Tube Loss-based Detection of Anomalies in Electricity Consumption Time Series Data

Shruti Mehta*, Pritam Anand*, and Yash Vasavada* (Senior Member, IEEE)
Dhirubhai Ambani University, Gandhinagar, 382007, India
Email: shruti_mehta@dau.ac.in, pritam_anand@dau.ac.in, yash_vasavada@dau.ac.in

***Abstract*—Anomaly detection in electricity consumption data is challenging due to its non-stationary and highly variable characteristics. In this work, we propose a Tube-based deep learning (DL) framework that learns an adaptive prediction interval (PI) without the constraint that the underlying data has a specific parametric form. The proposed method is based on a novel Tube loss function that can be integrated into various DL models to generate an adaptively optimized PI. Compared to existing Probabilistic Load Forecasting (PLF) models, the Tube Loss-based method (i) naturally yields a narrow PI since it explicitly minimizes the PI width in the optimization objective, and (ii) is simpler as it directly estimates the upper and lower bounds of the PI using a unified approach. Furthermore, for anomaly detection, our proposed approach is simpler, as it does not require labeled data; instead, a specific data point is classified as anomalous if it falls outside the PI determined using the Tube-based DL model. Empirical results on the LEAD dataset show that the proposed Tube-based method outperforms conventional baselines. Additionally, we analyze the effect of PI coverage, demonstrating its role in controlling detection sensitivity. The findings highlight the effectiveness of the proposed Tube-based learning for reliable anomaly detection in chaotic time series of electricity consumption.**

I. Introduction

Artificial Intelligence (AI) is increasingly being adopted for Electricity Load Forecasting (ELF) in Smart Grids. The ELF helps utility companies balance supply and demand in real time, facilitates predictive maintenance and grid optimization, and provides insights for decision-making processes [1]. In this paper, we propose a novel ELF method and apply it to detect anomalies in electricity consumption time series. Such anomalies may occur due to sensor malfunctions, grid overloads, unexpected power outages, electricity theft caused by unauthorized or illegal connections to the grid, etc. [2]. Detecting such anomalies is critical for (i) ensuring grid reliability, operational stability, and efficient energy management by utilities, (ii) enabling fair dynamic pricing through accurate consumption profiling, thereby promoting efficient energy use and potential cost savings for consumers [3], and (iii) reducing energy waste by avoiding unnecessary activation of inefficient and polluting peaker plants, thus supporting environmental sustainability.

* Equal contribution.

This project is funded by Smart Energy Learning Center, Dhirubhai Ambani University, Gandhinagar, under the CSR grant no. CSR-25/BSES/A8-YV/SELC.

Related Works: A major challenge in anomaly detection (AD) is the scarcity of high-quality annotated datasets. This leads to a heavy reliance on unsupervised or semi-supervised methods for AD [4]. Tukey's Fence method is a widely used (unsupervised) statistical method for detecting outliers [5] that uses interquartile range (IQR)-based thresholds. In [6], the authors used Tukey's Fence method to label the anomalies in Low Carbon London and Solar Home Electricity datasets, and used these 'pseudo' labels to train their proposed Feature Guidance Network (FGNet) for AD. Similarity-based measures, such as KNN (K Nearest Neighbors), detect anomalies based on the distance of a data point to its nearest neighbors. In [7], the authors apply KNN, SVM (Support Vector Machine), Logistic Regression, and MLP (Multilayered Perceptron) for AD in fuel consumption data. Authors in [8] extract micromoment-based features and apply an improved KNN algorithm for detecting anomalies in buildings. In [9], the authors use KNN, LOF (Local Outlier Factor), and IF as baseline methods to detect anomalies in large-scale commercial building energy consumption data. Another class of unsupervised methods includes Deep Learning (DL) based methods which are increasingly becoming popular due to their ability to handle the inherent complex patterns and long-term dependencies in time series data. Among them, AD with Auto-Encoders (AE) is a frequently used approach [10], [11], [12], [13], [14], [15]. In this paper, we have evaluated our proposed approach against two AD schemes with AE, specifically (i) a Convolutional Long Short-Term Memory Auto-Encoder (ConvoLSTM-AE) proposed in [10] for detecting anomalies in video sequences, learning both spatial and temporal patterns, (ii) a Variational AE (VAE) developed in in [12] to learn normal energy usage patterns of household appliances and flag deviations as anomalies.

Contributions:

1) We propose a novel Tube-based AD framework that leverages the DL models to capture temporal dependencies in the electricity consumption time series. Unlike traditional statistical approaches, the proposed method operates within a probabilistic forecasting paradigm, where anomalies are identified using adaptive prediction intervals (PIs) rather than point estimates. This enables explicit uncertainty quantification and more reliable detection under varying data conditions [16].
2) The proposed approach is distribution-free and dynamically adjusts to temporal variations, allowing context-

aware unsupervised AD. Unlike several existing methods, the AD is simpler and direct as it does not require labeled data or generate pseudo-labels [6].

3) We compare the performance of the Tube-based AD using LSTM, GRU, TCN, and Transformer with the existing benchmarks in the literature, specifically, the FG-Net [6], KNN [17], [9], LOF [18], [9], IF [19], [9], ConvoLSTM-AE [10] and VAE [12]. Results show that the proposed Tube-based models achieve consistent improvements in several performance metrics.

II. Proposed Approach

A. Estimation of PI using a Tube Loss-based Approach

For an electric load sequence $\{L_i\}_{i=1}^{t}$, let $\mathbf{L}_t \triangleq [L_t, L_{t-1}, \ldots, L_{t-w+1}]^\top \in \mathbb{R}^w$ denote a lag window of size w at time t. The point estimate $\hat{L}_{t+1} = \mathbb{E}[L_{t+1} \mid \mathbf{L}_t]$ of the electric load at time $t+1$ is replaced in the proposed Tube-based probabilistic load forecasting (PLF) approach by a PI $\mathcal{B}_{t+1} \triangleq [\hat{L}_\ell(\mathbf{L}_t;\boldsymbol{\theta}), \hat{L}_u(\mathbf{L}_t;\boldsymbol{\theta})]$, where $\hat{L}_\ell(\mathbf{L}_t;\boldsymbol{\theta})$ and $\hat{L}_u(\mathbf{L}_t;\boldsymbol{\theta})$ are the lower and the upper bounds of the estimated PI, and $\boldsymbol{\theta}$ represents a vector of the parameters of a DL network used to obtain $\mathcal{B}_{t+1}$. In contrast to conventional point-forecasting-based AD methods that rely on reconstruction errors and predefined thresholds to identify anomalies, the objective in the PLF is to estimate the PI, since it represents the normal behavior of the observed random process. An observation L_{t+1} lying outside $\mathcal{B}_{t+1}$ is flagged anomalous. An anomaly flag z_{t+1} at time $t+1$ is set to 0 if $L_{t+1} \in \mathcal{B}_t$, else z_{t+1} is set to 1.

Our proposal is obtain the PI using a novel Tube Loss-based network with the objective of meeting the following probabilistic coverage guarantee: $\mathbb{P}\left(L_{t+1} \in \mathcal{B}_{t+1} \mid \mathbf{L}_t\right) \geq 1-\alpha$, where $1-\alpha$ is the target confidence level, and $\mathbb{P}(x)$ denotes the probability of occurrence of an event x. The PI $\mathcal{B}_{t+1}$ is obtained through a mapping $f_{\boldsymbol{\theta}} : \mathbb{R}^w \to \mathbb{R}^2$ whose parameter vector $\boldsymbol{\theta}$ is obtained by optimizing (see Fig. 1) the following loss function:

$$\mathcal{C}(\boldsymbol{\theta}) \triangleq \sum_{i=t-w-1}^{t-1} \mathcal{T}_{(1-\alpha)}^{(r)}\left(L_{i+1}, \hat{L}_\ell(\mathbf{p}_i;\boldsymbol{\theta}), \hat{L}_u(\mathbf{p}_i;\boldsymbol{\theta})\right) + \delta \sum_{i=t-w-1}^{t-1} \left|\hat{L}_u(\mathbf{p}_i;\boldsymbol{\theta}) - \hat{L}_\ell(\mathbf{p}_i;\boldsymbol{\theta})\right|. \quad (1)$$

The first term in the above corresponds to the Tube Loss function with the skewness parameter $r \in [0,1]$. It ensures that the estimated PI $[\hat{L}_\ell(\mathbf{p_t},\boldsymbol{\theta}),\ \hat{L}_u(\mathbf{p_t},\boldsymbol{\theta})]$ satisfies the desired $(1-\alpha)$ coverage. This function is defined over four disjoint regions $\{I_k\}_{k=1}^4$ as follows:

$$\mathcal{T}_{(1-\alpha)}^{(r)}(x,\ell,u) = \begin{cases} (1-\alpha)(x-u), & \text{if } x \in I_1, \\ \alpha(u-x), & \text{if } x \in I_2, \\ \alpha(x-\ell), & \text{if } x \in I_3, \\ (1-\alpha)(\ell-x), & \text{if } x \in I_4, \end{cases} \quad (2)$$

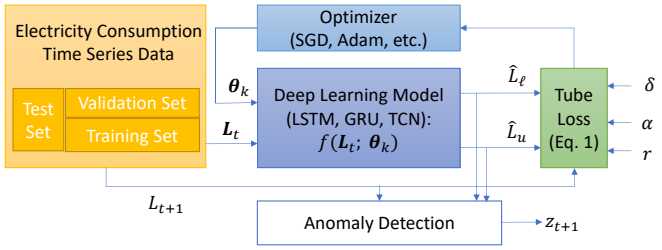

Fig. 1. Proposed learning based method for anomaly detection using Tube loss.

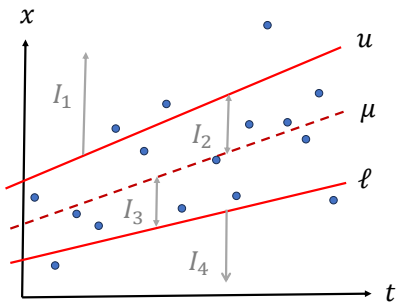

Fig. 2. Four intervals I_k, $k \in [1,\ldots,4]$, in (3).

where the regions $\{I_k\}$ are given as (see Fig. 2)

$$\begin{aligned} I_1 &= \{\, x : x > u \,\}, \\ I_2 &= \{\, x : x \leq u,\ x \geq \ell,\ x-\mu \geq 0 \,\}, \\ I_3 &= \{\, x : x \leq u,\ x \geq \ell,\ x-\mu < 0 \,\}, \\ I_4 &= \{\, x : x < \ell \,\}. \end{aligned} \quad (3)$$

In the above, $\mu \triangleq r\,\ell + (1-r)\,u$.

The first term in the objective function in (1) is the proposed Tube loss function designed to learn the PI to obtain $1-\alpha$ coverage. For the samples that lie outside the PI $[\ell, u]$ shown in Fig. 2, the loss is $(1-\alpha)(x-u)$ when $x > u$ (i.e., the sample lies in the region I_1), and it equals $(1-\alpha)(\ell-x)$ when $x < \ell$ (i.e., the sample lies in the region I_4). This ensures that the violations outside the interval $[\ell, u]$ are penalized proportionate to their excursion beyond the limits, and this penalty $1-\alpha$ is large (since α is typically a small value, e.g., $\in \{0.01, 0.1\}$). This encourages the learned interval to achieve a coverage level close to $(1-\alpha)$. Furthermore, the loss is *non-zero* even when $x \in [\ell, u]$, and its value depends on the weighted average $\mu = r\ell + (1-r)u$ of the lower and the upper bounds. Specifically, if $x \in I_2$ (i.e., $x \in [\ell, u]$ and $x-\mu \geq 0$), the loss is $\alpha(u-x)$, else if $x \in I_3$ (i.e., $x \in [\ell, u]$ and $x-\mu < 0$), the loss is $\alpha(x-\ell)$. The model applies comparatively a smaller penalty (since α is a small number) to the samples that fall within the PI compared to the samples that fall outside. However, even for the samples inside the PI, the penalty increases with $|x-\mu|$, thereby pulling the upper and the lower bounds u and ℓ closer to the sampled data x.

By making the penalty $(1-\alpha)$ for the samples x lying outside the PI $[\ell, u]$ larger than the penalty α for the samples

that lie inside the PI, the model ensures that a failure to cover the data is worse than having a somewhat wide PI. As an example, with $\alpha = 0.05$, the coverage is 95%, and the penalty for missing the Tube is 19 times greater than the penalty for the Tube being too wide. The skewness parameter r allows controlling the location of the weighted average μ inside the Tube. The penalty for the samples falling in the upper or the lower half of the Tube can be varied, allowing the PI to adapt to the real-world electricity consumption data that are often characterized by skewed distributions (e.g., sudden massive spikes in the load may be more frequent compared to large drops in consumption). Thus, the proposed Tube loss generalizes the pinball (quantile) loss. Here, the lower and upper bounds ℓ and u can be interpreted as learned quantiles, and the asymmetric penalties mimic quantile regression behavior on both sides. The loss encourages the learned interval to satisfy $\mathbb{P}\big(x \in [\ell, u]\big) \approx 1 - \alpha$, while also shaping the conditional distribution of x within the interval. However, unlike the Quantile-based DL schemes, that require training of a pair of deep learning models independently (e.g., for minimizing the Pinball Loss function for the estimation of the q^{th} and $(1-q)^{th}$ quantile functions), the proposed approach estimates the two PI bounds by training a single DL model such that a PICP (Prediction Interval Coverage Probability) greater than or equal to $1-\alpha$ is achieved with a minimum possible width, measured by the MPIW (Mean Prediction Interval Width).[1] In addition, the Tube Loss approach facilitates the recalibration through the δ parameter that significantly reduces the width of the PI on test set. The recalibration procedure begins by training with the initial value $\delta = 0$, followed by computing the PICP on the validation set $\{s_i\}_{i=t'+1:t}$. Let $\epsilon > 0$ denote a user-defined tolerance parameter and $\eta \geq 0$ a user-defined step size. If the validation PICP satisfies $\text{PICP} > 1 - \alpha + \epsilon$, the model is retrained with the updated parameter $\delta \leftarrow \delta + \eta$. Instead, if $\text{PICP} < 1 - \alpha$, the model is retrained with $\delta \leftarrow \delta - \eta$. This iterative adjustment continues until the resulting PICP computed on the validation set lies within the desired tolerance interval $[\,1-\alpha,\ 1-\alpha+\epsilon]$.

The second term $\delta \sum |u - \ell|$ in (1) imposes a penalty on the interval width by penalizing the total width of the PI regardless of value of x relative to u and ℓ, and thereby encourages sharper estimates. Through the optimization of (1), the network simultaneously estimates the lower and upper bounds, simplifying the overall forecasting process. The proposed Tube Loss-based AD scheme is summarized in Algorithm 1.

Algorithm 1: Tube Loss–Based AD

Require: Training data $\mathcal{S} = \{L_1, L_2, \ldots, L_t\}$, lag window w, confidence level $1-\alpha$, and margin parameter set Δ
Ensure: Anomaly label z_{t+1} for observation L_{t+1}

Training Phase
for each $\delta \in \Delta$ **do**
  Initialize model parameters $\boldsymbol{\theta}$
  Train model by minimizing:

$$\hat{\boldsymbol{\theta}}^{(\delta)} \leftarrow \arg\min_{\boldsymbol{\theta}} \mathcal{C}(\boldsymbol{\theta}; \delta)$$

  Evaluate performance on validation set
end for
Select optimal δ^* based on validation performance
Obtain final model parameters $\hat{\boldsymbol{\theta}} \leftarrow \hat{\boldsymbol{\theta}}^{(\delta^*)}$

Prediction Phase
Construct input vector $\mathbf{L}_t = [L_t, L_{t-1}, \ldots, L_{t-w+1}]^\top$
Compute PI bounds:

$$(\hat{L}_\ell, \hat{L}_u) \leftarrow f_{\hat{\boldsymbol{\theta}}}(\mathbf{L}_t)$$

AD Phase
if $L_{t+1} < \hat{L}_\ell$ OR $L_{t+1} > \hat{L}_u$ **then**
  $z_{t+1} \leftarrow 1$ (anomaly)
else
  $z_{t+1} \leftarrow 0$ (normal)
end if
return z_{t+1}

B. Pipeline

Dataset: In this study, we utilize a publicly available dataset, the Large-Scale Annotated Dataset for Energy Anomaly Detection (LEAD1.0), created by leveraging electricity meter data from the Great Energy Predictor III competition, specifically focusing on 1413 non-residential buildings [9]. The data comprises one year of hourly readings (from the year 2016 onward) collected from 16 different sites worldwide, totaling over 12 million data points, which are labeled to denote the point as well as sequential anomalies.

Data Preprocessing: Prior to model training, the meter readings are subjected to second-order differencing to remove deterministic trends and seasonal components, thereby promoting approximate stationarity of the time series. The differenced series is subsequently standardized via z-score normalization to stabilize the input distribution and facilitate efficient training.

Training and Hyperparameter Tuning: The dataset is divided chronologically into 80%, 10%, and the remaining 10% for training, validation, and testing. At each time step t, the model is trained using the input feature vector $\mathbf{L}_t \triangleq [L_t, L_{t-1}, \ldots, L_{t-w+1}]^\top$, constructed using a lag window of size w. As shown in Fig. 1, different deep autoregressive architectures, including LSTM, GRU, and TCN, integrated with Tube Loss, are trained to capture temporal dependencies in the data. The models are trained to estimate PIs, which are subsequently used for AD. The performance of the models depends on several hyperparameters, such as input window size, number of layers, number of hidden units, number of training epochs, batch size, and learning rate. These hyperparameters are optimized using a grid-search strategy based on their performance on the validation set. In addition, Tube-Loss-based models require tuning of the margin parameter δ for trading the PI sharpness with coverage (see (1)). For the Seq2Seq Transformer-based Tube model, additional hyperparameters such as, the number of attention heads, model dimension, feed-forward network size, and number of encoder–decoder layers are also tuned.

[1]The *PICP* and the *MPIW* are defined as $\text{PICP} = \sum_{i=1}^{k} \mathbb{I}\Big(\hat{L}_\ell(\mathbf{L}_{t+i}) \leq L_{t+i} \leq \hat{L}_u(\mathbf{L}_{t+i})\Big)/k$, (where $\mathbb{I}\{\cdot\}$ is the indicator function) and $\text{MPIW} = \sum_{i=1}^{k} \Big(\hat{L}_u(\mathbf{L}_{t+i}) - \hat{L}_\ell(\mathbf{L}_{t+i})\Big)/k$.

TABLE I
PERFORMANCE COMPARISON FOR BUILDING ID 1353 DATA IN [9].

Approach	Precision	Recall	F1
Proposed Approach			
TCN Tube	0.059	0.175	0.088
LSTM Tube	0.067	0.182	0.097
GRU Tube	0.047	0.188	0.076
Transformer Tube	0.042	0.156	0.066
Existing Baselines			
Tukey [5]	0.0284	0.370	0.053
FGNet [6]	0	0	0
ConvoLSTM-AE [10]	0	0	0
VAE [12]	0.0102	0.013	0.011
KNN [9]	0	0	0
LOF [9]	0.006	0.013	0.009
IF [9]	0.010	0.026	0.015

TABLE II
PERFORMANCE COMPARISON FOR BUILDING ID 293 DATA IN [9].

Approach	Precision	Recall	F1
Proposed Approach			
TCN TUBE	0.029	0.152	0.049
LSTM TUBE	0.031	0.217	0.054
GRU TUBE	0.036	0.174	0.060
Transformer Tube	0.034	0.152	0.055
Existing Baselines			
Tukey [5]	0.022	0.141	0.039
FGNet [6]	0	0	0
ConvoLSTM-AE [10]	0	0	0
VAE [12]	0	0	0
KNN [9]	0	0	0
LOF [9]	0.005	0.022	0.008
IF [9]	0.023	0.109	0.038

III. RESULTS AND DISCUSSION

Benchmarks used: In this study, a diverse set of baseline AD methods are employed, spanning statistical, classical ML, and DL paradigms. Among statistical approaches, Tukey's fence method is employed as a widely used baseline for AD. Additionally, we consider a neural network-based approach (FGNet) trained on labels generated using Tukey's method, following [6]. Among classical ML methods, we include KNN [17], [9], LOF [18], [9], and IF [19], [9], which detect anomalies based on distance, density, and isolation-based approaches, respectively. Further, DL approaches are also considered, including ConvoLSTM-AE [10] and VAE [12], which rely on reconstruction error for AD. These baselines provide a comprehensive benchmark to evaluate the proposed Tube-based AD framework across different methodological perspectives. The comparative evaluation is based on Precision, Recall, and F1-score (with "positive" class corresponding to anomalous instances, whereas the "negative" class denoting the normal observation) [4].

Experimental Test Setup: Experiments were conducted on a system with an Intel® Core™ i7-14700K processor, 32 GB RAM, and dual GPUs (Intel® UHD Graphics 770 and a 4 GB discrete GPU) using Python 3.11.9.

Comaprison against Benchmarks: Table I and Table II present the performance comparison of Tube-based AD models with various baseline methods for Building IDs 1353 and 293, respectively. In addition to the Tube-based DL models, a sequence-to-sequence (Seq2Seq) Transformer-based model [13] integrated with Tube Loss is also included to capture long-range temporal dependencies using attention mechanisms. It can be observed that Tube-based models achieve higher F1-scores than Tukey's fence method, with improvements of approximately 3.5%, 4.4%, and 2.3%. Similarly, for Building ID 293, improvements of approximately 1.0%, 1.5%, and 2.1% are observed. FGNet is trained on pseudo-anomalies generated by Tukey's fence method, following the approach of [6]. During testing, we evaluate FGNet on the actual ground-truth anomaly labels. As observed in the tables, for FGNet and several other baseline methods, the F1-score is zero due to the absence of true positive detections, indicating a failure to correctly identify any ground-truth anomalies. Transformer integrated with the Tube Loss also outperforms the baselines. These improvements indicate the efficacy of the Tube-based models in capturing the anomalous behaviour more accurately than the conventional statistical and reconstruction-based methods. Fig. 3 depicts the comparison of the proposed method with baselines on the basis of macro-F1 score. Here, Transf Tube refers to Transformer Tube. It can be observed that the Tube-based models consistently achieve higher macro-F1 scores than the baseline approaches. The Tube-based framework operates within a probabilistic forecasting setting, providing PIs, which serves as an added advantage over conventional baseline methods. Fig. 4 shows a comparison of PICP and MPIW for Tube-based models. As illustrated in

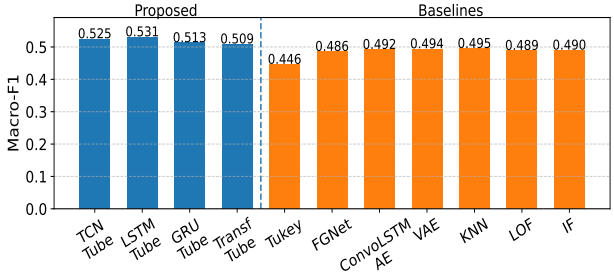

Fig. 3. Comparison of Macro-F1 score on Building ID 1353 data in [9].

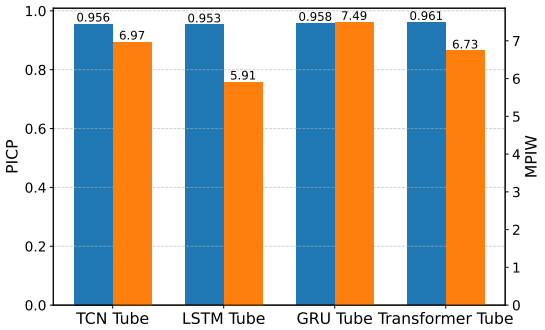

Fig. 4. Comparison of PI quality for Tube-based models on Building ID 1353 data in [9], evaluated in terms of PICP and MPIW.

TABLE III
PERFORMANCE OF LSTM TUBE FOR COVERAGE VALUES FOR BUILDING ID 1353 DATA IN [9].

Coverage	Precision	Recall	F1
0.99	0.052	0.078	0.063
0.95	0.067	0.182	0.097
0.9	0.048	0.253	0.08
0.85	0.041	0.312	0.073
0.8	0.036	0.357	0.066

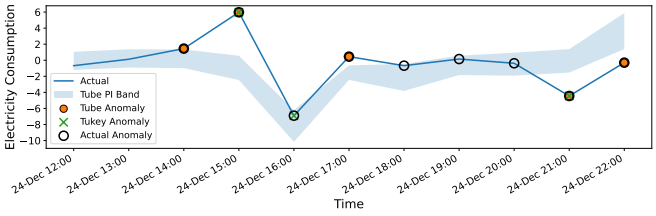

Fig. 5. Visualization of AD results showing actual consumption, PI generated by LSTM-Tube at 0.80 coverage, and detected anomalies compared with Tukey and ground truth labels for Building ID 1353 data in [9].

the plot, all models attain similar coverage levels, whereas variations in MPIW indicate differences in the tightness and efficiency of the learned PIs. LSTM-Tube achieves the lowest MPIW, indicating more efficient interval estimation, which can contribute to improved anomaly localization. The results highlight a fundamental limitation of statistical methods such as Tukey's fence method. Tukey's fence relies on fixed quantiles and implicit normality; hence, it struggles with skewed and dynamic anomalies. In contrast, Tube-based models use adaptive PIs, enabling more robust detection in non-stationary electricity load data.

Ablation Study: Table III presents the ablation study of the LSTM Tube for different target coverage values for building ID 1353. Among all configurations, a coverage of 0.95 achieves the best F1-score, indicating an optimal balance between precision and recall. It can be observed that the choice of coverage significantly affects the F1-score. At higher coverage levels, the PI widens, making the model more conservative, and lower coverage levels yield narrower PIs, leading to more aggressive AD. These results highlight the importance of selecting an appropriate coverage level to balance detection sensitivity and false alarm rates in Tube-based AD. From Fig. 5, it can be observed that the LSTM Tube can capture anomalous deviations outside the PI, approximately aligning with the ground truth anomalies.

IV. CONCLUSION

This paper proposes a Tube-based DL framework for AD in electricity consumption data, formulated as a probabilistic forecasting approach using a learned PI that adapts to temporal variability, and improves anomaly detection reliability. The approach is distribution-free and model-agnostic, making it robust to skewed and heavy-tailed data and easily adaptable across architectures. Experimental results show consistent gains over baselines, achieving average F1-score improvements of 2–3% (up to 4.4% in some cases), while several baseline methods fail to detect meaningful anomalies.

REFERENCES

[1] L. Zhang, G. Wang, and G. B. Giannakis, "Real-time power system state estimation and forecasting via deep neural networks," *IEEE Trans. Power Syst.*, 2018.

[2] O. Y. Maryasin and L. Tihomirov, "Using generative neural networks to detect point anomalies in energy consumption data," in *2024 6th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA)*. IEEE, 2024, pp. 636–641.

[3] Y. Chen *et al.*, "Machine learning-based anomaly detection for load forecasting under cyberattacks," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 1–11, 2019.

[4] X. Jia, P. Xun, W. Peng, B. Zhao, H. Li, and C. Shen, "Deep anomaly detection for time series: A survey," *Computer Science Review*, vol. 58, p. 100787, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574013725000632

[5] A. Páez and G. Boisjoly, *Exploratory Data Analysis*. Cham: Springer International Publishing, 2022, pp. 25–64. [Online]. Available: https://doi.org/10.1007/978-3-031-20719-8_2

[6] Y. Shi, L. Xu, P. Wang, X. Wu, L. Wang, and Y. Hou, "Real-time anomaly detection of electricity time series data based on future-guidance network," *IEEE Signal Process. Lett.*, 2025.

[7] J. Mulongo, M. Atemkeng, T. Ansah-Narh, R. Rockefeller, G. M. Nguegnang, and M. A. Garuti, "Anomaly detection in power generation plants using machine learning and neural networks," *Applied Artificial Intelligence*, vol. 34, no. 1, pp. 64–79, 2020.

[8] Y. Himeur, A. Alsalemi, F. Bensaali, and A. Amira, "Smart power consumption abnormality detection in buildings using micromoments and improved k-nearest neighbors," *International Journal of Intelligent Systems*, vol. 36, no. 6, pp. 2865–2894, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/int.22404

[9] M. Gulati and P. Arjunan, "Lead1.0: a large-scale annotated dataset for energy anomaly detection in commercial buildings," in *Proceedings of the thirteenth ACM international conference on future energy systems*, 2022, pp. 485–488.

[10] W. Luo, W. Liu, and S. Gao, "Remembering history with convolutional LSTM for anomaly detection," in *2017 IEEE International conference on multimedia and expo (ICME)*. IEEE, 2017, pp. 439–444.

[11] A. Takiddin, M. Ismail, U. Zafar, and E. Serpedin, "Deep autoencoder-based anomaly detection of electricity theft cyberattacks in smart grids," *IEEE Systems Journal*, vol. 16, no. 3, pp. 4106–4117, 2022.

[12] M. Castangia, R. Sappa, A. A. Girmay, C. Camarda, E. Macii, and E. Patti, "Anomaly detection on household appliances based on variational autoencoders," *Sustainable Energy, Grids and Networks*, vol. 32, p. 100823, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352467722001151

[13] J. Zhang, H. Zhang, S. Ding, and X. Zhang, "Power consumption predicting and anomaly detection based on transformer and K-Means," *Frontiers in Energy Research*, vol. 9, p. 779587, 2021.

[14] T. Mu, Y. Yu, G. Feng, H. Luo, and H. Yang, "Detecting anomalous electricity consumption with transformer and synthesized anomalies," *PeerJ Computer Science*, vol. 9, p. e1721, 2023.

[15] A. Francis and B. A. Jose, "Short term load forecasting and anomaly detection in residential power consumption," in *2024 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)*. IEEE, 2024, pp. 1–7.

[16] P. Anand, T. Bandyopadhyay, and S. Chandra, "Tube loss: A novel approach for prediction interval estimation and probabilistic forecasting," *arXiv preprint arXiv:2412.06853*, 2024.

[17] X. Wang and S.-H. Ahn, "Real-time prediction and anomaly detection of electrical load in a residential community," *Applied Energy*, vol. 259, p. 114145, 2020.

[18] M. Jesmeen, J. Hossen, and A. B. A. Aziz, "Unsupervised anomaly detection for energy consumption in time series using clustering approach," *Emerging Science Journal*, vol. 5, no. 6, pp. 840–854, 2021.

[19] Z. Huang, W. Mao, M. Chen, Q. Wu, B. Xiong, and W. Xu, "An intelligent operation and maintenance system for power consumption based on deep learning," in *IOP Conference Series: Materials Science and Engineering*, vol. 486, no. 1. IOP Publishing, 2019, p. 012107.