Anomaly Detection in Residential Electricity Consumption with GAN-CNN for Enabling Smart Grid Security and Efficiency Monitoring

Badari Narayana Palety [1,*], C. Mahalakshmi [1], and P. Nagasekhar Reddy [2]

[1] Department of Electrical and Electronics Engineering, Annamalai University, Chidambaram, Tamil Nadu, India
[2] Department of Electrical and Electronics Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, India
Email: badari.palety@gmail.com (B.N.P.); maha_c2008@yahoo.com (C.M.); pnsreddy04@gmail.com (P.N.R.)
*Corresponding author

***Abstract*—Efficient management of energy resources relies heavily on residential electricity usage. However, challenges such as defective smart meters, erratic consumption patterns, and energy theft pose serious threats to the stability of the smart grid. Addressing these irregularities requires robust anomaly detection methods. This study presents a novel approach to anomaly detection in smart grid systems by integrating Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs). The GAN generates synthetic electricity consumption data closely resembling actual usage patterns, leveraging its understanding of typical consumption behaviors learned from extensive historical data. Meanwhile, a CNN architecture extracts crucial features from consumption data, capturing spatial and temporal correlations. Through a rigorous training process, the GAN-CNN model learns to identify regular consumption patterns and detect anomalies effectively. Extensive testing using real-time consumption data demonstrates the model's remarkable accuracy, achieving an impressive accuracy rate of approximately 0.9976. This methodology proves robust, exhibiting high detection accuracy, low false positive rates, and resilience to noise and data volatility. Implementation of this GAN-CNN-based anomaly detection approach in smart grid systems holds significant implications for enhancing grid security, resource distribution efficiency, and customer service. It empowers utilities to swiftly address equipment issues and mitigate energy theft, thereby improving overall grid stability and operational effectiveness. Consequently, this research contributes substantially to the advancement of smart grid analytics and sets the stage for future innovations in the field.**

***Keywords*—Generative Adversarial Network (GAN), Convolutional Neural Network (CNN), anomaly detection**

I. Introduction

The increasing use of smart grid technologies, which aim to improve the efficiency, dependability, and environmental responsibility of energy distribution, is a result of an increasing need for consumption of energy in residential areas. These systems' essential components, smart metres, offer detailed, real-time tracking and measurement of the use of electricity. But with these systems becoming more complicated, new difficulties have emerged, especially with regard to maintaining their effectiveness and security. The effective operation of the smart grid can be disrupted by anomalies such as broken smart metres, unusual usage patterns, and unauthorized energy consumption, which may result in possible financial losses, operational shortcomings, and compromising energy security. The study of anomaly detection, which aims to spot abnormalities from expected or typical patterns in datasets, has attracted a lot of interest in response to these problems [1]. Techniques for detecting anomalies are essential for preserving the reliability of smart grids and defending against fraud or system breakdown. Traditional systems frequently rely on statistical thresholding or rule-based approaches, but these approaches can have trouble adjusting to complex and ever-changing abnormalities. Machine learning techniques have recently proven to be incredibly effective at handling problems involving anomaly identification. Deep learning techniques in particular, which automatically uncover complex patterns and representation from data, have demonstrated promise in a number of anomaly detection applications. The utilization of Generative Adversarial Networks (GANs) in combination with Convolutional Neural Networks (CNNs) for identifying anomalies in time-series data is one such method [2].

In order to address the issues raised by anomalies in home electricity usage within the framework of a smart grid system, this research suggests an innovative approach that combines the strength of GANs and CNNs. The main goal is to create an efficient system for detecting anomalies that can correctly pinpoint abnormal consumption patterns. The suggested GAN-CNN hybrid technique makes use of the advantages of both GANs and CNNs: CNNs for extracting pertinent geographical and temporal information from the consumption sequences, and GANs for producing artificial information that captures typical consumption behavior. Ullah *et al.* [3] intend to greatly improve the safety and efficiency

Manuscript received November 28, 2023; revised January 22, 2024; accepted April 3, 2024; published February 10, 2025.

tracking of the smart grid by utilizing these technologies. The main accomplishment of this study is the creation of a thorough framework for anomaly identification that has various advantages over current approaches. The GAN model learns the complicated distribution of typical consumption patterns by being trained on a sizable historical dataset of home power consumption. This knowledge is then put to use to provide artificial consumption data, which acts as a benchmark for detecting deviations. By capturing the subtle spatial and temporal relationships contained in the consumption sequences, the architecture of CNN further improves this process. The co-optimization of these models is facilitated by the combination of GANs and CNNs inside an adversarial training framework, improving the overall performance of anomaly detection [4].

The field of managing energy in home contexts has undergone a revolution due to the widespread use of smart grid technology. However, these systems' heightened complexity and interconnectedness have created new difficulties, particularly in maintaining their efficiency and security. The stability and dependability of the smart grid are seriously threatened by anomalies in home electricity consumption, which result from things like broken smart metres, irregular usage patterns, and unauthorised energy consumption. Maintaining the grid's financial stability, operational effectiveness, and general security depends on addressing these abnormalities. An essential component of smart grid management involves anomaly identification, which entails finding variations from anticipated consumption patterns in huge, dynamic databases [5]. Traditional anomaly detection techniques that rely on static criteria or statistical criteria frequently fail to detect complex and dynamic anomalies. In response, the field of machine learning has shown that it is capable of meeting these obstacles. Particularly adept at comprehending complicated patterns of data and representations are deep learning techniques. In this context, a convincing method for dealing with anomaly detection in time-series analysis data is the integration of GANs with CNNs.

In order to solve the anomaly detection problems that plague home electricity usage in the setting of smart grids, this research presents a novel approach combining GANs and CNNs. The main objective is to create a system that is efficient at identifying anomalous consumption patterns. The suggested GAN-CNN fusion makes use of the advantages of both systems: CNNs for extracting useful spatial and temporal information from consumption sequences, and GANs for creating synthetic data that captures typical consumption habits. Li *et al.* [6] aim to greatly improve the safety and efficiency monitoring capabilities of smart grid systems by integrating the technologies. The main contribution is a thorough framework for anomaly identification that outperforms traditional approaches in many ways. The model learns a detailed understanding of typical usage patterns by training a GAN on a large historical dataset of home power consumption. This knowledge forms the basis for the creation of synthetic consumption statistics, which acts as a reference point for identifying deviations.

Numerous benefits, including two-way communication and healing themselves capabilities, are brought about by the transformation of traditional electrical networks into Smart Grids (SGs). Due to the variety and vulnerability of SGs, this paradigm change also poses substantial cybersecurity issues. These dangers are exacerbated by the use of antiquated communication standards and the development of the IoT. While Secured Information and Event Management (SIEM), systems have become known as a potential cybersecurity solution, current SIEMs frequently neglect to take into account the unique features and historical communication protocols of SGs. The requirement for efficient anomaly detection as well as cybersecurity in SGs drives the investigation of cutting-edge strategies to fully handle these problems [7]. Zhang *et al.* [8] explore several SG security facets, such as theft of energy detection, finding anomalies, and SIEM system creation, while identifying difficulties and possible future paths. This process is advanced by the CNN architecture's incorporation by unravelling the complex spatial and temporal relationships inherent in consumption sequences. An adversarial training architecture, which promotes a mutually beneficial connection between GANs and CNNs, allows for comprehensive model optimization, increasing the overall effectiveness of anomaly detection.

The key contributions of this work are as follows:

- The study uses a novel GAN and CNN combo to find irregularities in residential electricity use.
- The suggested GAN-CNN model has excellent levels of accuracy in detecting a variety of anomalies, including broken smart metres, unusual usage patterns, and cases of energy theft.
- Real-time anomaly detection capabilities of the GAN-CNN model make it possible to spot abnormal patterns of consumption as they emerge. This aids in early identification and action, enabling utility providers to handle problems quickly and preserve grid stability.
- The GAN-CNN-based method exhibits robustness over noise and data fluctuations that are frequently found in real-world contexts. This flexibility improves the model's generalization skills and ensures reliable performance under a variety of data situations and consumption patterns.
- The study has important ramifications for monitoring smart grid efficiency and security. The suggested method equips utility firms to deal with issues like equipment malfunctions and energy theft by automated the detection of anomalous consumption patterns.

The rest of this paper is structured as follows: In Section II, the basic ideas of GANs and CNNs are introduced along with a summary of relevant work in anomaly detection. The issue statement, together with data pre-processing, model design, and training approach, are described in depth in Section III. The study's setup, dataset outline, and assessment measures are presented in Section IV. Section V presents the findings and discussions, while Section VI presents the conclusions and suggested next steps. By providing a solid solution for

anomaly identification in home electricity consumption, our research aims to increase the security and effectiveness of smart grids.

II. Related Works

Theft of electricity is the biggest problem facing smart grids. As a result of the reliance of current procedures on particular equipment, it cannot be easily identified. Additionally, the approaches' performance is constrained by their inability to extract useful information from highly dimensional data on electricity usage, as well as by an increase in the false positive rate. Additionally, inaccurate Electrical Theft Detection (ETD) utilizing data-driven algorithms is hampered by uneven data. In the literature, sampling approaches are utilized to overcome this issue. However, the outdated sampling methods produce inaccurate and insufficient data that lower the ETD rate. Aldegheishem *et al.* [9] proposed that two original ETD models were created. In the first model, a hybrid sampling strategy, an artificial minority the oversampling method with edited nearest neighbor, is presented. Additionally, dimensionality reduction and information extraction from data on electricity consumption are done using AlexNet. A light gradient enhancement model is then employed for classification. In the second model, the actual distribution of the data on electricity usage is captured using a conditional Wasserstein generated adversarial network with gradient penalty. To create more accurate data for the minority class, auxiliary provisional information is included in the formula. Furthermore, the dataset's dimensions are decreased by using the Google Net architecture. The classification of trusting and untrustworthy customers is done lastly using adaptive boosting. Utilizing actual power usage data from China's state grid business, both models were trained and evaluated. The effectiveness of the suggested models is assessed using a variety of performance criteria, including precision, recall, accuracy, F1-score, etc. The simulation results show that the proposed models perform better in terms of effective ETD than the currently used methods, including support vector machines, extreme gradient boost, convolution neural networks, etc. Our model's limitation in taking into account outside variables, like electricity prices, weather, and breaks, which can affect electricity usage patterns in real-world settings, is a flaw.

The Advanced Metres Infrastructure (AMI), a prominent early example of the Internet of Thing in the field of smart grids, offers real-time data from smart metres to grid operators and users, maximizing the possibility of demand response. Without security protection, the newly gathered data can, however, be deliberately manipulated and cause great harm. Yao *et al.* [10], propose a smart grid energy theft detecting system that protects energy privacy. We use coupled CNNs in particular to identify anomalous behavior in metering data from a long-term pattern observation. Additionally, we use the Paillier algorithm to safeguard the privacy of the energy. In this way, the transfer of the users' energy data is securely safeguarded, and data disclosure is kept to a minimum. Our security research shows that data confidentiality and identification are both accomplished by our system. According to experimental findings, our updated CNN model can accurately detect abnormal behaviors up to 92.67% of the time. The requirement to improve our system by lowering communications and computing overhead could be a disadvantage for future work.

Data from smart grids can be analyzed to find anomalies in a variety of areas, such as cyber security, fault finding, electricity theft, etc. Different factors, such as odd customer consumption habits, broken grid infrastructures, outages, outside cyberattacks, or energy fraud, may be to blame for the strange abnormal behaviors. The smart grid's anomaly detection has recently gained a lot of attention from investigators, and it has been utilized in many high-impact domains. The smart grid has a number of difficult difficulties, but one of the biggest is how to integrate effective anomaly detection for various types of anomalous behavior. Banik *et al.* [11] provided a scoping analysis of the study from the most recent developments in the field of smart grids anomaly detection. For a thorough understanding and examination of the research issues so far, we categorize our study based on a wide range of factors. The path forward for future studies on identifying anomalies in smart-grid networks has been briefly presented after analyzing the deficiencies in the evaluated study. The system's current lack of connection with various sources of data (substations, feeders, laterals, and customers) could potentially impede real-time insights and thorough anomaly identification in smart grid technologies, which is a limitation in our planned work.

To protect the safety of Cyber-Physical Systems (CPS), anomaly detection is essential. Traditional anomaly detection approaches, which must deal with an increasing amount of data and require domain-specific knowledge, are unable to properly address these issues due to the complexity of CPSs and the sophistication of attacks. Deep Learning-based Detection of Anomalies (DLAD) techniques have been developed to do this. Luo *et al.* [12] reviewed modern DLAD techniques in CPSs. To grasp the key characteristics of the present approaches, we provide a taxonomy based on the kinds of anomalies, tactics, implementation, and assessment measures. They also use this taxonomy to pinpoint and highlight novel traits and layouts within each CPS domain. Also, talk about the drawbacks and unresolved issues with these techniques. Additionally, we empirically investigate the features of common neural models, the process of DLAD techniques, and the performance while running DL models to provide users with insights into selecting appropriate DLAD methods in practice. Finally, we examine the shortcomings of DL methods, results, and potential paths to enhance DLAD techniques and stimulate more investigation. The survey's focus on the use of deep learning in CPS safety measures could have the unintended consequence of drawing attention away from non-deep learning strategies, which are as important in tackling urgent deployment challenges in CPS detection of anomalies.

The conventional electrical system has entered a new digital era known as the "Smart Grid (SG)", which offers

a number of advantages including two-way communication, ubiquitous control, and self-healing. Due to SG's varied and unsecure character, this new reality nevertheless creates considerable cybersecurity threats. In particular, SG depends on antiquated communication standards that were not developed with cybersecurity in mind. Additionally, the emergence of the IoT presents significant cybersecurity difficulties. A developing technology in the field of cybersecurity is Secure Information and Event Management (SIEM) systems, which can identify, normalize, and correlate a large number of security events. They are capable of managing all aspects of security in a smart environment like SG. However, the historical communication protocols and other specific SG quirks and features are not taken into consideration by the contemporary SIEM systems. Radoglou-Grammatikis *et al.* [13] presented the SG-focused Secure and Private Smart Grid (SPEAR) SIEM. The work's key contribution is the development and execution of a SIEM system that can identify, normalize, and correlate abnormalities and cyberattacks against a wide range of SG application-layer protocols. It is significant that genuine data from four actual SG use cases—a hydropower plant, (b) an electrical substation, (c) a power plant, and (d) a smart home—are used to show the SPEAR SIEM's detection capability. The integration of various intrusion and detection of anomalies models, mining association rules, and healing themselves mechanisms in the aforementioned future plans could result in complexity and resource demands, which could have an adverse effect on the effectiveness and practical execution of the Big Data Analytics Capabilities (BDAC) Evaluation Engine and SPEAR SIEM.

III. Research Framework

The suggested methodology begins with data preparation and employs Min-Max scaling to standardize the consumption data in order to improve smart grid security, efficiency monitoring, and anomaly identification in household electricity usage. This vital step makes sure that every consumption figure is converted into a dependable scale. In order to find anomalous patterns in home electricity consumption, it then applies the Isolation Forest algorithm, a potent technique for finding anomalies. After carefully separating the data into training and testing subsets, it could utilize one to train our anomaly detection model and the other to assess how well it performs. The core of the approach is the combination of a Convolutional Neural Network (CNN) with a Generative Adversarial Network (GAN). The CNN discriminator excels at differentiating between actual and synthetic data, effectively signalling anomalies, while the GAN's generator network learns to produce synthetic consumption data that closely reflects typical usage patterns. An antagonistic training approach is used to hone the skills of this dynamic combination. The installed GAN-CNN model integrates into the smart grid infrastructure and continually tracks real-time household electricity consumption. Any variations from anticipated patterns are immediately recognized and reported as potential anomalies, enabling fast action and maintenance. Additionally, methodology goes beyond than simple anomaly identification. It also makes it easier to increase energy efficiency by pointing out areas where it is being wasted and offering useful insights. In conclusion, the novel approach combines Min-Max scaling, Isolation Forest for anomaly identification, and GAN-CNN technology to build a strong system for protecting smart grids, monitoring effectiveness, and identifying abnormalities in home electricity consumption. This comprehensive strategy ushers in a more dependable and effective era for household energy management by strengthening the grid against abnormalities as well as promoting cost savings and sustainability. The suggested approach's workflow is depicted in Fig. 1.

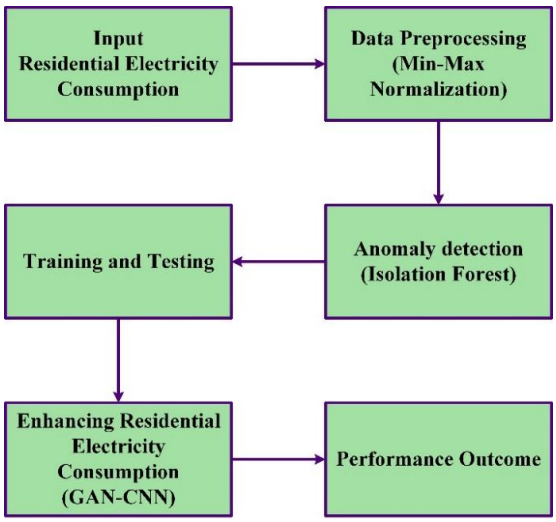


Fig. 1. Workflow of the suggested approach.

A. Data Collection

The data collected from the Pecan Street dataset, which comprises daily energy consumption records for 346 residential households spanning from January 1st to April 30th, 2016, requires expansion to address missing data, create additional time granularities, perform feature engineering, and potentially incorporate external data sources like weather information. These steps are essential for thorough analysis and to derive meaningful insights from this valuable dataset [14].

B. Data Pre-processing

When using cutting-edge methods like the GAN-CNN model for anomaly identification, data preparation is crucial to the analysis and modelling of electricity consumption data. Min-max normalization stands out as a crucial step among the essential data preparation techniques. In min-max normalization, each feature's minimum value is changed to 0 and its highest value is changed to 1. Depending on the original value in relation to the minimum and maximum values of the feature, all intermediate values are scaled to fall within the 0–1 range. Suppose you have an array of numbers $S = [m_1, m_2, \ldots, m_u]$. It will first find the minimum and maximum values of the array: min_s and max_s subsequently utilizing the min and max values, every

original value m_u is converted into a min-max normalized value $\hat{m}_u$ employing the following expression in Eq. (1):

$$\hat{m}_u = \frac{m_u - min_s}{max_s - min_s} \quad (1)$$

As a result, the training of machine learning models like the GAN-CNN is shielded from the influence of any one feature. Selecting the feature of interest, specifying the intended range, figuring out the dataset's minimum and maximum values, and then using the min-max normalization algorithm to rescale each data point are the simple steps involved. This method has a number of benefits, such as consistent scaling, preservation of data relationships, and improved model performance. Alternative techniques might be better appropriate for severely skewed or non-normally distributed data because it assumes a roughly linear distribution within the given range. Ultimately, min-max normalization ensures that the data on electricity consumption is scaled appropriately, enhancing the accuracy and dependability of the smart grid's anomaly detection system.

C. Isolation Forest

According to Sadaf and Sultana [15], the Isolation Forest was developed for locating outliers. The same mechanism is employed here. The process employs binary trees, a specific form of trees, to identify errors or issues. This feature allows it to complete the task swiftly and with limited memory usage, making it ideal for managing extensive information. Isolation Forests can be compared to Random Forests. They are made using decision trees. The reason why this model falls into the unsupervised category is that it does not possess any pre-established labels. Data that has been randomly selected is used in isolation forest. Randomly chosen attributes are utilized in the construction of this forest, which follows a tree structure. The likelihood of the samples being unusual or strange decreases because additional cuts were required for those located deeper within the tree. Similar to shorter branches, unusual findings are revealed, denoting observations that the tree identified as distinct from other samples with greater ease. As stated earlier, the technique used for identifying outliers, known as Isolation Forests, consists of multiple decision trees that make yes/no judgments. Isolation Trees (iTrees) are the name given to each tree in an isolation forest. The algorithm begins by creating Isolation Trees to train on the data. Here is a step-by-step breakdown of the entire algorithm:

- Randomly selecting a smaller set of data from a larger collection, they are organized into a unique type of tree structure with two branches or categories.
- An initial random feature (chosen from the set of all N characteristics) is chosen before the tree branches. The branching process is then based on a random threshold (any value between the minimum and maximum values of the chosen characteristic).
- A data point moves to the left branch or the right branch depending on whether its value is smaller than the chosen threshold. A node is thus divided into left and right branches.
- This step-by-step procedure is repeated until all of the data points are entirely isolated or, if a maximum depth is determined, until it is reached.
- To generate random binary trees, the aforementioned stages are repeated.

The training process is considered concluded after combining a set of Isolation Forests. Each trained tree is responsible for processing a data point during the scoring process. The "anomaly score" has been assigned to every individual data point. The depth of the tree directly impacts the score, signifying how far it had to search for the data. This score is determined by adding up the depths identified by each individual iTree. Anomalies are given a score of −1 and normal points are given a score of 1, based on the degree of unusual data considered.

IV. Proposed GAN-CNN Mechanism

A. Generative Adversarial Network

Based on its previous observations, a GAN can fabricate novel information. It examines instances of objects and aims to generate novel creations that bear resemblance. The recognition of patterns in the examples enables it to generate its own creations. GANs can assist in the training of models for generating novel things by utilizing a method called supervised learning. Two smaller models are encompassed within a GAN [16], which is a type of computer model. By combining these two models, the outcome is the production of objects with enhanced realism:

- The generator model is instructed to produce novel outcomes during its training process.
- The discriminator model differentiates between authentic and fraudulent inputs. Its objective is to determine whether an input originates from the original training data for the generator model.

To the point where the discriminator model is unable to tell the difference between genuine and generated inputs, this adversarial strategy helps to advance the generator model's capabilities.

B. GAN Architecture

The structure of a GAN consists of two primary components. The function of the discriminator is to verify the authenticity of the computer-generated data produced by the generator. Depending on its authenticity, a data instance is classified as true or false, likely to be part of the original training data). The generator attempts to deceive the discriminator through employing larger amounts of data and generating convincing outcomes. The generator and discriminator function as adversaries due to their conflict. The goal of the generator is to create realistic looking things, whereas the discriminator aims to identify items that are counterfeit. These two elements practice together, gradually enhancing their powers. It could pick up on and imitate complicated training material, such as voice, video, and picture. A whole GAN architecture is shown in the schematic in Fig. 2. A GAN employs the following fundamental process:

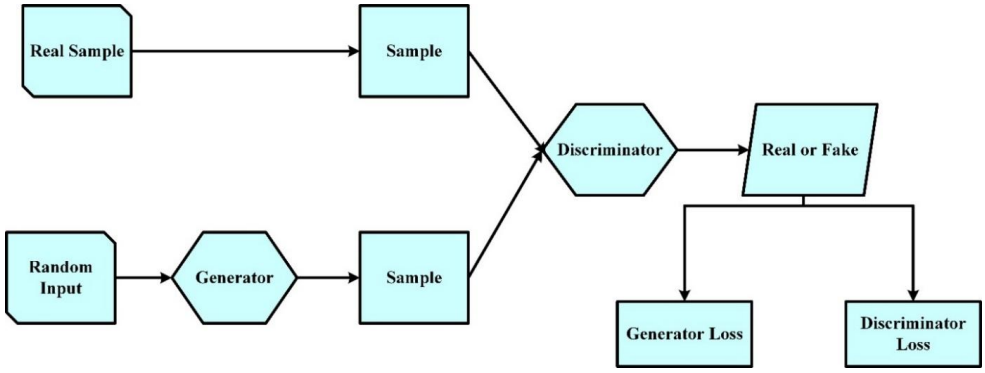


Fig. 2. GNN architecture.

- The generator accepts an input that contains random numbers.
- The generator transforms the input into an image.
- The discriminator takes in both the image produced by the generator and extra, genuine images.
- The discriminator compares the complete set of photos in an effort to distinguish between authentic and false ones.
- The discriminator provides a forecast for each image, expressing the likelihood of authenticity as a number between 0 and 1. A fraudulent image receives a score of 0, whereas a true image receives a score of 1.

A constant feedback loop is produced by this approach. The generator provides the discriminator with fresh, enhanced created images while the discriminator establishes the ground truth (empirical truth) for image inputs.

C. Convolutional Neural Network

CNNs, alternatively referred to as Convolutional Neural Networks, have been developed to scrutinize data utilizing a grid format. These tools excel in processing images and videos as they possess the ability to comprehend the connections between various elements within them. Convolutional layers contribute to the complexity of CNNs [17], differentiating them from other types of neural Convolution, a mathematical operation involving matrix multiplication, is employed to process the data input into networks. The convolution method utilizes small squares of input data to analyze image characteristics and retain pixel arrangement. A typical CNN design is shown in Fig. 3.

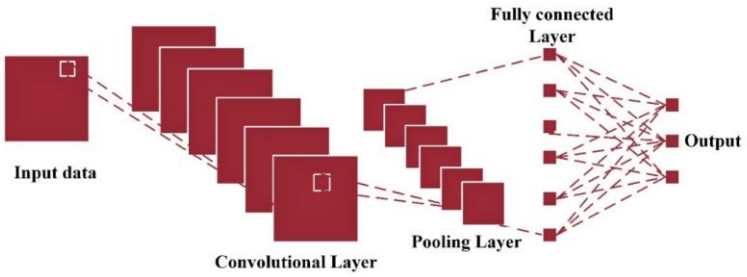


Fig. 3. Basic architecture of CNN.

1) Convolutional layers

Convolutional layers employ filters or kernels to examine and study the given input data. The purpose of filters is to detect particular patterns or features, such as lines, corners, or intricate shapes within deeper layers. When things in the picture move, the filters make a map that shows where they are. A specialized map displaying various attributes is produced by manipulating the image through a filter. It exists because of the convolutional layer. Through the utilization of convolutional layers in unison, we can construct more sophisticated models capable of extracting more specific data from photographs. It is the responsibility of the convolutional layers to eliminate particular characteristics from the input images. Various characteristics encompass aspects like the meeting point of two lines, the edges of an object, the tactile sensation of something, or even intricate patterns.

2) Pooling layers

Utilizing pooling layers aids in diminishing the size of the input and simplifying its handling. The convolutional layers are followed by the subsequent layers. To put it simply, the "spatial dimensions" in images refer to the image's width and height. The arrangement of small squares in rows and columns forms a grid through which an image could be observed. Through the implementation of pooling layers, the input space is reduced and the amount of information within a network is decreased. This technique prevents the model from closely matching the training data, resulting in faster training. Max pooling makes the model invariant to small transitions and reduces computational complexity by lowering the size of the feature map. Without max pooling, the network would not be able to identify features even little rotations or shifts. As a result, the model would be less resistant to changes

in object location inside the image, potentially reducing accuracy. Pooling consists of two main variants: max pooling and average pooling. The process of max pooling enables us to locate the highest numerical value within each feature map. For example, if the pooling window size is 22, it would select the pixel with the highest value out of a group of 22 pixels. In the area of pooling, max pooling is utilized to pinpoint the utmost crucial characteristic or feature. The average of all values within the pooling window is determined via average pooling. It offers a uniform, typical feature depiction.

3) Fully connected layers

Among the essential layers in a Convolutional Neural Network (CNN), the fully-connected layer holds a prominent place. In a fully connected layer, all neurons are interconnected with every neuron found in the layer beneath it. To enable predictions and labelling of inputs, Convolutional Neural Networks (CNNs) typically incorporate fully connected layers at the end, which utilize the features extracted from convolutional and max pooling layers. The high-dimensional output from the previous convolutional and pooling layers is flattened into a one-dimensional vector by fully connected layers. Instead of focusing on localized features, the network can now mix and integrate all of the retrieved information throughout the entire image. It aids in comprehending the image's larger context. The integrated features are mapped to the desired output, such as class labels in classification tasks, via the fully connected layers. They serve as the network's last decision-making component, figuring out what the features that were extracted signify in relation to the particular issue at hand. A hierarchy of features is produced by combining the Convolution layer with the max-pooling layer, then related sets. Simple patterns are picked up by the top layer, which is then built upon to pick up more complicated patterns.

D. GAN-CNN Mechanism

Leveraging a Generative Adversarial Network (GAN) combined with a Convolutional Neural Network (CNN) holds immense potential for advancing smart grid security and optimizing efficiency in monitoring residential electricity consumption while simultaneously bolstering anomaly detection capabilities. This innovative approach harnesses the power of GANs to generate synthetic data closely resembling real-world electricity consumption patterns. The GAN component consists of a generator network that creates synthetic consumption data and a discriminator network that evaluates the authenticity of these samples. Through an adversarial training process, the generator learns the intricate distribution of normal consumption patterns, enhancing its ability to mimic actual behaviors. Complementing the GAN, the CNN comes into play by processing the electricity consumption data, which could take the form of time-series data or spectrograms, depending on the application. The CNN's convolutional layers adeptly extract intricate features and subtle patterns from the input data, thereby enabling it to discern deviations from typical consumption behaviors. These deviations can encompass a range of anomalies, including sudden spikes, irregular consumption patterns, or unusual drops in usage. The real strength of this integrated GAN-CNN framework lies in its anomaly detection capabilities. The CNN-processed data is meticulously compared to established consumption profiles and predefined thresholds. Whenever deviations that might signify anomalies are detected, the system promptly triggers alerts or initiates predefined responses. These responses can include notifying grid operators for manual intervention, isolating the affected area of the grid, or implementing automated corrective actions. Furthermore, the system continually refines its models through reinforcement learning, ensuring adaptability to evolving consumption patterns and emerging security threats. In essence, the incorporation of a GAN-CNN framework into smart grid systems offers a robust solution for enhancing security, improving efficiency, and bolstering anomaly detection in residential electricity consumption. This approach not only safeguards the stability of residential electricity grids but also ensures a more optimized and resilient energy infrastructure capable of responding effectively to the ever-evolving demands and challenges of the modern world (Fig. 4).

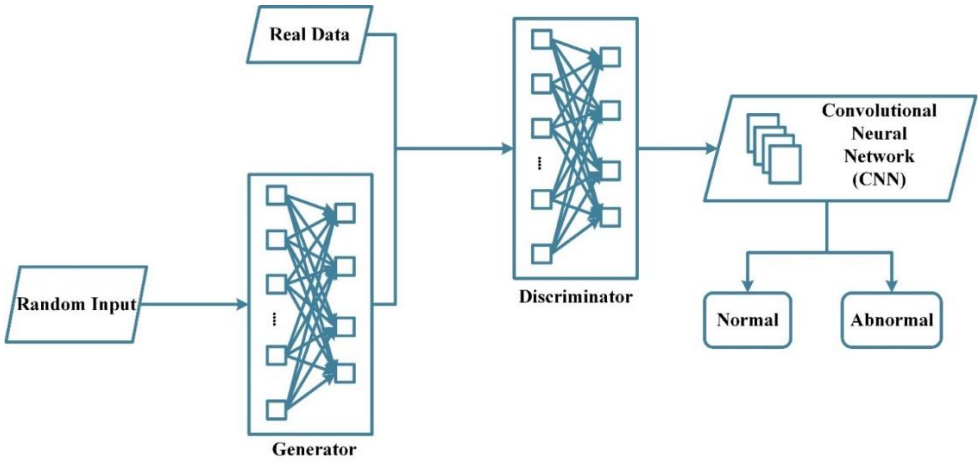


Fig. 4. Workflow of the GAN-CNN mechanism.

V. Result and Discussion

In this study, suggested a novel technique for detecting anomalies in home electricity consumption was put into practice by combining a Generative Adversarial Network (GAN), a Convolutional Neural Network (CNN), and the Isolation Forest algorithm. Python was used for this implementation, utilizing a number of machine learning libraries. The efficiency of this methodology in finding abnormalities in data on home electricity consumption was very high. A reliable method was found to be the combination of Isolation Forest for initial anomaly identification and GAN-CNN for additional analysis and confirmation. The findings of the experiments are presented below, along with a discussion of their implications.

A. Experimental Outcome

1) Daily mean electricity consumption

The daily mean power consumption is a crucial indicator for comprehending patterns of energy use, maximizing resource allocation, and assisting in the development of a more resilient and effective energy infrastructure. It is a crucial tool for utility companies, decision-makers, and customers who want to manage and conserve energy effectively. When you look at the chart, certain patterns become apparent. Consumption is typically higher during the weekdays, peaking in the late afternoon and early evening when homes and businesses prepare for various activities. On the other hand, consumption tends to be lower during the day and more constant throughout the day on weekends. Utility firms could discover these insights extremely useful since they can use them to distribute resources more effectively, anticipate periods of high demand, and improve system stability. This graphical representation highlights the need of tracking and managing energy usage for both utilities and consumers while also illuminating daily changes in electricity consumption. The daily mean electricity consumption based on original data is shown in Fig. 5.

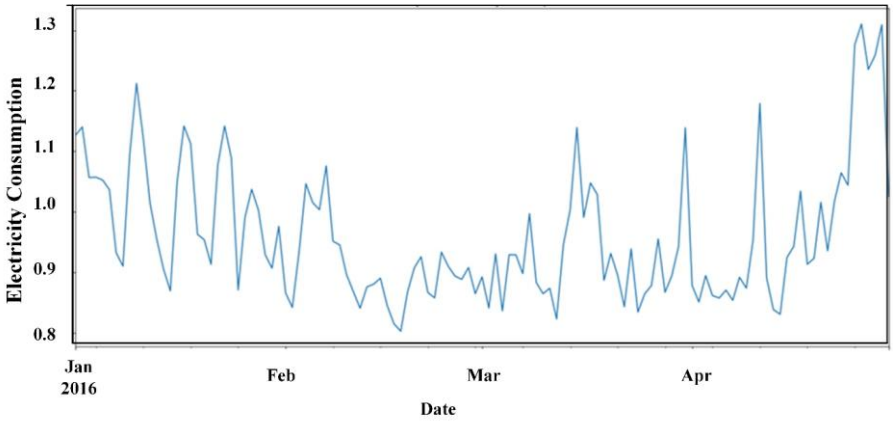


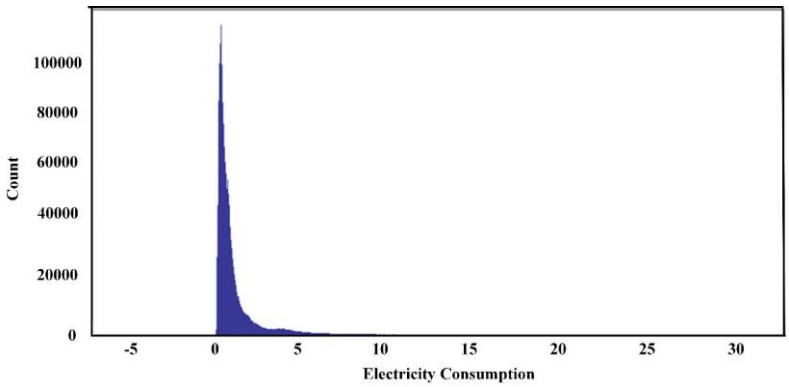
Fig. 5. Daily mean electricity consumption.

2) Distribution of electricity consumption in sample and original data

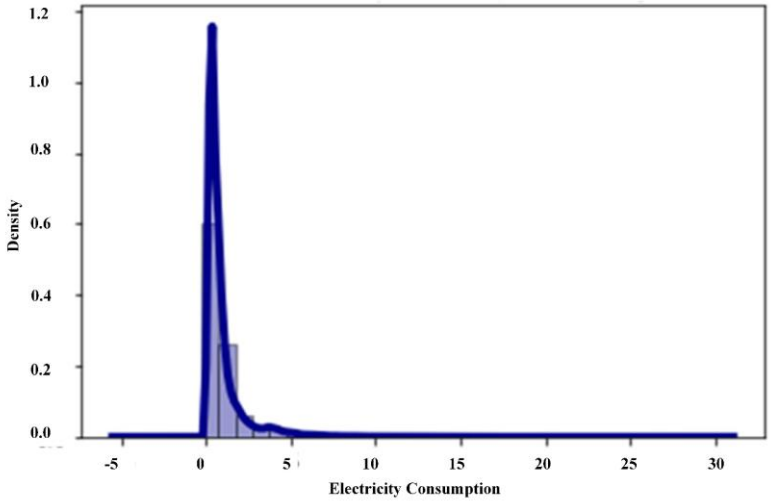
For energy management and planning, it is crucial to know how electricity consumption is distributed throughout a sample dataset. As illustrated by Fig. 6(a) and (b), both the sample and the original data's distribution of power use exhibit a fairly typical pattern, with a peak around the intermediate consumption values. Numerous household and business energy datasets share this trait. Most observations show moderate consumption, which suggests typical usage behaviors. However, there are anomalies with noticeably higher consumption values, which may be explained by a number of variables like severe weather, broken equipment, or specialized industrial facilities. It is vital for a number of reasons to analyze how electricity consumption is distributed in the sample data. Utility businesses could more effectively distribute resources and foresee peak demand periods by using it to identify the variety of consumption patterns within their service areas. Additionally, it helps in developing pricing strategies and programmes for energy saving that are specific to certain consumer groups.

3) Average electricity consumption

A thorough visual depiction of data that changes with time and day is offered by a heatmap that compares hours of the day to days of the week. The y-axis of this heatmap reflects the days of the week, usually from Sunday through Saturday, while the x-axis represents the 24 hours of a day, starting at midnight and going through one-hour intervals. Each heatmap cell represents a distinct hour and day combination, with the colour intensity representing the value or intensity of a variable at that particular time. A fast evaluation of temporal patterns is made possible by the colour scale, which uses lighter colours for lower values and deeper hues for higher values. This visualization technique is incredibly helpful for distinguishing between regular and abnormal trends, as well as for pinpointing peak consumption times, discovering anomalies, and tracking changes in patterns over time. The heatmap in Fig. 7 displays the typical electricity consumption.



(a)



(b)

Fig. 6. Distribution of electricity consumption in (a) Sample data and (b) original data.

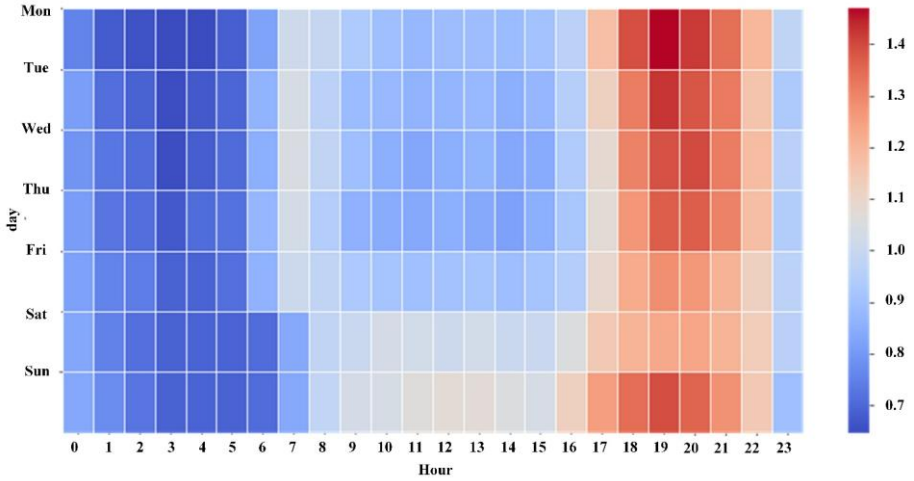


Fig. 7. Average electricity consumption (day vs. hour).

4) Rolling average electricity consumption

The graph below investigates the fluctuation in the annual average electricity use. The average monthly power usage is displayed on the vertical axis in units, with each point on the chart representing a particular month. The amount of power used appears to follow clear seasonal trends. Typically, more cooling systems and air conditioners are used during the summer, when demand is highest. On the other hand, during the milder spring and autumn seasons, consumption may decrease. By studying this data, utility companies might predict peak demand periods and allocate resources accordingly, leading to more efficient energy management. Fig. 8 shows graphically the suggested strategy's rolling average electricity consumption.

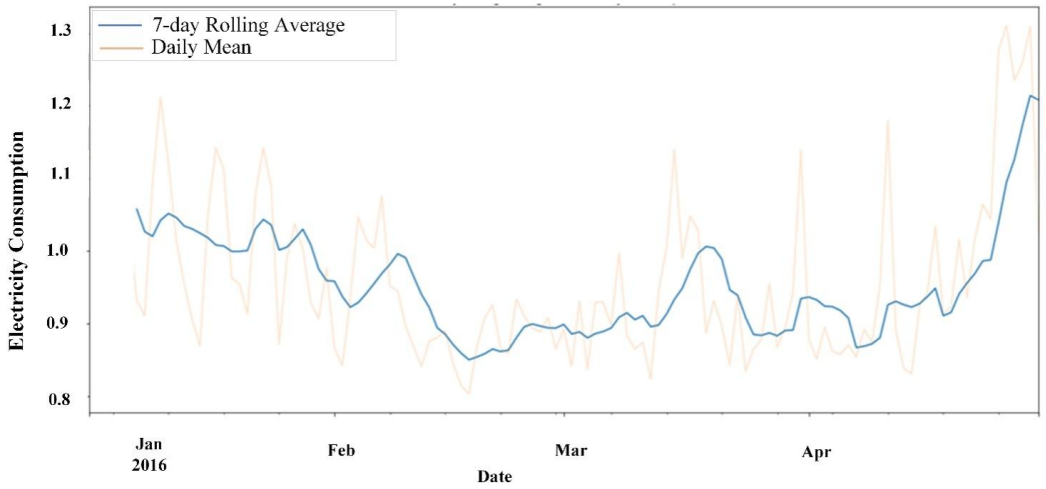


Fig. 8. Rolling average electricity consumption (7-day).

5) Electricity consumption by hour

The day-to-day variations in electricity consumption can be clearly seen in this representation. Common patterns could be increased usage during the day as homes and businesses power up for various activities and reduced consumption at night as demand tends to decline. Fig. 9 can also show any spikes in consumption or other unexpected patterns, which is useful for grid design and energy management. Utility firms can more efficiently distribute resources, foresee peak demand, and create initiatives to encourage energy saving during periods of high usage by understanding these consumption trends. By seeing chances for efficiency improvements, it also gives consumers the power to make knowledgeable decisions about their energy usage and perhaps even save money.

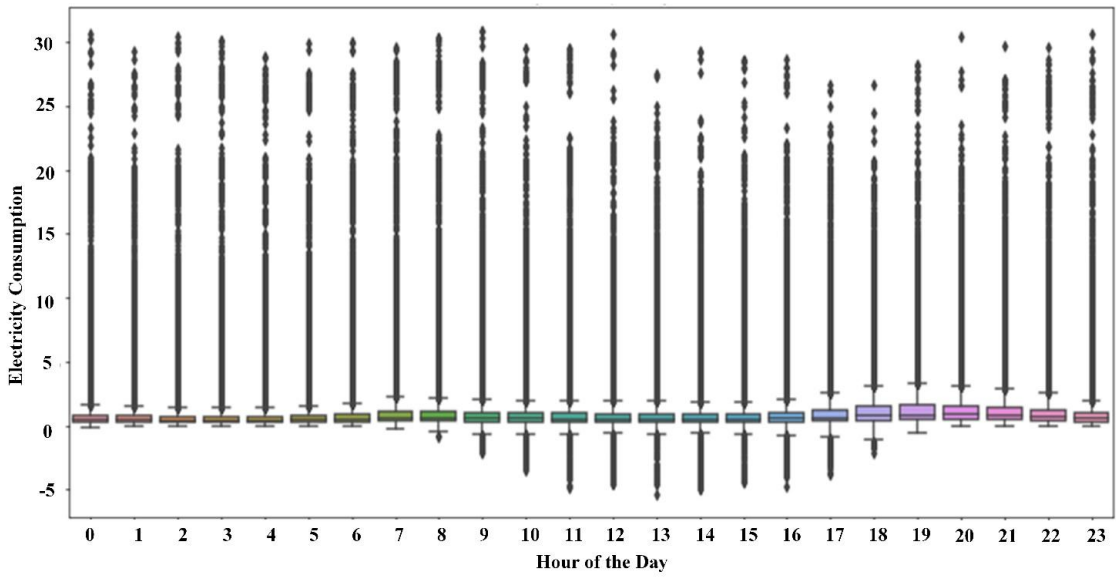


Fig. 9. Electricity consumption by hour.

6) Anomaly detection using Isolation Forest

A powerful method for locating odd patterns or outliers in a collection is anomaly detection utilizing the Isolation Forest. This technique works by identifying anomalies, or data points, which are those that differ considerably from the majority of the data. The Isolation Forest utilizes a tree-based strategy to effectively isolate anomalies, in contrast to conventional methods that rely on the density of data points, making it particularly effective in high-dimensional landscapes. The Isolation Forest is a versatile and well-liked technique in the field of machine learning and data analytics due to its capacity to find abnormalities without requiring in-depth prior knowledge of the data distribution. Fig. 10 illustrates the suggested technique's recognition of anomalies.

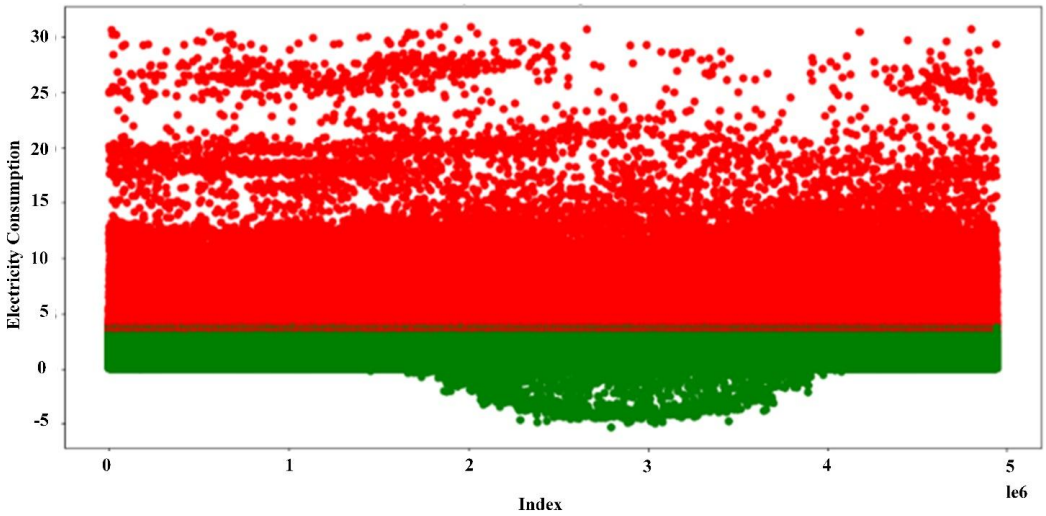


Fig. 10. Anomaly detection using isolation forest.

7) ROC curve

The effectiveness of a classification model is graphically depicted by a ROC curve (Receiver Operating Characteristic). At different classification thresholds, it demonstrates the trade-off between true positive rate (sensitivity) and false positive rate (1-specificity). ROC curves often have values between 0.0 and 1.0. However, a surprising case can be illustrated by a ROC curve with an AUC (Area Under the Curve) of 1.0. A classifier that completely distinguishes between positive and negative examples without making any mistakes has an AUC of 1.0. It delivers 100% sensitivity and 100% specificity across all categorization thresholds, in other words. The ROC curve is depicted graphically in Fig. 11.

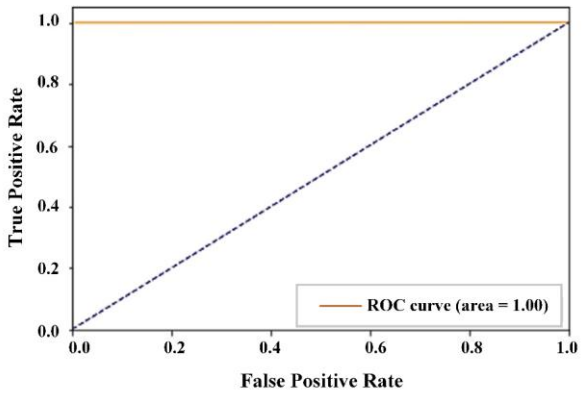


Fig. 11. Receiver Operating Characteristic (ROC) curve.

8) Confusion matrix

An essential tool for assessing the effectiveness of classification model performance is a confusion matrix. To distinguish between true positives, true negatives, false positives, and false negatives, it gives a structured summary of a model's predictions. The accuracy, precision, recall, specificity, F1-score, and false positive rate could be calculated using this 2×2 matrix, providing a thorough evaluation of a model's capacity to make accurate classifications and pointing out potential improvement areas in binary or multi-class classification problems. The confusion matrix for the suggested strategy is presented in Fig. 12.

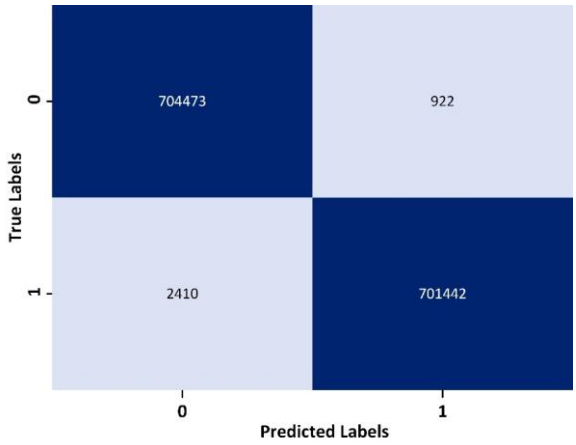


Fig. 12. Confusion matrix.

9) Model performance

The suggested Generative Adversarial Network with Convolutional Neural Network (GAN-CNN) model's performance metrics highlights the model's outstanding accuracy of 0.9976, which denotes a high percentage of accurate classifications. The model also shows a low loss of 0.0070, which indicates little uncertainty in its predictions. When taken as a whole, these metrics show how well the model performs data classification, indicating that it can make decisions that are accurate and dependable when used for the task at hand. To ensure its robustness and generalizability, a thorough examination should take into account other metrics and potential testing on unreleased data.

When training and assessing machine learning models, the accuracy and loss graph is a vital visualization. This dynamic chart illustrates how accuracy and loss change with each training period while tracking the model's performance over time. The model's capacity to make accurate predictions is demonstrated by the accuracy histogram, which is typically depicted as a blue bar. An upward trajectory indicates improved performance as

training goes on. The loss histogram, which is shown in orange, on the other hand, emphasizes the diminishing error between the model's predictions and the actual results. A falling loss indicates that the model is improving its prediction accuracy as training progresses These curves together shed light on the model's training dynamics. The characteristics of a well-trained model that successfully extracts patterns from the data are high accuracy and low loss. This visualization helps deep learning experts keep track of and optimize their models for the best outcomes.

B. Performance Evaluation of the Proposed Method

To perform an extensive performance assessment of the suggested anomaly detection approach for residential electricity consumption. It makes utilization of important indicators like Accuracy, Precision, Recall, and the F1-score to evaluate how well the strategy is working to improve smart grid security and efficiency monitoring.

1) Accuracy: The percentage of accurately detected instances (including true positives and true negatives) relative to all instances is known as accuracy. It gives a general evaluation of the method's capacity to distinguish between anomalies and non-anomalies as expressed in Eq. (2).

$$Accuracy = \frac{TP+TN}{Total\ Instances} \quad (2)$$

2) Precision: Precision determines the proportion of accurate positive forecasts to all positive predictions. It assesses the method's capacity to prevent false alarms by confirming that any anomaly that is flagged is actually an anomaly expressed in Eq. (3).

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

3) Recall: Recall, measured as the ratio of true positives to all actual positives, reflects the method's capacity to accurately identify all genuine abnormalities. It assesses how well the technique captures abnormalities is expressed in Eq. (4).

$$Recall = \frac{TP}{TP+FP} \quad (4)$$

4) F1-score: The harmonic average of recall and precision is known as the F1-score. When there is an imbalance between the positive and negative classes, it offers a fair assessment of the technique's effectiveness is expressed in Eq. (5).

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (5)$$

The proposed anomaly detection approach achieves high marks for accuracy, precision, recall, and F1-score, demonstrating its efficacy in boosting smart grid security and efficiency monitoring of residential consumption of electricity. The accuracy of anomaly detection is crucial for grid sustainability and reliability, and these results demonstrate the viability of our technique in real-world situations. Table I displays the preliminary evaluation of the suggested strategy.

TABLE I. PERFORMANCE ASSESSMENT

Metrics	Proposed GAN-CNN
Accuracy	0.99
Precision	0.98
Recall	0.95
F1-score	0.97

The model provides excellent accuracy of 0.99, which means that almost every instance in its evaluation is correctly classified. Its effectiveness in minimizing false positives is further highlighted by the precision score of 0.98, which shows that the bulk of the positive instances it predicts are in fact true. A significant number of the real positive cases were captured by the model, as seen by the recall score of 0.95. With a 0.97 F1-score, a standardized metric for precision and recall, the performance in handling both false positives and false negatives is well-balanced. These metrics demonstrate the model's powerful categorization abilities as a whole, but it's important to consider if these metrics are accurate given the task at hand and its particular context.

C. Findings Comparison between Suggested Work and Existing Works

The performance metrics for AlexNet, Super-Resolution Convolutional Neural Network (SR-CNN), K-Nearest Neighbours-Support Vector Machine (KNN-SVM), and a proposed Generative Adversarial Network-Convolutional Neural Network (GAN-CNN) are compared in Table II. Accuracy, precision, recall, and F1-measure seem to be the main metrics used to evaluate these models in a classification setting. It is noted that the proposed GAN-CNN model outstands with amazing accuracy of 0.99, indicating its exceptional capacity to correctly categorize examples when comparing with other models like AlexNet, SR-CNN and KNN-SVM. Additionally, it has a remarkable precision rate of 0.98, which shows a low percentage of false positives. However, it's important to remember that selecting the best model should be based on the requirements of a particular assignment, as different models may perform better in certain elements of classification performance.

TABLE II. COMPARISON ASSESSMENT

Metrics	Accuracy	Precision	Recall	F1-measure
AlexNet	0.91	0.89	0.92	0.94
SR-CNN	0.90	0.87	0.85	0.89
KNN-SVM	0.75	0.79	0.81	0.85
Proposed GAN-CNN	0.99	0.98	0.95	0.97

The suggested approach has the highest accuracy when measured against alternative methods. The graphical representation highlights the suggested approach's enhanced efficacy in correctly identifying cases, highlighting its potential as a workable solution in the existing situation.

D. Discussion

The GAN-CNN implementation on the Python platform has shown its potential for anomaly identification in residence electricity consumption. This technique, which makes utilization of deep learning approaches, provided a

solid means of improving anomaly detection. This system's efficacy is greatly influenced by its high accuracy and precision. It is essential for maintaining the system's dependability and trustworthiness, which are of the utmost importance in real-world applications, to be able to reliably detect true anomalies while minimizing false alarms. Isolation Forest was employed for preliminary detection and GAN-CNN for confirmation, using the best features of both methods. This hybrid strategy made the procedure of finding anomalies more thorough and precise. Since it enables quick detection of anomalies as they happen, the real-time monitoring capability is a key benefit. This ensures that prompt steps may be taken to safeguard the integrity of home electricity usage. It is particularly crucial for addressing problems like power theft, device faults, or cybersecurity risks. Additionally, the efficiency optimization component holds promise because it has the potential to produce significant advantages. Utilities could decrease costs and contribute to the sustainability of the energy system, aligning with more general environmental aims, by optimizing energy distribution and consumption. It's crucial to understand that the calibre and representativeness of the training data, for example, might have an impact on how well the methodology performs. As a result of this research, it is clear that employing GAN-CNN to detect anomalies in home electricity consumption has enormous potential to increase security and effectiveness. A more dependable, safe, and sustainable household energy infrastructure may be made possible by the integration of precise anomaly detection, real-time monitoring, and efficiency optimization.

VI. Conclusion

A number of significant findings have been drawn from the research on anomaly identification in home electricity usage utilizing a Generative Adversarial Network (GAN) with Convolutional Neural Network (CNN) to discover anomalous patterns and enable improved smart grid security and efficiency monitoring. The GAN-CNN method proved successful in identifying irregularities in residence electricity consumption trends. It could spot anomalous consumption patterns that could point to smart grid defects, malfunctions, or even security breaches by utilizing the power of deep learning and generative modelling. The study emphasizes the potential of the GAN-CNN model as a useful tool in reaching this goal, which is to increase smart grid security, which is of utmost importance.

It could help in spotting unusual activity that might be a sign of a cyberattack or grid tampering, improving overall security. The investigation highlights the significance of efficiency monitoring in the context of home electricity usage in addition to security. Unusual patterns of consumption could result in energy waste and inefficiency, which could be expensive and harmful to the environment. The GAN-CNN method helps to identify these inefficiencies. When it comes to identifying intricate patterns in data on electricity consumption, the GAN-CNN model excels. This is essential for correctly recognizing anomalies because traditional rule-based approaches could have trouble capturing the complexity and variety of real-world consuming behaviors. The capacity of the GAN-CNN system to carry out anomaly detection in real-time is a key benefit. As a consequence, abnormalities could be quickly identified and fixed, avoiding possible interruptions and maintaining the efficient operation of the smart grid. It's crucial to remember that the effectiveness of this strategy strongly depends on the calibre of the data utilized and the pre-processing procedures.

The effectiveness of the model depends on the accurate labelling of anomalies and careful handling of missing or noisy data. Scalability issues are critical for larger smart grids with more complicated consumption patterns. The GAN-CNN approach needs more investigation and testing to make sure it can handle the size and complexity of real-world smart grid data. The research shows the potential of combining a convolutional neural network with a generative adversarial network for anomaly identification in data on residence electricity consumption. This strategy helps in monitoring and enhancing energy efficiency as well as enhancing smart grid security by identifying potential threats. To effectively utilize the positive aspects of this technology in actual smart grid scenarios, implementation and scalability issues need to be taken into account.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Badari Narayana Palety conducted the research; C. Mahalakshmi analyzed the data; P. Nagasekhar Reddy wrote the paper; all authors had approved the final version.

References

[1] K. K. Zame, C. A. Brehm, A. T. Nitica, C. L. Richard, and G. D. Schweitzer III, "Smart grid and energy storage: Policy recommendations," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1646–1654, 2018.

[2] S. K. Rathor and D. Saxena, "Energy management system for smart grid: An overview and key issues," *International Journal of Energy Research*, vol. 44, no. 6, pp. 4067–4109, 2020.

[3] F. U. M. Ullah, A. Ullah, I. U. Haq, S. Rho, and S. W. Baik, "Short-term prediction of residential power energy consumption via CNN and multi-layer bi-directional LSTM networks," *IEEE Access*, vol. 8, pp. 123369–123380, 2019.

[4] M. Massaoudi, H. Abu-Rub, S. S. Refaat, I. Chihi, and F. S. Oueslati, "Deep learning in smart grid technology: A review of recent advancements and future prospects," *IEEE Access*, vol. 9, pp. 54558–54578, 2021.

[5] K. Kimani, V. Oduol, and K. Langat, "Cyber security challenges for IoT-based smart grid networks," *International Journal of Critical Infrastructure Protection*, vol. 25, pp. 36–49, 2019.

[6] Y. Li, X. Peng, J. Zhang, Z. Li, and M. Wen, "DCT-GAN: Dilated convolutional transformer-based gan for time series anomaly detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, 2021.

[7] K. Hameed, M. Barika, S. Garg, M. B. Amin, and B. Kang, "A taxonomy study on securing Blockchain-based Industrial applications: An overview, application perspectives, requirements, attacks, countermeasures, and open issues," *Journal of Industrial Information Integration*, vol. 26, 100312, 2022.

[8] Y. Zhang, Q. Ai, H. Wang, Z. Li, and X. Zhou, "Energy theft detection in an edge data center using threshold-based abnormality

detector," *International Journal of Electrical Power & Energy Systems*, vol. 121, 106162, 2020.
[9] A. Aldegheishem, M. Anwar, N. Javaid, N. Alrajeh, M. Shafiq, and H. Ahmed, "Towards sustainable energy efficiency with intelligent electricity theft detection in smart grids emphasising enhanced neural networks," *IEEE Access*, vol. 9, pp. 25036–25061, 2021.
[10] D. Yao, M. Wen, X. Liang, Z. Fu, K. Zhang, and B. Yang, "Energy theft detection with energy privacy preservation in the smart grid," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7659–7669, 2019.
[11] S. Banik, S. K. Saha, T. Banik, and S. M. Hossain, "Anomaly detection techniques in smart grid systems: A review," in *Proc. 2023 IEEE World AI IoT Congress (AIIoT)*, 2023, pp. 331–337.
[12] Y. Luo, Y. Xiao, L. Cheng, G. Peng, and D. Yao, "Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities," *ACM Computing Surveys (CSUR)*, vol. 54, no. 5, pp. 1–36, 2021.
[13] P. Radoglou-Grammatikis *et al.*, "Spear siem: A security information and event management system for the smart grid," *Computer Networks*, vol. 193, 108008, 2021.
[14] Pecan street electricity data. [Online]. Available: https://www.kaggle.com/datasets/zhitingzheng/pecan-street-electricity-data
[15] K. Sadaf and J. Sultana, "Intrusion detection based on autoencoder and isolation forest in fog computing," *IEEE Access*, vol. 8, pp. 167059–167068, 2020.
[16] X. Xia *et al.*, "GAN-based anomaly detection: A review," *Neurocomputing*, vol. 493, pp. 497–535, Jul. 2022. doi: 10.1016/j.neucom.2021.12.093
[17] W. Ullah, A. Ullah, I. U. Haq, K. Muhammad, M. Sajjad, and S. W. Baik, "CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks," *Multimed Tools Appl*, vol. 80, no. 11, pp. 16979–16995, May 2021. doi: 10.1007/s11042-020-09406-3