanced approaches for intelligent energy monitoring and stability prediction.

Study	Year	Method	Strengths	Limitations
[19]	2025	Attention-Based Deep Learning	Enhanced instability detection capability	Limited to voltage stability analysis
[20]	2022	Hybrid Attention-Based Deep Learning	Improved temporal feature representation	Focused mainly on renewable forecasting
[21]	2026	LSTM-Transformer with Attention	Improved long-range temporal dependency learning	Limited early warning capability
[22]	2026	Transformer-Based Hybrid Framework	Effective multivariate temporal modeling	Focused mainly on forecasting tasks

2.3. Research Gap and Motivation

Although recent studies have demonstrated the effectiveness of deep learning, attention-based architectures, and Transformer models for power system stability analysis, several important research gaps remain. First, most existing approaches primarily focus on improving prediction accuracy or forecasting performance, while providing limited interpretability regarding the operational factors influencing stability decisions. Consequently, these models often operate as black-box systems, limiting their practical adoption in safety-critical IoE-enabled smart city energy applications.

Second, previous studies generally employ either attention mechanisms or explainability techniques independently. Attention-based approaches improve contextual feature learning but rarely provide comprehensive post hoc explanations of model decisions. Conversely, studies employing explainable AI often analyze conventional machine learning models rather than attention-based deep learning architectures.

Third, although Transformer-based models have recently been introduced for energy applications, they are mainly designed for forecasting and estimation tasks rather than intelligent power stability classification using multivariate operational features. Moreover, comprehensive investigations combining Multi-Head Attention with SHAP explainability for transparent power stability monitoring remain very limited.

Accordingly, the proposed framework addresses these research gaps by simultaneously providing accurate stability classification, contextual feature interaction learning, robust SHAP-based explainability, and reproducible model interpretations for IoE-enabled smart city energy systems.

3. Proposed Framework and Methodology

This section presents the proposed XAI-driven deep learning framework developed for intelligent power stability monitoring in IoE-enabled smart city environments. The proposed architecture integrates Multi-Head Attention mechanisms with SHAP-based explainability analysis to enhance contextual feature representation, nonlinear dependency learning, and interpretable instability monitoring under highly dynamic urban energy conditions.

The overall framework consists of multiple sequential stages, including data acquisition, preprocessing and normalization, attention-based deep learning modeling, intelligent stability classification, and explainability analysis. Initially, multivariate operational energy data are collected from the Electrical Grid Stability Simulated Dataset, which represents diverse operational behaviors under varying stability conditions. The collected data are subsequently preprocessed through data cleaning and Min-Max normalization to improve data consistency, numerical stability, and learning convergence during model training.

Following preprocessing, the normalized operational parameters are provided to the proposed attention-driven deep learning architecture. The Multi-Head Attention mechanism enables the model to capture complex feature interactions and selectively emphasize influential operational dependencies associated with stability behavior. Unlike conventional deep learning architectures that feature relatively uniform importance, the proposed attention-based framework dynamically prioritizes critical operational patterns, thereby improving contextual learning capability and monitoring effectiveness within dynamic IoE-enabled environments.

After feature interaction learning and nonlinear representation extraction, the framework performs intelligent classification of operational conditions into stable and unstable states. To improve transparency and decision interpretability, SHAP-based explainability analysis is incorporated to quantify the contribution of individual operational parameters to model predictions. The explainability component provides both global and local

interpretability insights, enabling comprehensive analysis of influential stability-related operational factors and supporting trustworthy intelligent decision-making within safety-critical smart city energy infrastructures.

Furthermore, the proposed framework is comparatively evaluated against multiple intelligent classification models using several performance evaluation metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. Figure 1 illustrates the overall workflow of the proposed explainable attention-based deep learning framework for intelligent power stability monitoring in IoE-enabled smart cities.

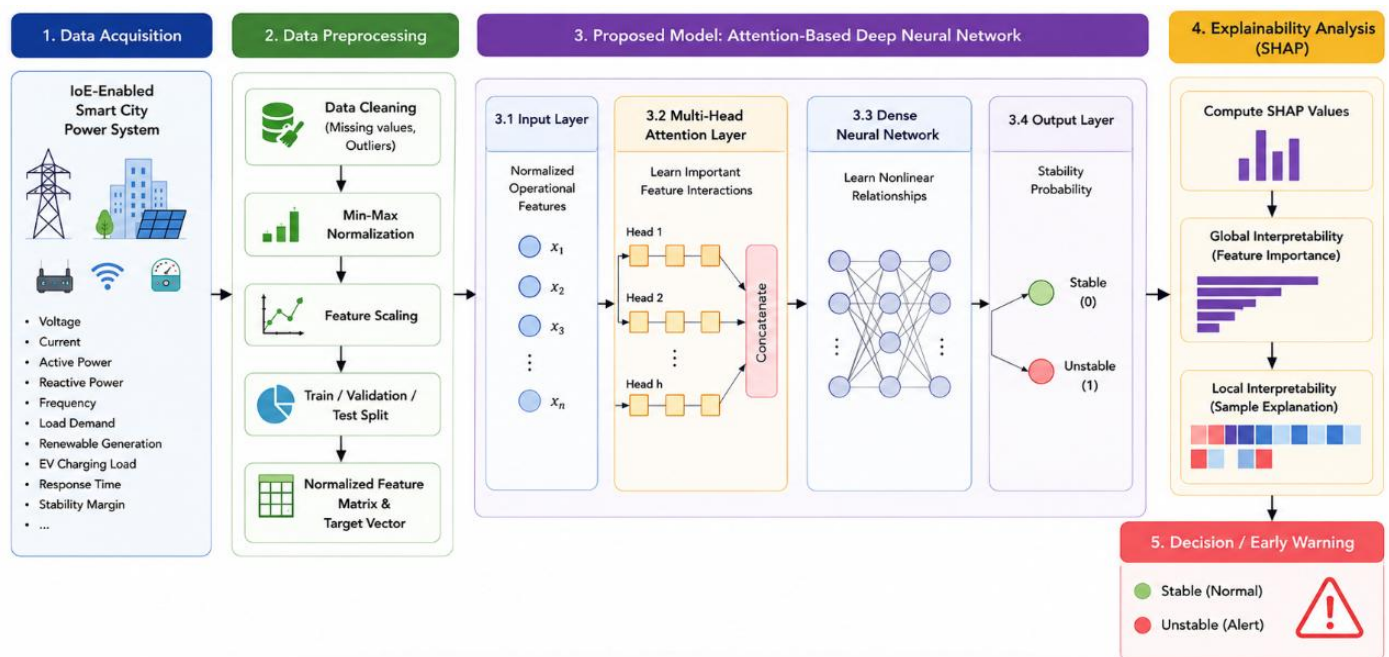


Figure 1. Overall architecture of the proposed explainable deep learning framework integrating Multi-Head Attention and SHAP for intelligent power stability monitoring in IoE-enabled smart city environments.

3.1. Dataset Description and Data Collection

The experimental evaluation of the proposed framework is conducted using the Electrical Grid Stability Simulated Dataset, which is publicly available through the Kaggle platform [23]. The dataset represents operational behaviors of modern intelligent energy systems under varying load demand, power distribution, and dynamic consumption conditions. Due to its multivariate operational structure and stability-oriented characteristics, the dataset has been widely utilized in intelligent power stability prediction and monitoring studies.

The dataset contains multiple operational parameters associated with energy generation, consumption behavior, reaction time, and dynamic power system interactions. These attributes collectively represent the operational state of the energy environment under both stable and unstable operating conditions. Furthermore, the dataset includes a binary stability label indicating whether the operational condition corresponds to a stable or unstable system state [24].

From an IoE-enabled smart city perspective, the dataset effectively reflects the dynamic operational complexity commonly observed in modern urban energy infrastructures, where renewable energy integration, distributed energy resources, electric vehicle charging activities, and heterogeneous consumption patterns continuously influence system stability behavior. Consequently, the dataset provides a suitable benchmark environment for evaluating intelligent monitoring frameworks under realistic multivariate

operational conditions [25]. Table 3 summarizes the Electrical Grid Stability simulated dataset used in this study.

The utilized dataset consists of approximately 60,000 operational samples with multiple numerical features describing power-related characteristics, including reaction time parameters, generated and consumed power values, and grid balancing attributes. Let the operational dataset be mathematically represented as:

$$X = \{x_1, x_2, x_3, \dots, x_N\} \quad (1)$$

where N denotes the total number of operational samples, $x_i \in \mathbb{R}^d$ represents a multivariate feature vector, d denotes the number of operational features.

Each operational sample is associated with a binary stability label defined as:

$$y_i \in \{0, 1\} \quad (2)$$

where 0 represents a stable operational condition, 1 represents an unstable operational condition.

The multivariate structure of the dataset enables the proposed framework to learn nonlinear operational relationships and contextual feature interactions associated with stability behavior. In addition, the dataset provides sufficient variability for evaluating attention-based deep learning architectures under highly dynamic smart city energy environments.

Although the Electrical Grid Stability Simulated Dataset is a benchmark dataset and does not fully capture the complexity of real-world IoE-enabled smart city energy systems, it provides a standardized and reproducible evaluation environment that enables fair comparison with existing studies, which was one of the primary motivations for selecting this dataset. Therefore, it is well suited for validating the proposed framework under controlled operational conditions.

Table 3. Summary of the Electrical Grid Stability Simulated Dataset.

Property	Description
Dataset Name	Electrical Grid Stability Simulated Dataset
Source	Kaggle
Application Domain	Power Stability Monitoring
Data Type	Multivariate Operational Energy Data
Number of Features	12 Features
Number of Samples	Approximately 60,000
Output Classes	Stable/Unstable
Main Attributes	Power Consumption, Reaction Time, Grid Parameters
Target Variable	Stability Label

3.2. Data Preprocessing

Data preprocessing represents a critical stage in intelligent power stability monitoring due to the heterogeneous nature of operational energy data and the sensitivity of deep learning architectures to feature scale variations and noisy operational patterns. In IoE-enabled smart city environments, operational parameters are continuously generated from distributed energy resources, intelligent sensors, communication infrastructures, and dynamic consumption behaviors, thereby requiring robust preprocessing mechanisms to improve data quality and learning effectiveness.

Initially, the operational dataset was inspected for missing values, duplicated samples, and abnormal numerical inconsistencies to ensure data integrity prior to model training. Since the dataset primarily consists of numerical operational attributes, preprocessing focused mainly on feature normalization and numerical scaling to improve convergence stability during the deep learning optimization process.

To reduce scale imbalance among operational parameters, Min-Max normalization was applied to transform all input features into a unified numerical range between 0 and 1. Feature normalization is particularly important for attention-based deep learning architectures because large numerical disparities may negatively affect attention weight learning and nonlinear representation extraction [26].

After normalization, the dataset was separated into input feature vectors and target labels. Let the normalized feature matrix be represented as:

$$X \in \mathbb{R}^{N \times d} \quad (3)$$

where N denotes the number of operational samples and d represents the number of operational features. The target vector is defined as:

$$Y = \{y_1, y_2, \dots, y_N\} \quad (4)$$

where each $y_i \in \{0, 1\}$ corresponds to either stable or unstable operational conditions.

Furthermore, preprocessing operations were designed to preserve the underlying operational relationships among multivariate energy parameters while improving numerical stability during attention-based learning. The normalized feature representations subsequently serve as input to the proposed Multi-Head Attention-based deep learning architecture for intelligent stability classification and contextual operational analysis.

3.3. Proposed Multi-Head Attention-Based Deep Learning Model

The core component of the proposed intelligent monitoring framework is the Multi-Head Attention-based deep learning architecture designed to capture complex nonlinear operational relationships and contextual feature interactions associated with power stability behavior in IoE-enabled smart city environments. Unlike conventional deep learning architectures that process input features with relatively uniform importance, the proposed attention-driven model dynamically emphasizes influential operational parameters contributing to stability and instability conditions [27].

It should be noted that the proposed Multi-Head Attention mechanism is applied to multivariate feature representations within each independent operational sample rather than to sequential time-series data. Since the utilized benchmark dataset does not provide temporal ordering among samples, positional encoding is not incorporated into the proposed architecture. Consequently, the attention mechanism is designed to capture contextual feature interactions instead of temporal dependencies.

The proposed architecture consists of multiple sequential processing layers, including an input representation layer, Multi-Head Attention mechanism, dense neural learning layers, output classification layer, and SHAP-based explainability analysis. Initially, normalized operational features are provided to the input layer, where multivariate feature representations are generated for contextual operational analysis. Let the normalized operational feature matrix be represented as:

$$X = [x_1, x_2, x_3, \dots, x_n] \quad (5)$$

where $X \in \mathbb{R}^{N \times d}$, N denotes the number of operational samples, d represents the number of operational features.

To improve contextual dependency learning, the Multi-Head Attention mechanism transforms the input features into query (Q), key (K), and value (V) representations using trainable projection matrices [28]:

$$\begin{aligned} Q &= XW^Q \\ K &= XW^K \\ V &= XW^V \end{aligned} \quad (6)$$

where W^Q , W^K , and W^V denote trainable weight matrices associated with the attention mechanism.

The attention operation is computed using scaled dot-product attention as follows [29]:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (7)$$

where d_k represents the dimensionality of the key vectors, Q denotes the query matrix, K denotes the key matrix, V denotes the value matrix, and QK^T computes pairwise similarity scores between query and key representations.

After training the proposed Multi-Head Attention-based model, SHAP explainability was implemented using the GradientExplainer algorithm to provide post hoc interpretation of the learned predictions. A representative background dataset was supplied to the explainer to estimate feature attributions, and SHAP values were subsequently computed for the evaluation samples to generate both global and local explanations. No modifications were made to either the SHAP algorithm or the proposed model architecture, as the explainability analysis was directly applied to the trained network. This implementation enabled reliable interpretation of the learned feature contributions while preserving the original prediction performance of the proposed framework.

To enhance contextual interaction learning, multiple attention heads are employed simultaneously. Each attention head independently learns distinct operational dependencies and influential feature relationships associated with stability behavior. The outputs of all attention heads are subsequently concatenated and linearly projected to generate the final contextual representation [30]:

$$\text{MHA}(X) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^O \quad (8)$$

where head_i denotes the output of the i^{th} attention head, h represents the total number of attention heads, W^O denotes the output projection matrix.

Following attention-based contextual learning, the extracted feature representations are processed through fully connected dense neural layers to learn nonlinear operational relationships associated with stable and unstable energy conditions. Rectified Linear Unit (ReLU) activation functions are utilized to improve nonlinear learning capability and accelerate convergence during optimization. nonlinear learning capability, accelerate convergence, and mitigate gradient vanishing issues during optimization. The ReLU activation function is mathematically defined as [31]:

$$\text{ReLU}(x) = \max(0, x) \quad (9)$$

where x denotes the input value to the activation function, the output equals x when $x > 0$, otherwise, the output becomes 0.

The use of ReLU enhances computational efficiency and improves deep neural representation learning for complex operational stability patterns within dynamic IoE-enabled smart city environments.

To reduce overfitting risk and improve model generalization capability, dropout regularization layers are incorporated between dense neural layers. The final output layer employs a sigmoid activation function to estimate the probability of instability occurrence [32]:

$$\hat{y} = \sigma(Wz + b) \quad (10)$$

where z denotes the learned feature representation, W and b represent trainable parameters, $\sigma(\cdot)$ denotes the sigmoid activation function.

The predicted operational state is subsequently determined as follows:

$$\begin{aligned} \hat{y} \geq 0.5 &\rightarrow \text{Unstable} \\ \hat{y} < 0.5 &\rightarrow \text{Stable} \end{aligned} \quad (11)$$

To improve model transparency and interpretability, SHAP-based explainability analysis is integrated into the proposed framework. The SHAP mechanism quantifies the contribution of individual operational parameters to model predictions and provides both global and local interpretability insights for intelligent instability analysis [33].

The proposed attention-driven architecture enables efficient learning of contextual operational dependencies and influential feature interactions within highly dynamic IoE-enabled smart city environments. Furthermore, the integration of Multi-Head Attention and SHAP significantly improves intelligent monitoring capability, interpretability, and decision reliability for real-time power stability analysis.

The detailed architecture of the proposed Multi-Head Attention-based deep learning model is illustrated in Figure 2, while the overall intelligent monitoring procedure is summarized in Algorithm 1.

Algorithm 1. Workflow of the proposed explainable Multi-Head Attention-based intelligent monitoring framework for contextual stability analysis in IoE-enabled smart city environments.

Input:

Electrical Grid Stability Simulated Dataset D

Feature matrix X

Stability labels Y

Output:

Stability prediction $\hat{Y}$

SHAP-based explainability analysis

Start:

1. Load the Electrical Grid Stability Simulated Dataset.

Perform data preprocessing:

2. Data cleaning

3. Min-Max normalization

4. Feature scaling

5. Split the dataset into training and testing subsets.

6. Feed normalized operational features into the input layer.

7. Generate Query (Q), Key (K), and Value (V) matrices.

Apply Multi-Head Attention Mechanism:

8. Compute scaled dot-product attention.

9. Learn contextual feature interactions.

Concatenate attention heads and generate contextual feature representations.

10. Pass learned representations through dense neural layers:

11. ReLU activation

12. Dropout regularization

Perform binary stability classification using sigmoid activation:

13. Stable (0)

14. Unstable (1)

Evaluate model performance using:

15. Accuracy, Precision, Recall, F1-score, and ROC-AUC

Apply SHAP explainability analysis:

16. Compute feature importance

17. Generate global interpretability

18. Generate local interpretability

19. Produce intelligent stability monitoring and early warning decisions.

End

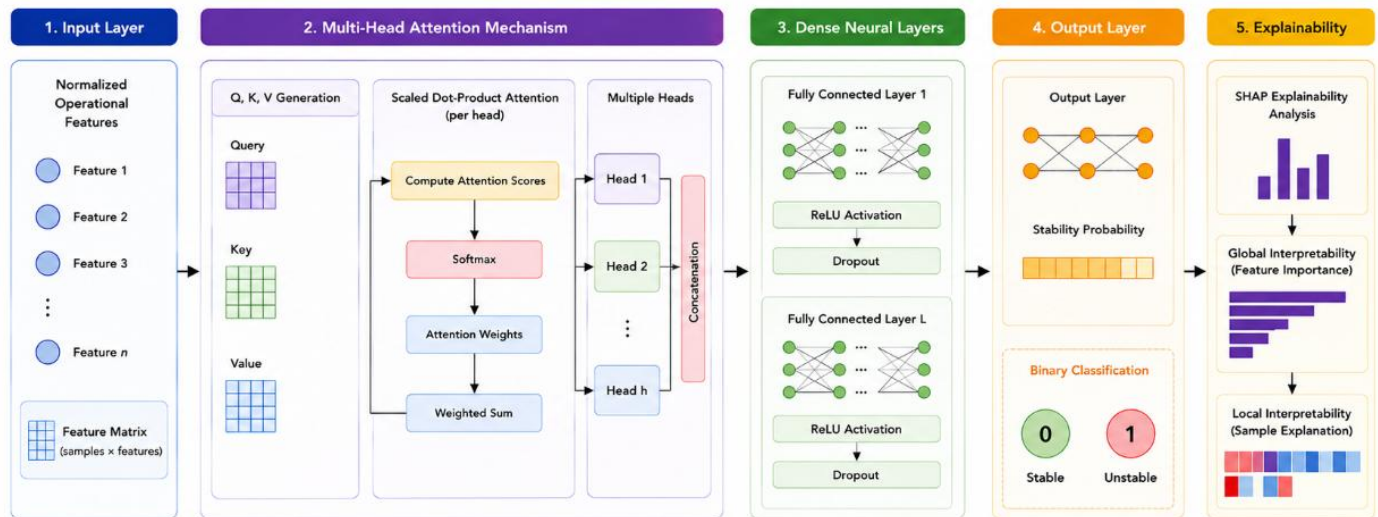


Figure 2. Detailed architecture of the proposed XAI-driven Multi-Head Attention deep learning framework for intelligent power stability classification and explainability analysis in IoE-enabled smart cities.

3.4. Model Training and Optimization

The proposed Multi-Head Attention-based deep learning framework was trained to perform intelligent binary classification of operational stability conditions in IoE-enabled smart city environments. The training process focused on optimizing nonlinear feature representation learning, improving contextual operational dependency extraction, and enhancing classification robustness under dynamic energy conditions.

Following data preprocessing and normalization, the dataset was divided into training and testing subsets using an 80:20 ratio. The training subset was utilized for parameter optimization and contextual feature learning, while the testing subset was reserved for independent model evaluation and generalization analysis.

During the training phase, the proposed architecture learns the nonlinear mapping function between operational input features and stability labels through iterative optimization of trainable parameters. Let the predicted output probability be represented as:

$$\hat{y}_i = f(x_i; \theta) \quad (12)$$

where x_i denotes the input operational feature vector, θ represents the trainable model parameters, $\hat{y}_i$ denotes the predicted probability of instability.

To optimize the classification process, Binary Cross-Entropy (BCE) loss was employed as the objective function due to its effectiveness in binary classification problems. The BCE loss function is mathematically defined as [34]:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (13)$$

where N denotes the number of training samples, y_i represents the ground-truth label, $\hat{y}_i$ denotes the predicted probability.

The Adam optimization algorithm was utilized for parameter updating because of its adaptive learning capability and computational efficiency in deep learning optimization tasks. Adam combines momentum-based optimization and adaptive learning rate adjustment to improve convergence stability during training. The parameter update process can be represented as [35]:

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (14)$$

where η denotes the learning rate, $\hat{m}_t$ and $\hat{v}_t$ represent bias-corrected first and second moment estimates, ϵ is a small numerical stability constant.

To improve generalization capability and reduce overfitting risk, dropout regularization layers were integrated into the dense neural architecture. Furthermore, validation monitoring was employed during training to evaluate model convergence behavior and learning stability across epochs.

The proposed framework was implemented using Python-based deep learning libraries within the Google Colab environment. The model training process was executed using multiple epochs and mini-batch learning to improve computational efficiency and optimization stability. The overall training process aims to minimize classification loss while maximizing monitoring accuracy and contextual operational learning capability within highly dynamic IoE-enabled smart city energy environments. All experiments, including the reported inference time measurements, were conducted using the Google Colab environment with TensorFlow/Keras under the same experimental settings. Table 4 summarizes the hyperparameter configuration utilized for training and optimizing the proposed Multi-Head Attention-based deep learning framework.

Table 4. Hyperparameter configuration of the proposed Multi-Head Attention-based deep learning framework.

Hyperparameter	Value
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Number of Epochs	30
Activation Function	ReLU
Output Activation	Sigmoid
Loss Function	Binary Cross-Entropy
Dropout Rate	0.3
Attention Mechanism	Multi-Head Attention
Number of Attention Heads	4
Data Normalization	Min-Max Normalization
Classification Type	Binary Classification

4. Experimental Results and Discussion

This section presents the experimental evaluation and performance analysis of the proposed XAI-driven Multi-Head Attention-based deep learning framework for intelligent power stability monitoring in IoE-enabled smart city environments. The experimental analysis focuses on evaluating classification performance, contextual operational learning capability, model generalization, and effectiveness explainability under dynamic operational conditions.

The proposed framework was comparatively evaluated against multiple intelligent classification approaches, including Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), XGBoost, and Gated Recurrent Unit (GRU) models. Comparative evaluation was conducted using several performance metrics to ensure comprehensive assessment of monitoring capability and classification robustness.

4.1. Evaluation Metrics

To evaluate the effectiveness of the proposed intelligent monitoring framework, several widely adopted classification performance metrics were utilized, including accuracy, precision, recall, F1-score, and Receiver Operating Characteristic Area Under the Curve (ROC-AUC) [36]. Accuracy measures the overall classification correctness of the

intelligent monitoring framework. Precision evaluates the reliability of instability predictions. Recall measures the capability of the framework to correctly identify unstable operational conditions. The F1-score represents the harmonic meaning between precision and recall.

The ROC-AUC metric evaluates the discriminative capability of the classification model across different decision thresholds. Higher ROC-AUC values indicate stronger separation capability between stable and unstable operational conditions [37].

The utilization of multiple evaluation metrics provides comprehensive assessment of monitoring reliability, instability detection capability, and operational robustness within highly dynamic IoE-enabled smart city environments.

4.2. Comparative Performance Analysis

Table 5 summarizes the comparative classification performance of the proposed framework and baseline intelligent learning models. Figures 3 and 4 present the comparative performance analysis of the evaluated intelligent monitoring models in terms of classification accuracy and ROC-AUC performance.

Table 5. Comparative performance evaluation of intelligent monitoring models.

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression (LR)	0.81	0.76	0.69	0.73	0.785
Random Forest (RF)	0.947	0.94	0.91	0.93	0.939
XGBoost	0.962	0.96	0.93	0.95	0.955
SVM	0.980	0.98	0.97	0.97	0.976
GRU	0.970	0.93	0.99	0.96	0.998
Proposed Model	0.973	0.98	0.95	0.97	0.997

The experimental results demonstrate that the proposed Multi-Head Attention-based framework achieves highly competitive monitoring performance compared with conventional machine learning and advanced deep learning approaches. In particular, the proposed framework achieved an accuracy of 97.35% and a ROC-AUC score of 0.997, indicating excellent discriminative capability between stable and unstable operational conditions.

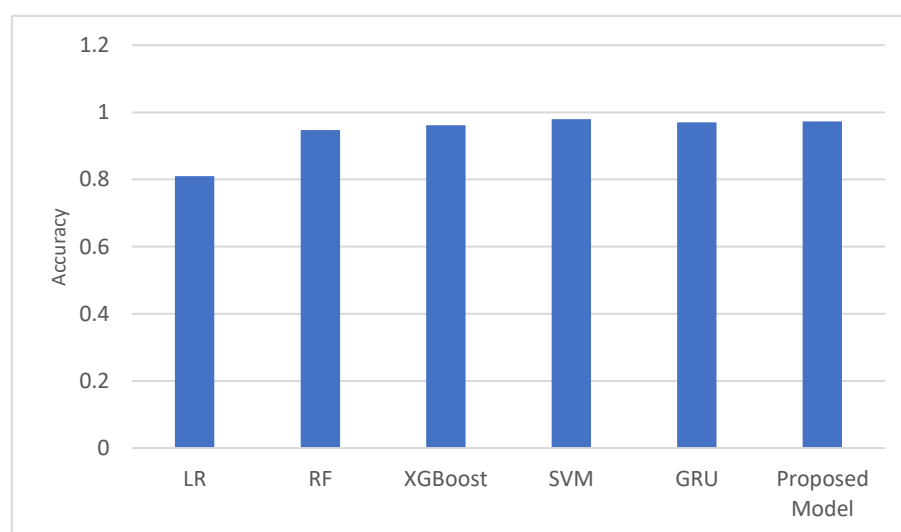


Figure 3. Comparative accuracy analysis of intelligent monitoring models for power stability classification in IoE-enabled smart city environments.

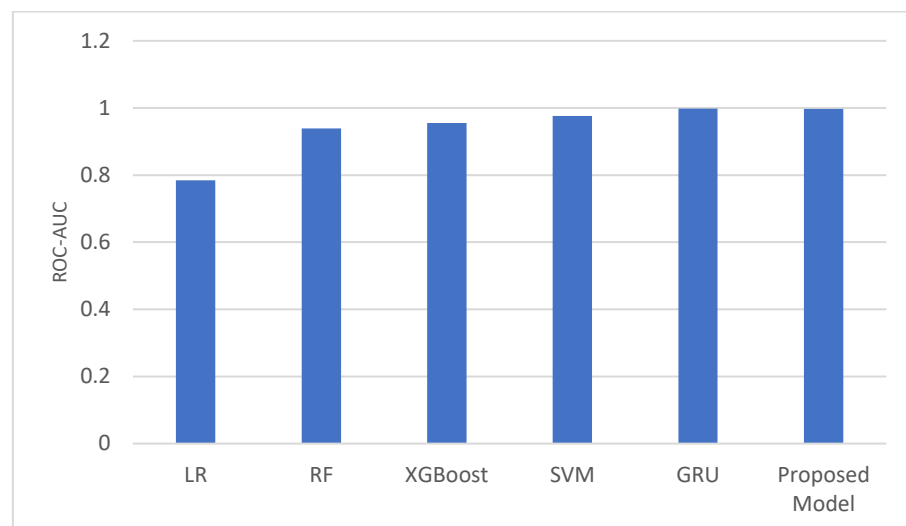


Figure 4. Comparative ROC-AUC performance of intelligent monitoring models for stability discrimination under dynamic operational conditions.

Compared with traditional machine learning approaches such as Logistic Regression and Random Forest, the proposed framework demonstrated substantially improved capability in learning nonlinear operational relationships and contextual feature interactions associated with dynamic stability behavior. Furthermore, the integration of Multi-Head Attention mechanisms significantly enhanced contextual operational dependency learning and instability sensitivity under fluctuating energy conditions.

Although the GRU model achieved a marginally higher ROC-AUC score, the proposed Multi-Head Attention-based framework demonstrated more balanced overall monitoring capability by simultaneously providing high classification performance, contextual operational dependency learning, and SHAP-driven interpretability analysis. Unlike GRU architectures that primarily emphasize sequential dependency modeling, the proposed framework dynamically prioritizes influential operational features through attention mechanisms while offering transparent global and local explainability for intelligent decision-making in safety-critical IoE-enabled smart city environments. Therefore, the proposed framework provides a more practical and trustworthy solution for real-world intelligent power stability monitoring applications.

The comparative analysis also indicates that attention-based contextual learning mechanisms can significantly improve intelligent monitoring capability in highly dynamic IoE-enabled urban energy environments characterized by distributed operational interactions and continuously fluctuating stability conditions.

Although the GRU model achieved comparable predictive performance, the proposed Multi-Head Attention-based framework offers additional advantages beyond classification metrics. Specifically, the Multi-Head Attention mechanism captures contextual interactions among multivariate operational features, while the integrated SHAP analysis provides transparent global and local explanations of the prediction process. Moreover, the computational performance analysis presented in the following section demonstrates that these benefits are achieved with a lightweight architecture and low inference latency, supporting the practical deployment of the proposed framework for real-time IoE-enabled smart city energy monitoring.

To further assess the practical deployment capability of the proposed framework, its computational performance was evaluated in terms of model complexity and inference efficiency. As summarized in Table 6, the proposed framework contains only 9985 trainable parameters with an approximate model size of 117 KB, indicating a lightweight architecture. Furthermore, inference on the complete test set of 12,000 samples required only

0.0408 s, corresponding to an average latency of 0.0034 ms per sample and a throughput of approximately 293,940 samples/s. These results indicate that the proposed framework achieves efficient inference while maintaining high predictive performance, supporting its potential applicability for real-time power stability monitoring in IoE-enabled smart city environments.

Table 6. Computational Performance of the Proposed Multi-Head Attention Framework.

Metric	Value
Trainable Parameters	9985
Model Size	117.02 KB
Test Samples	12,000
Total Inference Time	0.0408 s
Average Latency per Sample	0.0034 ms
Throughput	293,940 samples/s

To further evaluate the classification capability of the proposed Multi-Head Attention-based framework, confusion matrix analysis was conducted to examine the distribution of correct and incorrect predictions under stable and unstable operational conditions. Confusion matrix evaluation provides detailed insights into the monitoring reliability and instability detection capability of intelligent classification systems operating in highly dynamic IoE-enabled smart city environments. Figure 5 illustrates the confusion matrix generated by the proposed framework during testing evaluation.

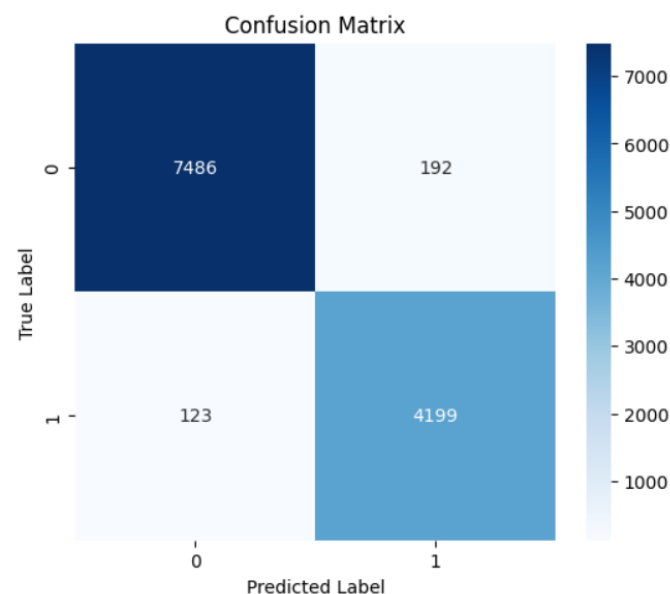


Figure 5. Confusion matrix of the proposed Multi-Head Attention-based intelligent power stability monitoring framework.

The confusion matrix results indicate that the proposed framework achieves highly reliable classification performance for both stable and unstable operational conditions. The model correctly classified 7486 stable samples and 4199 unstable samples, while only a limited number of misclassifications were observed. Specifically, 192 stable samples were incorrectly classified as unstable, whereas 123 unstable samples were classified as stable.

The obtained results demonstrate the strong discriminative capability of the proposed attention-driven architecture in identifying operational instability patterns under

fluctuating energy conditions. Furthermore, the relatively low false-positive and false-negative rates indicate that the framework maintains robust monitoring reliability and balanced classification performance across both operational classes.

From an intelligent energy management perspective, accurate instability identification is essential for preventing abnormal operational conditions and supporting reliable decision-making within critical smart city infrastructures. The confusion matrix analysis therefore confirms the effectiveness of the proposed framework in achieving dependable stability monitoring and intelligent operational assessment under dynamic IoE-enabled urban energy environments.

To further investigate the contribution of the Multi-Head Attention mechanism, an ablation study was conducted by comparing the proposed framework with a baseline deep neural architecture without the attention mechanism. The objective of this analysis is to evaluate the effectiveness of contextual feature prioritization and attention-driven operational dependency learning in improving intelligent power stability monitoring performance within dynamic IoE-enabled smart city environments. Table 7 presents the comparative ablation study results between the baseline deep neural architecture and the proposed Multi-Head Attention-based framework. Figure 6. Comparative ablation analysis between the baseline DNN architecture and the proposed Multi-Head Attention-based framework.

Table 7. Ablation study analysis of the proposed Multi-Head Attention-based framework.

Configuration	Precision	Recall	F1-Score	Accuracy
DNN without Attention	0.97	0.96	0.97	0.970
Proposed Multi-Head Attention Framework	0.98	0.95	0.97	0.973

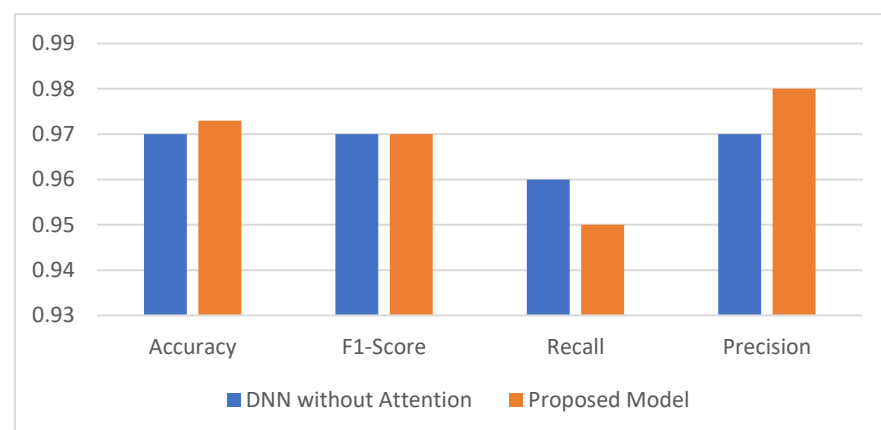


Figure 6. Comparative ablation analysis between the baseline DNN architecture and the proposed Multi-Head Attention-based framework.

The ablation study results confirm the effectiveness of the integrated Multi-Head Attention mechanism in improving contextual operational dependency learning and intelligent stability classification capability. The proposed framework achieved improved overall classification performance compared with the baseline deep neural architecture without attention integration.

Although the performance improvement is relatively moderate, the obtained results demonstrate that the Multi-Head Attention mechanism enhances contextual feature interaction learning and enables more effective prioritization of influential operational parameters associated with stability behavior. Furthermore, the proposed framework additionally provides SHAP-driven explainability capability, thereby improving operational

transparency and interpretable intelligent decision-making within safety-critical IoE-enabled smart city environments.

4.3. Training Performance Analysis

To evaluate the learning behavior and convergence stability of the proposed framework, training and validation performance curves were analyzed during the optimization process. Monitoring the evolution of training accuracy, validation accuracy, training loss, and validation loss provides important insights into model generalization capability, optimization stability, and overfitting behavior. Figures 7 and 8 illustrate the training and validation accuracy and loss curves obtained during model training.

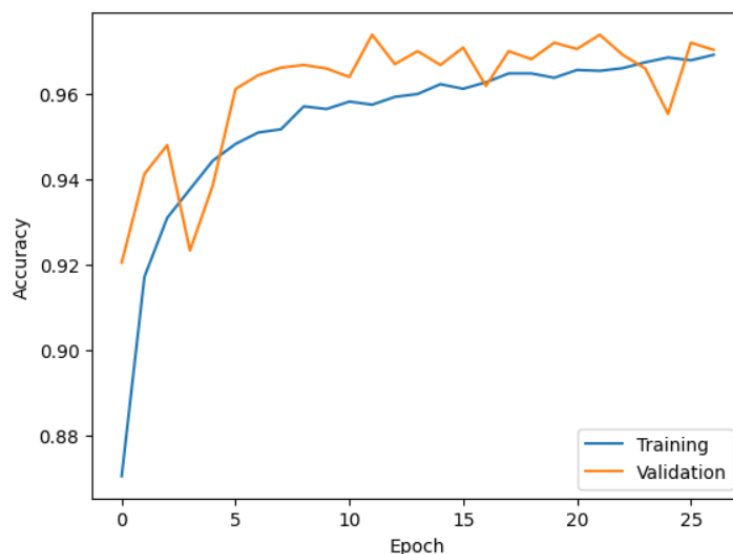


Figure 7. Training and validation accuracy curves of the proposed Multi-Head Attention-based framework.

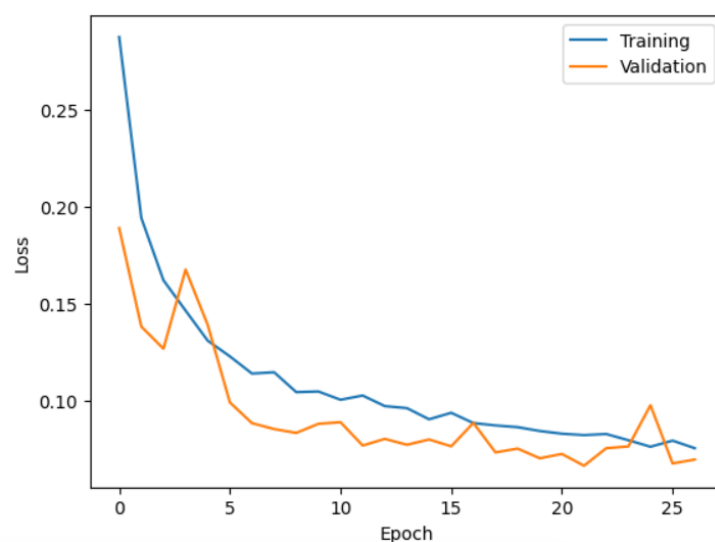


Figure 8. Training and validation loss curves of the proposed Multi-Head Attention-based framework.

The training performance results indicate satisfactory convergence throughout the optimization process. Although the validation accuracy curve exhibits minor oscillations across some training epochs, the overall trend demonstrates progressive learning and convergence without sustained performance degradation or divergence. The training

accuracy consistently improves as the learning process advances, while the validation accuracy remains within a relatively narrow range despite these expected fluctuations. Such behavior is commonly observed in mini-batch optimization and does not necessarily indicate unstable learning. Furthermore, the training and validation loss curves exhibit an overall decreasing trend, confirming effective parameter optimization throughout the training process. The relatively small gap between the training and validation performance suggests that the proposed framework achieves good generalization capability without evidence of significant overfitting. Overall, the obtained learning curves demonstrate that the proposed Multi-Head Attention-based framework converges satisfactorily and provides reliable classification performance for intelligent power stability monitoring in IoE-enabled smart city environments.

The observed convergence behavior further demonstrates the suitability of the proposed attention-based architecture for intelligent monitoring applications in dynamic IoE-enabled smart city environments. The integration of Multi-Head Attention with dropout regularization contributes to robust contextual feature learning while maintaining good generalization performance under varying operational conditions.

4.4. SHAP Explainability Analysis

To improve transparency and interpretability of intelligent monitoring decisions, SHAP-based explainability analysis was conducted to evaluate the contribution of operational parameters toward power stability prediction within IoE-enabled smart city environments. The integration of XAI mechanisms is particularly important for safety-critical urban energy infrastructures, where reliable and interpretable decision-making is required to support intelligent operational management and trustworthy instability analysis.

The proposed framework employs SHAP to provide both global and local interpretability insights for the Multi-Head Attention-based deep learning model. SHAP analysis enables comprehensive evaluation of feature importance by quantifying the contribution of each operational parameter toward model predictions under varying stability conditions.

Global explainability analysis was performed to identify the most influential operational parameters affecting stability behavior across the entire dataset. Figure 9 presents the SHAP summary plot generated by the proposed framework.

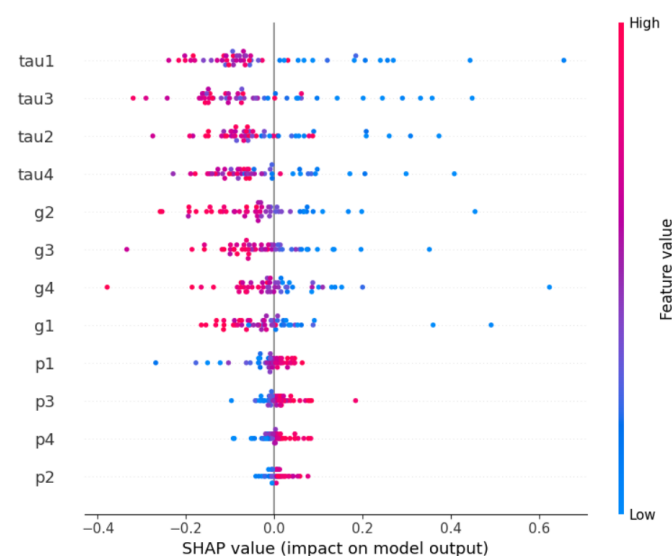


Figure 9. Global SHAP summary illustrates feature importance and operational impact on power stability behavior.

The global SHAP analysis demonstrates that temporal operational parameters, particularly the τ -related features, exhibit the greatest influence on intelligent stability prediction behavior. Features associated with operational response time and dynamic system interaction contribute significantly toward instability classification decisions.

From an engineering perspective, these findings are consistent with established power system stability principles. The reaction-time parameters (τ_1 – τ_4) characterize the dynamic response of generators and system components following operational disturbances. Faster dynamic responses generally improve the ability of the electrical grid to maintain synchronous operation, whereas slower responses increase the likelihood of instability. Likewise, the power-related variables (p_1 – p_4) reflect the balance between generated and consumed power, which is a fundamental determinant of power system stability. The generator-related parameters (g_1 – g_4) also influence the dynamic operating conditions of the electrical grid. Therefore, the SHAP results not only identify statistically important features but also provide explanations that are physically meaningful and consistent with established power system behavior.

Furthermore, the SHAP summary visualization indicates that higher feature values in several temporal operational parameters tend to increase instability probability, whereas lower values contribute toward stable operational conditions. This behavior confirms the capability of the proposed Multi-Head Attention mechanism in effectively capturing contextual operational dependencies and nonlinear feature interactions associated with dynamic urban energy environments.

The obtained results also demonstrate that the proposed explainable framework successfully identifies influential operational patterns contributing to instability behavior, thereby improving monitoring transparency and supporting interpretable intelligent energy management within IoE-enabled smart city infrastructures.

To further investigate prediction-level interpretability, local SHAP analysis was conducted to explain individual operational stability predictions. Figure 10 illustrates the SHAP local interpretability visualization generated for representative operational samples.

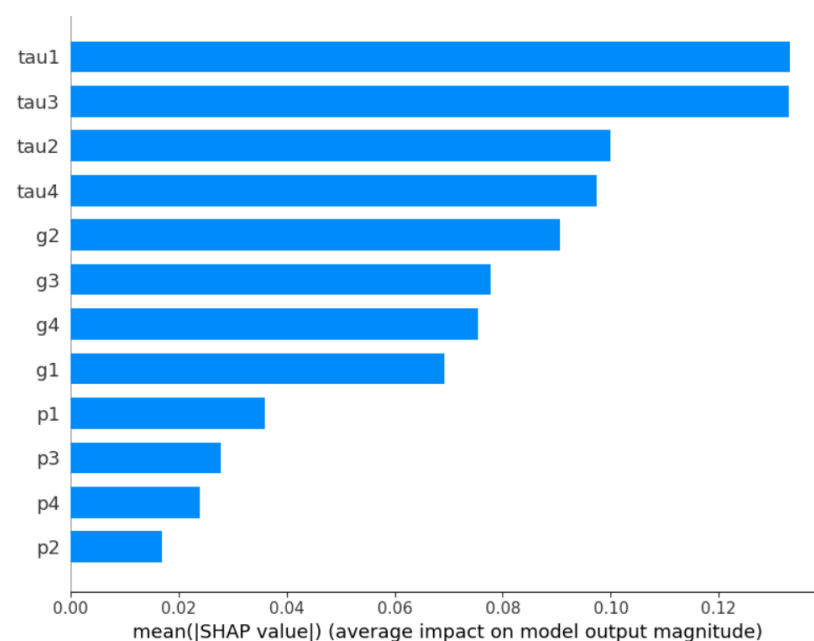


Figure 10. Local SHAP explainability analysis illustrates feature-level contributions toward individual stability predictions.

The local interpretability analysis provides detailed insights into how specific operational parameters influence individual classification decisions. Positive SHAP contributions increase the probability of instability prediction, whereas negative SHAP values support stable operational classification outcomes.

The obtained local explanations improve transparency in intelligent monitoring decisions and enable detailed operational diagnosis for abnormal stability conditions. From an intelligent energy management perspective, local explainability analysis can support early warning systems, operational fault analysis, and interpretable decision support within highly dynamic smart city energy environments.

Overall, the integration of SHAP-based explainability significantly enhances the transparency, trustworthiness, and operational interpretability of the proposed intelligent monitoring framework. The combined utilization of Multi-Head Attention and XAI mechanisms therefore provides an effective solution for intelligent, interpretable, and reliable power stability monitoring in future IoE-enabled smart city infrastructures. Figure 11 (Local SHAP waterfall) illustrates feature-level contributions toward an individual power stability prediction.

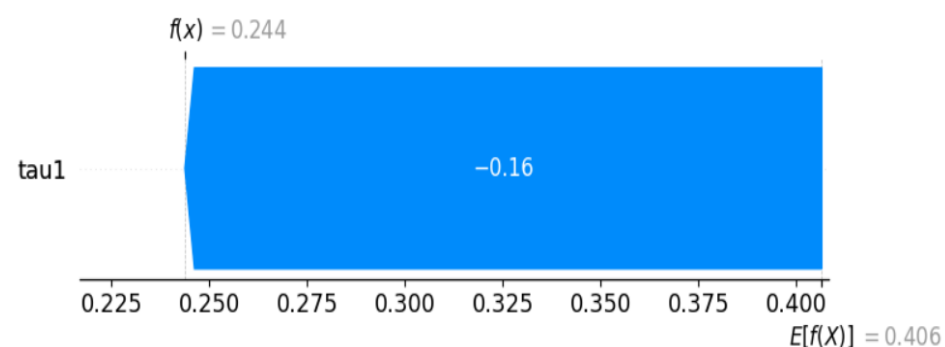


Figure 11. Local SHAP waterfall plot illustrating feature-level contributions toward an individual power stability prediction generated by the proposed Multi-Head Attention-based framework.

The local SHAP explainability analysis provides detailed insight into the operational factors contributing to individual stability predictions. Positive SHAP contributions indicate operational features that increase instability probability, whereas negative SHAP values contribute toward stable operational behavior. The obtained visualization demonstrates that temporal operational parameters exert substantial influence on prediction outcomes, thereby confirming the capability of the proposed framework to provide transparent and interpretable intelligent decision-making for IoE-enabled smart city energy infrastructures.

The SHAP analysis further reveals that the identified influential features are consistent with the expected behavior of electrical power stability systems. In particular, the reaction-time parameters (τ_1 – τ_4) received the highest SHAP importance scores, indicating that the dynamic response of system components plays a dominant role in determining stability. This finding is consistent with established power system principles, where delayed responses to load variations and operational disturbances increase the likelihood of instability. Similarly, the power-related variables (p_1 – p_4) and grid parameters (g_1 – g_4) also contributed to the prediction process, although with relatively lower importance, reflecting their supporting role in maintaining stable operating conditions.

From a practical perspective, these explainability results provide valuable insights for energy operators and decision-makers. Rather than functioning as a black-box predictor, the proposed framework identifies the operational parameters that most strongly influence stability decisions, enabling operators to focus on the most critical factors affecting system reliability. These insights can support the prioritization of system monitoring,

facilitate the early detection of abnormal operating conditions, and assist engineers in developing more effective preventive control and maintenance strategies for IoE-enabled smart city energy systems. Consequently, the proposed framework not only achieves high predictive performance but also provides actionable and trustworthy explanations that enhance confidence in real-world operational decision-making.

4.5. SHAP Sensitivity Analysis

To further evaluate the robustness of the SHAP explanations, a sensitivity analysis was conducted by retraining the proposed Multi-Head Attention-based framework using five different random seeds (1, 10, 25, 42, and 100). For each training run, SHAP values were recomputed using the same experimental configuration, and the resulting feature importance rankings were compared to assess the consistency of the explainability results across different model initializations.

The robustness of the SHAP explanations was quantified using pairwise Spearman rank correlation coefficients between the feature importance rankings obtained from different training runs. As summarized in Table 8, the correlation coefficients ranged from 0.909 to 0.958, indicating a high degree of consistency in feature importance rankings despite variations in random initialization. These results demonstrate that the SHAP explanations are stable and reproducible, providing additional confidence in the interpretability of the proposed framework.

The consistently high Spearman rank correlation coefficients further indicate that the SHAP explanations are not sensitive to random model initialization, thereby demonstrating the robustness, reliability, and reproducibility of the proposed explainability framework.

Table 8. Pairwise Spearman rank correlation coefficients of SHAP feature importance rankings obtained from models trained with different random seeds.

	Seed_1	Seed_10	Seed_25	Seed_42	Seed_100
Seed_1	1.000	0.944	0.930	0.958	0.937
Seed_10	0.944	1.000	0.930	0.909	0.944
Seed_25	0.930	0.930	1.000	0.944	0.916
Seed_42	0.958	0.909	0.944	1.000	0.916
Seed_100	0.937	0.944	0.916	0.916	1.000

5. Discussion

The experimental results demonstrate that the proposed Multi-Head Attention-based deep learning framework achieves highly competitive monitoring performance for intelligent power stability classification within IoE-enabled smart city environments. The integration of attention mechanisms enables effective contextual feature interaction learning and improves sensitivity toward influential operational dependencies associated with instability behavior.

Compared with conventional machine learning approaches, including Logistic Regression, Random Forest, and XGBoost, the proposed framework demonstrates superior capability in capturing nonlinear operational patterns and dynamic contextual relationships within multivariate energy environments. In addition, the attention-driven architecture exhibits strong generalization capability and stable convergence behavior during optimization.

Although recurrent architectures such as GRU achieved competitive performance, the proposed framework provides additional advantages through contextual feature prioritization and integrated SHAP-based explainability analysis. Unlike conventional black-

box deep learning models, the proposed framework offers transparent operational interpretation and feature-level decision analysis, thereby improving trustworthiness and practical applicability in safety-critical smart city infrastructures.

The explainability analysis further confirms that temporal operational parameters play a dominant role in stability prediction behavior. This finding aligns with the dynamic characteristics of modern IoE-enabled urban energy systems, where operational response time and distributed energy interactions significantly affect stability conditions under fluctuating load environments.

Overall, the obtained results confirm the effectiveness of integrating Multi-Head Attention and SHAP-based explainability within a unified intelligent monitoring framework for transparent, adaptive, and reliable power stability analysis in future smart city environments.

6. Conclusions

This paper presented an XAI-driven deep learning framework integrating Multi-Head Attention and SHAP for intelligent power stability monitoring in IoE-enabled power systems supporting smart city energy applications. The proposed framework was developed to address the increasing operational complexity of modern power systems characterized by renewable energy integration, distributed energy resources, interconnected IoT devices, electric vehicle charging infrastructures, microgrids, and highly dynamic operating conditions.

The proposed Multi-Head Attention-based architecture demonstrated strong capability in learning contextual feature interactions and nonlinear operational dependencies associated with power stability behavior. Furthermore, the integration of SHAP-based explainability analysis enhanced model transparency by providing comprehensive interpretation of the operational parameters influencing stability classification, thereby improving the interpretability and trustworthiness of the proposed framework.

Experimental evaluation using the Electrical Grid Stability Simulated Dataset demonstrated that the proposed framework achieves highly competitive classification performance compared with multiple intelligent learning approaches. The proposed model achieved an accuracy of 97.35% and a ROC-AUC score of 0.997, indicating strong discriminative capability between stable and unstable operating conditions. In addition, the SHAP analysis revealed that temporal operational parameters have the greatest influence on stability classification while also providing physically meaningful insights into the factors affecting power system stability.

The experimental findings demonstrate that integrating Multi-Head Attention mechanisms with Explainable Artificial Intelligence techniques enhances intelligent stability monitoring, operational transparency, and decision support for IoE-enabled power system applications, including renewable energy integration, distributed energy resources, microgrids, electric vehicle charging infrastructures, and smart building energy management systems. Moreover, the proposed framework provides a scalable and interpretable solution for real-time stability assessment within these application domains.

Although the present study focuses on binary stability classification, the proposed framework establishes a foundation for future intelligent decision-support systems in IoE-enabled power networks. Future research will extend this work toward stability margin estimation and early warning prediction using real-time sequential power system data, thereby improving its applicability to practical operational decision-making. Additional research directions include real-time streaming-based monitoring architectures, online adaptive learning mechanisms, uncertainty-aware deep learning models, federated learning for distributed IoE environments, and the integration of graph neural networks and

advanced Transformer-based architectures to further enhance intelligent stability analysis in large-scale IoE-enabled power systems.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the author.

Acknowledgments: During the preparation of this manuscript/study, the author used [GPT, 5.5] for the purpose of improving the clarity of writing, enhancing language refinement, and supporting consistency in figure presentation. The author has reviewed and edited the output and takes full responsibility for the content of this publication.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IoE	Internet of Energy
XAI	Explainable Artificial Intelligence
SHAP	SHapley Additive exPlanations
IoT	Internet of Things
DNN	Deep Neural Network
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Networks
BiLSTM	Bidirectional Long Short-Term Memory
GRU	Gated Recurrent Unit
ReLU	Rectified Linear Unit
BCE	Binary Cross-Entropy
LR	Logistic Regression
RF	Random Forest
SVM	Support Vector Machines
DT	Decision Trees
XGBoost	Extreme Gradient Boosting

References

1. Bashiir, A.A. Smart Cities and IOT for Sustainable Urban Development. *J. Biol. Appl. Sci.* **2024**, *3*, 23–27.
2. Hossain, E.; Khan, I.; Un-Noor, F.; Sikander, S.S.; Sunny, M.S.H. Application of Big Data and Machine Learning in Smart Grid, and Associated Security Concerns: A Review. *IEEE Access* **2019**, *7*, 13960–13988. <https://doi.org/10.1109/ACCESS.2019.2894819>.
3. Tripathi, S.; Verma, P.K.; Goswami, G. A review on SMART GRID power system network. In Proceedings of the 2020 9th International Conference on System Modeling and Advancement in Research Trends, SMART 2020, Moradabad, India, 4–5 December 2020; pp. 55–59. <https://doi.org/10.1109/SMART50582.2020.9337067>.
4. Fathollahi, A.; Noor Nateq Alfaisaly, E.A.; Velaga, K.S.; Guo, Y.; Yu, W. Edge AI for Smart Cities: Foundations, Challenges, and Opportunities. *Int. J. Recent Innov. Trends Comput. Commun.* **2024**, *11*, 268–276. <https://doi.org/10.3390/smartcities8060211>.
5. Kumari, N.; Tran, B.; Sharma, A.; Alahakoon, D. A Comprehensive Review on Stability Analysis of Hybrid Energy System. *Sensors* **2025**, *25*, 2974. <https://doi.org/10.3390/s25102974>.
6. Ali, J.S.; Qiblawey, Y.; Alassi, A.; Massoud, A.M.; Mueen, S.M.; Abu-Rub, H. Power System Stability With High Penetration of Renewable Energy Sources: Challenges, Assessment, and Mitigation Strategies. *IEEE Access* **2025**, *13*, 39912–39934. <https://doi.org/10.1109/ACCESS.2025.3546491>.
7. Liu, X.; Liu, R.; Zhang, X.; Abate, A.G.; Nielsen, P.S.; Keles, D. IoT-enabled interoperability for energy-mobility integration: A systematic review of protocols, platforms, and standards. *Appl. Energy* **2026**, *411*, 127611. <https://doi.org/10.1016/j.apenergy.2026.127611>.
8. Aysel, H.I.; Cai, X.; Prugel-Bennett, A. Explainable Artificial Intelligence: Advancements and Limitations. *Appl. Sci.* **2025**, *15*, 7261. <https://doi.org/10.3390/app15137261>.

9. Santos, M.R.; Guedes, A.; Sanchez-Gendriz, I. SHapley Additive exPlanations (SHAP) for Efficient Feature Selection in Rolling Bearing Fault Diagnosis. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 316–341. <https://doi.org/10.3390/make6010016>.
10. Attia, H. Deep Learning-Based Stability Prediction for Electrical Grids: A Data-Driven Approach. In Proceedings of the 17th International Conference on Development in eSystem Engineering (DeSE), Khorfakkan, United Arab Emirates, 6–8 November 2024; pp. 574–578. <https://doi.org/10.1109/DeSE63988.2024.10912018>.
11. Zhang, M.; Li, J.; Li, Y.; Xu, R. Deep Learning for Short-Term Voltage Stability Assessment of Power Systems. *IEEE Access* **2021**, *9*, 29711–29718. <https://doi.org/10.1109/ACCESS.2021.3057659>.
12. Alazab, M.; Khan, S.; Krishnan, S.S.R.; Pham, Q.V.; Reddy, M.P.K.; Gadekallu, T.R. A Multidirectional LSTM Model for Predicting the Stability of a Smart Grid. *IEEE Access* **2020**, *8*, 85454–85463. <https://doi.org/10.1109/ACCESS.2020.2991067>.
13. Omar, M.B.; Ibrahim, R.; Mantri, R.; Chaudhary, J.; Selvaraj, K.R.; Bingi, K. Smart Grid Stability Prediction Model Using Neural Networks to Handle Missing Inputs. *Sensors* **2022**, *22*, 4342. <https://doi.org/10.3390/s22124342>.
14. Divya Singhal, A.S.; Ahuja, L. Forecasting Smart Grid Stability Using a Hybrid CNN Bi-LSTM Approach. *SN Comput. Sci.* **2024**, *5*, 2–8. <https://doi.org/10.1007/s42979-024-02879-1>.
15. Alsirhani, A.; Mujib Alshahrani, M.; Abukwaik, A.; Taloba, A.I.; Abd El-Aziz, R.M.; Salem, M. A novel approach to predicting the stability of the smart grid utilizing MLP-ELM technique. *Alex. Eng. J.* **2023**, *74*, 495–508. <https://doi.org/10.1016/j.aej.2023.05.063>.
16. Liu, H.; Wang, Z.; Liu, Y.; Zhou, J.; Chen, C.; Ma, H.; Huang, X.; Wang, H.; Ji, X. A Transformer-Based Hybrid Neural Network Integrating Multiresolution Turbulence Intensity and Independent Modeling of Multiple Meteorological Features for Wind Speed Forecasting. *Energies* **2025**, *18*, 4571. <https://doi.org/10.3390/en18174571>.
17. Salman, D.; Ali, A.F.; Ali, S.A.; Mohamed, A.S. Enhancing Power Grid Stability through Reactive Power Demand Forecasting Using Deep Learning. *SSRG Int. J. Electr. Electron. Eng.* **2024**, *11*, 170–185. <https://doi.org/10.14445/23488379/IJEEE-V11I12P116>.
18. Ćosić, S.; Vokony, I. Deep Learning-Based Dynamic State Estimation for Frequency Stability Monitoring in Power Systems with High Penetration of Renewable Generation †. *Eng. Proc.* **2023**, *41*, 16. <https://doi.org/10.3390/engproc2023041016>.
19. Al, J.R.T. Attention-Based Deep Learning for Voltage Stability in EV-Integrated Grids. In Proceedings of the 2025 International Conference on Power Electronics and Energy (ICPEE), Bhubaneswar, India, 14–16 December 2025; pp. 1–5. <https://doi.org/10.1109/ICPEE65973.2025.11410980>.
20. Ma, Z.; Mei, G. A hybrid attention-based deep learning approach for wind power prediction. *Appl. Energy* **2022**, *323*, 119608. <https://doi.org/10.1016/J.APENERGY.2022.119608>.
21. Swarnkar, M.S.V. Deep Learning Driven Hybrid LSTM-Transformer Approach for Reliable Stability Prediction in Modern Microgrids. In Proceedings of the 2026 Sixth International Conference on Advances in Electrical, Computing, Communications and Sustainable Technologies (ICAECT), Bhilai, India, 8–9 January 2026; pp. 1–6. <https://doi.org/10.1109/ICAECT68478.2026.11426129>.
22. Ertürk, M.; Emeç, M.; Turhan, M. Hybrid Deep Learning and Transformer-Based Framework for Multivariate Electricity Consumption Forecasting. *Appl. Sci.* **2026**, *16*, 2760. <https://doi.org/10.3390/app16062760>.
23. Breviglieri, P.C. Smart Grid Stability Dataset Kaggle. Available online: <https://www.kaggle.com/datasets/pcbreviglieri/smart-grid-stability> (accessed on 15 July 2026).
24. Franović, B.; Baressi Šegota, S.; Anđelić, N.; Car, Z. Decentralized Smart Grid Stability Modeling with Machine Learning. *Energies* **2023**, *16*, 7562. <https://doi.org/10.3390/en16227562>.
25. Satu, M.S.; Riyad, M.M.H.; Khan, M.I. Machine learning approaches to predict the stability of smart grid. *Frankl. Open* **2026**, *15*, 100542. <https://doi.org/10.1016/J.FRAOPE.2026.100542>.
26. Almarshdi, R.; Nassef, L.; Fadel, E.; Alowidi, N. Hybrid Deep Learning Based Attack Detection for Imbalanced Data Classification. *Intell. Autom. Soft Comput.* **2023**, *35*, 297–320. <https://doi.org/10.32604/iasc.2023.026799>.
27. Elwhishi, A.; Younis, A.A.; Akhunzada, A. Attention-Enhanced Hybrid Architecture for Efficient Intrusion Detection in Industrial IoT. *IEEE Open J. Commun. Soc.* **2026**, *7*, 1330–1347. <https://doi.org/10.1109/OJCOMS.2026.3661768>.
28. Al-Tawil, B.; Jung, M.; Hempel, T.; Al-Hamadi, A. Multi-Head Attention-Based Framework with Residual Network for Human Action Recognition. *Sensors* **2025**, *25*, 2930. <https://doi.org/10.3390/s25092930>.
29. Litman, E. Scaled-Dot-Product Attention as One-Sided Entropic Optimal Transport. *arXiv* **2025**, arXiv:2508.08369.
30. Oikonomou, V.P. MultiHeadEEGModelCLS: Contextual Alignment and Spatio-Temporal Attention Model for EEG-Based SSVEP Classification. *Electronics* **2025**, *14*, 4394. <https://doi.org/10.3390/electronics14224394>.
31. Alshede, H.; Nassef, L.; Alowidi, N.; Fadel, E. Ensemble Voting-Based Anomaly Detection for a Smart Grid Communication Infrastructure. *Intell. Autom. Soft Comput.* **2023**, *36*, 3257–3278. <https://doi.org/10.32604/iasc.2023.035874>.

32. Xu, S.; Li, X.; Xie, C.; Chen, H.; Chen, C.; Song, Z. A high-precision implementation of the sigmoid activation function for computing-in-memory architecture. *Micromachines* **2021**, *12*, 1183. <https://doi.org/10.3390/mi12101183>.
33. Yuan, C.; Cai, Y.; Zhang, Z. Application of SHAP-Based Explainable Machine Learning in Remaining Useful Life Prediction for Aircraft Engine Systems. 2025; p. 48. Available online: <https://www.researchsquare.com/article/rs-7802765/v1> (accessed on 15 July 2026).
34. Chen, Y.; Li, L.; Li, W.; Guo, Q.; Du, Z.; Xu, Z. Fundamentals of neural networks. In *AI Computing Systems*; Morgan Kaufmann: San Francisco, CA, USA, 2024; pp. 17–51. <https://doi.org/10.1016/B978-0-32-395399-3.00008-1>.
35. Shao, Y.; Wang, J.; Sun, H.; Yu, H.; Xing, L.; Zhao, Q.; Zhang, L. An Improved BGE-Adam Optimization Algorithm Based on Entropy Weighting and Adaptive Gradient Strategy. *Symmetry* **2024**, *16*, 623. <https://doi.org/10.3390/sym16050623>.
36. Alshede, H.; Jambi, K.; Nassef, L.; Alowidi, N.; Fadel, E. FedAvg-P: Performance-Based Hierarchical Federated Learning-Based Anomaly Detection System Aggregation Strategy for Advanced Metering Infrastructure. *Sensors* **2024**, *24*, 5492. <https://doi.org/10.3390/s24175492>.
37. Powers, D.M.W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv* **2011**, arXiv:2010.16061.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.