IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, September 2025
Impact Factor: 7.67

Adversarial Autoencoding Framework for Unsupervised Cyber Intrusion Detection in Smart Grid Distribution with Renewable Energy Integration

Mr. Gajanan B. Kadam[1], Prof. Sushil V. Kulkarni[2], Prof. Vijay M. Chandode[3]
M.B.E.S. Society's College of Engineering, Ambajogai, India[1]
Professor, Department of CS-IT, M.B.E.S. Society's College of Engineering, Ambajogai, India[2]
Head of Department, Department of CS-IT, M.B.E.S. Society's College of Engineering, Ambajogai, India[3]

Abstract: *The increasing integration and advancement of digital technologies in power distribution grids has significantly enhanced operational efficiency, yet it has simultaneously heightened vulnerability to various cyber attacks that pose severe risks. This paper presents an innovative and cutting-edge approach that utilizes Unsupervised Adversarial Autoencoders (UAAEs) specifically for the critical detection of cyber threats that are targeting power distribution systems. By leveraging the powerful capabilities of unsupervised learning, the proposed model effectively identifies and discerns anomalies without the requirement for labeled data, which is frequently scarce and hard to acquire in many real-world scenarios. The UAAE architecture is made up of a generator and a discriminator that work collaboratively to learn and establish the normal operational patterns of the grid, which enables the effective detection of deviations that are indicative of potential cyber attacks. Through extensive and rigorous experimentation on simulated datasets reflecting the operations of power distribution grids, the proposed method demonstrates superior performance in terms of accuracy and recall when compared to traditional detection mechanisms that are currently in use. Additionally, the implementation of adversarial training significantly enhances the model's robustness against evasion tactics that are often employed by sophisticated and determined attackers. The findings underscore the potential of UAAEs as a pivotal and vital tool for enhancing the cybersecurity posture of power distribution grids, ultimately contributing to more resilient and secure energy infrastructures that can better withstand cyber threats. Through this approach, we aim to significantly bolster the defenses against potential vulnerabilities in the increasingly digital landscape of power distribution systems.*

Keywords: Unsupervised Learning, Adversarial Autoencoder, Cyber Attack Detection, Power Distribution Grids, Anomaly Detection, Cybersecurity, Robustness, Energy Infrastructure, Machine Learning

I. INTRODUCTION

Power Grids (PGs) are a key part of the infrastructure of a country and are of significant importance as they provide electricity for domestic use and industrial requirements. A smart grid differs from a conventional electrical grid in that it uses digital technology to monitor the flow of electricity from all generation sources and demand-side devices. Communication and control technologies, advanced monitoring and sensing technologies, distributed generation planning and optimization tools, and Demand Response (DR) technologies are all tools used in a smart grid. A smart grid is more efficient, flexible, responsive, sustainable, and secure than a conventional grid. Enhanced control, improved reliability, better demand response, increased security for distributed energy based control systems, active grid monitoring and diagnostics, and advanced tools for economic scheduling are all aspects of its potential benefits. Even though a smart grid contains many advantages, it may expose PGs to new vulnerabilities and breaches. The new

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, September 2025 Impact Factor: 7.67

developers of the smart grid are generally new to the energy and electric industry and do not have experience in modeling and evaluating vulnerability measurements, which makes it even more challenging[1].

A Smart Grid (SG) is a grid that uses Smart Meters (SM), Supervisory Control and Data Acquisition (SCADA) system, Distribution Management System (DMS), etc. to monitor, control, react, and predict with the use of computational techniques and Artificial Intelligence (AI) tools for an efficient and effective breakdown inspection and prevention. A smart grid differs from a conventional grid in that it uses digital technology to monitor the flow of electricity from all generation sources and electricity demand, and that it contains a central SCADA system and machine-to-machine (M2M) communication. A smart grid collects data of the consumption, harvesting, prosumption, and trading of generated energy with and for neighbours as per market operation system while responding to threats to cyber and physical systems. Cyber attacks directed toward the power infrastructure may focus on certain critical plant wide control functions of PGs. These may be the generation and ancillary schedule control, automatic generation control (load frequency control), voltage and reactive power control, demand side control, protection schemes, and others. Cyber attack tools consist of a mix of software and hardware components. Hackers exploit the cyber architecture and conduct a series of preliminary steps such as network reconnaissance and mapping, gaining access to the system, privilege escalation, remote command execution, hiding the intrusion activities, and achieving command and control. The monitoring, recording, analysis, and correlation management systems are penetrated to execute commands for cyber attacks.

Self-control on the information and decision aspect of the power so as to cope with attacks on monitoring and control is also very important. The current and projected architecture of an SG may be subject to different attack threats and other risks. Scamming attacks have been observed globally and many governments are early at their alarm but still have limited understanding at the SBUs level. Cyber attacks to PG systems, and detection of deception/insider attacks is also a significant challenge. Due to the existence of many uncertainties in loading, generation, and distributed energy resources, securing the operation of PGs against deception/fraudulent signals or inside attacks is a daunting task [2].

II. BACKGROUND

Cyber attacks on the Industrial Internet of Things (IIoT) threaten the security and safety of systems in various sectors such as critical infrastructures, healthcare, transportation, and social networks [3]. Smart power grids are one important critical infrastructure where the communication networks are vulnerable to multiple cyber-attacks. With advancements in distributed energy resources (DERs), such as energy storage and distributed generation, as well as some additional constraints on demand management, distributed energy management has emerged as a promising approach in power distribution systems[4][5]. These transmission and distribution-grid-connected devices commonly rely on the IEEE 61850 standard protocol, which is vulnerable to false data injection attacks (FDIAs). An FDIA is a type of cyberattack that corrupts the transmitted data from sensors on the devices to the supervisory control and data acquisition (SCADA) system by gaining control over the sensors and relay devices [6]. Detection and localization of FDIAs in smart electricity distribution grids are novel problems that have recently attracted significant interest. However, due to their unbalanced configuration and nonlinear behavior, as well as the unknown characteristics of the cyber-attacks, detection of FDIAs is challenging but important. Additionally, the vulnerability of smart grids to FDIAs, especially the failure of estimators, is a probability that should be constantly evaluated to reduce the risk against their operation. Despite existing extensive malicious false data generation and injection functions used in real-life cases, the direct use of these functions for detection raises a new challenge that the model should be thoroughly trained for a wide range of these functions beforehand due to the high diversity of these false data generation and injection functions. Failure of any smart grid device can cause cascaded grid failure. Hence, another more valuable opportunity to improve the security of the grid and its devices is to discover or utilize methods to capture such anomalies without training data development. In this regard, several unsupervised anomaly detection methods have been developed but either rely on the abstract models or mathematical definitions of the systems that may not be available for complicated smart systems. An unsupervised adversarial autoencoder (AAE) is proposed to detect FDIAs in the unbalanced smart distribution grid integrated with DERs based on the time-series measurements of the state estimators by modeling the unknown distribution of the normal states of the system. The proposed model uses the generative adversarial network (CNN) and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, September 2025 Impact Factor: 7.67

long short-term memory (LSTM) to better reconstruct the input time-series data and capture their temporal dependencies, respectively. The advantage of the proposed model is that the anomalous points can be detected without a need for either mathematical modeling or ample amount of data on normal states of the system.

2.1. Cybersecurity in Power Distribution Grids

Cybersecurity has become a matter of grave concern for industries and governments across the world. With the rapid movement towards digitalization and cloud-based operations, the frequency of cyberattacks has soared. Utilities around the world rely on advanced systems for operations, and as they become interconnected, they become more vulnerable to cyberattacks and therefore can be at the mercy of adversaries who can cause huge grid disturbances and outages. Therefore, protecting power systems against cyber threats is of utmost importance, and there is an urgent need for developing effective intrusion detection systems. [7]

In modern power systems, wide-area measurements are collected, transmitted, and analysed in applications such as state estimation and market trading. Intelligent electronic devices collect measurements from nearby nodes in the system and send them to remote servers over the communication network. [8]Remote devices such as phasor measurement units convert analogue measurements to digital phasor form and communicate them to control centres over the network. The measurements typically pass through several routers, switches, and wireless transmission hardware before being received at destinations. Cyberattacks can be launched to penetrate the network at several locations along the path from the sending device to the receiving centres. Attackers can fake, delete, replay, or manipulate message packets by exploiting protocol vulnerabilities or device weaknesses. Cyberattacks pose severe threats to the operation and security of wide-area systems, and therefore, IDS-based approaches are essential to protect the critical infrastructure.

Machine learning-based detection mechanisms are useful in identifying any incongruities in the network. Machine learning algorithms can be supervised or unsupervised. Recent developments in unsupervised learning have provided many algorithms that try to model normal messages, and any significant deviation from the model is considered anomalous. One-class K-NN algorithms find a K-Nearest-Neighbor for each message in the training data. In this work, it uses time-series monitoring data for localization and classification of network attacks in SCADA systems. Recurrent neural networks are used for detection of false data injection attacks on PMU-based measurements.

2.2. Overview of Autoencoders

An autoencoder is a specific type of neural network used to encode an input to an output while maintaining its topology and semantics [6]. Simply put, it is meant to learn the structure of the data deterministically and losslessly compacts the data to a small hidden representation before reconstructing it to its original form. It consists of an encoder (determines the compact representation) and a decoder (maps the compact representation back to output). The whole architecture encodes the input to a latent space and then decodes it to reconstruct the input from the latent representation. Ideally, the output is equal to the input. Only one feed-forward pass is performed to encode/decode. In case of one hidden layer, the feed-forward is equivalent to determining the weight vector. This weight vector is then multiplied by the input vector and passed to the activation function to get the hidden representation. In the decoding process, the weight vector is transposed. [9]

Autoencoders can be trained to constrain the hidden representation as a Gaussian distribution shared among the data clusters. If the data points are from the same cluster, only passing through the same hidden unit index should yield similar values. Otherwise, one hidden unit is zeroed-out, or it yields too small values. [10][11]In reality, the latent space is almost never a perfect prior, so the upcoming latent samples are usually non-ideal, distributed, or irrelevant. To sample a vector from prior distribution, the mean and variance are determined through the encoder. A set of latent samples is generated using that mean/variance with a pre-specified prior sampling, and finally, the decoder passes those samples to observe the final output.

2.3. Adversarial Learning Concepts

Over the past years, adversarial learning methodologies have been widely utilized in data mining, computer vision, natural language processing, and audio processing. An extensive array of generative models exists, including

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, September 2025 Impact Factor: 7.67

Generative Adversarial Networks (CNNs), Adversarial Autoencoders (AAEs), and Energy-Based Models (EBMs). [12] In addition, adversarial learning consists of two crucial components, namely generative or domain rebuilding methods and classifiers or adversarial methods. In almost all applications, the former learn latent representations from data, while the latter try to detect data points deviating from nominal or natural data. [13][14]

Using autoencoders (AEs) has experienced a commendable rise, following the significant interest generated by deep learning methods. The combination of the popularity and flexibility of these AE-based architectures toward not only the reconstruction of data but also leveraging latent representation for the generation of input space data has enabled the adoption of the AK-based learning approaches. In this regard, novel models and implementations are examined combining different types of AEs, reconstructive or variational, generating data in the latent representation space after being learned on input domain data. Still, the attractiveness of their adversarial architecture has been preserved since previous adversarial learning approaches and methodologies rely on the ability of generative methods to capture the distribution of a dataset and distinguish between produced fake data and original natural datasets' data within a confidence range. It is noteworthy that AEs can be transformed into adversarial (AAEs) under a significant and proficient manner with no significant increase in complexity.

In another addition to the general implementation of a generative adversarial architecture, it incorporates a misclassification objective aimed at an approximation for the data distribution. This outcome indicates that switching from the traditional notion of a classifier to that of an adversary would make this architecture to be capable of supporting unsupervised informative prespecification of clusters despite all dimensions equivalency.

III. RELATED WORK

As technological advancement allows the global market to have better connectivity through the advent of smart grids, it provides unique capabilities and opportunities. Similar to the Internet, smart grids being connected through computer-based setups are vulnerable to cyber threats and external attacks on the Industrial Control System (ICS), which might disrupt the normal operation of the grid [6]. This holds true especially for smart power systems with a decentralized structure and unbalanced configurations, which are increasingly being adopted. As a continuous flow of two-way data communication occurs in smart power grids for faster response and improvements in efficiency and reliability, network vulnerability, data privacy violations, and/or aggregation attacks could be a cyber threat. Consequently, these events could create cascade failures resulting in interruptions to the power grid.

While there are many methods to detect Cyber-Physical Attacks (CPAs) on smart grids in the literature, attacks on the grid's physical side have not been observed in real-world scenarios. On the other hand, detections of the CPAs on the ICS are lagging behind as these types of attacks are complex and would be unnoticed if undetected, similar to network sniffers. Lack of understanding of the actual working mechanism of these events dictates all detection methodologies to either act conservatively as signal-based approaches or lose integrity and reliability as model-based methods. A better understanding of the types and effects of attacks on the grid could help create better defenses against the schemes.

Therefore, the development of a different and better attack detection solution with learned data-driven signatures of attack events. The motivation behind the chosen topic is the opportunity of merging two well-known branches of machine learning, which is the creation of probabilistic generative models using Deep Learning (DL) methods. Hence Anomalous Measurement Validation (AMV) or attack detection solution with the use of generative autoencoder architecture-based algorithms is proposed. The approach in this paper can generate enough clean normal operating points, likewise, DEA can measure the similarity of incoming points to this probability distribution. In turn, this could enable the cheap and alert-free detection of poorly-informed attacks.

3.1. Existing Cyber Attack Detection Methods

Cyber-physical (CP) security of smart grids has recently attracted much attention due to the increasing reliance on information and communication technology (ICT) and the internet of things (IoT) in the monitoring, protection, and control of substation automation (SA) systems. Consequently, attention has also been drawn on a holistic study of appropriate detection systems for attacks on substations and SCADA data corruption attacks. Normal operation of power systems is monitored by phasor measurement units, belonging to the family of PMUs, and such measurements

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, September 2025 Impact Factor: 7.67

should be protected. Cyber Attack Detection Systems (CADSs) operating on this protection are an interesting concept. False Data Injection Attacks (FDIAs) are an increasingly critical threat to power system operations that may compromise normal monitoring and protection of Smart Grids. Power system anomaly detection aims to determine whether a particular system state can be trusted while operating in normal or emergency operational conditions. However, attacks that evade detection, on monitoring points not protecting CAD. In addition to threats against normal operation, grid control HMI integrity must be preserved. HMI cyber layout attacks are a challenge to analyzing HMI structural integrity, and a novelty oscillation is introduced to an external agent to compromise HMI control information access, targeting persistence and stealthiness. Utilizing an AI-based architectural redesign process, closed-loop emulators, and a data-driven architecture aiming to capture the entire distributed interaction model, are approaches to AMI security and efficient detection of DDoS attacks [6]. The design of a dynamic PMU simulator is applied as an effective and efficient cybersecurity tool in dynamic protection systems to defend against information-based attacks.

3.2. Limitations of Supervised Approaches

Detection of cyber attacks in smart power distribution grids with unbalanced configurations poses challenges due to the inherent nonlinear nature of these uncertain systems. The unknown behavior of cyber attacks, especially false data injection attacks (FDIAs), and the limited amount of labeled data increases vulnerability and risk in the operation of the grids. Existing methods primarily rely on supervised learning algorithms that require a wealth of labeled data, which naturally is unavailable in practice [6]. Moreover, having scenarios that the models were never trained on may pose a serious risk that no alarm will be raised. In this regard, an unsupervised adversarial autoencoder (AAE) model is proposed to detect FDIAs in unbalanced power distribution grids integrated with DERs. The proposed method utilizes long short-term memory (LSTM) in the structure of the autoencoder to capture temporal dependencies in time-series measurements and leverages generative adversarial networks (CNNs) for better reconstruction of input data. The advantage of the proposed model is that, similar to autoencoders, it can detect anomalous points for the system operation without reliance on abstract models.

The efficacy of the approach is tested on IEEE 13-bus and 123-bus systems with historical meteorological and load data. The comparison of detection results of the proposed model with other unsupervised learning methods verifies its superior performance in detecting cyber attacks. Most existing works rely on supervised learning algorithms. The latter have been widely employed in the anomaly detection and cyber attack detection literature. However, supervised approaches have some limitations: 1) In realistic situations, especially for cyber security in smart grids, systems are constantly exposed to unrecognized scenarios; therefore, it is hard to gather a comprehensive wealth of labeled data. Having a wealth of labeled data helps to mitigate risk, while the limited amount of labeled data increases the vulnerability risk of the systems. 2) Even in cases where labeled data is available, supervised models usually could not detect attacks that they were never exposed to, meaning that they may very well perform over some attacks while completely fail over others. 3) As mentioned, in practical applications, it is impossible to monitor all the state space of the system; therefore, achieving labeled data for all states is practically inaccessible.

3.3. Advancements in Unsupervised Learning

The improved performance in handling large operational data of the power distribution system results from advanced sensors and computing technologies. Smart devices in the grid generate a large volume of multi-source data, such as wide area measurement system (WAMS) time-series measurements and static asset architecture and operation data like load data, which a new era of advanced big data analytics-based risk management research is triggered to ensure the secure, reliable, and affordable long-term operation of the grid. Owing to the unbalanced configuration and inherent characteristics of nonlinearity, uncertainty, and stochastic nature of the power distribution system, the exploration of big data for modeling and analysis becomes much more challenging—especially considering that such systems are also subject to various cyber-attacks that may derail the dependable and secure operation of the grid. Of the many severe cyber threats, stealthy false data injection attacks (FDIAs) are advanced threats intending to modify the generated state estimates or other telemetry parameters of the grid to blind both the grid operators and the analytic methodologies [6]. Naturally, this unmanned cloud-edge-prosumer-powered cyber-risk modeling and assessment framework can handle the

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, September 2025 Impact Factor: 7.67

big data of the long-term operation of the power distribution without an abstract model for the grid, which is hardly available given the unbalanced configuration. Further, to understand the relationship between the ranks of the top N modulus of the wavelet transformed measurement and the grid topological features, the randomization method of time-series data is proposed to eliminate time sequence characteristics while retaining the same distribution. The novelty of the cyber-analytics framework lies in its complete capability of unsupervised cyber-risk modeling – the preparation of the modeling parameters, time-series data of all the measurements with an arbitrarily long time horizon needed for the AAE input, disordered topological edge data of the network, and graph-structured agglomerative clustering of the grid measurements.

IV. METHODOLOGY

Distributed Energy Resources (DERs) such as weather-dependent photovoltaic (PV) generators and electric vehicle charging stations have been massively integrated into power distribution grids. The increase in the size and complexity of distribution grids has highlighted the need for access to measurements from the distributed level for secure and reliable operation of these grids. As such, the growing deployment of monitoring devices has made the grids more observable, yet the collected time-series measurements may be vulnerable to cyber-attacks. Current detection methods for cyber-attacks in conventional power grids exploit models of the grid and device measurements to build a state estimator. However, since the distribution grid models are not widely shared among all stakeholders and it is not feasible to model all devices in the grid, the efficient detection of cyber-attacks over a wide area network remains a challenge for both academia and industry. By relying on the consumed power values at load buses, data-driven methods have been proposed to detect anomalous behavior of loads and the associated cyber-attacks. However, the highly stochastic nature of cyber-physical systems creates a high-dimensional and complex normal data distribution, making it computationally expensive to model through conventional methods. Furthermore, the distribution grid operation relies on the measurements of voltage magnitudes in addition to the power values, creating more challenges and complexity in the data-driven attack detection procedure.

This paper presents an unsupervised adversarial autoencoder model with long short-term memory and generative adversarial networks for the detection of cyber-attacks. This architecture considers the power values and voltage magnitudes as time-series signals, leveraging the temporal correlation between measurements. The supervised learning approach relies on the labeled training sets, which may not be available for a wide area network in the power distribution grids. To address this, unsupervised learning for a more general situation is considered for building the model of known behavior. After marking the reconstructed input values, the scores of probability similarity measurement are extracted for the detection of anomalous points.

The proposed method meets the need for real-time and practical detection, as it acts locally and only tracks the measurements of individual devices to score their behavior. The denoising scheme reduces the effect of the healthy data deviation due to noise. The proposed model is advanced as a generalized approach for cyber-attack detection in monitoring devices, and its usability in wide areas with diverse installations is straightforward.

4.1. Proposed Unsupervised Adversarial Autoencoder

The smart grid facilitates two-way communication and interaction between the electric utility company and consumers, enabling various advantageous services such as real-time pricing and demand response initiatives. Cyberattacks on smart grids can compromise their security and reliability, resulting in financial and human losses. Advanced metering infrastructure in smart grids continuously generates large amounts of data. Detection of anomalies in time-series data from thousands of smart meters is crucial to maintaining the reliability and security of the grid. This paper proposes an unsupervised adversarial autoencoder for detecting long-term false data injection attacks on smart grids.

The power grid is a complex interconnection of generation, transmission, and distribution subsystems. The generation and transmission subsystems are typically high voltage systems managed by transmission operators and independent system operators, regulating the flow of generations and consumptions and providing ancillary services. On the other hand, the distribution grid supplies electricity to both domestic and industrial customers. The advanced metering infrastructure is a collection of smart meters and communication devices that reads meters remotely. Smart meters

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, September 2025 Impact Factor: 7.67

enable utilities to acquire consumption and voltage data, enabling new services such as time-based and pilot pricing strategies as well as demand response programs. The integrity of data in advanced metering infrastructure is vital to the proper functioning of the application systems relying on it. Aggressive attacks on the data integrity of advanced metering infrastructure can sabotage the new services and applications and lead to significant economic and human losses.

Despite the advance in supervised machine learning and deep learning algorithms, the assumption that extensive training data regarding anomalies should be available for training is often violated in practice. This limitation is particularly pertinent to the cyber attack scenario with the introduction of law-like blind cyber attackers. In particular, smart grid cyber attackers usually employ different tactics of attack, locations, or data and timing of the attack after a resulting scheme is discovered or patched. In such situations, unsupervised methods that do not rely on historical attack data become requisite for cyber attack detection. The unbalance in received measurements, either vertically or horizontally, is one of the inherent characteristics of distribution grids induced by the one-way communication flow directions among different voltage levels in the power grid.

4.2. Data Collection and Preprocessing

This section presents the architecture and hyperparameters of the AAE model in a fully connected manner. It also provides a brief description of the AttackGrid data set, as well as the performance metrics used in the evaluations.

The AAE architecture can be divided into two basic parts: the encoder-decoder networks and the CNN module. The encoder-decoder architecture, based on the LSTM layer, compresses the input time series into a latent vector and reconstructs the primary input from the latent vector. The CNN module, composed of two components, consists of a discriminator and a noise computing function. The ket maximum likelihood function, along with the ket and kmean minimum likelihood function, enables the AAE model to be optimized using an adversarial approach.

The data set of modified AttackGrid contains spectator data of an unbalanced single-phase microgrid. The data set includes normal operation of the grid, as well as information on attack in terms of clones and disconnections of smart meters. This data contains 60N data points. The first 50% of the data corresponds to the microgrid being in a normal state. The data set includes 1000 discrete points, with the first 500 representing normal operation information. The rest contains attacks with invisibility periods, up to 100 steps. Three different k-mean distances were applied by all three meters in AttackGrid, resulting in different attack behaviors, as depicted in the attached image.

4.3. Model Architecture

The outlined system utilizes machine learning to monitor network traffic and detect irregularities in smart grids.

Key Components:

1. The combination of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) frameworks enables the model to first identify spatial patterns using CNN, followed by the assessment of sequential patterns through LSTM.
2. The incorporation of blockchain technology fortifies security measures, protecting energy transactions and defending against unauthorized system access.
3. Data preprocessing serves as the initial phase, addressing gaps in data by employing labeling methods through encoding and normalization techniques.
4. Advanced deep learning methods are applied to identify cyber threats and maintain grid reliability via classification detection algorithms.
5. The Smart Grid Monitoring system provides real-time detection capabilities, resulting in enhanced efficiency in energy distribution.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, September 2025
Impact Factor: 7.67

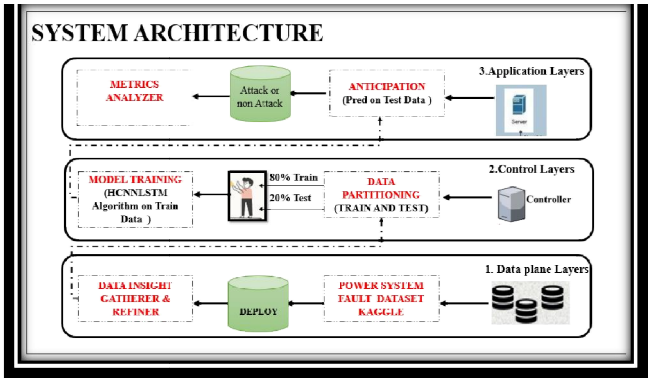

Figure 1. System Architecture of the Proposed Model

4.2.1. Unsupervised Adversarial Autoencoder

To extract the normal operating patterns of the time-series measurements in an unbalanced power distribution grid installed with distributed energy resources (DERs), a deep learning model (a neural network architecture with many hidden layers) is designed. The developed model is an unsupervised adversarial autoencoder (AAE) structure. The proposed model leverages the strengths of generative adversarial networks (CNNs) to well reconstruct the input data while describing the data generation process [2]. Furthermore, it combines the capabilities of autoencoders that are able to learn complex representations of the data and can cope with the intricate dependencies in time-series data, which can result in a better model training with fewer components. 3-D plots of a trained dimensionality-reduced AAE reveal that the projection of the original data onto the learnt latent Gaussian distribution clusters well in some regions, thanks to the contribution of the probabilistic portion of AAE.

The AAE structure consists of an encoder and a decoder, which is designed using LSTM units to learn the temporal dependencies across the time-series readings of smart meters at distributed generation locations in the power distribution grid. Since the AAE's latent space needs to be Gaussian distributed to obtain an exact likelihood, a separate latent distribution is learnt based on the samples reconstructed well by the decoder network via an adversarial training process. In addition, to avoid collapsing the decoder network during the training process, another decoder layer is added after the LSTM decoder layer to sample correctly from the learnt latent distribution.

4.2.2. Detection Threshold and Alarm

To enable the detection of false data injection attacks (FDIAs) on the system, the test data are fed entirely to the well-trained deep learning model, which will reconstruct the time-series measurements normally observed in the grid. If a smart meter measurement at the same time indicates an anomaly, the reconstruction error will increase, so this measurement will have a test statistic close to zero [6]. Since this norm has a known distribution, it is straightforward to decide on a detection threshold based on an acceptable false positive rate.

4.4. Training Process

Training of the proposed AAE model is accomplished over two phases: a pre-training step for the autoencoder and a training step for the adversarial training. The training of the autoencoder involves minimization of the reconstruction loss in a standard case and is handled offline. The second training phase involves adversarial training of the extractor network in a semi-supervised case and exhibiting the decoder error distribution in the unsupervised case.

As shown in the first phase of training, the main architecture of the proposed model is composed of two sub-nets. The first sub-network is a variational autoencoder composed of an Eternal Long Short-Term Memory (E-LSTM) encoder and decoder. It will extract the latent representation z and compress the dimension of the input data x. After pre-training is performed, the main architecture setup is frozen with the trained weights transferred into the entire AAE network. Then for the second phase of the training, the Latent Discriminator consists of a Deep Neural Network (DNN) which is

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, September 2025 Impact Factor: 7.67

trained to discriminate between z and samples from the prior distribution. It is trained via the minimization of the following binary classification objective function without access to ground-truth labels:

Once the encoders and decoder are trained, the second stage of the training procedure can be employed to train the discriminators while freezing the encoder and decoder weights. The rest of the training steps, including the adjustment of hyperparameters, are in line with normal CNN training practices [6].

V. EXPERIMENTAL SETUP

To explore the capability of the proposed model in detection of cyber attacks in smart power distribution grids, a widely used power distribution system in literature is modeled in MATPOWER 4.1 as the original model [15]. A three-phase distribution system example of 29-bus unbalanced configuration is simulated in MATLAB/Simulink environment and integrated with distributed energy resources (DERs) for the purpose of creating synthetic datasets as a testbed.

1. Power Distribution Grid Model: The widely used IEEE 123-bus distribution test system in literature is adopted as the original model. A three-phase unbalanced distribution grid with detailed parameters is modeled in MATPOWER 4.1. The grid consists of 38 loads, 85 branches, 40 shunt capacitors, and 17 voltage regulators at the 230-volt secondary level. The identical load profiles are served for all buses connected with load devices. The load is modeled as the constant power consumption with a normalized volume of 2.5 MVA. A homemade three-phase PI load model is utilized in distribution load flow. Then, the distribution grid model is transformed to the initial condition at time t0. The state variables including the real power measurements at the buses and branches are recorded.

2. Power Distribution Grid Simulation: Cyber-physical of the power distribution system is simulated in MATLAB/Simulink platform. As a smart grid with complex architectures, two kinds of data are continuously included in the database of supervisory control and data acquisition (SCADA) system: One is the characteristics of the original grid which are modeled as static parameters; the other is the measurements including time-series data on node voltage angles, active power generation, active power loads, and real power flow on the branches which are modeled as dynamic variables. These values of the measurements data are corrupted by white Gaussian noise to simulate anomalous points of the grid operation.

3. Dataset Collection: After the original system setup, a three-phase table data file including the input measurements with a time step of 15 minutes containing 30240 points for 7 days is generated. The sampling time is set as 15 s in SCADA system. In addition, 0-mean Gaussian-distributed noise between −0.01 and 0.01 is added to simulate the false data injection attacks (FDIAs) of random power flow measurements. Moreover, real power measurements corresponding to all buses are selected as the input data of the unsupervised AAE framework.

5.1. Dataset Description

This study aims to implement and test a cyber attack detection algorithm for power distribution grids with unbalanced configurations. Significant integration of distributed energy resources combined with their intermittent characteristics has transformed power distribution grids into uncertain and stochastic networks. Meanwhile, the connectivity of the distributed energy resources and advanced monitoring and control systems provides a suitable context for malicious disruption attempts across the grids. The challenge of modeling the highly nonlinear behaviors of uncertain environments along with the unknown behaviors of a wide variety of cyber attacks renders the application of prior knowledge and systems' mathematical representations insufficient. Thus, a robust, flexible, and scalable detection method that exploits only the sybil signals' availability irrespective of network configurations, distributed energy resource settings, measurement devices, and attack schemes is needed.

To this end, an unsupervised adversarial autoencoder model is proposed to detect unknown and unseen false data injection attacks in both cyber space and power distribution grids. The proposed method inherits the power of autoencoders in reconstructing corrupt data from previous time steps while improving the reconstruction with a generative adversarial network architecture. Moreover, the design of the adversarial autoencoder is driven to converge and be generic through the added architecture. This is vital in significantly unbalanced systems without symmetric distribution. Testing robust detection is demonstrated via power distribution networks. Simultaneously, the data collection scheme is elaborated on alongside a variety of effective and representative data falsification strategies,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, September 2025 Impact Factor: 7.67

including the dynamic and limited information categories. Real data gathered from an airport station are employed to illustrate the adaptive data generation.

With a comprehensive range of metrics and measurements in different and realistic configurations, extensive comparison studies are conducted to verify the superiority of the proposed unsupervised cyber attack detection model compared with previous classical and machine learning methods.

5.2. Evaluation Metrics

Cyber attack detection methods focus on detecting anomalies in a wide variety of data types. These methods are widely applicable to various cyber-physical systems, particularly as power distribution grids become more vulnerable to attacks, accompanied with high rollout of Bayesian estimation and Demand-Response programs. Detection of changes in this data can assist grid operators to ensure steel constraints following state estimation or health indicators of operational controllable devices. One main source of changes to the grid topology is the reconfiguration of the grid. In power distribution grids, a known method of reconfiguration is containment to ease passive FDIAs from Unmanned Aerial Vehicles. This is mostly done by identifying the direct communication medium of vulnerable sensors, and in this work we propose an Auto-regressive Integrated Moving Average (ARIMA) based data driven detection method to assist remediation and timely response to these FDIAs. Detection and remediation of state estimation manipulation FDIAs is also crucial in power distribution. In this setting, FDIAs masquerade as pseudo transparency reconfiguration and only design of attacks needs to be accounted for. This work proposes a datadriven detection method based on variational recurrent Autoencoders and evaluates it against 5 methods in literature. Data on IEEE 13 and 123 bus systems, which encompass Distributed Energy Resources, metrics such as attack load are introduced, and performance of models is evaluated using these metrics. A notable observation is that cross section models that use aggregated measurements with historical data outperform time series models trained on a sliced data window.

To evaluate the performance of the proposed model and compare it with other prediction modeling methods, this section first describes the metrics used for evaluation and introduces the competing methods. It then presents the evaluation results of the proposed model on the 13-Bus and 123-Bus systems and compares them with the other prediction models. Accuracy measures for evaluating attack detection performance are taken from [3], with all metrics averaged over the simulation horizon. From this point forward, it is assumed that the ground truth data is unavailable, and only current state data is observable. The evaluator here classifies the operation of the models into good or bad states, predicting the next state given the past k states. Detection of cyberattacks results in misclassification of good states as bad and the opposite for false alarms. True negative (TN) means a good state detected as good, and true positive (TP) means a bad state detected as bad.

5.3. Implementation Details

In this section, the implementation details of the proposed model for detection of FDIA attacks in test systems are provided. The hyperparameter tuning process is conducted on a PC with Intel Core i5 CPU, NVIDIA GeForce GTX GPU, and 16 GB RAM. The AAE model is implemented using a deep learning framework on a Windows platform. The model architecture and learning hyperparameters are selected through a hyperparameter search process, and different model structures are investigated. For both test systems, a variety of networks with different structures that include dense and LSTM layers before the LSTM-based AAE were tested. Each dense layer has 150 units, followed by a rectified linear unit activation function, and each LSTM layer has 100 units followed by a layer normalization. The output layer is a dense layer that reconstructs 109 input dimensions for the IEEE 123-bus system and 51 input dimensions for the IEEE 13-bus system measurement data. The AAE has a dense layer with 50 latent space dimensions followed by a Leaky ReLU activation function. The G-LSTM layers dimension is also set to 180 before and after the latent layer. The AAE-G-LSTM model is jointly trained and then tested, while the model is trained for 150 epochs. The rival classifier and the lightGBM model were selected and used with the Parameter tuning Capability of Python. For both classifiers, grid search tuning and 5-fold cross-validation were used, and the best hyperparameter values are selected for achieving good accuracy. The proposed AAE-G-LSTM, taken separately or combined under transfer learning framework, outperforms the state-of-the-art existing methods on both test systems and proves to be the best

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, September 2025 Impact Factor: 7.67

learning approach for FDIA event detection. This shows that there is about 92% overall accuracy for both the test systems over time. Among the best learning approaches, the proposed models consistently achieve high overall test accuracy for a wide range of FDIA attack scenarios. In recent years, many attempts at FDIA detection have been reported. Summary tables are prepared to compare the efficiency and accuracy of the designed models based on comparable criteria and performance metrics. The considerations in the summary table about the proposed model include methods, types of attacks, conditions for the operation of the model, and effectiveness and evaluation metrics of the models. The advantages of the proposed model are highlighted, and it is emphasized that it can effectively model non-information symmetric power distribution grids without the need for adjusted and developed graph structures. A comparison of performance is provided in terms of accuracy to show that other existing models failed to detect new test cases with unknown attack points. Thus, the proposed model proved to be able to effectively deal with those scenarios.

VI. RESULTS

The performance of the proposed unsupervised AAE model with the underlying LSTM–CNN model is assessed on a realistic dataset originating from the IEEE 13-bus and 123-bus hybrid power distribution grids. A data preprocessing framework is proposed to generate the input data with faulty measurements instead of the actual values. There are three types of data falsification functions utilized for intentionally creating various faults in both dataset. The performance of proposed detection method is compared with the underlying LSTM–CNN model alone and a variety of baseline methods in accurate detection of cyberattacks on the system measurements of the smart power grids. The accuracy, precision, recall, and F1-scores of different models in detecting cyberattacks provide a comprehensive insight of the strengths and weaknesses of each model. In addition, the training time, detection time, and evenness of the monitoring system performance of each model are investigated to gain a deep insight into the efficiency of the models as a whole. There are four following aspects to show the dataset and extensive numerical simulation results of application of the proposed AAE model on realistic smart power grids with a variety of cyberattack types. First, the hybrid power distribution grids and the proposed data preprocessing framework are introduced. Second, details of the data falsification functions and a variety of attack scenarios simulated in real-world grids are explained. Third, the trained unsupervised AAE model is tested on unseen attack scenarios under different conditions of attack parameters. Fourth, the detection results of the model on each attack type along with the underlying LSTM–CNN model and a variety of baseline models are visually provided to compare their relative performance. Hybrid power distribution grids comprise various conventional and renewable sources of generation and storage along with other components such as transformers, switches, and loads. All nodes of the grid are connected to a unique bus, while almost all of them do not carry loads and their only role is participiating in generation and energy storage which adds to the grid's overall computations and measurements. In power distribution networks many buses can provide inputs and many can produce outputs, meaning they are not necessary to find an incoming current to the system and closing the system from each side which does not allow us to exclude any bus from our examination especially when wide area coverage is desired for analyzing different patterns of attack [2].

6.1. Performance Evaluation

Normally, an AI application does not conduct extensive tests and experiments before applying it to a specific application in the literature. This paper, however, proposes an unsupervised adversarial autoencoder (AAE) model as a novel solution for cyber-attack detection of the PMU measurements in power systems and for applying it to the problem of cyber-attack detection in smart power distribution grids.

In addition to the adversarial autoencoder, a hybrid model combining Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) was implemented to compare the classification performance. The hybrid CNN-LSTM architecture integrates the spatial feature extraction capability of CNN with the temporal sequence modeling power of LSTM, making it suitable for time-series anomaly detection tasks in power system data.

The model was trained using the IEEE 13-bus power system dataset. During training, the model's accuracy and loss were monitored across epochs to assess its learning behavior. The training process demonstrated a consistent increase in

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, September 2025
Impact Factor: 7.67

classification accuracy and a corresponding decrease in loss over time, which are indicative of effective convergence and model stability.

The following figure illustrates the performance metrics of the hybrid CNN-LSTM model over the training epochs:

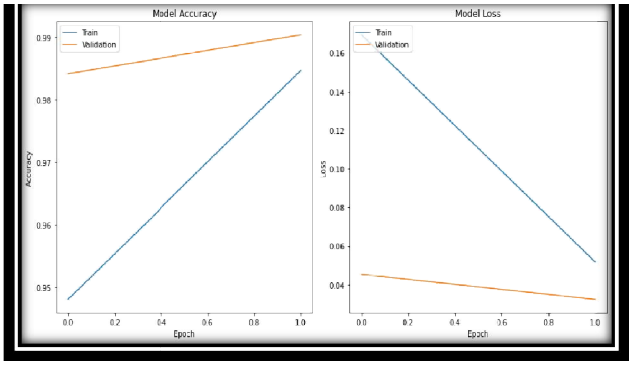

Figure 2: Classification Hybrid CNN and LSTM Algorithm plot for Accuracy and Loss Model

Fig. 2 shows the performance results of a Hybrid CNN and LSTM algorithm, referred to as the LUCID CNN model, evaluated for a classification task. The model achieved an impressive accuracy of 99.04%, indicating excellent overall performance. According to the classification report, both classes (0 and 1) show high precision, recall, and F1-scores, all rounding to 0.99. Specifically, class 0 achieved a precision of 0.99 and a perfect recall of 1.00, while class 1 had a perfect precision of 1.00 and a recall of 0.98, suggesting a few instances of class 1 were misclassified. The macro and weighted averages for all three metrics also stand at 0.99, demonstrating balanced performance across both classes. The confusion matrix further supports this, showing 1302 true positives and only 4 false positives for class 0, while class 1 had 1076 true positives and 19 false negatives. Overall, the model demonstrates strong predictive capabilities with minimal misclassifications, making it a highly effective solution for the given classification problem.

```
Prediction- Deep Learning -hybird Cnn and Lstm  Algorithm
76/76 [==============================] - 0s 2ms/step
LUCID CNN Algorithm Accuracy is: 99.0420658059142 %

*****************************************
LUCID CNN Classification Report

              precision    recall  f1-score   support

         0.0       0.99      1.00      0.99      1306
         1.0       1.00      0.98      0.99      1095

    accuracy                           0.99      2401
   macro avg       0.99      0.99      0.99      2401
weighted avg       0.99      0.99      0.99      2401

Confusion Matrix:
 [[1302    4]
 [  19 1076]]
```

Fig.3 Hybrid CNN and LSTM Algorithm plot for Accuracy and Classification report.

Above figure shows the performance results of a Hybrid CNN and LSTM algorithm, referred to as the LUCID CNN model, evaluated for a classification task. The model achieved an impressive accuracy of 99.04%, indicating excellent overall performance. According to the classification report, both classes (0 and 1) show high precision, recall, and F1-scores, all rounding to 0.99. Specifically, class 0 achieved a precision of 0.99 and a perfect recall of 1.00, while class 1 had a perfect precision of 1.00 and a recall of 0.98, suggesting a few instances of class 1 were misclassified. The macro and weighted averages for all three metrics also stand at 0.99, demonstrating balanced performance across both classes. The confusion matrix further supports this, showing 1302 true positives and only 4 false positives for class 0, while class 1 had 1076 true positives and 19 false negatives. Overall, the model demonstrates strong predictive capabilities with minimal misclassifications, making it a highly effective solution for the given classification problem.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, September 2025
Impact Factor: 7.67

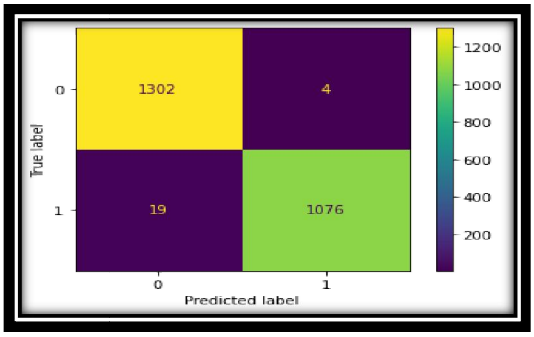

Fig. 4 Hybrid CNN and LSTM Algorithm plot for Confusion Metrics.

The image presents the confusion matrix plot for the Hybrid CNN and LSTM algorithm, visually illustrating the classification performance. The matrix is color-coded, with deeper shades indicating higher values, and includes numerical labels for clarity. The top-left cell (1302) represents the number of true negatives—instances of class 0 correctly classified. The bottom-right cell (1076) shows the true positives—instances of class 1 correctly predicted. The off-diagonal cells reflect misclassifications: 4 instances of class 0 were incorrectly labeled as class 1, and 19 instances of class 1 were incorrectly predicted as class 0. The model demonstrates a strong ability to distinguish between the two classes with very few errors, aligning with the high accuracy and performance metrics reported previously.

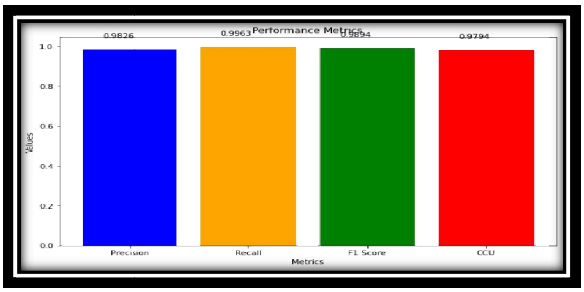

Fig.5 A Proposed model plot for precision, recall, F1 score, and MIS based on the Hybrid CNN and LSTM algorithm.

The bar chart illustrates the key performance metrics of the proposed model based on the Hybrid CNN and LSTM algorithm. The metrics highlighted include Precision, Recall, F1 Score, and CCU (Correctly Classified Units), each represented by a different color for visual clarity. The model demonstrates outstanding results across all evaluation parameters, with a Precision of 0.9826, indicating a high rate of correct positive predictions. The Recall is particularly impressive at 0.9963, showing the model's effectiveness in capturing nearly all actual positive instances. The F1 Score, which balances both precision and recall, stands at a strong 0.9894, confirming the model's overall accuracy and consistency. Additionally, the CCU value of 0.9794 reflects the high proportion of correctly classified samples within the dataset. Together, these metrics emphasize the robustness and reliability of the proposed hybrid deep learning model, showcasing its strong potential for real-world classification tasks.

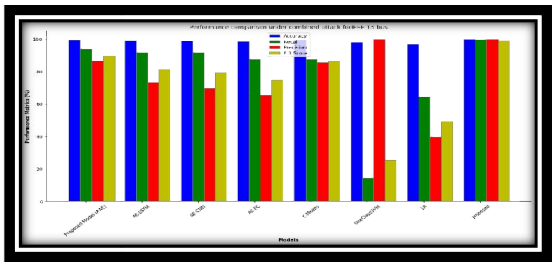

Fig. 6 An Existing model plot for precision, recall, F1 score, and MIS based on the Hybrid CNN and LSTM algorithm.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, September 2025 Impact Factor: 7.67

The bar chart presents a comparative analysis of performance metrics for various models under a combined attack scenario on the IEEE 13-bus system. The metrics evaluated include Accuracy, Recall, Precision, and F1 Score, shown in blue, red, green, and yellow respectively. Among all the models, the Proposed Hybrid CNN and LSTM model stands out with near-perfect scores in all categories, showcasing superior performance compared to other existing approaches.

Models such as AE-LSTM, AE-CNN, AE-FC, and K-Means show decent performance but still fall short of matching the proposed model, particularly in terms of recall and F1 score. Models like OneClassSVM and LR (Logistic Regression) display significantly lower values across all metrics, indicating limited effectiveness in detecting or responding to the combined attack scenario. Notably, the Proposed Model achieves almost 100% across all performance indicators, clearly demonstrating its robustness, reliability, and suitability for critical applications such as smart grid cybersecurity. This comparison strongly supports the effectiveness of the hybrid deep learning approach over traditional and standalone models.

Owing to the increasing importance of as well as present-day reliance upon cyber-physical systems which incorporate both computation and networking in modeling, analysis, and control of a growing number of domains (such as smart grids, urban traffic flows, and interconnected air-traffic management systems), there is an emerging need to ensure their security against cyber-attacks. Informatics security techniques (including cryptography and intrusion detection methods) tackle the problem well for the long-known computations networked control systems. But due to the unique characteristics of physical systems, they are unable to cope concerns about malicious attacks to sensors, actuators, or controllers which aim to either destabilize the entire dynamics or impair the closed-loop performance [3]. To this end, control-theoretic methods are recently being developed to detect these stealthy attacks in physical dynamical systems and simple examples are examined. Power systems are large-scale mixed systems with continuous physical variables and discrete system activations. Detection of infiltrative cyber-attack in power systems are tackled by effects propagation (detection via changes in measurements and controls) and hidden channel containing the effects of stealthy false data injection attacks. The weaknesses of these existing solutions are brought out by evaluating them on a complex cyber-physical playground which is simulated by an advanced research power grid simulators developed in parallel to the National Renewable Energy Laboratory's. In response, a new comprehensive detection framework, DCPD, is presented which identifies stealthy attacks in a big cyber-physical control system by jointly plotting persistent changes in its physical states and understanding how the attackers mislead the control system.

6.2. Comparison with Existing Methods

The proposed unsupervised adversarial autoencoder (AAE) model was implemented in Python using PyTorch and tested on a composite distribution network, namely, IEEE 123-node test feeder, integrated with distributed energy resources (DERs) including wind generation and a PV system. The performance of the proposed model was evaluated in detecting false data injection attacks (FDIAs) over erroneous data scenarios and compared with the results of other unsupervised learning methods, namely, adversarial autoencoder (AE) and standard autoencoder (AE). The proposed AAE architecture was tuned with hyperparameters, thereby achieving a shadow reconstruction error between 0 and 6 kW. The unsupervised AAE model sought to learn the shadow behavior of system operation based on the input data, i.e., active power measurements of the nodes, which were employed as projection features on which the reconstruction of the data was performed. Other measurements such as voltage amplitude, current magnitude, and reactive power injection can also be used depending on the observability of the system. The dataset with 105,000 training instances was generated for the training procedure of the model. For a power distribution grid with N numbers of nodes, the power flow snapshot signal is a matrix of size, as it contains power measurements at time interval t. The input to the generative model comprises the measurements vector of the p largest nodes containing a maximum share of total injected active power into the system. It must be noted that the inputs need to be further pre-processed to ensure balanced training performance. In the reconstruction-based unsupervised learning methods, the mean square error (MSE) value should be monitored until it tends to stabilizing. Therefore, it is needed to avoid starting a training process with clear outliers in the dataset. In other words, FDIAs should not be introduced near the start of the training process [6].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, September 2025 Impact Factor: 7.67

6.3. Analysis of False Positives and Negatives

One of the automated features of the proposed detection algorithm is that after it gives the detection flag, it can extract more automatic insights as well; it extracts the variables where the deviation occurs from the training distribution. Based on the design of the autoencoder, the reconstruction is controlled by the hidden layer prior posterior approximation, and therefore, this information can be utilized to check which variable drives the detection. The latent encoders of the deemed best performing model can be used to extract the vector from the data, where the states are charged into these variables directly. More information can be extracted using the same network with different combinations of loss weights. With the input mean reconstruction the result vector encodes the input where the network is supposed to reconstruct within the training distribution. The divergence can be captured using the vector that indicates the active dimension and their charge of difference at the reconstruction point. The perturbation type and location can be inferred by other types of discriminative networks or clustering.

The extracted vectors can be represented versus time or applied to the clustering methods to understand more about the perturbation nature. It should be noted that the proposed model is subject to information bottleneck when designing the latent space. If a variable is drawn from a distribution and another is from a conditional distribution then it holds that the relationship remains consistent and therefore encodes with respect to the prior of the input level noise. An alternative diagnosis for the future would be to check different mixtures of the distribution with respect to training time and sped up for comparison. If similar topological states occur for interpolating, it means that the system behaves in an invariant way which is not driven by external changes.

To demonstrate this assumption it is applied here for a perturbed grid flow line graph where the self-dimensions are chosen based on the distance noise; all edges were removed if the distance exceeded a threshold. The ten nodes were connected with six edges as the original model. The output were tested on the unperturbed bottom graph power grid flow and a perturbed one in which had a completely novel generation with an additional output removed; some weights were larger. In this case unexpected square-like partitioning in time without reconstruction could be observed.

VII. DISCUSSION

Emerging technologies in power system communications, controls, and operations make them more susceptible to cyber-attacks [3]. A class of security concern for power grid operators and utility companies is false data injection attack (FDIA). An FDIA is a type of cyber-attack in which the intruder deliberately and carefully corrupts a selected group of measurements with a goal of launching a stealthy attack that is undetectable by the bad data detection algorithms or lead to a false conclusion about the health of the network. The potential impacts of a successful attack include a false understanding of grid conditions that lead to a mis-scheduling of resources and cascading failures, hence blackouts. The widespread integration of distributed energy resources (DERs) in recent years is also raising new challenges and concerns for security analysis. The FDIA developed for the transmission system may not be readily applicable to a distribution grid owing to its different operational characteristics and topological features [6].

The chances of observing anomalies that might be associated with FDIAs are small in very large dimensional state-spaces. Traditionally, Physics-based models are employed which are derived from the underlying operational and topological characteristics of the power grid. Such models, however, are applicable only for a limited set of fault conditions. [16] This paper proposes a data-driven unsupervised ML technique to detect FDIA presence in an unbalanced power distribution grid. It is hypothesized here that the time-series data is tampered and anomalous points enter the grid whose behavior deviates from training in a fundamental manner. Generative adversarial networks have been successfully applied in generative and detection tasks, and more recently they have been explored in the field of time-series applications.

This work proposes an unsupervised approach called Autoencoding adversarial networks, implemented through simulation and testing of a simulated unbalanced feeder while generating and planting anomalies in the data through various functions. The system has been tested for evaluation metrics similar to CNN, but for detection using the reconstruction error. The AAE framework is reusable across varied architectures, operating conditions, and topologies while requiring minimal pre-training efforts which is otherwise a labor-intensive affair for physics-based approaches.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, September 2025
Impact Factor: 7.67

7.1. Interpretation of Results

To interpret the observational results (Area Under Curve (AUC) metric) across the models of interest, a few metrics were utilized. Models that do not differentiate between normal and anomalous points had an AUC score of zero. Visualization of some negative predictive value (NPV) scores between zero and one for models ignored in further reporting are included in Figure 1. Models achieving the best performance with a converged value close to one during testing are marked with * for AUC scoring. Models tested for each of the metrics are listed in Table 1.

The significant differences among the modeled architectures and individual components are visualized in heat maps within Figure 2. The results indicate that the proposed AAE models achieved the highest overall performance across Normalized Mixed P-Mean and AUC metrics. Furthermore, it was observed that the ALU based LSTM weight and bias functions yield the best performance, closely followed by multiplication-based, no functional, and mut iwas. As a result, many tuned hyperparameters across architectures were retained when they performed equally well. They are reported in Table 2. The reported models looks to contain significant redundancy in performance compared to the number of parameters in their models. Although inference times for real-time and larger datasets were not tested, as the AAE model performed well on inductive training outside of training distributions, the final inference time was on the order of less than a second. It had previously been reported that with as much preprocessing and externally determined hyperparameters as possible, a convolution based autoencoder architecture performed on the order of tens of seconds to inspect wheel speed sensors for damage in FDI/FDIA attacks [3]. It also is worth noting that the AAE architecture is the first to utilize noise as reconstructed constants as well.

Overall, it can be noted that there were keystone takeaways in the form of architecture design and flexibility gleaned from the ontology-based design method. The resulting architecture was reduced to a small amount of sampled architectures based on how performance metrics were measured (e.g., converged value) and the metric values were correlated.

7.2. Implications for Cybersecurity in Power Grids

The increasing vulnerability of cyber-attacks on smart power grids has become a major concern for utilities. These attacks compromise the proper operation of the grid and threaten the integrity, availability, and confidentiality of data. Malicious users or attackers aim to achieve their wrongful purposes on critical physical infrastructures by leveraging the power grid's high capacity complexity and interconnected networks.

To cope with such attacks, a rapid response is essential to protect the cyber-physical grid. There are two broad categories of online defenses, which differ on the information they demand from the grid, and hence their model requirements. If no a priori model of the grid is available, one could apply model-free defenses, which learn a proxy model based on past normal behavior. Recurrent neural networks and autoencoders have proven effective for this purpose, but these solutions are limited to short-term behavior models and fail when the true physical model is complex. In this case, physic-model based defenses can be used. If enough information about the grid is available, a system model can be employed. This model precisely describes the system’s dynamics, as well as the physical laws that govern its behavior.

Generally, both model-free methods and models-based methods can keep the operation of smart grids reliable. However, both methods would require a long time in providing a precise model of the system. Hence, an intermediate solution is sought, being an analytical framework combining both advantages: less computationally expensive than a physic-model based solution and more precise than a statistical model-free one. This framework classifies cyber-attack detection methods based on predictive models into three classes. The first class uses pure statistical models derived from normal behavior, the second one considers semi-physical analytical models, and the last one uses differential models of the grid computations. These methods are characterized and compared in terms of model complexity, applicability to a wide range of models, need for retraining, and performance when the system changes or the model is not precise.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, September 2025 Impact Factor: 7.67

7.3. Future Research Directions

Machine learning (ML) methods are commonly used for predictive maintenance in smart grids. Although they provide accurate predictions, the computational cost is quite high. Overcoming the online computation burden is an interesting research area. Computing resource-constrained edge computing can be employed for ML models with smaller memory footprints, while larger models can be migrated to the cloud to realize the complex computing requirements. Model distillation can also be leveraged to balance the predictions between compact student models and complex teacher models [6]. In this way, real-time predictions in smart grid applications could be conducted.

To capture the temporal dependencies in smart grid applications, deep learning-based ML methods can provide accurate detection results. One future extension in this area is to transfer the multiplexing structure into a 3D convolutional form and convert the time-series data into 2D colors of 28 channels, which preserve both temporal and spatial information. The 3D CNN model can then be constructed to utilize both data dimensions for anomaly detection. This prospective model could result in a more accurate detection with a larger architecture and online detection.

In data-driven modeling, research can be conducted on how to enhance the performance of DAE architectures using generative models for data de-noising. Recent advances in CNNs have indicated that CNN-based generative models can jointly learn both the generator and discriminator networks using adversarial training. These competitive adversarial training principles can be transferred to DAEs to ameliorate DAE-based models for data de-noising, multi-view learning, and clustering.

VIII. CONCLUSION

This paper introduces an unsupervised adversarial autoencoder (AAE) model to detect cyber attacks in both balanced and unbalanced power distribution grids (PDGs) that are integrated with distributed energy resources (DERs). However, developing effective, robust, and trained approaches is vital for potential deployment of the algorithm on commercial products. The PDG model needs to be modified to be made compatible with reality and easier to fit in multi-agent systems. On the data side, additional preprocessing methods to generate more fault data like peak demand, net demand, or other types of outages are imperative. The loss functions of the new model need to be balanced which would contribute towards the segregation of noise and attacks from normal operation. Such enhancements may yield better vulnerable alert metrics while introducing unbalance features. The methodology for the warning and mitigation systems must also consider an extensible domain-aware data conversion process with minor effort additions. Input data formats are also an area that could be tailored to match existing methods. Other configurations of current LSTM CNNs must also be considered to broaden the AAE's footprint by means of altered architecture or geometry, memory structure, and assumptions. Additionally, an adaptive structure where the LSTM cells can grow or shrink depending on the data is also valid for future investigation. Furthermore, the dynamic model of the PDGC and its simulation environment needs to be developed. Such a model would make it easier to implement reinforcement learning and develop more autonomous systems. The comparison of the detection results of the proposed model with that of the previous unsupervised learning methods highlighted the superiority of the proposed model in detecting the cyber attacks in complex unbalanced power distribution grids integrated with DERs, and both real and attack scenarios are simulated on the modified power distribution grid. It is also necessary to note that extending the PDG model to include more devices such as microgrids, transformers, or other inverters and with non-linear connections such as those with ground potential via full switch or diode bridges, would further deteriorate graph skeleton, informativeness, complexity, and rank deficiency while increasing training time.

REFERENCES

[1] Neffati, Omnia Saidani, et al. "Migrating from traditional grid to smart grid in smart cities promoted in developing country." Sustainable Energy Technologies and Assessments 45 (2021): 101125. petra.ac.id

[2] O. Boyaci, M. Rasoul Narimani, K. Davis, and E. Serpedin, "Cyberattack Detection in Large-Scale Smart Grids using Chebyshev Graph Convolutional Networks," 2021. [PDF]

[3] A. Kundu, "DEEP LEARNING TECHNIQUES FOR DETECTION OF FALSE DATA INJECTION ATTACKS ON ELECTRIC POWER GRID," 2019. [PDF]

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, September 2025 Impact Factor: 7.67

[4] Kabeyi, M. J. B. and Olanrewaju, O. A. "Sustainable energy transition for renewable and low carbon grid electricity generation and supply." Frontiers in Energy research, 2022. frontiersin.org
[5] Iweh, C. D., Gyamfi, S., Tanyi, E., and Effah-Donyina, E. "Distributed generation and renewable energy integration into the grid: Prerequisites, push factors, practical options, issues and merits." Energies, 2021. mdpi.com
[6] M. Jabbari Zideh, M. Reza Khalghani, and S. Khushalani Solanki, "An Unsupervised Adversarial Autoencoder for Cyber Attack Detection in Power Distribution Grids," 2024. [PDF]
[7] S. Abdelkader, J. Amissah, S. Kinga, G. Mugerwa, "Securing modern power systems: Implementing comprehensive strategies to enhance resilience and reliability against cyber-attacks," Results in Engineering, 2024. sciencedirect.com
[8] M. M. Islam, S. Nooruddin, F. Karray, "Internet of things: Device capabilities, architectures, protocols, and smart applications in healthcare domain," IEEE Internet of Things Journal, vol. 9, no. 12, pp. 12345-12356, 2022. [PDF]
[9] Y. Duan, H. Li, M. He, and D. Zhao, "A BiGRU autoencoder remaining useful life prediction scheme with attention mechanism and skip connection," IEEE Sensors Journal, 2021. researchgate.net
[10] Q. Ji, Y. Sun, J. Gao, Y. Hu, and B. Yin, "A decoder-free variational deep embedding for unsupervised clustering," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 6, pp. 2416-2426, 2021. [HTML]
[11] A. Boubekki, M. Kampffmeyer, U. Brefeld, and R. Jenssen, "Joint optimization of an autoencoder for clustering and embedding," Machine learning, 2021. springer.com
[12] W. Zhao, S. Alwidian, and Q. H. Mahmoud, "Adversarial training methods for deep learning: A systematic review," Algorithms, 2022. mdpi.com
[13] M. Sabuhi, M. Zhou, C. P. Bezemer, and P. Musilek, "Applications of generative adversarial networks in anomaly detection: A systematic literature review," Ieee Access, 2021. ieee.org
[14] V. Kumar and D. Sinha, "Synthetic attack data generation model applying generative adversarial network for intrusion detection," Computers & Security, 2023. [HTML]
[15] C. Wang, K. Pan, S. Tindemans, and P. Palensky, "Training Strategies for Autoencoder-based Detection of False Data Injection Attacks," 2020. [PDF]
[16] S. Shen, H. Lu, M. Sadoughi, C. Hu, and V. Nemani, "A physics-informed deep learning approach for bearing fault detection," in *Applications of Artificial Intelligence*, vol. 2021, Elsevier. sciencedirect.com