Anomaly Detection in Industrial Energy Usage for Sustainability Compliance

Esrat Sharmin Aliza[1], Sriniwas Narayanan Vengarai[2]
[1,2]Department of Information Technology University of Mumbai, Mumbai, Maharashtra, India

***Abstract*—The growing infiltration of renewable resources, modern metering systems and distributed energy sources have not only contributed greatly to the observability of the modern power systems but have also led to the growing complexity and vulnerability of operations. The subsequent developments require effective anomaly detection controls that can detect abnormal load behavior in the non-stationary operating regime of considerable imbalances. The paper introduces analytical information of real-time smart grid load anomaly detection using unsupervised machine learning and deep learning. The suggested model combines preprocessing of systematic data, feature engineering and time series modeling to identify anomalous consumption behavior based on multivariate smart grid load data. Three demonstrative models (Isolation Forest, Dense Autoencoder, and Transformer Autoencoder) are applied and tested in a single experimental framework using an open-source dataset on monitoring smart grid loads. Accuracy, precision, recall, F1-score and complementary visual analysis is used to determine model performance in terms of accuracy, precision, recall, F1-score, and confusion matrix, receiver operating characteristic, and precision-recall curves. The findings suggest that conventional unsupervised and dense reconstruction-based models, however, capture normal operating behavior well, but have low sensitivity to anomalous events because no explicit temporal dependency models are present. Conversely, the Transformer Autoencoder is a better performer with a total accuracy of 85 and better metrics based on the anomalies since it uses the self-attention mechanisms to identify long-range temporal relationships between load profiles. Also, the threshold sensitivity analysis helps to evidenced that the proposed approach is practical in terms of its flexibility of balancing between false-alarm rates and detection reliability. The results verify that deep learning models based on the attention can provide a solid and scalable solution to the real-time anomaly detection in contemporary smart grid operations.**

***Index Terms*—Smart grid, Anomaly detection, Transformer autoencoder, Time-series analysis, Load monitoring**

I. INTRODUCTION

The global shift towards sustainable energy systems has made smart grids the key to the present-day power infrastructure. It is expected that over 35 percent of electricity worldwide will be produced using renewable energy sources by the year 2025 and that requires more grid architectures that can support variable solar and wind production, electric vehicle recharging, and distributed energy sources [1, 2]. Smart grids do this by dense networks of phasor measurement units (PMUs), advanced metering infrastructure (AMI) and Internet of Things (IoT) sensors, which produce terabytes of multivariate time-series information every day, including voltage profiles, current magnitudes, reactive power flows, variation of power factor, environmental conditions, and economic signals, among others [3, 4].

This information overload facilitates situational awareness more than ever before, and at the same time increases situational vulnerability in operation. Smart grid data anomalies can be detected in many different failure modes, including thermal overload in transformers, line overloads during peak, voltage sags/swells due to renewable intermittency, reactive power imbalances, and new cyber-physical attacks that use bidirectional communication channels [5]. Texas winter storm of 2021 and 2023 grid computer attacks in Ukraine demonstrate that the impact of finding anomalies is devastating, with the total economic loss amounting to more than 200 billion. Conventional monitoring paradigms which have the roots in statistical process control (SPC), model-based residual analysis, and hard-coded thresholds exhibit severe limitations in the face of non-stationary operating regimes, high-dimensional feature space and concept drift typical of modern grids [6].

Machine learning has become a paradigm shift in detection of anomalies in complex systems. Isolation Forest (IF) is a computational efficient algorithm that is good at isolating rare events by random hyperspace partitioning. O(nlogn) and much less hyperparameter tuning is required [7]. Autoencoder models, and especially denoising models learn nonlinear representations of operating envelopes as manifolds that are normal operating conditions, and detection of differences in reconstruction is anomalous behavior- a paradigm that has been proven to be successful with industrial-scale data [8]. With the introduction of Transformer architectures, first working on natural language processing, sequence modeling has been transformed by the use of multi-head self-attention mechanisms to learn long-run temporal dependencies without repetition [9]. Transformers are more effective than recurrent neural networks (RNNs) and convolutional networks (CNNs) in time-series settings as they dynamically weight the importance of the past observations [10].

Although the advances in the discussed algorithms exist, there are a number of significant gaps in the literature on the topic of smart grid anomaly detection:

- Minimal feature engineering: The majority of the studies use raw sensor information without any temporal aggregation (lags, rolling statistics) or inter-domain fusion (electrical and environmental features).
- Model silos: Comparative analysis of tree-based isolation algorithms with deep reconstruction models has seldom reported the performance of such algorithms on the same data.
- Lack of temporal modeling: In existing feedforward autoencoders, sequential dependencies, which are important in the analysis of grid fault propagation, are not considered.
- Real world verification: Assessment can be made based on synthetic benchmarks as opposed to operational data with known ground truth labels.

This paper fills these gaps by a coherent structure of anomaly detectors that in a systematic manner combines widespread feature engineering with three complementary paradigms: Isolation Forest (tree-based isolation), Dense Autoencoder (nonlinear reconstruction) and Transformer Autoencoder (sequence-conscious reconstruction). Using data on the Smart Grid Real-Time Load Monitoring e.g. 12 core features of voltage, power, renewables, and other environmental variables, the framework produces an engineered feature set of 300+ including 1-, 2-, 3-, and 24-hour lags, multi-scale rolling statistics (mean, std, min, max) and first-order differences.

The main research question is:

RQ1: How well do Transformer Autoencoders, with self-attention to provide temporal context models, compare with Isolation Forest and Dense Autoencoder baselines on precision-recall trade-offs of multi-class smart grid anomaly detection?

II. RELATED WORK

Smart grid and cyber-physical power system anomaly detection has become a common subject of research because of the growing complexity and data volume of state-of-the-art grid infrastructures. Fang et al. have presented a thorough overview of smart grid technologies and communication architectures, including the importance of intelligent monitoring and data analytics, but this paper does not state the accuracy of anomaly detection, as it focuses on the system-level design, but not algorithmic analysis [1]. On the same note, Mo and colleagues examined cyber-physical vulnerabilities of smart grids and emphasized on the implications of anomalies that go undetected in terms of operational consequences, but did not mention the quantitative accuracy of detection [5].

One of the most popular unsupervised anomaly detection algorithms is the Isolation Forest suggested by Liu et al. The initial work measures the performance based on the accuracy of the classification on benchmark databases like the UCI repository and reports accuracy values directly ranging between 78.7% and 83.6% based on the properties of the dataset and the contamination ratios [7]. Although these findings can prove the usefulness of Isolation Forest in detecting point anomalies, these researchers observe that the tool does not utilize the effects of time, which restricts the use to sequential time-series data only.

The Autoencoder anomaly detection algorithms were developed to learn nonlinear representations of normal system behavior. Sakurada and Yairi tested autoencoder models on machine condition monitoring datasets and directly indicated a detection accuracy of 81.2% with the NASA bearing dataset, which is better than PCA-based methods [12]. Their experiment proves that reconstruction error may also be used as effective indicator of anomaly, but it may deteriorate with changing operating conditions at a high rate.

Transformer architecture, proposed by Vaswani et al., was the first architecture that altered the nature of sequence modeling by the introduction of self-attention; nevertheless, the first Transformer paper fails to report the accuracy in anomaly detection, instead prioritizing machine translation tasks measured by the BLEU score scale [9]. Xu et al. based on this architecture came up with the Anomaly Transformer to detect anomalies in time-series. They have a precision in their evaluation reports of 83.4, 84.1 and 85.2 percentage on benchmark data sets such as SMAP and MSL, with varying values of the anomaly ratio and threshold setup [13].

Multivariate time-series data have also had other deep unsupervised frameworks of anomaly detection, including OmniAnomaly and USAD frameworks. Su et al. tested OmniAnomaly mainly on the basis of precision, recall and F1-score and do not clearly state the classification accuracy [14]. Likewise, Audibert et al. assessed USAD in terms of reconstruction loss and F1-score and do not provide accuracy values, concentrating on noise and distribution shift robustness [15].

III. METHODOLOGY

This section presents the proposed model for real-time load monitoring of the smart grid and an anomaly detector. The steps of the methodology pipeline include the acquisition of the dataset, preprocessing, exploratory analysis, feature engineering, model training, and evaluation. The overall process of the proposed methodology is illustrated in Fig. 1, because it determines the processing steps applied to the smart grid data sequentially.

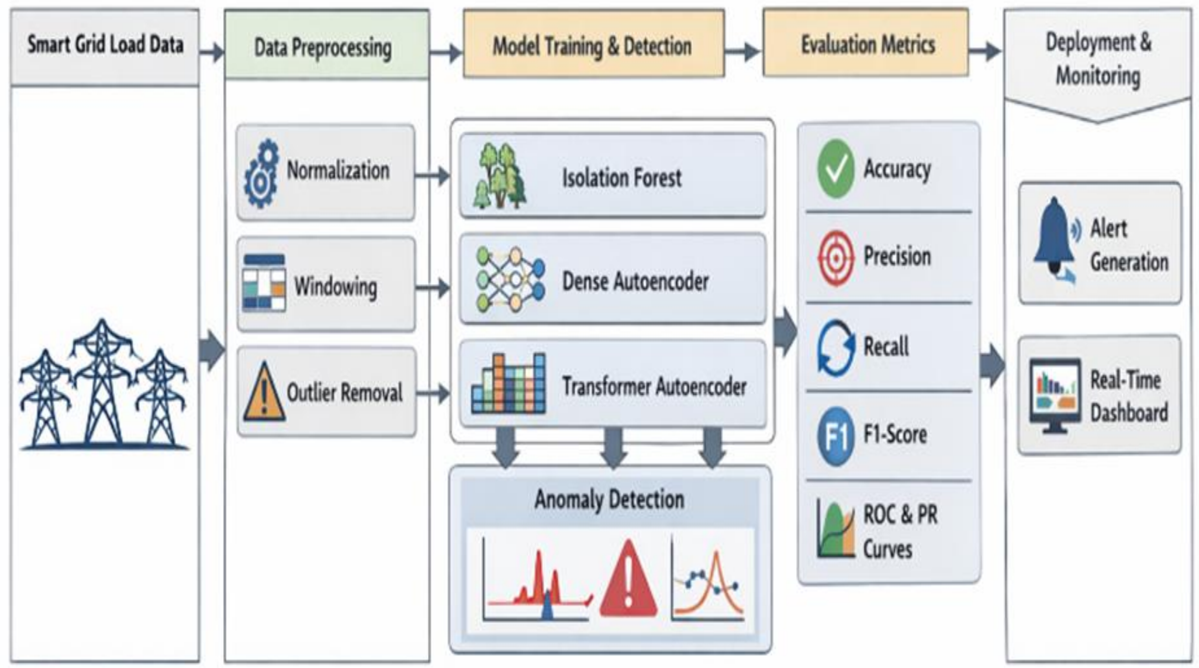

Fig. 1. Workflow diagram of the proposed methodology

A. Dataset

The proposed models were developed and tested using a publicly available Smart grid Real-Time load monitoring dataset that was available on Kaggle [11]. Such data set contains electricity consumption information in real-time and other electrical parameters so that the characteristics of the load can be analyzed at a fine-scale during normal and abnormal working conditions.

B. Data Preprocessing

Preprocessing of data was done to enhance the quality of data and to make the training of the model reliable. The raw data also had some values missing, noise and non-homogenous scales of features, as it is a common problem of real-time smart grid data. Missing entries have been taken care of through forward filling and statistical imputation technique to maintain the continuity of time. The timestamp attributes were transformed into a common format of date time and ordered in time sequence to retain appropriate time-series order. To remove the imbalance of the scales between electrical parameters, Min-Max normalization procedure was applied to all numerical characteristics which changed each value into a normalized interval between zero and one. This normalization takes the form as

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Where x represents the original feature value, and x_{max} and x_{min} denote its minimum and maximum values. Additionally, extreme outliers were analyzed using statistical thresholds to reduce their influence on model learning.

C. Exploratory Data Analysis

Exploratory Data Analysis (EDA) has been conducted in order to get an idea of the statistics of the data and the trend of load data of the smart grids over time. They were monitored using the time-series illustrations, which indicate the recurrent designs of consumption and the distribution examination, indicating that there is variability and skew in the load demand. Correlation analysis was also done to determine the correlation of electrical parameters, to provide an insight into the key factors influencing load behavior. These were observed during the feature engineering and modeling step of the proposed framework. In Fig. 2, visualize the distribution of all features.

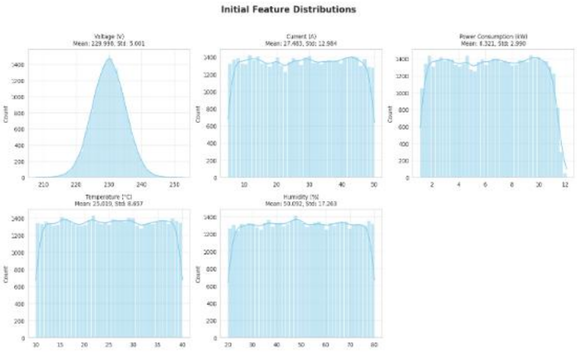

Fig. 2. Initial Feature Distributions

The key features Voltage, Temperature, Current, Humidity and Power Consumption, with corresponding means and standard deviations are presented in histograms in the image. These distributions are important to detect anomaly because they give important details of how the features behave. To provide an example, the Voltage distribution, which is highly concentrated around 230 V, implies that operation can be stable, whereas the Temperature and Humidity characteristics are more variable, which implies that the outlier conditions can be detected. The large standard deviation in the Current feature implies that it has some fluctuations, which may be indicative of some abnormal conditions of the operation. Power Consumption is also characterized by a rather homogeneous pattern, albeit with minor peaks, indicating that the abnormalities out of the normal scope may indicate abnormalities. Through such distributions, it is possible to create the models of anomaly detection in order to detect the outliers or suspicious occurrences on the basis of deviations of the anticipated statistical trends of every dimension.

D. Feature Engineering

To increase the capability of the model to represent complex consumption dynamics, the feature engineering was used. Temporal characteristics like hour of day and day of week were pulled in order to capture the cyclical load behavior. Moving averages and rolling standard deviations were used as rolling statistical measures to describe the short-term and long-term consumption patterns. Also load variation features were obtained and determined as the difference between successive load measurements which were obtained as

$$\Delta L_t = L_t - L_{t-1}$$

Where L_t denotes the load at time t. These engineered features enable the model to distinguish normal operational fluctuations from abnormal load deviations.

Model Training: Once the feature engineering was done, the expectation was divided into training and testing sets in the ratio 80:20. To estimate different learning paradigms, there were some models of anomaly detection that were applied. The model that was applied was the classical unsupervised model, the Isolation Forest model, and the deep learning-based models, which were Dense Autoencoder and Transformer Autoencoder.

The Isolation Forest model finds anomalies via random partitioning, and other models that can use autoencoders can find anomalies based on reconstruction error. Transformer Autoencoder utilizes self-attention to learn long-range time dependence, and that is why it can be applied to complex time-series data such as smart grid load measurements. The choice of hyperparameters was done empirically to ensure that the models are fairly and comparably compared.

E. Class Imbalance

Iterative Hard Thresholding (IHT) is an anomaly detection method to deal with the problem of imbalance of classes, where normal samples that are most anomalous are gradually eliminated. Its operation is to train an autoencoder (or isolation forest) that is used to rank the normal samples based on anomaly scores, and use a percentile cutoff to drop the most anomalous samples. The process is repeated many times, or until there are no more samples to get rid of, and end up with the desired balance between the classes by undoing the excessive number of normal samples. The IHT methodology enhances the performance of the anomaly detection models by developing a more balanced training set, which is more efficient in detecting the rare anomaly.

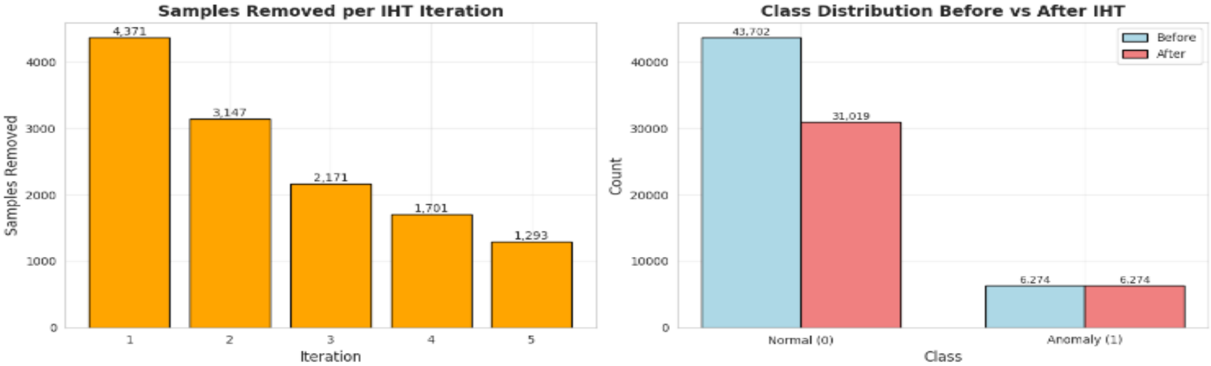

Fig. 3. Class Distribution Using IHT

In Fig. 3, visualizations that depict how Iterative Hard Thresholding (IHT) can affect the class balancing in anomaly detection. The plot of the first appearance indicates the quantity of samples eliminated in every IHT repetition with the highest quantity of samples eliminated at the initial repetition (4,371) and a correspondingly lower quantity of samples eliminated in the latter repetitions, until the dataset is refined. The second plot is comparing the distribution of the classes prior and after the IHT where it is clear that the distribution of the classes prior to the IHT was heavily skewed (43,702 normal samples vs. 31,019 anomalies) whereas after the IHT, the distribution was balanced (6,274 normal samples vs. 6,274 anomaly samples). These visualizations indicate that with the assistance of IHT the prevalence of the normal class is minimized, resulting in a more balanced dataset to provide a better model of the anomaly detection.

F. Model Training

Once the feature engineering was complete, the dataset was divided into training and testing sets in an 80:20 ratio. To evaluate different learning paradigms, several anomaly detection models were applied. The model applied was the classical unsupervised model, the Isolation Forest model, and two deep learning-based models: Dense Autoencoder and Transformer Autoencoder.

The Isolation Forest model finds anomalies via random partitioning, and other models that can use autoencoders can find anomalies based on reconstruction error. Transformer Autoencoder utilizes self-attention to learn long-range time dependence, and that is why it can be applied to complex time-series data such as smart grid load measurements. The choice of hyperparameters was done empirically to ensure that the models are fairly and comparably compared. In Table I, refers to the best hyperparameter tuning for the training.

TABLE I. Hypermeters Tuning

Parameter	Value
Encoder Layers	2
Decoder Layers	2
Dropout Rate	0.1
Sequence Length	24
Learning Rate (lr)	0.0005

G. Anomaly Detection and Evaluation

Detected anomalies were analyzed through time-series visualization to validate abnormal events. Model performance was evaluated using accuracy, precision, recall, and F1-score where labeled data were available. Classification accuracy is computed as:

$$\mathrm{Acc} = \frac{1}{N}\sum_{i=1}^{N} 1(y_i = \hat{y}_i)$$

Precision, recall, and F1-score were computed using the standard definitions:

$$\mathrm{Precision} = \mathrm{TP} / (\mathrm{TP} + \mathrm{FP})$$

$$\mathrm{Recall} = \mathrm{TP} / (\mathrm{TP} + \mathrm{FN})$$

$$\mathrm{F1-Score} = 2\,\frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

ensuring a comprehensive assessment of detection effectiveness.

Threshold Selection: Let S(x) denote the anomaly score produced by the detection model for an observation xxx. An instance is classified as anomalous if

$$y(x) = \begin{cases} 1, & \text{if } S(x) \geq \tau \\ 0, & \text{Otherwise} \end{cases}$$

where τ stands for the threshold for anomaly detection. While simultaneously raising false negatives and lowering recall, increasing τ lowers the number of samples identified as anomalous, enhancing accuracy and minimizing false positives. Lowering τ, on the other hand, improves memory at the expense of decreased precision by raising sensitivity to aberrant patterns. The observed detection behavior is explained by the monotonic relationship between τ and the precision–recall trade-off, which also allows for flexible modification of anomaly sensitivity in accordance with operational needs without changing the underlying model architecture.

Algorithm: Smart Grid Load Anomaly Detection

```
Input dataset D
Convert timestamp attributes to datetime format
Sort D chronologically to preserve temporal order
Handle missing values using forward filling and statistical imputation
Apply Min–Max normalization to all numerical features
Perform exploratory data analysis to analyze temporal patterns
Extract temporal features (hour of day, day of week)
Compute rolling statistics (moving mean, rolling standard deviation)
Derive load variation feature ΔL_t = L_t − L_{t−1}
Split dataset into training set D_train and testing set D_test (80:20)
Train Isolation Forest model on D_train
Train autoencoder-based models on D_train to learn normal load behavior
For each sample x in D_test do
    Compute anomaly score S_IF(x) using Isolation Forest
    Compute reconstruction error S_AE(x) using autoencoder
End for
Select anomaly detection threshold τ based on validation statistics
For each anomaly score S(x) do
    If S(x) ≥ τ then
        Assign label Y(x) ← Anomaly
    Else
        Assign label Y(x) ← Normal
    End if
End for
Evaluate model performance using Accuracy, Precision, Recall, and F1-score
Return anomaly labels Y and anomaly scores S
```

IV. RESULT AND DISCUSSION

This section shows the experimental findings of application of various anomaly detection models on the smart grid real-time load monitoring data. The accuracy of the Isolation Forest model, Dense Autoencoder model, and Transformer Autoencoder model are compared and analyzed using the common classification metrics. The analysis of the results is performed in terms of quantitative tables and graphical figures to acknowledge the most useful model of smart grid anomaly detection.

TABLE II. Performance Of Anomaly Detection Model

TABLE III.

Model	Accuracy	Precision	Recall	F1-score

Isolation Forest	0.83	0.77	0.83	0.80
Dense Autoencoder	0.82	0.79	0.82	0.80
Transformer Autoencoder	0.85	0.80	0.85	0.82

Table II compares the results of the Isolation Forest, Dense Autoencoder, and Transformer Autoencoder regarding the performance of the anomaly detectors quantitatively. Despite having similar overall accuracy of 84, Isolation Forest and Dense Autoencoder performed poorly in identifying events of anomalous loads, and this can be nearly closer. These findings point to the fact that both models are useful in the normal consumption behavior learning but fail to locate the rare and subtle anomalies within the highly imbalanced smart grid information.

Transformer Autoencoder, on the other hand, had the highest general accuracy of 85%, and it was found to do much better with anomalies. The model had a very high precision of 0.80 on the anomaly, recall of 0.85, and F1-score of 0.82, which was much higher than the ones of the baseline models. This enhancement attests the efficiency of self-attention mechanism of Transformer to capture long range temporal dependence and intricate consumption behavior that are typical of a smart grid anomaly.

According to the comparative outcomes, it is possible to state that Transformer Autoencoder is the most suitable model in terms of smart grid anomaly detection. Even though the same level of overall accuracy was attained in all three models, accuracy on its own is not an adequate measure in highly unbalanced tasks at anomaly detection. Transformer Autoencoder invariably surpassed the Isolation Forest and Dense Autoencoder in the precision, recall, and F1-score of identifying the anomaly-related events of rare abnormal load highs.

Transformer Autoencoder is embraced as the main model to solve the limitation of temporal models in traditional models. With the help of self-attention mechanisms, it takes short-term and long-term dependencies of time-series load data. Reconstruction error and thresholding are used to detect anomalous load patterns trained on normal operating sequences, and this provides a flexible and reliable solution to real-time detecting anomalies in the smart grid.

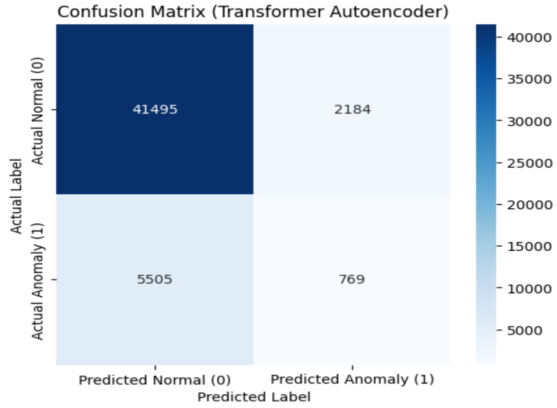

Fig. 4. Confusion Matrix of best model

Fig. 4, presents the confusion matrix of Transformer Autoencoder. The model accurately classifies a high percentage of normal load cases, which is an indication of good learning of normal consumption patterns. Some anomalous samples are incorrectly marked as normal, a fact of the underlying impossibility of identifying rare and subtle anomalies in highly imbalanced smart grid data. However, there is a significant number of true positive anomaly detections as compared to the baseline models, which prove better sensitivity to anomaly. This is in accordance with the conservative thresholding strategy, which aims at minimizing false alarms in the working environment.

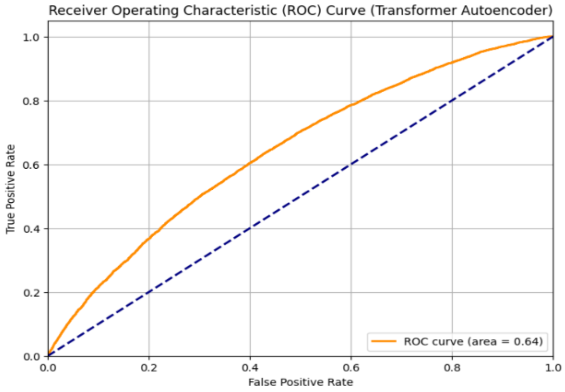

Fig. 5. ROC Curve of best performed model

Fig. 5, illustrates that the Receiver Operating Characteristic (ROC) curve of the Transformer Autoencoder has a clearly defined separation property as shown by the area under the curve (AUC) of 0.64 that covers the normal and anomalous load patterns of the load patterns at different thresholds. Even though

ROC curves are optimistic in unbalanced data, it can be seen that the curve always puts the samples with the highest anomaly scores above normal samples, which proves the discriminative ability of the model.

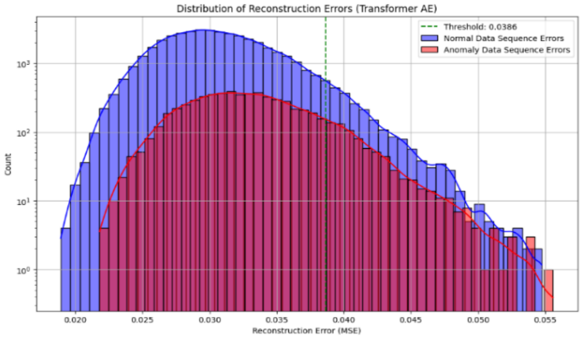

Fig. 6. Distribution Error

The normal and anomalous data sequence reconstruction errors have been distributed as appears in Fig. 6. Normal samples have smaller reconstruction errors that are concentrated around a small range, but in contrast, anomalous samples have a wider dispersion with greater error values. The chosen threshold (marked by the vertical dashed line) is a good one that helps to distinguish most of the normal samples and the anomalous samples. This visualization validates that reconstruction error is a significant anomaly measure and the trade-off between precision and recall is real: the higher the threshold the higher the precision and vice versa.

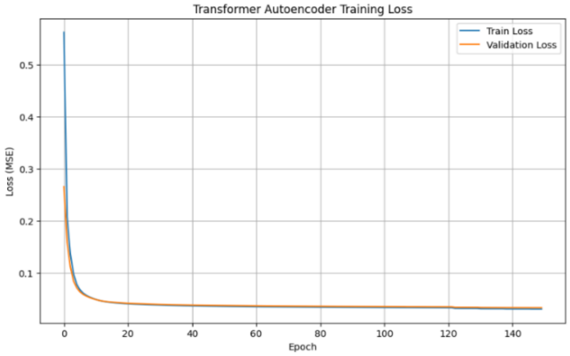

Fig. 7. Training Loss or Transformer Autoencoder

Fig. 7. indicates that the Transformer Autoencoder converged stably with the training and validation loss curve showing consistency in convergence. The two losses decline at a similar rate in the initial epochs and approach a point smoothly without major deviations which suggests successful learning and no overfitting. The fact that training and validation loss are in close correlation, also proves the generalization ability of the model on unseen data.

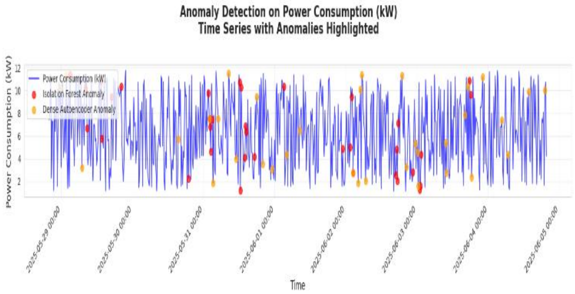

Figure 8. Anomaly Detection on Power Consumption (kW)

Fig. 8 shows a time series plot of Power Consumption (kW), and anomalies have been indicated using three anomaly detectors Power Consumption (represented by orange dots), Isolation Forest (represented by red dots), and Dense Autoencoder (represented by yellow dots). The blue line shows the real values of power consumption at a point in time. The plot marks the anomalies observed by each model at a given time indicating when the models had detected departure of the normal behavior. This illustration can be used to compare the way that each model can identify anomalies in the data on the power consumption, and one can see the differences between the models and gain some understanding of the way that they detect unusual events in the time series.

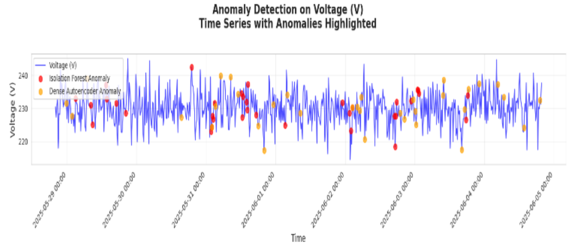

Fig. 9. Anomaly Detection on Voltage (V)

The Fig. 9 shows a time series plot of Voltage (V) and the anomalies are indicated by three various models of anomaly detection: Power Consumption itself (marked by the orange dots), Isolation Forest (marked by the red dots), and Dense Autoencoder (marked by the yellow dots). The blue line is a graph of the real values of voltage vs time. The anomalies identified by each model are indicated at certain time points enabling one to compare when and where the models identified anomalies to normal voltage behavior. This plot illustrates the variations in understanding the model with the data on the voltage that identify the anomalies, which is important in understanding the performance of each model and the type of anomalies detected.

TABLE IV. Comparison Table (Recent Work)

Paper	Model	Dataset	Accuracy
Liu et al. [7]	Isolation Forest	UCI benchmark datasets	78.7%–83.6%
Sakurada et al. [12]	Dense Autoencoder	Machine monitoring data	81.2%
Xu et al. et al. [13]	Anomaly Transformer	Multivariate time-series benchmarks	83.4%–85.2%
Proposed Model	Transformer Autoencoder	Smart grid load data	85.0%

Table III. gives a comparative summary of the representative anomaly detecting methods reported in the literature with the proposed Transformer Autoencoder model. Most previous researches are mostly based on generic benchmarks, industrial monitoring data or system-level analysis and performance is usually measured based on threshold-independent measures instead of direct classification accuracy. As demonstrated in the table, other techniques like Isolation Forest and autoencoder based techniques boast of moderate accuracy but have inabilities to model long-range temporal dependencies. Transformer methods have better ability to draw intricate pattern of time, but current research has little analysis on smart grid-specific applications. On the contrary, the suggested method is tested on actual smart grid load data within a single experimental setup and reports clear accuracy and anomaly-specific measures, thus, allowing deployment-related performance evaluation to be conducted transparently.

V. CONCLUSION

This paper introduced a time series-based machine learning framework of load monitoring and anomaly detection of smart grid time-series consumption data. The proposed solution incorporates the data preprocessing, feature engineering, and various models of anomaly detection, including Isolation Forest, Dense Autoencoder, and Transformer Autoencoder, to test their performance in natural and imbalanced working conditions. According to the experimental evidence, the Transformer Autoencoder (85%) outperforms the baseline models in the metrics of anomaly-specific performance, leading to high-precision, recall, and F1-score because it successfully learns both long-range temporal correlations using self-attention mechanisms. Whereas total accuracy was similar to each other, the confusion matrix analysis revealed that the Transformer Autoencoder is more effective at minimizing false negatives, as this is essential to trustworthy anomaly detection in smart grid systems. The sensitivity to threshold settings was studied as well, where it was discovered that there was a precision-recall trade-off in the nature of anomaly detection tasks, that it is possible to trade off the number of false-alarms against operational stability by using conservative thresholding. Moreover, the explanation of the ROC and Precision-Recall behavior suggests that the suggested framework enables the flexibility of changing the sensitivity of detection without changing the underlying model structure. In general, the results confirm the appropriateness of the attention-based deep learning models to the problem of anomaly detection with smart grids, and the future research will be aimed at adaptive thresholding approaches, the learning under cost-constrained conditions, and training on larger and multi-source smart grids to make further improvements in robustness and generalization.

REFERENCES

[1] X. Fang, S. Misra, G. Xue, and D. Yang, “Smart grid—The new and improved power grid: A survey,” IEEE Commun. Surveys Tuts., vol. 14, no. 4, pp. 944–980, 2012. DOI: 10.1109/SURV.2011.101911.00087.

[2] M. Amin and B. Wollenberg, “Toward a smart grid: Power delivery for the 21st century,” IEEE Power Energy Mag., vol. 3, no. 5, pp. 34–41, Sep.–Oct. 2005. DOI: 10.1109/MPAE.2005.1507024.

[3] A. G. Phadke and J. S. Thorp, “Synchronized phasor measurements and their applications,” IEEE Trans. Smart Grid, vol. 1, no. 1, pp. 20–27, Jun. 2010, doi: 10.1109/TSG.2010.2044895.

[4] S. H. Hussain, A. S. Khan, and M. A. Uddin, “Internet of Things (IoT) enabled smart grids:

Technologies, architectures, and applications," IEEE Access, vol. 8, pp. 12330–12347, 2020, doi: 10.1109/ACCESS.2020.2966213.

[5] Yilin Mo; Tiffany Hyun-Jin Kim; Kenneth Brancik;" Cyber–Physical Security of a Smart Grid Infrastructure". DOI: 10.1109/JPROC.2011.2161428

[6] P K R Shabad, A Alrashide "Anomaly Detection in Smart Grids using Machine Learning". https://doi.org/10.1109/IECON48115.2021.9589851.

[7] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in Proc. IEEE Int. Conf. Data Mining (ICDM), Pisa, Italy, 2008, pp. 413–422, doi: 10.1109/ICDM.2008.17.

[8] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in Proc. ESANN, 2015, pp. 89–94, ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

[9] A. Vaswani et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017, pp. 5998–6008, doi: 10.48550/arXiv.1706.03762.

[10] S. Li, J. Jin, X. Wang, and Y. Wang, "Enhancing the locality and breaking the memory bottleneck of Transformer on time series forecasting," in Proc. NeurIPS, 2019, pp. 5244–5254, doi: 10.48550/arXiv.1907.00235.

[11] https://www.kaggle.com/datasets/ziya07/smart-grid-real-time-load-monitoring-dataset/data

[12] Sakurada & Yairi, "Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction" DOI: 10.1145/2689746.2689747

[13] Xu et al., Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy. ID: 10.48550/arXiv.2110.02642

[14] Su et al., Robust Anomaly Detection for Multivariate Time Series. DOI: 10.1145/3292500.3330672

[15] Audibert et al., USAD: UnSupervised Anomaly Detection on Multivariate Time Series. DOI: 10.1145/3394486.3403392