International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246.IJETCSIT-V4I3P120
Eureka Vision Publication | Volume 4, Issue 3, 189-197, 2023

Original Article

Anomaly Detection in AMI and Smart Meter Data for Electricity Theft, Outage, and Equipment Fault Identification: A Comprehensive Review

Krishna Gandhi[1], Pankaj Verma[2]
[1]Illinois State University, 100 N University St, Normal, IL 61761, United States.
[2]Indian Institute of Management, Bangalore (IIM-Bangalore), Bannerghatta Road, Bengaluru, Karnataka, India.

Abstract - Advanced Metering Infrastructure (AMI) and smart meters have become important elements of the contemporary power distribution system, providing the ability to monitor finely, have two-way communication, and make operational decisions based on the data. Although these systems have significant positive aspects in regard to efficiency, reliability and customer interactions, they also pose new threats in regards to data integrity, system security and operational anomalies. Non-technical losses, including electricity theft, equipment faults, and operational events, including outages and restoration events may be the reason behind the abnormal patterns in the smart meter data. Such anomalies should be prevented through proper and efficient tracking of them in order to guarantee grid stability, as well as minimizing financial costs, and consumer confidence. The paper includes an extensive literature review of anomaly detection methods used on AMI and smart meter data but in relation to the methodology of detecting theft, outages, and equipment-related faults. Traditional statistical methods, machine learning methods, and deep learning methods are systematically analyzed, as well as data preprocessing techniques, feature engineering techniques, and metrics. Practical issues are also addressed, including imbalance of data, privacy, scalability and interpretability. The paper is an attempt to give a systematic source to the researcher and practitioners wishing to learn the current status of anomaly detection in smart metering systems and its application in intelligent power distribution management.

Keywords - Advanced Metering Infrastructure, Smart Meters, Anomaly Detection, Electricity Theft, Outage Detection, Equipment Faults, Machine Learning, Review.

1. Introduction

Advanced Metering Infrastructure (AMI) has to a large extent facilitated the transformation of the traditional power distribution network to a smart, data-driven system. Smart meters, which are the major feature of AMI, they have high-resolution measurements of electricity consumption, voltage levels, power quality indicators, and event logs. Contrary to the traditional electromechanical meters, smart meters enable two-way communications between utilities and consumers enabling the automated billing and demand response programs as well as the near real-time detection of grid conditions.

Alongside these benefits, the prevalence of smart meters has brought about new loopholes and hurdles in the operations. Metering data abnormalities may be caused by numerous causes, such as deliberate manipulation of the data to steal electricity, non-deliberate metering data error due to hardware degradation, communications failures or due to actual system events (load switching, outages). It is not easy to distinguish between these causes and especially in the large-scale distribution systems where millions of meters are generating data at any given time.

The problem of electricity theft is still one of the significant concerns of utilities all over the world and a significant contributor to non-technical losses and make power systems less financially viable. Globally, the financial impact of electricity theft is substantial, with annual losses estimated at close to USD 89 billion, a burden that falls most heavily on emerging economies. In several regions, non-technical losses arising from practices such as meter tampering, illegal connections, and billing manipulation can represent as much as 40 % of the electricity distributed, significantly weakening both revenue streams and the overall stability of power systems. Evidence from country-level studies further demonstrates that this issue is widespread rather than isolated. For instance, utilities in Brazil have reported that roughly 15 % of supplied electricity is lost due to theft, while in Colombia, non-technical losses have been observed at approximately 14.7 % of total energy generation [1]. Similarly, major utilities in South Africa have reported losses amounting to hundreds of millions of dollars annually as a result of electricity theft. These persistent and large-scale financial losses underscore the pressing need for advanced analytical approaches capable of accurately identifying and mitigating abnormal consumption behavior within modern power distribution networks. The conventional methods of detection are through manual inspection and billing audit which is labour intensive, expensive and not always effective. Likewise, it is essential to ensure that the service reliability and customer dissatisfaction is minimized by detecting outages and equipment faults in a timely manner. The data delivered by a smart meter is providing unparalleled possibilities to identify the presence of such events automatically; nevertheless, the amount, speed, and fluctuation of such data require sophisticated methods of analysis [2].

Anomaly detection has become one of the important analytical activities in AMI data analysis, with the aim of detecting the patterns that are deviating [3]. These anomalies can be contextual, collective or point based in accordance to the cause and the way they are manifested in the data. This paper will give an in-depth overview of anomaly detection methods as applied to smart metering systems and especially three broad areas of application electricity theft detection, outage detection, and equipment fault diagnosis.

2. Advanced Metering Infrastructure and Smart Meter Data Characteristics

AMI systems are made of smart meters, communication, data concentrators and head-end systems that gather and handle metering data. Smart meters generally log the energy used every few seconds to a few hours and can also log many other parameters, such as the voltage, current, power factor and event flags.

Figure 1. Advanced Metering Infrastructure

2.1. Data Types and Resolution

Smart meter data can be broadly categorized into:

- Interval consumption data, representing energy usage over fixed time intervals
- Event data, indicating occurrences such as power loss, restoration, or tamper alerts
- Power quality measurements, including voltage sags, swells, and interruptions
- Metadata, such as meter location, customer class, and installation details

The high temporal resolution of interval data enables detailed analysis of consumption patterns but also increases data volume and noise sensitivity.

2.2. Data Quality Issues

One of the most important issues in the analysis of AMI and smart meter data is data quality because the reliability of the anomaly detection algorithms is necessarily conditional on the accuracy and completeness of values behind the algorithms. On real-life deployments, the smart meter data is often impacted by numerous flaws caused by the technical aspects and operational elements [4]. These flaws can greatly distort consumption patterns and unless taken care of can create wrong interpretation of anomalies.

Missing or incomplete data is one of the most frequent data quality problems. Missing values can happen because of some communication issues, network congestion, temporary metering failure or planned maintenance. The missing data can be limited in nature in few intervals in certain cases and continuous in other cases. These gaps may interfere with continuity of time-series data, and cloud actual behavior patterns, and it may be hard to learn to differentiate between natural variability in consumption and abnormal occurrence like outages or theft.

Delay in communication and loss of packets aggravates the AMI data analysis. Latency and data loss are inevitable since smart meters usually use wireless or power-line communication networks, especially in large systems or systems that are geographically distributed. Late readings can come in the wrong order causing gaps in time, which can provide false signals to a sequence-based anomaly detection model [5]. Temporal alignment is also necessary to ensure that there are no false readings labeled as anomalies by the models.

Figure 2. Data Charactersitics of a Smart Meter

Other issues of concern are time synchronization errors. Even though the smart meters are expected to work with synchronized time mechanisms, they may experience discrepancies due to clock drift, or synchronization errors, or aggregation issues on the gateway level. Even the small delay effects can significantly affect the high-resolution data analysis, particularly when the anomalies are detected by the abrupt variation in data or by the correlations of various meters.

Poor quality of data is also caused by measurement noise and sensor error. The sources of noise can be old hardware, physical environment, electricity and electromagnetic signals, or calibration errors. This noise may be in the form of random variations or slow variations in consumption or voltage values. Although these deviations do not necessarily mean that the system is misbehaving, they might be misconstrued to represent abnormal behavior when they are not addressed in a proper filter. These data quality issues make the generation of anomalies more difficult since anomalies in the readings of meters can be as a result of the flaws in the data and not actual events that could include electricity theft, outages, or equipment problems [6]. Therefore, the effective anomaly detection frameworks should include the extensive data validation and preprocessing policies. The kinds of techniques that are usually used to improve the reliability of the data are missing data imputation, temporal realignment, noise filtering and consistency checks among neighboring meters. Early-stage handling of data quality is thus a critical requirement towards having the correct, credible, and operationally significant anomaly detection in AMI-based systems.

3. Anomalies in Smart Meter Data

Deviations in the usual consumption or operation patterns in the power distribution system are manifested as anomalies in the data of Advanced Metering Infrastructure (AMI) and are typically indicators of underlying technical or non-technical issues in the power distribution system. These irregularities may be caused by both intentional human activities or system-level incidents and by equipment failures, and they need to be identified in a timely manner to make sure that the grid remains reliable, the revenues are preserved and the operations are efficient. Depending on their origin and impact on the system, the anomalies in smart meter data can be classified into electricity theft and non-technical losses, outage events, and equipment faults.

3.1. Electricity Theft and Non-Technical Losses

Electricity theft is a deliberate conduct that employs the intention to minimize, or avoid the registered energy use, which leads to high levels of non-technical losses to the utilities. The frequent ones are meter tampering, unlawful connections, magnetic impairment and circumventing metering devices. The activities in smart meter data can frequently appear in the form of sharp or gradual reductions in consumption, non-uniform load profiles, atypically flat profiles, or disagreements between historical consumption of a consumer and the consumption of similar meters. Theft behaviors are dynamic unlike technical faults and change over time as the consumers struggle to avoid detection systems. This dynamic character renders detection of theft especially difficult, so methods of analysis able to detect long-term and insidious deviations should be utilized instead of focusing on sudden deviations.

3.2. Outage Events

Outage incidents can be defined as the partial or complete power outage to a customer and are usually indicated in AMI records as zero energy consumption, voltage drop or loss messages and communication disconnection. Although outages can be confused with the absence of data due to the failure of communication, proper recognition is necessary by using the relationships between meter data and the occurrence of events in order to correlate the network-level data. Multi meter spatial and time consistency can be one of the main indicators of real outage events [7]. Proper outage identification with the help of smart meter data facilitates quicker fault localization, better restoration planning, and situation awareness among the operators of the distribution system.

3.3. Equipment Faults

The anomalies that are caused by equipment are as a result of fault or degradation in the smart meters, transformers, or other parts of the distribution. These errors can cause aberrant readings, slow measurement drift, loss of data periodically or uneven voltage and current readings. In contrast to theft-related aberrations, equipment failures tend to have long-lasting or continuous trends that become more severe with time. Preventive maintenance is heavily reliant on the timely detection of the anomalies because the equipment failures that have gone unnoticed may result in incorrect billing, lower quality of power, and, in the worst-case scenario, a domino effect of the damaged equipment. Using smart meter data to identify faults enables the utilities to adopt not only reactive maintenance measures, but also predictive and condition-based maintenance measures as well.

4. Data Preprocessing and Feature Engineering

Effective anomaly detection of smart meter data is based on data preprocessing and feature engineering. AMI data is normally large, noisy and heterogeneous, where measurements are made at different resolutions and subject to communication limitations. Consequently, without proper preprocessing, it is not possible to use raw meter readings in order to reliably detect anomalies. Preprocessing quality has a direct effect on the model robustness, generalization ability and detecting accuracy especially in machine learning-based and data-driven methods.

4.1. Data Cleaning and Normalization

An initial step like data cleaning is required to overcome flaws that are usually present in smart meter datasets. Loss of values will most likely be a result of communication error, planned maintenance, or failure of the devices. Such gaps are normally addressed by interpolation methods (linear or spline interpolation) or by statistical and model-based imputation methods which use temporal or spatial correlations between meters [8]. Outliers due to transmission errors or sensor glitches need to be marked and eliminated because they can greatly bias learning algorithms when they go unblemished.

Normalization is done to have similar feature scale especially when using data related to customers with varying levels of consumption. To enhance model convergence and stability, the use of techniques like minmax scaling, or z-score normalization is common. Besides this, when combining the resulting data of several meters, temporal alignment is required, so that readings are at the same time interval. The appropriate cleaning and normalization can be used to maintain meaningful consumption patterns and remove the artifacts that can cause a false anomaly detection.

4.2. Feature Extraction

Instead of using the raw energy readings formulated, anomaly detection models usually use higher-level features that are more representative of consumption behavior. Commonly extracted daily and weekly load profiles are used to observe the habits of usage and variations of the activity that can be related to abnormality. Mean, variance, skewness, and kurtosis are statistical elements that contain concise descriptions of consumption patterns within given time periods.

Time frequency characteristics of signal processing methods such as Fourier or wavelet transforms are usually used to represent transient and periodic phenomena [9]. The given features are especially useful when it comes to sudden changes that can be connected to theft, outages, or equipment problems. Also contextual features like temperature, humidity, day of the week or holiday are also added to differentiate between actual behaviour change and external effects. The techniques of feature selection and dimensionality reduction are then used to select the most discriminative information and minimize the computational complexity so as to improve the detection performance and scalability.

5. Conventional Statistical and Rule-Based Approaches

Prior to the popularity of machine learning methods, statistical analysis and rule-based frameworks were used as the primary methods to detect anomalies in smart meter and AMI data. These methods are based on pre-formulated assumptions regarding the normal consumption behavior and apply to analytic or heuristic rules to point to the deviations. They were simple, transparent, and did not need much computing, which made them appealing to initial AMI applications [10]. Nevertheless, they are in many cases ineffective in contemporary power systems with a wide range of consumption behaviour, dynamic user behaviour, and growing infiltration of distributed energy resources.

5.1. Threshold-Based Methods

One of the most commonly used and simplest methods of detecting anomalies is to use threshold based methods. Under these strategies, anomalies are reported when the meter readings or calculated features surpass the set upper or lower thresholds. The thresholds could be determined by historical averages, contractual limits or engineering judgment, or by being adaptive, constantly changing over time based on the observed consumption patterns [11]. The simpleness of implementation and high interpretability of threshold-based methods is usually employed in simple theft detection, outage detection, and the generation of alarms. Though these are the benefits, threshold-based methods are very sensitive to the choice of threshold. Poor thresholds may lead to either false alarms or missed anomalies especially where consumption variation is high. Normal consumption patterns can be influenced by seasonal changes, weather influences, and behavioral shifts that can cause the patterns to run across the static thresholds, which lowers their reliability. Subsequently, such techniques do not suffice in most cases when applied in solitude and they tend to be increasingly integrated with more sophisticated methods of analysis.

5.2. Statistical Modeling

The goal of statistical modeling techniques is to describe normal consumption behavior in terms of probabilistic or time-series models and deviant behavior in the form of deviation in terms of expected patterns. Autoregressive (AR) and autoregressive moving average (ARMA) models are methods that have been used to deal with the temporal dependence of meter readings. The control chart techniques such as the Shewhart, CUSUM and exponentially weighted moving average (EWMA) charts are common when it comes to tracking slow changes or sudden changes in consumption behavior [12]. The frameworks of hypothesis testing also allow the formal statistical analysis of whether the observed deviations are statistically significant.

The methods work well in fairly stable and stationary conditions, with consumption pattern with predictable time structures. Nevertheless, the AMI data in the real world usually show non-stationary characteristics dictated by the effects of lifestyle change, renewable integration, and changes in consumer behavior. In this case, the traditional statistical models are unable to scale to the situation and result in inferior performance at detection. This low capability of capturing nonlinear relationships and nonlinear dynamics has led to the replacement of nonlinear methods of detection of anomalies with data-driven and learning-based methods.

6. Machine Learning Approaches

The popularity of machine learning based methods in detecting anomalies in smart meter and AMI data has grown recently because of their capability to directly learn complex patterns of data. Machine learning models do not make rigid assumptions about the behavioral patterns during consumption, unlike rule-based and statistical-based frameworks, and can be adjusted to different and changing usage patterns. Such ways are mostly classified as supervised, unsupervised, and semi-supervised learning, as they relate to the existence of labeled data.

6.1. Supervised Learning

The applications of supervised learning methods are strongly examined in the detection of electricity theft, classification of outages and identification of equipment fault. Usually, the support vector machines (SVMs), decision trees, random forests, K-nearest neighbors (KNN), and logistic regression are popular models. Such models are learnt on labeled data, in which normal and abnormal consumption patterns are explicitly known. Under the condition of good labels, it is possible to attain good performance in classification using supervised methods with good decision boundaries.

But in real-life AMI systems, labeled data on anomalies are usually limited, expensive to acquire and extremely skewed because anomalous behavior is much rarer than normal behavior [13]. This imbalance on classes can skew the training of the model and lower the accuracy of detection of rare occurrences. In addition, controlled models can have difficulties in generalizing to novel or novel anomaly types, which restricts their performance in dynamically powered systems in the long-term.

Table 1. Summary of Machine Learning Approaches for Anomaly Detection in AMI and Smart Meter Data

Category	Representative Algorithms	Learning Paradigm	Key Strengths	Limitations	Typical Applications in AMI
Support Vector Machines (SVM)	SVM, One-Class SVM	Supervised / Semi-supervised	Effective in high-dimensional spaces; robust to overfitting with proper kernel selection	Requires careful kernel and parameter tuning; limited scalability for very large datasets	Electricity theft detection; abnormal consumption classification
Decision Tree-Based Models	Decision Trees, Random Forests, Gradient Boosting	Supervised	Interpretable decision rules; handles nonlinear relationships; resistant to noise	Requires labeled data; may struggle with evolving consumption behavior	Theft identification; customer behavior classification
k-Nearest Neighbors (k-NN)	k-NN	Supervised	Simple to implement; no explicit training phase	Computationally expensive for large datasets; sensitive to feature scaling	Consumption anomaly detection; pattern similarity analysis
Naïve Bayes Classifiers	Gaussian NB, Multinomial NB	Supervised	Computationally efficient; works well with limited training data	Assumes feature independence; lower accuracy for complex patterns	Preliminary theft screening; anomaly pre-filtering
Clustering	k-Means,	Unsupervised	No labeled data	Cluster interpretation	Detection of

Algorithms	DBSCAN, Hierarchical Clustering		required; useful for exploratory analysis	may be ambiguous; sensitive to distance metrics	unusual load profiles; consumer segmentation
Dimensionality Reduction Methods	PCA, ICA	Unsupervised	Reduces noise and redundancy; highlights dominant consumption patterns	Limited capability to capture nonlinear relationships	Feature extraction; anomaly visualization
Isolation-Based Methods	Isolation Forest	Unsupervised	Designed specifically for anomaly detection; scalable to large datasets	Less effective when anomalies resemble normal data	Detection of rare abnormal consumption events
Ensemble Learning Models	Bagging, Boosting, Hybrid Ensembles	Supervised	Improved robustness and accuracy; reduces model bias	Increased computational complexity; reduced interpretability	Large-scale theft detection; fault classification
One-Class Classification Models	One-Class SVM, SVDD	Semi-supervised	Effective when only normal data is available	Sensitive to training data quality; limited adaptability	Outage detection; equipment fault identification

6.2. Unsupervised Learning

Unsupervised learning methods solve the drawbacks of the availability of the labeled data by detecting the anomalies based only on the deviation of the normal patterns that are learned. Clustering algorithms, principal component analysis (PCA), autoencoders, and isolation forests are also typical methods of this category. The strategies of these methods are that normal consumption behavior has regular patterns, and anomalies are represented by outliers or deviations in the feature space.

The unsupervised techniques are especially applicable to large-scale systems of AMI when the labels of the anomalies are not available or reliable. They are useful in detecting new or changing patterns of anomalies even without prior information and they are useful in the early detection of anomalies and exploratory analysis [14]. Nevertheless, the methods tend to be non-interpretive and can be parameter-dependent, potentially implying bias in the performance of the methods under highly changing consumption conditions.

6.3. Semi-Supervised Learning

Semi-supervised learning offers an intermediate between supervised and unsupervised methods and makes use of a small portion of labeled data, usually as normal behavior, and a much larger collection of unlabeled observations. The most frequently used in this classification are one-class classification models, including one-class SVMs and autoencoders, which are used to both learn a succinct description of normal consumption patterns and detect irregularities as anomalies.

They are also very well adapted to the AMI anomaly detection case where there is a large amount of normal operating data, and a small amount of anomalous labels, or they may not be known. The semi-supervised models are superior in generalization to the highly supervised models and specificity to the unsupervised models. Because of this, they are finding use as a practical anomaly detector in smart grid monitoring in large scale and reliable systems.

7. Deep Learning-Based Anomaly Detection

The use of deep-learning-based systems has found great interest in AMI and smart-meter anomaly detection because such systems are highly effective in modeling complex and nonlinear relationships in high dimensional data. Compared to conventional machine learning techniques which often use manual features, deep learning models are able to automatically extract hierarchical representations using raw or slightly preprocessed meter reads. This renders them especially viable in the detection of small, changing, and high-impact anomalies like electricity theft, intermittent faults and aberrant load behavior.

7.1. Autoencoders

In smart-meter data, autoencoders constitute some of the commonly utilized deep learning models in anomaly detection. These models learn a small latent representation of the input data by training to recreate normal consumption patterns. The anomalies are detected when the error of reconstruction surpasses a specified level meaning that an irregularity has been detected in the normal behavior learned. A number of autoencoders have been implemented in AMI applications. Sparse autoencoders reinforce discriminative features that the model learns through sparsity constraints, whereas variational autoencoders (VAEs) induce probabilistic latent representations, making them less sensitive to noise and uncertainty. The models have been proven to be very efficient in identifying non-technical losses and abnormal usage patterns especially with the pairing of anomaly data that is limited.

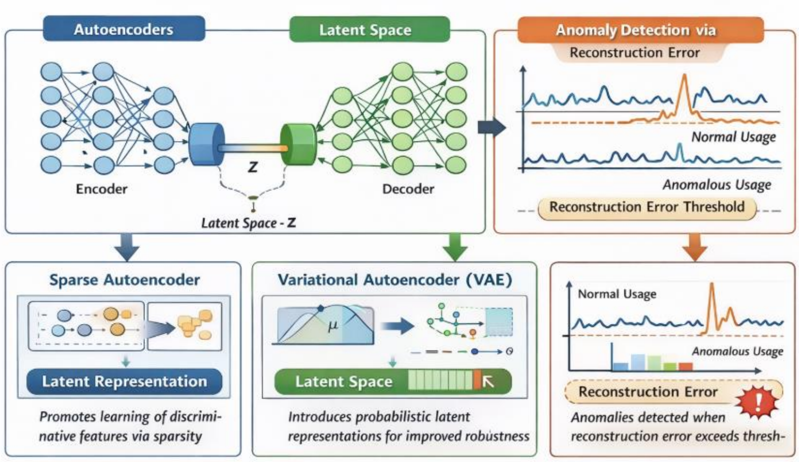

Figure 3. Autoencoders for Anamoly Detection in Smart Meter Data

7.2. Recurrent Neural Networks

Recurrent neural networks (RNNs) particularly the use of long short-term memory (LSTM) and gated recurrent unit (GRU) models are very good at the sequential and temporal modeling of smart-meter data. Such models are able to contain long term dependencies and seasonal consumption trends thus constructive detection of anomalies which occur over time slowly. RNN-based models in AMI applications have been applied in the detection of electricity theft, identification of an outage, and monitoring a fault in equipment. This can be attributed to their capacity to maintain the historical context, meaning that they can discriminate normal fluctuations and true anomalous events. Nevertheless, RNNs can be computationally expensive and can prove to be sensitive to tuning to operate at scale.

7.3. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) represent a type of artificial intelligence that can identify and understand the words within an image and are a form of artificial intelligence that is able to recognize and comprehend words in an image. Convolutional neural networks (CNNs) have been progressively used to detect anomalies by converting time series data in one dimensional form into two dimensional analysis, including the load profile image or recurrence plot. This allows CNNs to learn the local patterns, correlations as well as structural anomalies in consumption behavior.Models that are built on CNN are especially useful in detecting localized anomalies, including the occurrence of sudden load variations, or the recurrence of abnormal patterns of usage. Space invariance properties enable them to be resistant to noises and small changes in consumption data. Consequently, CNNs have demonstrated excellent results in the realization of electricity theft and abnormal load signatures in massive AMI data sets.

Hybrid deep learning models integrate several parts of the neural networks in order to leverage their complementary advantages. Examples of these models are CNN-LSTM and CNN-GRU [15], where the convolutional layer acquires spatial/local features, and the recurrent layer is used to simulate temporal dependencies. Such hybrid models are being investigated extensively to perform complex anomaly detection tasks on smart grids since they enable a more insightful view of the consumption behavior. Through the ability to simultaneously capture spatial, temporal and contextual features, hybrid architectures have been shown to be more accurate and robust at detection than single-model based systems, especially in the situations where meter data is highly dynamic and heterogeneous.

8. Evaluation Metrics and Benchmarking

The importance of accurate assessment of the anomaly detection models in the data of the smart meter and AMI is especially significant, as these systems have a class imbalance. Such anomalies include electricity theft, outages, or equipment malfunctions, which are very rare compared to the normal consumption, and which standard accuracy measures are not appropriate to measure performance meaningfully. To make the right evaluation, the metrics should indicate the capability to recognize the rare events and the cost of misclassification. Precision is the percentage of correctly detected anomalies of all the cases reported as anomalous:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \qquad (1)$$

High precision indicates that flagged anomalies are likely to be genuine, reducing the operational burden associated with investigating false alarms.

Recall (or sensitivity) quantifies the proportion of actual anomalies that are correctly detected:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \qquad (2)$$

High recall ensures that few anomalies are missed, which is crucial for preventing financial losses or safety risks in power systems. The F1-score provides a balanced measure of precision and recall, calculated as:

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (3)$$

It is also practical when distributions in the classes are unequal, and it punishes the models that are doing well on one measure and bad on the other.

Another commonly used approach to assessing model discrimination capability is the Receiver Operating Characteristic (ROC) curves and the Area under the Curve (AUC). The ROC curves are curves that give the true positive rates against the false positive rates at various threshold settings whereas AUC gives a summary of the total capability of the model to distinguish anomalies and normal observations. The higher the AUC values of a model, the better will be the generalization of a model to different operating conditions.

Besides these measures, cost-sensitive assessment is significant in AMI applications. False positives (indicating normal consumption as abnormal) can raise the cost of operational investigations, whereas false negatives (Missing of real abnormalities) can lead to loss of revenue, safety risks or service interruptions. Thus, performance measurement must include both quantitative and application cost factors in order to make sure that models are correct and operational.

The model detection anomalies are normally benchmarked on IEEE standard test systems (e.g., 14-bus, 39-bus, and 118-bus systems) or large scale synthetic AMI data [16]. Limited real-world datasets are being utilized increasingly to test the validity of generalizability and robustness, which offer a more realistic test of model performance in heterogeneous and dynamic settings.

9. Practical Challenges and Limitations

Although the AMI and smart-meter data anomaly detection methods have made significant progress, there are still a number of practical concerns that limit their use. These issues are of paramount importance to credible, scalable and deployable implementation in real world power systems.

9.1. Information Asymmetry and Data Sparing.

Abnormalities of the AMI datasets are rare by nature, therefore leading to highly skewed data distributions. Outages, equipment faults or electricity thefts represent only a minor part of the total dataset. This imbalance is a major problem to supervised learning methods, which are based on adequate labeled examples to learn decision boundaries effectively. In addition, high-standard labels are usually expensive and hard to transport especially in cases of theft or minor equipment malfunctions. Labels can also in most instances be noisy or inconsistent thus making model training more difficult and lowering the accuracy of detection. This has been demonstrated to be partially addressed by techniques like oversampling and synthetic anomaly generation and semi-supervised techniques, but it is hard to generalize to unknown patterns of anomaly.

9.2. Privacy and Security Issues.

The information on smart meters records the finesse consumption patterns of the household that may expose sensitive data about inhabitants behavior, lifestyle, and occupancy. This poses significant privacy challenges and more so in the context of data sharing or cloud analysis. The regulatory regulations including GDPR and local utility policies require AMI data to be handled carefully. Such methods as data anonymization, aggregation, and secure multiparty computation are used more and more often to find the golden mean between the necessity of anomaly detection and consumer privacy. Also, cyber-security attacks such as false data injection may undermine the performance of anomaly detection and should be considered during model design and deployment.

9.3. Scalability and Deployment

Utilitarian AMI implementations are characterized by millions of smart meters producing high resolution, near real-time data streams. Computational and storage requirements are high, especially with the deep learning based models, when it comes to processing such large amounts of data. To achieve timely detection of anomalies, there is a need to deploy strategies that are efficient such as distributed computation, edge processing and incremental or online learning algorithms. Operational applications, like outage management and theft prevention are also important with real-time or near-real-time detection, in which delayed detection can cause financial losses or reduced system reliability.

9.4. Interpretability

Furthermore, most current anomaly detectors, especially deep learning models, including autoencoders, LSTM, GRU, and hybrid CNNRNN models, are black box models, and thus have low interpretability. This non-transparency may become an obstacle to the credibility of utility operators because decisions on stealing investigation or equipment repairs usually need explanations. Recent studies highlight the significance of explainable AI (XAI) methods which give information on the reasons why a particular reading or pattern is detected as anomalous. The fact that interpretability can be integrated to help enhance operational decision-making also enables regulatory compliance and acceptance of advanced analytics in the utility industry.

10. Conclusion

The AMI and smart meter data anomaly detection has become an important part of the modern power distribution system as it allows detecting the presence of electricity theft, outages, and equipment malfunctions early. Analytical methods, both traditional, relying on the use of traditional statistical tools, and advanced, based on the application of deep learning models, have been studied to solve this issue. Although machine learning methods and deep learning techniques provide better detection performance, their practical use requires proper preprocessing, feature engineering and proper consideration of practical constraints. Further research and development on the front is necessary to improve the reliability of grids, decrease the number of losses, and facilitate the shift to intelligent energy systems.

References

[1] Carr, D., & Thomson, M. (2022). Non-technical electricity losses. Energies, 15(6), 2218.

[2] Wang, Y., Chen, Q., Hong, T., & Kang, C. (2018). Review of smart meter data analytics: Applications, methodologies, and challenges. IEEE Transactions on smart Grid, 10(3), 3125-3148.

[3] Guerrero-Prado, J. S., Alfonso-Morales, W., Caicedo-Bravo, E., Zayas-Pérez, B., & Espinosa-Reza, A. (2020). The power of big data and data analytics for AMI data: A case study. Sensors, 20(11), 3289.

[4] Wang, Y., Chen, Q., Hong, T., & Kang, C. (2018). Review of smart meter data analytics: Applications, methodologies, and challenges. IEEE Transactions on smart Grid, 10(3), 3125-3148.

[5] Kim, D. S., Chung, B. J., & Chung, Y. M. (2020). Analysis of AMI communication methods in various field environments. Energies, 13(19), 5185.

[6] Chin, Y. S. (2019). Anomaly detection frameworks for identifying energy theft and meter irregularities in smart grids (Doctoral dissertation, University of Malaya (Malaysia)).

[7] Dunn, L. N., Sohn, M. D., LaCommare, K. H., & Eto, J. H. (2019). Exploratory analysis of high-resolution power interruption data reveals spatial and temporal heterogeneity in electric grid reliability. Energy Policy, 129, 206-214.

[8] Baddoo, T. D., Li, Z., Odai, S. N., Boni, K. R. C., Nooni, I. K., & Andam-Akorful, S. A. (2021). Comparison of missing data infilling mechanisms for recovering a real-world single station streamflow observation. International Journal of Environmental Research and Public Health, 18(16), 8375. https://doi.org/10.3390/ijerph18168375.

[9] Silik, A., Noori, M., Altabey, W. A., Ghiasi, R., & Wu, Z. (2021). Comparative analysis of wavelet transform for time-frequency analysis and transient localization in structural health monitoring. Structural Durability & Health Monitoring, 15(1), 1.

[10] Tshivhase, N., Hasan, A. N., & Shongwe, T. (2020). Proposed fuzzy logic system for voltage regulation and power factor improvement in power systems with high infiltration of distributed generation. Energies, 13(16), 4241.

[11] Chae, Y., Katenka, N., & DiPippo, L. (2019, September). An adaptive threshold method for anomaly-based intrusion detection systems. In 2019 IEEE 18th International Symposium on Network Computing and Applications (NCA) (pp. 1-4). IEEE.

[12] Mahmood, T., Balakrishnan, N., & Xie, M. (2021). The generalized linear model-based exponentially weighted moving average and cumulative sum charts for the monitoring of high-quality processes. Applied Stochastic Models in Business and Industry, 37(4), 703-724.

[13] Han, S. Y., No, J., Shin, J. H., & Joo, Y. (2016). Conditional abnormality detection based on AMI data mining. IET Generation, Transmission & Distribution, 10(12), 3010-3016.

[14] Hu, T., Guo, Q., Shen, X., Sun, H., Wu, R., & Xi, H. (2019). Utilizing unlabeled data to detect electricity fraud in AMI: A semisupervised deep learning approach. IEEE transactions on neural networks and learning systems, 30(11), 3287-3299.

[15] Shiri, F. M., Perumal, T., Mustapha, N., & Mohamed, R. (2023). A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU. arXiv preprint arXiv:2305.17473.

[16] Salem, C. (2020). Machine Learning Based Detection of False Data Injection Attacks in Wide Area Monitoring Systems (Doctoral dissertation, Concordia University).

[17] P. Verma and K. Gandhi, “Drilling optimization and ROP prediction with hybrid ML models,” Eur. J. Technol., vol. 5, no. 2, pp. 1–19, 2021.