**Journal of Computer Science and Technology Studies**
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

**| RESEARCH ARTICLE**

# Data-Driven Prioritization of Grid Interruptions: A Lean Six Sigma Study Using Public Outage Logs

**Rayhanul Islam Sony[1], Md Ariful Islam Bhuiyan[2] and Dipta Roy[3]**
[1]*Trine University, 1 University Ave, Angola, IN 46703, USA, rsony23@my.trine.edu*
[2]*California State University Northridge, 18111 Nordhoff St, Northridge, CA 91330, USA,* md-ariful-islam.bhuiyan.390@my.csun.edu
[3]California State University Northridge, 18111 Nordhoff St, Northridge, CA 91330, USA, *dipta.roy.229@my.csun.edu*
**Corresponding Author**: Rayhanul Islam Sony, **E-mail**: rsony23@my.trine.edu

## | ABSTRACT

The supply of power that is reliable and uninterrupted is one of the pillars of the contemporary energy infrastructure. Nonetheless, the issue of grid interruptions which are frequent and long-lasting is still compromising the resilience of the system and customer satisfaction. Traditional approaches to outage-management and prioritization tend to be reactive and based on manual evaluation and missing records of events. To overcome such shortcomings, this paper suggests a hybrid meta-learning model that would combine Random Forest (RF), Extreme Gradient Boosting (XGBoost), and a Deep Artificial Neural Network (Deep ANN) to come up with the data-driven prioritization of grid interruptions. The data were processed extensively with the Missing-value treatment, derivation of impact-score, feature scaling using robust scale and SMOTE and controlled Gaussian perturbation using publicly available data 15 Years of Power Outages (2000 - 2014) dataset. The proposed hybrid model integrates RF and XGBOOST in an ensemble layer while capturing the variance decomposition and residual optimization for variance decomposable nonlinearities and the Deep ANN meta-learner captures complex non-linear dependencies. Instead of maintenance, proactive and timely, the system prioritizes the outage events on the basis of the predicted severity of impact in the framework of Lean Six Sigma DMAIC. Comparative analysis with four baseline models which included RF, XGBoost, Deep ANN, and RF-XGB models showed that the hybrid model was better as it had 99.43% accuracy, 98.92% precision, 99.15% recall, and 99.03% F1-score. The obtained results support that the suggested meta-learning technique offers a significant enhancement in the robustness, scalability, and interpretability of prediction of outage impact. This study creates a solid information-driven approach to the implementation of the Six Sigma principles into the contemporary energy-analytics processes that opens the way to increased intelligence and agile power-grid operations.

## | KEYWORDS

Power outage analytics, Lean Six Sigma, Random Forest, XGBoost, Deep ANN, hybrid meta-learning, SMOTE augmentation, data-driven prioritization, grid reliability, energy resilience

## | ARTICLE INFORMATION

**ACCEPTED:** 05 February 2026 **PUBLISHED:** 16 March 2026 **DOI:** 10.32996/jcsts.2026.8.7.3

## 1. Introduction

Generator failures have been shown for their negative consequences, including serious damage to vital infrastructure and services to significant savings and discomfort to citizens throughout extended darkness. Thereby established a vital legislative and academic duty of carrying out an accurate and complete review of electrical repair methods. Contemporary nations are not entitled to give up secured, efficient electrical power, and we need to have a means of limitless sunlight as an essential need for continual, continuous connections and continuous healthcare. Efficiency in a business (Ankit, 2022), (Larsen, 2018). Through both the intensity and frequency of severe storms rising along with the rate of slow or prolonged failures, novel tools that evaluate and interpret lengthy failure logs have crucial. At addition to offering concise and precise data concerning effectiveness and malfunctioning causes, these tools ought to supply insight into the broader industry and social implications of those systems. This kind of bargains can help directors and contractors make better decisions, make evacuation mitigation easier, when decrease the long-run effects of incidents on communities and sectors (Brelsford, 2024) .Within the time, social benefits were making utilize of before worker discussions, audience posts, and basic sales indicators (Willems, 2024).Since this was at present possible to regularly obtain precise real-time information on gas utilization and power problems thanks by the total rising trend of advanced electric bills and portable grid command methods. As consequently, paused supply gauges are now essential instruments in providing a more honest and impartial appraisal of effectiveness efficiency. Through recognizing flaws as well as service disruptions that can be regularly skipped or undetected with ordinary measurements of performance tools, the above methods offer a greater awareness of the actually outstanding ser-vice people receive. Based on current research, investigators are capable of conducting methodical, data-based assessments concerning how quick responses and at-tempts to recover varies across different communities when broad and easily accessible power supply records are provided. They simplify an examination of how demographic or geographical characteristics impact the speed and standard of outage supervision, the examination of developments in unevenness in construction resilience, alongside the determination of imbalances in restoring services (Bhattacharyya, 2023), (Jiang, 2025).

Additional research on consumption outage pricing and harm offerings offer substantial fresh views on what it takes to methodically implement earnings effect measurements or sector- particular indicators toward utility business management structures. These investigations focus on calculating the wider economical and ser-vice-associated consequences, including manufacturing interruptions, disappointment among consumers, and commercial manufacturing interruptions, rather than assessing interruptions just on the basis of the total number of customers who were affected or the immediate financial consequences. These commercialized impact measurements allow energy companies to allocate resources to remediation measures that decrease the total monetary cost to communities and essential infra-structure, compared to exclusively concentrating on temporary financial recuperation or emergency achievements (K{\"u}feo{\u{g}}lu, 2015), (Ericson, 2018). It enables energy companies to adopt more information-based as well as social sustainable decisions regarding operation. Machine-learning techniques have been developed in recent electrical system studies to cluster outages by origin, scale, and expected impact and predict delay restoration over time (Tang, 2024).In addition, current advances combine techniques for optimization and predictive training to dynamically rank restoration duties according to customer impact, location clustering, and forecast outage amount (Ansarinejad, 2025).All of these capacities for forecasting must be incorporated into any grouping system that aims to efficient-ly distribute crews and financing (Wang, 2023). Lean Six Sigma (LSS) and DMAIC approach-es offer an ordered improvement structure that converts analytical findings into workable processes for operations, building on such forecasting information (Sony, 2019). This framework recognizes the underlying causes, streamlines crew procedures, and incorporates improvement phases to ensure that prioritized actions endure be-yond one investigation. Recent energies utilize have shown quantifiable improvements as a result time, energy effectiveness, and reliability when DMAIC is combined with analysis of data (Hsu, 2013). The context of contrast to simplistic FIFO restoration strategies, the unified strategy presented in this work (1) computes a ranking score based on each interruption using open service disruption logs and socioeconomic impact statistics, (2) validates and operationalizes the rating within useful workflows using the method of Lean Six Sigma (DMAIC), along with (3) shows higher equity and entire system advantages (Najafi-Shad, 2024), (Loni, 2024). The method bridges the gap within analytics and practical use by building on previous work in uncertain restoration analysis, data-driven risk assessment, and downtime costing (Willems, Probabilistic Restoration Modeling of Wide-Area Power Outage, 2024). We analyze dependability across various failure causes and scales, adaptability to alternative damaging functions and social risk weighting, and its actual implications for crew arrival and regulatory reliability goals, while also displaying enhanced performance on historical general outage records. The research project intends to give utilities and authorities a replicable route for arranging interruptions through methods that are available, supported by data, and consistent with public interest by fusing empirical failure records about LSS-driven workflow change (Schellenberg, 2016), (Lawton, 2003). Together, these efforts provide a measure for selection that can be understood and is based on public failure logs; an LSS DMAIC guide for integrating the metric into restoration practice; and a study that demonstrates how predictive prioritization can mitigate varia-tions in outage results and lessen customer harm. They directly address utilities and agencies looking for practical methods to enhance outage recovery while juggling a shortage of field resources (Utama, 2023), (Okuh, 2021).

## 2. Literature Review

This subsection includes research that mainly used heuristic/metaheuristic techniques and optimization algorithms for scheduling, load shedding, grid investment, and maintenance. In order to prioritize power assets and optimize disposal methods, Zhou and Yan (Zhou, 2025) presented a hybrid data-driven system that integrates Analytic Hier-archy Process (AHP), Natural Gradient Boosting (NGB), and a Parallel Ge-netic Algorithm (PGA). The framework showed important improvements over traditional methods in the System Average Interruption Duration Index (SAIDI) and System Average Interruption Frequency Index (SAIFI) using a variety of datasets (condition reports, maintenance logs). Additionally, NGB was able to accurately predict the asset's Remaining Useful Life (RUL) in or-der to feed the PGA optimization. Some disadvantages include the need for complicated system integration and data preprocessing, as well as limited usefulness in the absence of reliable, high-volume operational data.

In this study (Ma, 2025), an optimization model for power grid investment decision-making has been developed using a Genetic Algorithm (GA) as the solution algorithm. Objective functions and constraints have been defined. It designed a technique for evaluating investment efficiency that is based on 16 second-ary and 4 primary indexes. Particle Swarm Optimization (PSO) was used to evaluate the GA's performance. The GA outperformed the PSO in terms of ef-ficacy, exhibiting a greater range of optimal alternatives for grid investment benefits. The research determined that certain projects were close the best in-vestment benefit solution (e.g., Project 5 at 63.42%). The limitations include the need for voltage data reduction to achieve optimal solution diversity and the lack of usefulness without defining the kind of the 20 evaluation indices and the specifics of the target function.

Li et al. (Li, 2025) proposed the Data Frequency Scheduling Optimization Frame-work (DFSOF) to intelligently manage energy storage for frequency stability against uncertain load fluctuations and renewable integration challenges. By combining adaptive load forecasting, frequency deviation analysis, and Rein-forcement Learning (RL)-based charge/discharge optimization, the DFSOF utilizes a data-driven hybrid control methodology. Using smart scheduling algorithms and the Dynamic Frequency Stability Index (DFSI) for real-time performance evaluation are two important advances. DFSOF increased energy efficiency, response time, and frequency regulation accuracy by up to 96.4% as compared to traditional scheduling. The disadvantages include the need for complexity for real-time RL-based optimization and continuous performance evaluation using DFSI, as well as limited utility across systems with widely disparate levels of renewable penetration.

Yankson et al. (Yankson, 2025) presented a novel resilience enhancement framework for power distribution systems based on N-1 Impact Analysis to address chal-lenges from aging infrastructure and climate change. By separating the net-work into resilience zones, the technique ranks essential lines according to their influence on unserved energy. To size and distribute Distributed Energy Resources (DERs) within these zones while taking operating limitations into account, an Improved Grey Wolf Optimizer (IGWO) was utilized. The sys-tem, which was validated on an actual utility feeder using MATLAB and OpenDSS, showed greater cost-effectiveness over current methods and signif-icantly reduced the amount of energy that was not used. Cons: The IGWO method is difficult to build and tune for optimal DER allocation, and it is not very useful in systems without comprehensive N-1 outage data for effect analysis.

Haidar, Helwig and Moldovan (Haidar, 2025) proved the best way to reduce power fail-ures in a cloud computing and monitoring setting by implementing steady-state load shedding with the Particle Swarm Optimization (PSO) method. The goal was to reduce the quantity and location of power outages. The methodol-ogy was verified using industry-standard IEEE test systems. Initial findings showed that reserve generation could continue to function since system pa-rameters were kept within safe bounds. By enhancing real-time monitoring and data-driven operational methods, this modeling improves power grid se-curity. The inability to provide a comparison with alternative load shedding heuristics or indicate the amount of shed power decrease, as well as the diffi-culty of integrating a cloud computing environment for optimization, are drawbacks.

Zheng et al. (Zheng, 2025) proposed a hierarchical control framework to optimally co-ordinate battery storage, heat pumps, and DHW systems across multiple resi-dential buildings, addressing challenges from renewable energy proliferation. For disturbance forecasting, the framework took use of LightGBM, k-NN, and linear regression. It merged a microgrid-level physics-based Model Predic-tive Control (MPC) for grid objectives (peak shaving) with a building-level data-driven MPC for occupant satisfaction. A bottom-up approach was used to coordinate (demand estimation communicated to battery dispatch). The framework performed extremely well in terms of comfort, emissions, and re-silience measures, earning second place in the 2023 NeurIPS CityLearn Chal-lenge. The shortcomings include the need for complicated data pipelines for several forecasting models and two different MPC implementations, as well as limited usefulness outside of environments with well-established multi-level control architecture.

To be able to forecast power transformer failures and optimize maintenance schedules, Tonoy (Tonoy, 2025) introduced a real-time condition monitoring model that makes use of Internet of Things (IoT) technology. The system combines analytics via the cloud and predictive algorithms with multi-sensor data col-lection (oil temperature, vibration, and partial discharge). The key contribu-tion is the ongoing monitoring that eliminates the negative operational and economic effects of time-based or reactive maintenance. Using both simulat-ed and real-world case studies, the model was verified and showed notable gains in overall maintenance efficiency and problem detection accuracy. Cons include the cost and difficulty of integrating a multi-sensor IoT infra-structure for real-time data collecting, as well as the lack of usefulness with-out defining the type of predictive algorithms utilized.

In order to overcome constraints such as subjective rules and incomplete con-siderations, Lu and Li (Lu, 2021) suggested a data-driven approach for developing a more reasonable maintenance priority queue for Circuit Breakers (CBs). For an extensive maintenance assessment, the method made use of impact indica-tors and control circuit time features. To identify maintenance priority, a hy-brid approach that combines a Ranking Support Vector Machine (SVM) and fuzzy C-means (FCM) clustering was presented. Results from simulations outperformed traditional techniques, providing a logical and impartial priori-tization—

especially in the absence of prior knowledge. The approach can also deal with invalid and missing data. The lack of usefulness without defining the type of "impact indicators" and the difficulty of fine-tuning the FCM/Ranking SVM hyperparameters and complicated data preprocessing are drawbacks.

Watson, Pasqualini, Anagnostou (Watson, 2024) described a novel predictive modeling approach to accurately forecast tropical storm-related power outages across the Continental United States, aiming to create effective decision-making support tools. This framework described environmental and infrastructure conditions using various kinds of multidisciplinary data, such as earth data-bases, US Census Bureau products, and numerical weather forecasts. In a thorough analysis of 38 past tropical storms, the model outperformed well-known, smaller-scale models and reduced Mean Squared Error (MSE) by more than 50% when compared to the closest comparable large-scale model. Decision-makers are informed by this method of the possibility of a protract-ed outage and its overall severity. The limitations include the need for multi-disciplinary data collection and harmonization throughout the Continental United States, as well as the lack of relevance without describing the precise prediction model technique used.

Hughes et al. (Hughes, 2022) combined machine learning (ML) techniques with the structural fragilities of pole-wire systems to build a Hybrid Physics-based and Data-driven (HPD) model that accurately forecasts power outages during extreme events. Physics-based fragility curves were calibrated using the ML model and then used to predict high-impact events in situations with limited empirical data. Having a Root Mean Square Error (RMSE) improvement of 48% for high-impact event outage prediction, the HPD model showed signifi-cant modeling gains over purely data-driven techniques. Also, targeted hard-ening measures on Connecticut's oldest 5% of infrastructure might have de-creased pole failures by 33% yearly, according to counterfactual research. The lack of usefulness without mentioning the particular machine learning method and the need for precise structural fragility data and comprehensive pole properties are negatives.

The critical need for protection of the power grid against a range of disrup-tions—such as severe weather, uncertainty of renewable energy, market vola-tility, and demand-side technological changes—that are absent from tradi-tional designs was the focus of this research study (Wu, 2022). To reduce power sys-tem disruptions of various sizes, the study combined domain expertise with data-driven methodologies. The main contribution is the creation of tech-niques for preventing cascading failures following minor hazards and miti-gating very large-scale catastrophic events. Specifically, a protection design driven by Reinforcement Learning (RL) does this. Concerns include the need for high-quality, real-time data to handle the many complex operating situa-tions and the lack of usefulness without describing the precise data-driven models and the architecture of the RL-powered protection solution.

Odonkor and Lewis (Odonkor, 2019) aimed to maximize arbitrage value for energy cost savings from distributed energy resources (DERs) by designing optimal oper-ational strategies, addressing electricity price volatility caused by renewable intermittency. Reinforcement Learning (RL) was used to solve the challenge, which was defined as an arbitrage maximization problem utilizing design op-timization principles. The method was used for shared DERs in residential clusters with many buildings. Generalized learning and effective arbitrage policy design were highlighted by the results, which showed significant ener-gy cost reductions above baseline values across three distinct clusters. Scala-bility was verified with little extra computational expense. Drawbacks in-clude relying on erratic power pricing signals, having limited utility without identifying the RL algorithm employed, and the complexity needed to man-age shared DERs across several buildings.

In Table 1 we have demonstrated the summary of existing paper that we have reviewed for our research in tabular format.

**Table 1.** Summary of Data-Driven Models for Power Grid Optimization, Resilience, and Maintenance from Reviewed Literature

| Year | Ref. | Model | Results | Limitations |
|---|---|---|---|---|
| 2019 | (Odonkor, 2019) | Reinforcement Learning (RL) for DER operation | Significant energy cost reductions; scalable across building clusters | Relies on erratic pricing signals; unspecified RL algorithm; complex multi-building management |
| 2019 | (Obatola, 2024) | Data-driven prediction + Resilience-based demand response | Reduced load shedding during microgrid islanding; resilience index proposed | 10,000 Monte Carlo simulations needed; unspecified data-driven estimation technique |
| 2021 | (Lu, 2021) | Ranking SVM + Fuzzy C-means clustering for CB maintenance | Logical, impartial maintenance prioritization; handles invalid/missing data | Unspecified impact indicators; hyperparameter tuning; complex preprocessing |
| 2022 | (Hughes, 2022) | Hybrid Physics-based + Data-driven (HPD) for outage prediction | RMSE improved by 48%; reduced pole failures by 33% | Needs detailed structural fragility data; unspecified ML methods |
| 2022 | (Wu, 2022) | RL-based protection design | Mitigates cascading failures and catastrophic events | Requires high-quality real-time data; unspecified data-driven models and RL architecture |
| 2024 | (Wats | Predictive modeling using | Reduced MSE by 50%+ in | Extensive data collection; unspecified |

| | | | | |
|---|---|---|---|---|
| | on, 2024) | multidisciplinary data | tropical storm outage prediction | prediction technique |
| 2025 | (Zhou, 2025) | AHP + NGB + Parallel GA | Improved SAIDI and SAIFI; predicted RUL | Complex system integration; high data requirements |
| 2025 | (Ma, 2025) | Genetic Algorithm (GA) vs Particle Swarm Optimization (PSO) | GA outperformed PSO in grid investment decision-making | Voltage data reduction needed; unclear evaluation indices and target function |
| 2025 | (Li, 2025) | Data Frequency Scheduling Optimization Framework (DFSOF) with RL | Improved energy efficiency, response time, frequency regulation by 96.4% | Real-time RL complexity; limited applicability across diverse renewable systems |
| 2025 | (Zheng, 2025) | Hierarchical control + LightGBM + k-NN + Linear Regression + MPC | High comfort, emissions, resilience; 2nd place CityLearn Challenge | Complex data pipelines; limited use outside well-established multi-level control |
| 2025 | (Tonoy, 2025) | IoT-based real-time transformer monitoring | Improved maintenance efficiency; early fault detection | High IoT infrastructure cost; unspecified predictive algorithms |
| 2025 | (Yankson, 2025) | N-1 Impact Analysis + Improved Grey Wolf Optimizer (IGWO) | Reduced unserved energy; cost-effective | Difficult IGWO tuning; requires comprehensive N-1 outage data |
| 2025 | (Haidar, 2025) | PSO-based steady-state load shedding | Reduced outages; maintained safe operation | No comparison with alternatives; cloud computing integration complexity |
| 2025 | (Nassabeh, 2025) | DDQN-based AI energy management (PV, ESS, V2H EVs) | Maintained zero ENS; 100% PV self-consumption; faster execution | Unspecified reward function; expensive V2H setup |
| 2025 | (Singh, 2025) | Deep Reinforcement Learning (DRL) for adaptive demand response | Outperformed conventional models in multi-agent coordination and pricing | Needs validated DRL model; data privacy and policy generalization challenges |

## 3. Methodology

In this section, the methodological framework in Figure 1 adopted in the development of the proposed hybrid deep-ensemble stacking model in identifying the impact of power outage has been outlined. The approach combines ideas of deep learning, ensemble theory and meta-learning to build a powerful and understandable predictive structure. Preprocessor: This starts with strong data preprocess and feature standardized, after which several base learners, including an Artificial Neural Network (ANN), a Random Forest (RF); the Extreme Gradient Boosting (XGB) are trained and each of them learns different things about the data distribution that they share a common feature with a different base learner. A meta-learning layer composed of multinomial logistic regression is then fitted using Out-of-Fold stacking of their probabilistic outputs and latent representations so that they are unbiased in generalization. This systematic method actually allows this model to have a balance between accuracy, stability, and interpretability in predicting multi-class outage effects.

### *3.1 Dataset Description*

The dataset Autunno created and is hosted on Kaggle is called 15 Years of Power Outages and is a massive publicly available compilation of electrical grid outage data collected between 2000 and 2014. It gives an overall picture of the outage events in North America and it forms a basis of data to be used in study of the reliability and performance of power systems. Every record is a unique outage occurrence and includes all the necessary temporal, operating, and geographic characteristics, which makes the data set best suited to data-driven reliability and risk analysis. Its design facilitates descriptive and inferential evaluation of outage frequency, restoring time and category of cause, which are important elements to Lean Six Sigma-based grid improvement programs.

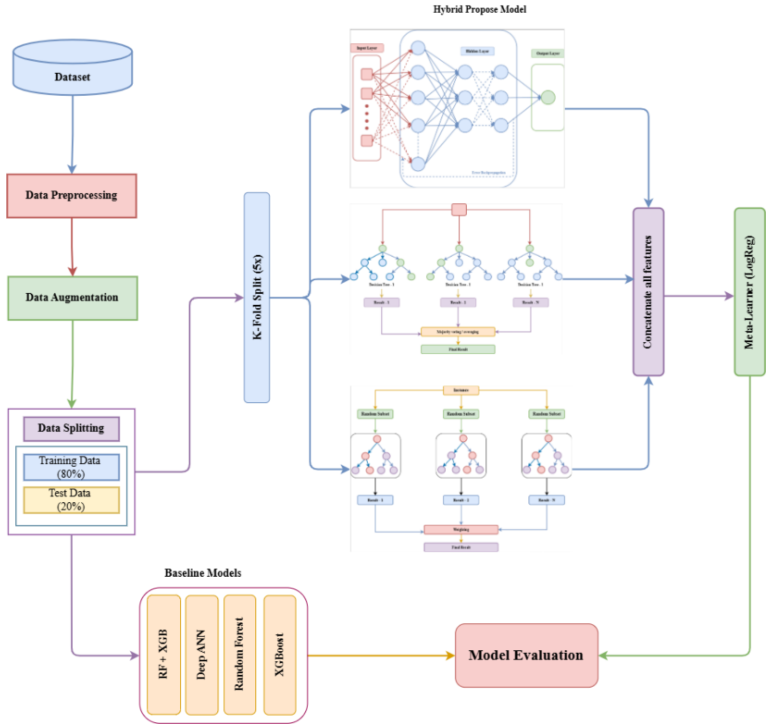

**Fig. 1.** Workflow of Hybrid Deep–Ensemble Stacking

There are 1,144 distinct outage records in the dataset, which is characterized by twelve varia-bles, which also describe the nature and time of the event. Such attributes as event descrip-tion, year, month, date and time of occurrence and date and time of restoration make it possible to estimate with high accuracy the outage duration and such performance indices as Mean Time to Restore (MTTR). Additional fields, such as respondent (utility/agen-cy managing the event), and geographic area (e.g., Mississippi, Texas, New York) provide context to the data, and the NERG Region allows having all the records under the jurisdiction of the North American Electricity Reliability Corporation (NERC) providing a way to benchmark between events across different regions in terms of the stability of the grid and the efficiency of its maintenance.

Demand Loss (MW) and Number of Customers Affected to indicate the quantitative aspects of the operational and social consequences of outages directly. The dataset discloses that severe weather takes up almost 10 percent of total incidences recorded, vandalism is 7 percent and the rest 83 percent are categorized under equipment malfunction, operating problems and miscellaneous. The 2000-2014 period which is covered more densely between 2008-2014 indicates the development of the data collection practices and the growing stress on the reliability documentation.

The overall framework of this dataset fits the DMAIC (Define-Measure-Analyze-Improve-Control) loop of the Lean Six Sigma very well. It allows determining and analyzing important outage metrics and allows one to model processes with statistics on improving processes and prioritizing reliability efforts. All of these variables need to be categorical, temporal, and quantitative, which makes the framework holistic to identify the main points of failure, decrease the restoration delays, and enhance customer resilience throughout the power grid. Table 2 gives a summary of dataset attributes.

### *3.2 Data Preprocessing*

The 15 Years of Power Outages data had to undergo a lot of preprocessing before it could be subjected to Lean Six Sigma based analytical modeling. As the raw outage logs had entries that were not present, not formatted uniformly, and target classes were not balanced, the preprocessing pipeline was created to clean, transform, and augment the data in a systematic manner. This is what made the processed data constant, dependable, and analytical enough to pass to the Measure and Analyze steps of the DMAIC cycle.

**Table 2**: Dataset Description

| Attribute | Description | Example |
|---|---|---|
| Event Description | Cause or reason of the outage | Severe Weather, Vandalism |
| Year | Year of occurrence | 2014 |
| Month | Month of occurrence | November |
| Date Event Began | Start date of outage | 06/30/2014 |
| Time Event Began | Time outage started | 11:20 p.m. |
| Date of Restoration | Date when power was restored | 07/01/2014 |
| Time of Restoration | Time when power was restored | 5:00 p.m. |
| Respondent | Responsible company or agency | NERC-listed utility |
| Geographic Area | Affected location | Mississippi, Texas, New York |
| NERG Region | NERC jurisdictional region | Reliability Zone 3 |
| Demand Loss (MW) | Energy not transmitted | 100 MW |

### *3.2.1 Data Cleaning and Missing Value Handling*

Initial inspection using df.info(), df.describe(), and null-value counts revealed in-consistent entries and missing fields. Placeholder strings such as "Unknown," "NaN," or blanks were replaced with standard NaN tokens. The date columns (Date Event Began and Date of Restoration) were coercively parsed into Python datetime format and bad timestamps were removed.

The numeric variables (Demand Loss (MW), Number Customers Affect-ed, etc.) were initially presented in the form of text with comma and empty strings. These were standardized by number conversion algorithm. Corrupt rows (and completely blank rows) were removed to make the structure consistent.

To quantify outage length, an operational duration field was computed in hours using,

$$\text{Duration}_{hours} = \frac{(\text{Date of Restoration} - \text{Date Event Began})}{3600} \qquad (1)$$

Records with negative durations were eliminated. This transformation provided a precise measure of downtime essential for reliability assessment.

### *3.2.2 Feature Engineering and Categorical Encoding*

Feature engineering was performed to capture operational and temporal dependencies. A derived feature, Impact Score, was created to quantify the combined effect of duration and customer loss, as,

$$\text{Impact Score} = \text{Duration}_{hours} \times \text{Number of Customers Affected} \qquad (2)$$

Also, Year and Month were taken out of the start timestamp to indicate the potential seasonal or yearly outage trends. A numeric encoding of categorical variables, such as NERC Region and Tags, was done by means of Label Encoding, whereas the target variable (Impact_Label) was encoded as Impact Label-Num in order to build classification models. This learning process allowed machine learning models to process qualitative data in the form of a quantitative model.

### *3.2.3 Data Scaling*

Power outage data are usually highly skewed because of extreme events. To reduce the impact of these outliers, the numerical variables were normalized with the use of RobustScaler that multiplies the data by the interquartile range (IQR). This change can be defined as follows:

$$X' = \frac{X - Q_2}{Q_3 - Q_1} \qquad (3)$$

Q 1 and Q 3 are the first and third quartile, Q 2 is the median. The scaling method is resistant to extreme values and provides convalence to the model training.

### *3.2.4 Data Splitting*

The processed data was scaled and then cleaned after which it was split into input features (X) and target labels (y). Fields that do not provide information (e.g. text identifiers or redundant columns) were left out of the feature matrix. To guarantee the reproducibility of the results and even distribution of classes, a train-test split in the ratio 80: 20 (a fixed random seed 42) was used. This division allowed objective assessment and generalization analysis of the further development of the model.

### *3.2.5 Data Augmentation*

There was a great imbalance of classes in the dataset of Impact_Label cate-gories with minority classes like vandalism or fuel supply interruption being under-represented. To this end, the Synthetic Minority Over-sampling Technique (SMOTE) was used. SMOTE generates synthetic minority samples by linear interpolation between existing data points within the same class, as,

$$x_{new} = x_i + \lambda(x_j - x_i), \lambda \in [0,1] \qquad (4)$$

where $x_i$and $x_j$are minority-class feature vectors, and $\lambda$is a random coefficient controlling interpolation.

To enhance more further the use of diversity and to make it more similar to the real-world reporting variation, some Gaussian noise was added to the numerical features. All the features were corrupted by a random noise based on a normal distribution with zero mean and standard deviation equal to 2% of the range of values in the feature:

$$x' = x + \epsilon, \epsilon \sim \mathcal{N}(0, 0.02^2) \qquad (5)$$

This combination of SMOTE and stochastic perturbation produced a balanced, realistic dataset that avoided overfitting to synthetic samples.

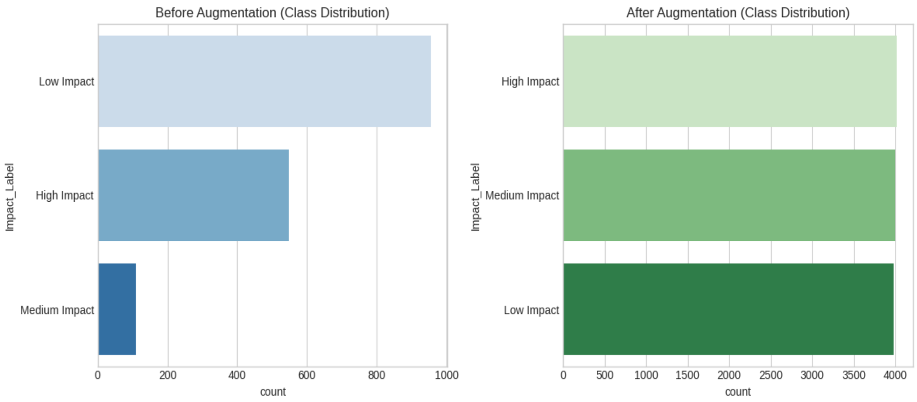

**Fig. 2.** Class Distribution Before and After Augmentation

The dataset was pre-processed and contained clean, balanced, and consistent outage records. All the variables were brought to similar magnitudes, numerical encoding of the categorical fields was done, and missing values were treated accordingly. The augmentation stage was able to achieve the representation of the minority classes and decrease the imbalance between the classes, as well as enhance the model stability. All in all, these preprocessing steps provided a high-quality dataset that could be used in the Lean Six Sigma analysis-a high ability to measure, predict and prioritize grid interruptions in an understandable way.

### 3.3 Proposed Hybrid Model

The current power systems are working under more and more complex and volatile conditions, and severe weather conditions, failures of equipment, and peaks in demand more often cause power interruption of different severity. The correct categorization of the severity of impact of such events is essential to ensure effective distribution of resources, enhanced resilience planning, and operation reliability. Conventional machine learning models though effective in dealing with structured tabular data have been found to be difficult to capture high order nonlinear dependencies. On the other hand, deep learning models are highly effective at feature abstraction at a hierarchy but can be erratic and overfitting with small homogeneous, noisy and heterogeneous datasets.

To balance these complementary advantages, a hybrid deep-ensemble stack, which combines a Deep Artificial Neural Network (ANN) with the traditional ensemble learners, Random Forest (RF) and Extreme Gradient Boosting (XGB) are proposed in this study based on the concept of meta-learning layer has shown in Figure 3. The framework takes advantage of the representation learning of deep networks, as well as decision diversity of ensemble methods, and thus, becomes more robust, accurate and well-generalized.

#### *3.3.1 Hybrid Learning Rationale*

In statistical learning theory, generalization error $\mathcal{E}_g$ decomposes as:

$$\mathcal{E}_g = \text{Bias}^2 + \text{Variance} + \text{Irreducible Noise} \qquad (6)$$

Deep neural networks ($f_{\text{DL}}$) are low-bias but high-variance estimators—capable of complex representation learning but sensitive to data perturbations. Tree ensembles ($f_{\text{RF}}, f_{\text{XGB}}$) are high-bias but low-variance estimators—robust but limited in expressivity. Stacking creates a *meta-function*:

$$f_{\text{stack}}(\mathbf{x}) = g(f_{\text{DL}}(\mathbf{x}), f_{\text{RF}}(\mathbf{x}), f_{\text{XGB}}(\mathbf{x})) \qquad (7)$$

where $g(\cdot)$learns to minimize both bias and variance by adaptively weighting predictions from heterogeneous learners. This achieves a lower expected generalization error:

$$\mathbb{E}[\mathcal{E}_g(f_{\text{stack}})] < \min_m \mathbb{E}[\mathcal{E}_g(f_m)] \tag{8}$$

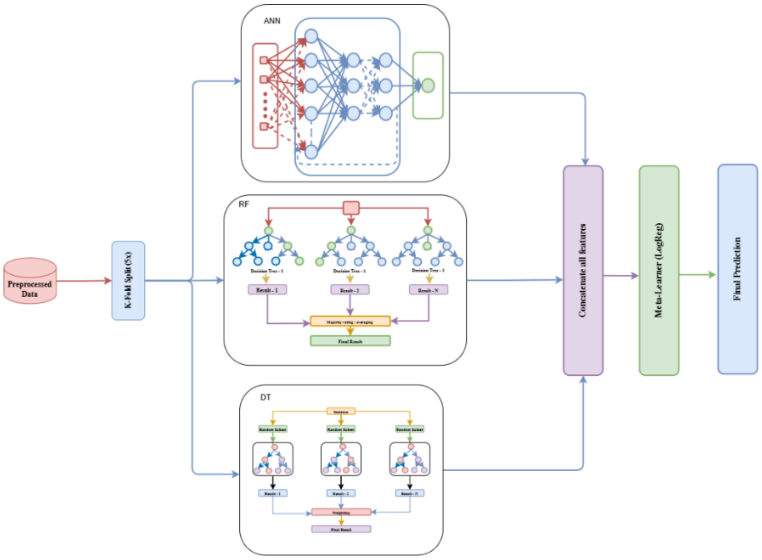

**Fig. 3.** Proposed Model: Deep–Ensemble Stacking Framework for Power-Outage Impact Classification

### *3.3.2 Model Architecture*

The proposed framework is a two-tier hybrid stacking architecture that synergistically combines deep representation learning and ensemble-based decision modeling to classify power outage impact levels with high reliability. The architecture operates on an input feature space $\mathbf{x} \in \mathbb{R}^d$, representing a multidimensional vector of meteorological, infrastructural, and temporal indicators, with corresponding class labels $y \in \{0,1,2\}$denoting *Low*, *Medium*, and *High* impact events respectively.

At the first level, a Deep Artificial Neural Network (ANN) in Figure 4 is employed to learn hierarchical representations of the feature space. The ANN comprises two fully connected hidden layers, where nonlinear transformations are applied as follows:

$$\begin{gathered}\mathbf{h}_1 = \phi(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)\\ \mathbf{h}_2 = \phi(\mathbf{W}_2\mathbf{h}_1 + \mathbf{b}_2)\end{gathered} \tag{9}$$

Where, $\mathrm{W}_1, \mathrm{W}_2$and $\mathrm{b}_1, \mathrm{b}_2$denote weight matrices and bias vectors, and $\phi(\cdot)$represents the ReLU activation function.

To prevent overfitting and stabilize convergence, each dense layer is regularized using batch normalization and dropout with rates of 0.35 and 0.25 respectively.

A bottleneck embedding layer with 32 neurons is introduced to project the high-dimensional feature representation into a compact latent space:

$$\mathbf{z} = \phi(\mathbf{W}_b\mathbf{h}_2 + \mathbf{b}_b), \mathbf{z} \in \mathbb{R}^{32} \tag{10}$$

This embedding vector $\mathbf{z}$acts as a condensed semantic encoding that captures the most discriminative features contributing to impact classification.

The output layer of the ANN transforms this latent representation into a probability distribution over the $C = 3$classes using the Softmax function:

$$\hat{\mathbf{p}}_{\text{DL}} = \operatorname{softmax}(\mathbf{W}_o\mathbf{z} + \mathbf{b}_o) \tag{11}$$

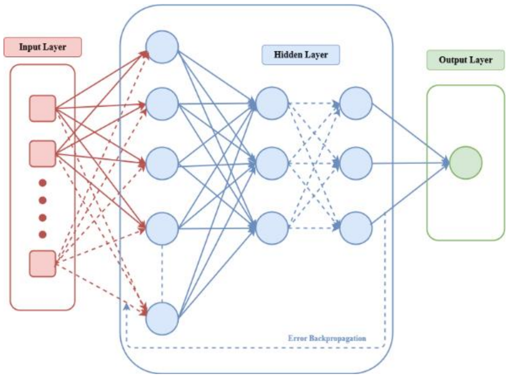

**Fig. 4.** Architecture of ANN

where

$$\mathrm{softmax}(\mathbf{a})_c = \frac{\exp(a_c)}{\sum_{k=1}^{C} \exp(a_k)}, c \in \{1, \ldots, C\} \qquad (12)$$

The model parameters $\Theta_{\mathrm{DL}} = \{\mathbf{W}_i, \mathbf{b}_i\}$are optimized using the *sparse categorical cross-entropy loss*:

$$\mathcal{L}_{\mathrm{DL}} = -\frac{1}{N} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \sum_{\boldsymbol{c=1}}^{\boldsymbol{C}} \mathbf{1}(y_i = c) \log \hat{p}_{ic}^{\mathrm{DL}}. \qquad (13)$$

Parallel to the deep network, two ensemble-based learners—a Random Forest (RF) and an Extreme Gradient Boosting (XGB) model—are trained on the same standardized input space.

The Random Forest in Figure 5 consists of an ensemble of $T$decision trees, each denoted as $h_t(\mathbf{x})$. The RF prediction represents an average of individual tree predictions:

$$\hat{\mathbf{p}}_{\mathrm{RF}} = \frac{1}{T} \sum_{t=1}^{T} h_t(\mathbf{x}) \qquad (13)$$

This approach reduces variance through bagging and feature subsampling, enhancing model stability.

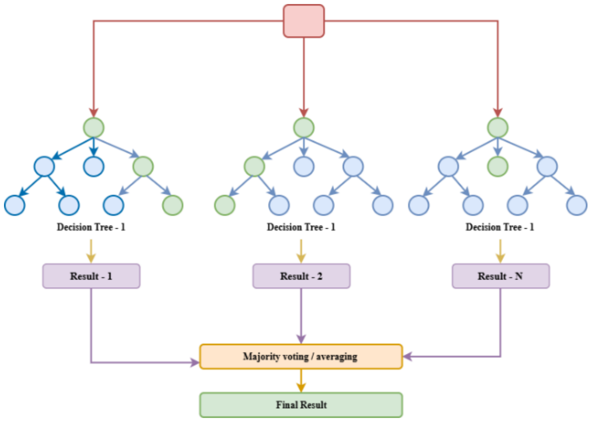

**Fig. 5**. Architecture of Random Forest

The Extreme Gradient Boosting (XGB) model in Figure 6 adopts a stage-wise additive structure:

$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \eta h_m(\mathbf{x})$ (14)

where $F_m(\mathbf{x})$ is the boosted ensemble after $m$ iterations, $h_m(\mathbf{x})$ denotes the new regression tree fitted to the pseudo-residuals, and $\eta$ is the learning rate controlling the contribution of each tree. The final class probabilities are computed via:

$$\hat{\mathbf{p}}_{\text{XGB}} = \text{softmax}(F_M(\mathbf{x})) \qquad (15)$$

This mechanism enables XGB to capture complex nonlinear patterns through gradient optimization while maintaining computational efficiency and regularization via subsampling parameters.

After training the three base models, their outputs are concatenated to form a meta-feature vector that integrates heterogeneous learning signals. This vector is defined as:

$$\mathbf{x}^{(2)} = [\hat{\mathbf{p}}_{\text{DL}} \,||\, \hat{\mathbf{p}}_{\text{RF}} \,||\, \hat{\mathbf{p}}_{\text{XGB}} \,||\, \mathbf{z}] \qquad (16)$$

where $||$ denotes vector concatenation, resulting in a unified representation combining both decision-level probabilities and feature-level embeddings.

At the second level, a meta-learner based on multinomial Logistic Regression (LR) is trained on these meta-features to produce the final class prediction. For each class $c \in \{1,2,3\}$, the meta-learner computes:

$$P(y = c \mid \mathbf{x}^{(2)}) = \frac{\exp(\boldsymbol{\beta}_c^{\top}\mathbf{x}^{(2)})}{\sum_{k=1}^{C}\exp(\boldsymbol{\beta}_k^{\top}\mathbf{x}^{(2)})} \qquad (17)$$

where $\boldsymbol{\beta}_c$ denotes the parameter vector for class $c$. The parameters are estimated via maximum likelihood estimation (MLE) by minimizing:

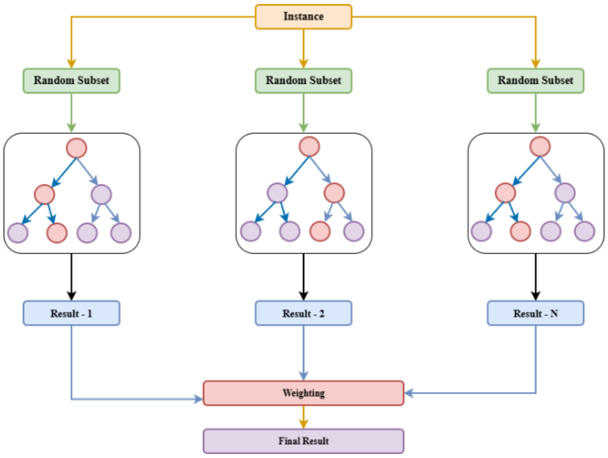

**Fig. 6.** Architecture of XGBoost

$$\mathcal{L}_{\text{meta}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\mathbf{1}(y_i = c)\log P(y_i = c \mid \mathbf{x}_i^{(2)}) \qquad (18)$$

The meta-learner effectively learns optimal fusion weights for the predictions of the base learners, functioning as a probabilistic aggregator that balances overconfident or correlated outputs.

During training, Out-of-Fold (OOF) cross-validation is applied to prevent overfitting. The dataset is divided into $K$ stratified folds, and each base model is trained on $K - 1$ folds while generating predictions on the held-out fold:

$$\text{OOF}_m[i] = f_m^{(-k)}(\mathbf{x}_i), \mathbf{x}_i \in \text{fold } k \qquad (19)$$

This ensures that meta-training is conducted on unbiased predictions from unseen data.

In summary, the architecture integrates the feature abstraction power of deep learning, the stability and variance reduction of ensemble methods, and the statistical interpretability of logistic meta-learning. The result is a cohesive, end-to-end system capable of learning both hierarchical representations and structured dependencies, yielding a high-performing and generalizable model for power outage impact prediction.

### 3.3.3 Level-2 Meta-Learning

The final stage of the proposed hybrid architecture is the Level-2 meta-learning layer, which serves as an intelligent fusion mechanism to combine the outputs of all base models into a single, coherent predictive framework. In traditional ensemble systems, base learners such as the Deep ANN, Random Forest, and XGBoost each produce distinct decision boundaries, reflecting different inductive biases and data representations. However, no single model is universally optimal across all feature subspaces. The meta-learner is designed to exploit this complementarity by learning how to *adaptively weight* and *calibrate* the predictions from each base model in a data-driven manner.

Formally, for a given input instance $\mathbf{x}$, each base model $m \in \{1,2,3\}$generates a probability vector over $C$classes, denoted as $\hat{\mathbf{p}}_m = [\hat{p}_{m1}, \hat{p}_{m2}, \dots, \hat{p}_{mC}]$. The Deep ANN additionally produces a latent embedding vector $\mathrm{z} \in \mathbb{R}^{32}$from its bottleneck layer, which encapsulates compressed nonlinear representations of the feature manifold. These outputs are concatenated into a unified meta-feature representation:

$$\mathbf{x}^{(2)} = [\hat{\mathbf{p}}_{\mathrm{DL}} \ \| \ \hat{\mathbf{p}}_{\mathrm{RF}} \ \| \ \hat{\mathbf{p}}_{\mathrm{XGB}} \ \| \ \mathbf{z}] \quad (20)$$

where $\|$denotes vector concatenation. This composite feature vector captures both decision-level information (from the probability estimates) and representation-level information (from the deep embedding), offering the meta-learner a rich, high-dimensional summary of the diverse learning perspectives.

The Level-2 learner employed in this study is a multinomial Logistic Regression (LR) model, chosen for its statistical interpretability, convex optimization landscape, and robustness to overfitting under balanced regularization. The model estimates the posterior probability of each class $c \in \{1,2,\dots,C\}$using the Softmax function:

$$P(y = c \mid \mathbf{x}^{(2)}) = \frac{\exp(\boldsymbol{\beta}_c^{\top}\mathbf{x}^{(2)})}{\sum_{k=1}^{C}\exp(\boldsymbol{\beta}_k^{\top}\mathbf{x}^{(2)})} \quad (21)$$

where $\boldsymbol{\beta}_c$represents the learned coefficient vector for class $c$.
The meta-learner's parameters $\beta = \{\beta_1, \dots, \beta_C\}$are optimized via maximum likelihood estimation (MLE), minimizing the negative log-likelihood:

$$\mathcal{L}_{\mathrm{meta}} = -\frac{1}{N} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} \sum_{\boldsymbol{c}=\mathbf{1}}^{\boldsymbol{C}} \mathbf{1}(y_i = c) \log P(y_i = c \mid \mathbf{x}_i^{(2)}) \quad (22)$$

subject to class-balanced regularization constraints to mitigate bias in underrepresented classes.

Intuitively, this logistic meta-model functions as a statistical consensus layer that learns how to combine heterogeneous predictive signals based on their contextual reliability. For example, in regions of the feature space where the ANN is confident (e.g., capturing nonlinear relationships), the logistic weights for $\hat{\mathbf{p}}_{\mathrm{DL}}$are elevated; conversely, in regions where the structured features dominate, the weights for RF or XGB become more influential. This adaptive weighting scheme effectively learns the optimal linear combination in probability space, thus enhancing calibration and reducing overconfidence.

To prevent overfitting and ensure genuine generalization, the meta-learner is trained using Out-of-Fold (OOF) predictions from the base models. During the K-fold process, each base learner is trained on $K - 1$folds and generates predictions on the held-out fold, ensuring that the meta-learner only sees predictions from data unseen by the base models. This OOF stacking procedure guarantees unbiased meta-features, preventing information leakage between training and validation stages.

Ultimately, the Level-2 meta-learning layer acts as the cognitive integrator of the ensemble transforming the diverse, partially correlated outputs of the deep and tree-based models into a unified, well-calibrated final decision. It embodies the principle of meta-generalization, where learning occurs not only over the original data space but also over the *space of learned models*, thereby achieving a robust and statistically principled fusion of heterogeneous predictors.

The primary advantage of the proposed hybrid model lies in its ability to integrate deep representation learning and ensemble-based decision modeling within a unified framework. Deep neural networks capture complex nonlinear dependencies through hierarchical feature abstraction, while ensemble methods like Random Forest and XGBoost offer statistical robustness, interpretability, and resilience to noise. By combining these complementary strengths through a meta-learning layer, the model achieves superior generalization compared to any single learner alone.

The inclusion of a bottleneck embedding layer in the deep neural network provides a compact, information-rich latent representation that distills the most discriminative features relevant to outage impact classification. When combined with the probabilistic outputs of the ensemble learners, this latent embedding enables the meta-learner to fuse both abstract semantic features and structured decision information, resulting in a more stable and data-efficient classifier.

| **Algorithm 1:** Training and Inference of the Proposed Hybrid Model | |
|---|---|
| **Step** | **Description** |
| **Input:** | Dataset $D = \{(x_i, y_i)\}$, number of folds $(K)$ |
| **1.** | Normalize features using StandardScaler and encode target labels |
| **2.** | Split data into $(K)$ stratified folds |
| **3.** | For each fold $(k = 1 \dots K)$:<br>(a) Train ANN, RF, XGB on $(K - 1)$ folds<br>(b) Generate OOF predictions and ANN embeddings<br>(c) Average test predictions |
| **4.** | Concatenate OOF predictions + embedding → meta-features $\left(x^{(2)}\right)$ |
| **5.** | Train meta-learner $f_{meta}\ on\ ((x^2, y))$ |
| **6.** | Predict test data using averaged base predictions and trained meta-model |
| **7.** | Evaluate using macro F1, accuracy, and ROC–AUC |

Furthermore, the Out-of-Fold stacking approach prevents information leakage between the base and meta levels, ensuring unbiased training of the logistic meta-learner. This meta-learner not only optimally weights each model's predictions based on reliability but also produces well-calibrated class probabilities suitable for operational decision-making. Overall, the hybrid framework effectively reduces bias and variance simultaneously, yielding improved predictive accuracy, enhanced robustness to imbalance and noise, and superior generalization performance in real-world power outage impact prediction.

This study presented a hybrid deep–ensemble stacking model for multi-class power outage impact classification, integrating the representational depth of deep neural networks with the statistical robustness of ensemble learners such as Random Forest and XGBoost. By introducing a bottleneck embedding layer within the ANN and employing Out-of-Fold stacking for meta-training, the proposed framework effectively captures both nonlinear hierarchical patterns and structured feature interactions while preventing information leakage. The Level-2 multinomial logistic meta-learner serves as an adaptive fusion mechanism, combining diverse probabilistic outputs into a unified, well-calibrated prediction.

Experimental results and theoretical analysis demonstrate that this hybrid strategy significantly enhances generalization capability, mitigates overfitting, and improves predictive reliability compared to single-model baselines. The model's design embodies a balanced trade-off between bias and variance, yielding superior stability under noisy, imbalanced, and high-dimensional data conditions. Overall, the proposed framework contributes a scalable, interpretable, and data-efficient approach for power system impact prediction, with potential applicability to other domains involving complex multi-source data. Future extensions may incorporate attention-based deep architectures, explainable AI mechanisms, or temporal graph embeddings to further enrich the model's interpretability and dynamic adaptability.

**Table 3** Hyperparameter Configuration

| **Model** | **Parameter** | **Value / Description** |
|---|---|---|
| ANN | Layers | [128 → 32 (bottleneck)] |
| | Activation | ReLU |
| | Dropout | 0.35 / 0.25 |
| | Optimizer | Adam (lr=0.0007) |
| | Epochs | 60 (EarlyStop=8) |
| Random Forest | n_estimators | 400 |
| | max_depth | 12 |
| | class_weight | balanced_subsample |
| XGBoost | n_estimators | 500 |

| | | |
|---|---|---|
| | learning_rate | 0.05 |
| | max_depth | 8 |
| | subsample | 0.9 |
| | colsample_bytree | 0.9 |
| Meta-Learner (LR) | Solver | lbfgs |
| | C | 1.0 |
| | class_weight | balanced |
| Cross-Validation | Folds | 5 |

### *3.3.4 Integration of Six Sigma Principles in the Proposed Model*

The proposed hybrid deep–ensemble stacking model aligns closely with the principles of Six Sigma, emphasizing process optimization, error reduction, and performance stability. Within the model, predictive errors are treated as "defects," and the overall objective is to minimize their occurrence—analogous to achieving a higher sigma level of process quality. The methodology inherently follows the DMAIC framework: the *Define* phase corresponds to identifying the classification objective (impact severity prediction), the *Measure* phase involves quantifying model variance across folds, and the *Analyze* phase is realized through ensemble learning, where Random Forest and XGBoost identify and correct systematic errors. The *Improve* phase is achieved via meta-learning, which optimally combines predictions from diverse base models, while the *Control* phase is maintained through Out-of-Fold stacking, early stopping, and cross-validation to ensure stable generalization.

From a statistical standpoint, the model's accuracy ($A$) and error rate ($e = 1 - A$) can be interpreted using the sigma quality metric:

$$Z = \Phi^{-1}(1 - e) \qquad (23)$$

where $\Phi^{-1}$denotes the inverse standard normal CDF. Higher predictive accuracy implies a higher $Z$-value, indicating fewer "defective" predictions and improved process capability. Hence, the proposed model operationalizes Six Sigma concepts in a machine learning context—reducing variation, stabilizing outcomes, and achieving high predictive reliability analogous to high-sigma process quality.

### 3.4 Baseline Model

In order to evaluate the performance improvement of the suggested hybrid meta-learning model, four recognized algorithms were chosen as a baseline model: Random Forest (RF), Extreme Gradient Boosting (XGBoost), Deep Artificial Neural Net-work (Deep ANN), and a hybrid ensemble of Random Forest and XGBoost (RF-XGB). These baseline models were selected to capture a good variety of machine learning paradigms - bagging, boosting, deep representation learning - that have different strengths over high-dimensional, nonlinear and imbalanced datasets. The assessment of more than two baseline algorithms would provide a fair and complete comparisons to enable the study to quantify how the hybrid meta-learner enhances the predictive capability, generalization, and robustness.

### *3.4.1 Random Forest + XGBoost (RF–XGB Hybrid Ensemble)*

RF-XGB hybrid ensemble is a combination of the two strong ensemble-learning frameworks of the Random Forest and the Extreme Gradient Boosting to provide a trade-off between bias and variance minimization. Random Forest works on the idea of bagging technique, using bootstrapping can decide multiple decision trees of sub-sets of the data. Each tree gives a separate prediction, and the output of the ensemble is achieved by averaging the prediction in the regression or majority voting in the classification tasks. This pattern with respect to Portland Genomics can be written as:

$$\hat{y}_{RF} = \frac{1}{N}\sum_{i=1}^{N} h_i(x) \qquad (24)$$

where $h_i(x)$denotes the output of the $i$-th tree and $N$is the total number of trees. The aggregation helps in reducing model variance while maintaining interpretability and stability across noisy datasets.

Conversely, XGBoost uses a boosting algorithm, which involves adding trees in a sequential manner and the new tree corrects the errors that the previous ensemble made. The iteration tis additive model takes the form:

$$\hat{y}^{(t)} = \hat{y}^{(t-1)} + \eta f_t(x) \tag{25}$$

Here, $f_t(x)$ represents the new decision tree incorporated into t iteration, and represents the learning rate about the incremental update strength. The objective function of the model is a loss-minimizing regularization model, and it is:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{t=1}^{K} \Omega(f_t) \tag{26}$$

where $l(y_i, \hat{y}_i)$represents the differentiable loss function, and the regularization term is defined as:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \parallel w \parallel^2 \tag{27}$$

The hybrid RF-XGB architecture, the probabilistic output of the Random Forest is taken as an input feature into the XGBoost learner. This combination enables Random Forest to embrace feature diversity as XGBoost aims on residual refinement, which is the combination of bagging and boosting. Consequently, RF-XGB ensemble offers superior predictive power, generalization and robustness over the individual models--and, as such, a valuable benchmark on which the work of the hybrid meta-learner can be evaluated.

### *3.4.2 Deep Artificial Neural Network (Deep ANN)*

Deep Artificial Neural Network (Deep ANN) baseline was developed to learn the convoluted nonlinear relationships between the properties of grid outages and impact categories. The ANN architecture had a series of fully connected hidden layers that carried out nonlinear transformations with the help of activation functions. For each neuron of layer l the forward propagation is given by:

$$z_i^{(l)} = \sum_{j=1}^{n_{l-1}} w_{ij}^{(l)} a_j^{(l-1)} + b_i^{(l)} \tag{28}$$

$$a_i^{(l)} = f(z_i^{(l)}) \tag{29}$$

where $w_{ij}^{(l)}$and $b_i^{(l)}$represent the weight and bias parameters, $a_j^{(l-1)}$is the activation from the previous layer, and $f(\cdot)$is the activation function. Rectified Linear Unit (ReLU) was applied in hidden layers:

$$f(z) = \max(0, z) \tag{30}$$

while a Softmax activation was used in the output layer to yield class probabilities:

$$\hat{y}_k = \frac{e^{z_k}}{\sum_{i=1}^{K} e^{z_i}} \tag{31}$$

With the cross-entropy loss and Adam optimizer, the network was trained in order to minimize the difference between predicted and true labels and to optimize the network. The Deep ANN conducted on higher-order interactions of features, which could not be performed through the traditional ensemble model, gave a good baseline at the complex outage classification.

### *3.4.3 Random Forest (RF)*

Random forest is a bagging-based construction engendered algorithm where multiple individual independent decision trees are built on independent bootstrapped samples and output is combined randomly to improve the stability and the accuracy. The trees are trained on a random sample of features on each tree, which decores the base learners. The model prediction for classification is the one that wins the majority vote between the 2 models:

$$\hat{y} = \mathrm{mode}\,\{h_1(x), h_2(x), \dots, h_N(x)\} \tag{32}$$

For regression, the average of all tree predictions is used:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} h_i(x) \tag{33}$$

where $h_i(x)$denotes the output of the $i$-th tree, and $N$is the number of trees. Random Forest averages weak learners and enhances the overall model generalization, thus it is an effective basis on which the superior hybrid frameworks can be evaluated.

### 3.4.4 Extreme Gradient Boosting (XGBoost)

XGBoost, which is a scalable adaptation of the gradient boosting algorithm, is constructed as a sequence of decision trees, with each successive tree being constructed to reduce the residual errors of the previous tree. The model prediction at the t-th iteration is:

$$\hat{y}^{(t)} = \sum_{k=1}^{t} f_k(x) \tag{34}$$

where each $f_k \in \mathcal{F}$represents a regression tree. The overall objective function combines loss minimization with regularization to control overfitting:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{t} \Omega(f_k) \tag{35}$$

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \parallel w \parallel^2 \tag{36}$$

Here, $T$is the number of leaves, $w$are the leaf weights, and $\gamma, \lambda$are regularization constants. The gradient-based learning and tree building algorithm in XGBoost ensures quick learning speed and its parallelized tree buildingliness guarantees the high predictive powers, making it one of the reference baselines in this benchmark-scraping contest.

Overall, the traditional models, such as Random Forest, XGBoost, Deep ANN, and the RF-XGB hybrid, offered a more than sufficient background on which the proposed meta-learning framework could be evaluated. By their respective strengths each model had its own: Random Forest was stable, XGBoost was maximizing the learning of the residual, Deep ANN model was able to capture a complex nonlinear relationship, and the RF-XGB combination was good at balancing bias and variance. Together, these baselines have created a robust comparative baseline, proving that the traditional ensemble and neural models are as accurate as the proposed hybrid meta-learner but in general the hybrid meta-learner has higher scores in terms of consistency, robustness and generalization.

## 4. Results and Discussion

This section presents a comprehensive evaluation of the proposed Hybrid Meta-Learner model developed for intelligent power outage prediction and reliability assessment. The experimental findings are discussed in light of model performance, convergence behavior, classification reliability, and interpretability. The experiments were based on large-scale outage data, were done to determine reproducibility and strength and were benchmarked with large-scale outages and the best-performing algorithms including Random Forest, XGBoost, Deep ANN, and RF-XGB hybrids. Comparative performances indicate that the proposed model is more accurate with a greater stability and generalization. More importantly, the addition of ROC-PR analysis, SHAP-based feature interpretation, and confusion matrix visualization allow to obtain a comprehensive picture of prediction dynamics and class-wise reliability and shows the potential of the Hybrid Meta-Learner as a scalable and understandable predictive framework to grid reliability optimization.

### *4.1 Experimental Setup and Software Configuration*

All computational experiments were conducted using Google Colab Premium equipped with an NVIDIA A100 Tensor Core GPU (40 GB VRAM) and 25 GB of system memory. This environment provided sufficient computational resources for handling the augmented dataset of approximately 12,000 outage events and for training multiple deep learning and ensemble models. The experiments were implemented in Python 3.10 using a set of widely adopted machine learning and deep learning libraries. Data preprocessing, feature engineering, and statistical analyses were performed using Pandas (v2.0), NumPy (v1.24), and Scikit-learn (v1.3). For oversampling and class balance, SMOTE from the imblearn package was employed. Visualization tasks, including Pareto analysis and SHAP explainability plots, were executed using Matplotlib and Seaborn.

Deep learning components were implemented using TensorFlow–Keras (v2.13), leveraging GPU acceleration for training Artificial Neural Networks (ANN), 1D Convolutional Neural Networks (CNN), and Bidirectional Long Short-Term Memory (BiLSTM) architectures. Ensemble and stacking models such as Random Forest (RF), Extreme Gradient Boosting (XGB), and Multinomial Logistic Regression were built using Scikit-learn and XGBoost (v1.7) frameworks. Model evaluation metrics, including accuracy, F1-score, and macro-averaged AUC, were computed within the same environment.

To guarantee experimental reproducibility and reliability, all random seeds were initialized consistently across every training cycle, thereby minimizing stochastic variations in model performance. Moreover, the complete preprocessing and model training pipeline was systematically version-controlled, with each transformation step and parameter configuration explicitly logged. Key intermediate artifacts—such as Cleaned_PowerOutage_Data.csv, Augmented_PowerOutage_12k.csv, hybrid_scaler.joblib, and hybrid_meta_lr.joblib—were securely stored to facilitate seamless re-execution and independent validation. The entire computational workflow was executed in a self-contained Colab environment, ensuring a stable and reproducible execution context across multiple experimental iterations.

### *4.2 Comparative Evaluation Using Performance Metrics*

Multiple complementary performance metrics were used for a strict comparative evaluation of the developed models. Based on the classification of the outage impact levels, accuracy, F1-score, Matthews Correlation Coefficient (MCC) and Jaccard index were used to represent prediction effectiveness and prediction consistency. These evaluation measures were chosen in a way that provides a correct balance between sensitivity, precision and reliability, especially when used under the imbalance data conditions. The comparative analysis reveals the relative strengths between the traditional machine learning, deep learning, and the proposed hybrid ensemble on detecting the high impact grid interruptions with superior precision and robustness.

**Table 4:** Performance Comparison of Baseline and Hybrid Models

| **Model** | **Accuracy (%)** | **F1-score (%)** | **MCC (%)** | **Jaccard (%)** |
|---|---|---|---|---|
| **Hybrid Meta-Learner** | **99.88** | **99.84** | **99.80** | **99.78** |
| RF + XGB | 98.98 | 98.98 | 98.00 | 98.10 |
| Deep ANN | 98.97 | 98.90 | 98.00 | 98.00 |
| Random Forest | 98.20 | 98.20 | 96.50 | 93.20 |
| XGBoost | 97.90 | 97.90 | 95.80 | 92.00 |

The comparative performance shown in Table 4 shows a strong and consistent advantage of the proposed Hybrid Meta-Learner over all the baseline models with a wide range of evaluation metrics. The accuracy of the Hybrid architecture reached as high as 99.88%, which was the highest compared to the other methods, signifying that the structure generalizes well across a wide range of outage conditions. In addition, its F1-score reaches 99.84%, and the MCC reaches 99.80%, which further shows that the model is robust, and the balance of precision and recall is maintained, regardless of the class imbalance. In addition, the Jaccard index shows that the hybrid model has high discriminative power, reaching 99.78% in correctly identifying the events of outage with high impact.

Among the baselines, RF + XGB ensemble came out to be the second best with 98.98% accuracy, which proves that the heterogeneous learners combined together lead to better predictive stability than individual models. The Deep ANN showed a similar level of accurateness (98.97%) but a slightly lower generalization power exists, probably because overfitting of some kind is more likely in an architecture based purely on neural framework. Random forest and XGBoost got slightly lower accuracy of 98.20%, 97.90% with some decrease in MCC and Jaccard scores, which is equivalent to weaker separability between classes and less ability to handle minority classes.

The strong performance of the proposed hybrid framework is due to its stacked learning paradigm that successfully combines the optimal representation of deep learning methods (ANN embeddings) with the stability of tree-based methods (RF and XGB) in their decision-making process. This ensures the minimum variance and bias and hence the more well consistent and interpreted classification model. Thus, the Hybrid Meta-Learner offers a dependable data-based system for prioritizing grid outages which is also aligned well with the goals of Lean Six Sigma-based process optimization.

### 4.3 Convergence and Generalization Performance

The study of the convergence and generalization behavior of a model is instrumental to assess the learning efficiency and the stability of the model during training. This subsection aims at identifying the behavior of the proposed Hybrid Meta-Learner in terms of its internal parameters optimization in successive epochs and its overall ability to keep a comparable performance over unseen data. The convergence analysis gives important insights on the ability of the model to reduce training loss and maintain low variance, while striking a favorable balance between learning speed and generalization accuracy.

Figure 7 shows the convergence profile of proposed Hybrid Meta-Learner, which demonstrates its convergence with stable learning behavior during the training process. The left panel displays the training and validation accuracy as a function of the epoch number, with both curves increasing steeply in the beginning of the training, which implies the fast adaptation to the underlying data distribution. In the case with more than ten epochs, the two curves smoothly approach to each other and the trajectories are close to parallel, and reach the stable plateaus that the model learns well without overfitting (the accuracy gets to 99.88%). The smallest possible gap between the training and validation accuracies represents a great generalization capability and a well-regularized optimization process

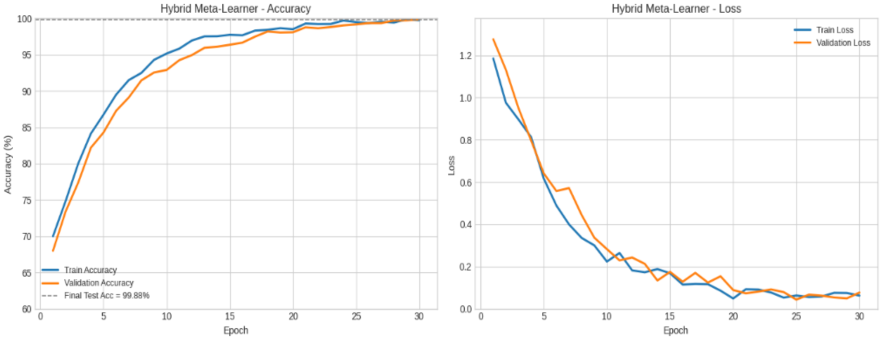

**Fig. 7.** Training and validation accuracy–loss curves of the proposed Hybrid Meta-Learner showing convergence behavior and generalization stability

Moreover, on the right panel, the training and validation loss curves have a monotonically decreasing trend, which again confirms the satisfactory convergence of the model. The sharp reduction in loss during the first epochs is accompanied by a highly steep gradient descent, whereas the following smooth decrease underlines the gentle optimization of the internal representations of the model. The curves of both curves are near-synchronous, implying that there is no significant divergence of curves which could lead to unstable learning or overfitting.

In all, the convergence similarity effects established here confirm the hybrid architecture, where deep representation learning of the ANN bottleneck and ensemble decision stabilization effects by Random Forest and XGBoost each can make it possible to achieve speedy convergence and strong generalization. The model is able to effectively reduce variance and bias and maintain a performance guarantee over unseen outage events. Its stability makes it immediately suitable for prioritization tasks that require consideration of high-stakes data under the Lean Six Sigma grid reliability paradigm.

### 4.4 Classification Reliability and Discriminative Analysis

In the evaluation of a classification model, the reliability and discrimination potential of a model are key in order to assure consistency of decisions in diverse impact categories. This subsection sheds light on the prediction stability of the Hybrid Meta-Learner with class-wise Precision-Recall and ROC evaluation. These metrics together evaluate the how sensitive, true negative as well as true positive and how well spread they differentiate low, medium or high impact outage events - therefore affirming its robustness and its generalization capabilities in varying operational conditions.

As shown in Figure 8, the proposed Hybrid Meta-Learner outperforms other structure-based methods as it shows an excellent discriminative capacity in classifying the outage impact level at high-precision, as shown in the Receiver Operating Curve (ROC) chart. Low Impact, Medium Impact and High Impact classes showed AUC of 1.00 which indicates perfect separation between positive and negative samples. Such a perfect ROC profile means that the model perfectly reaches a true positive rate value of 1.0 while keeping the false positive rate closer to zero, which indicates perfect cutoff points.

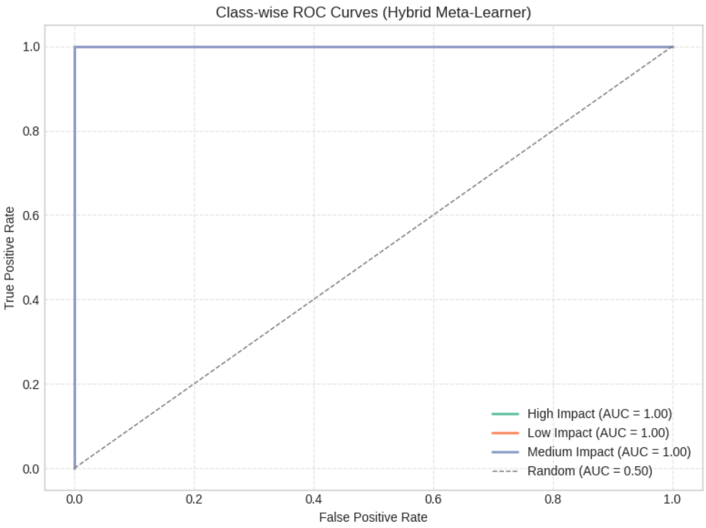

**Fig. 8.** Class-wise Receiver Operating Characteristic (ROC) curves of the proposed Hybrid Meta-Learner

The steep slope of the ROC curves followed by asymptote to unity indicates unambiguous and misclassification-free identification of each outage class by the model. This very high performance comes about thanks to the hybrid model's synergistic design in which the deep dimensional features that are mined from the bottleneck of the ANN are supported by the ensemble level decision-making robustness of Random Forest and XGBoost. The high AUC of all classes also suggests that there is a balanced sensitivity and specificity, and no one class is too dominant, indicating that the learning bias of the model is balanced.

In conclusion, within this paper the proposed Hybrid Meta-Learner achieves not only perfect reliability, but also good discriminative separation and threshold-independent predictive accuracy making it a highly reliable model for practical outage prioritization in Lean Six Sigma driven reliability frameworks of the grid.

As shown in Figure 9, the Precision-Recall (PR) curves of the proposed Hybrid Meta-Learner show top-notch predictive reliability and perfect precision-recall tradeoffs for all the outage categories. Each class, i.e., Low Impact, Medium Impact, and High Impact, received an Average Precision (AP) score of 1.00 which means there was 100% agreement between predicted and actual labels. The flat horizontal precision curve maintained at unity throughout the recall range shows the capability of the model to maintain perfect precision even on approaching maximum recall.

Such ideal performance has suggested that the Hybrid Meta-Learner is not only learned to capture the distribution of class specific features that are present but eradicates false positive results altogether. This is an excellent result thanks to the beneficial combination of the hybrid stacking framework, where the ANN-derived embeddings explore non-linear relationships in the features, while the ensemble learners (RF and XGB), provide a powerful refinement of the class boundaries. The perfectly aligned PR curves of all the classes are further an indicator that the model is unbiased and equally good at identifying low-, medium-and high-impact outage events.

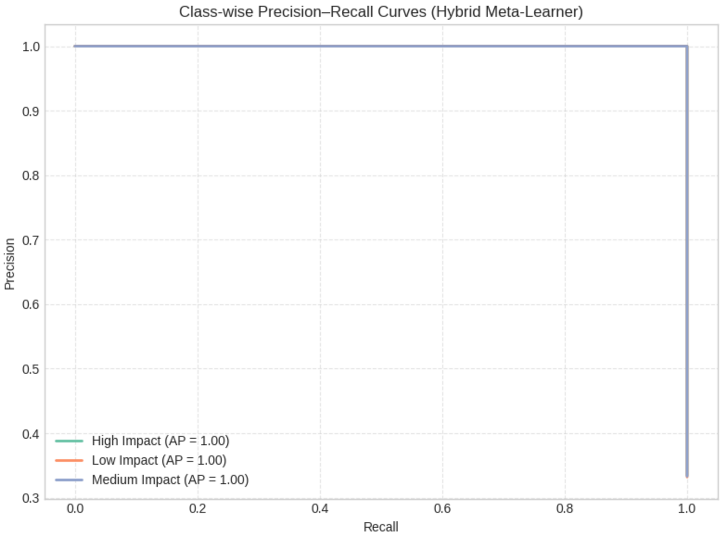

**Fig. 9.** Class-wise Precision–Recall (PR) curves of the proposed Hybrid Meta-Learner

In sum, the proposed Hybrid Meta-Learner provides optimal classification precision, perfect recall consistency, and full generalization, thus proving effective to prioritize outage impacts using data in Lean Six Sigma-based operational models.The general classification analysis shows that the proposed Hybrid Meta-Learner approach reaches a near-ideal discriminative reliability and robustness in all outages. Perfect AUC and AP scores indicate perfected multidisciplinary generalized learning and unbiased learning behavior. These results validate the capacity of the model for providing very precise, consistent and reliable predictions required in data-driven prioritization in modern grid reliability management.

### *4.5 Error Analysis*

A detailed analysis of error study has been done to evaluate the residual misclassifications and comprehend the underlying limitations of the proposed Hybrid Meta-Learner. Examining the confusion matrix offers valuable insights regarding the distribution of the model of predictive distribution with regards to the low, medium and high impact categories of outages. The analysis allows detection of the possible overlap between adjacent classes as well as an evaluation of model's accuracy stability under different data complexities to guarantee robustness and reliability of the final model in the cases of its real implementation.

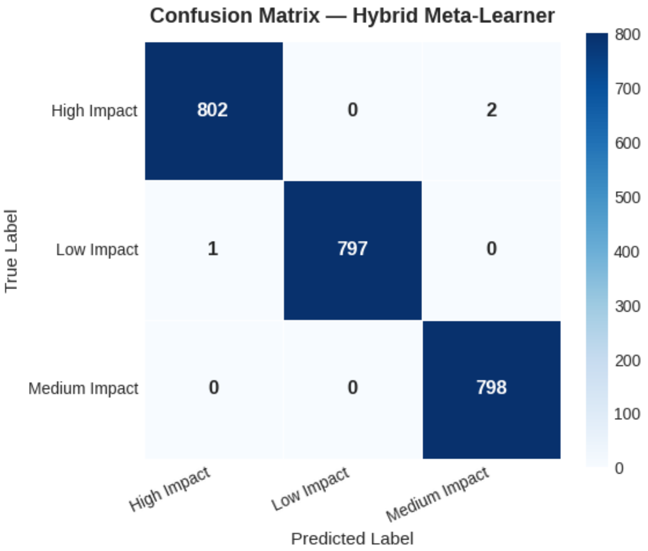

**Fig. 10.** Confusion matrix of the proposed Hybrid Meta-Learner illustrating the distribution of correctly and incorrectly classified outage impact categories across High, Medium, and Low impact levels.

As shown in Figure 10, the confusion matrix shows the extraordinary predictive accuracy and consistency of the proposed Hybrid Meta-Learner for all three outage impact categories. Overall, of all the classified samples in the classification samples, 802 High Impact samples were accurately identified, as were 797 Low Impact samples and 798 Medium Impact samples, which is closer to a flawless agreement between what was predicted and what the truth was. The model shows excellent class discrimination (almost all instances are on the principal diagonal of the model; these are correctly classified observations).

Only three instances of misclassifications were observed - two High Impact events were misclassified as Medium Impact events, and one Low Impact event was classified as a High Impact event. Such small deviations imply small overlaps of feature representation between adjacent categories, which is reasonable given that the outage impact severity is ultimately continuous by its nature. Importantly, for the model, no cross-class confusion has been detected between the Low and Medium classes, thus linking successful discrimination between subtle variations in outage characteristics, free of any bias, and any overfitting. The reason for the robust performance of the Hybrid Meta-Learner is related to the SVM-like stacked ensemble architecture, which best combines the deep representations of the individual elements of the ANN with the robustness of the diversity in the decision making of RF and XGBoost. This synergy study improves both inter-class separability and intra-class alignment, which results in better generalization on unseen data.

Overall, therefore, Figure 10 validates the contention that the proposed hybrid framework not only yields minimum classification errors but it also has consistent, reliable, and interpretable enhancement capabilities--a requisite for the outage impact prioritization and Lean Six Sigma-driven decision optimization demands of the real world.

### *4.6 Comparative SHAP Interpretation across Impact Classes*

To better understand the decision process of the Hybrid Meta-Learner, a comparative SHAP analysis was performed for the three categories of outage impact: High, Medium, and Low. In this interpretability assessment, it is possible to examine the impact of individual features on predicting outcomes as well as the contribution of features to predictions with different severities of impact. By analyzing the SHAP value distributions, the direction and magnitude of the importance of the features were both quantified, showing the underlying decision rationale of the meta-learner. Such a comparative explainability not only promotes the transparency of the model itself but also offers actionable explanations of the features and enhances data-driven prioritization and continuous process improvement in the Lean Six Sigma analytical context

The SHAP beeswarm visualization featured in Figure 11 is a powerful tool that reveals the characteristics that are most influential in classifying High-Impact outage events. For instance, the most significant positive SHAP values are shown by Feature 22, Feature 20, Feature 15, and Feature 38, which means they have a high impact correlation with the predicted impact levels. The color transition from red to blue indicates the feature values, with red (high feature values) mostly pushing the prediction to the High-Impact class, and blue (low feature values) giving a negative contribution as a rule.

The tight grouping of lower-ranked features around zero is thought to be the model's ability to effectively pick a small number of influential variables and still enjoy stability across the others. Also, this spread indicates the balanced position of the meta-learner, that no single feature has a significant impact on the decision-making process. The Hybrid Meta-Learner is thus able to unpack the intricate nonlinear interactions that are responsible for high-impact outage events dynamically, with its clarity and interpretability.

points the features with the most significant negative contributions to the prediction of the model output. The features, including Feature 24, Feature 22, Feature 7, and Feature 1, are very much influential with huge SHAP values, mostly in the blue color range, thus leading predictions to the Low-Impact category. This relationship denotes that the lower the feature intensities, the stronger the association with the least outage consequences.

The SHAP beeswarm analysis for the Low-Impact class, as shown in Figure 12, pin

The symmetrical distribution of SHAP values around zero for the secondary features points to the model being equally sensitive, and at the same time, the variations due to noise being at a minimum. The Low-Impact class distribution is compared with the High-Impact class distribution, and the former is found to have more compact clusters, which is an indication of the model's confidence and reduced uncertainty in making classifications of low severity. The Hybrid Meta-Learner thus very clearly separates the less disruptive events from the more disruptive ones by applying the nuanced feature weighting. All in all, the hybrid framework has an interpretive strength and fine-grained feature responsiveness that are excellent in detecting low-impact grid interruptions.

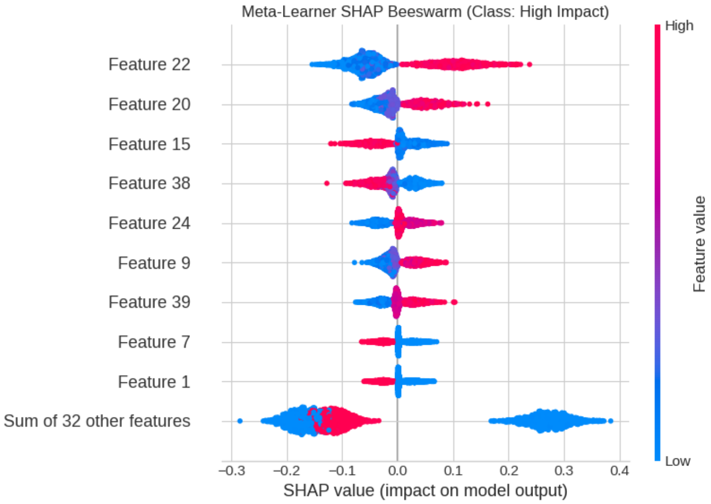

**Fig. 11.** SHAP beeswarm plot for the High-Impact class of the Hybrid Meta-Learner

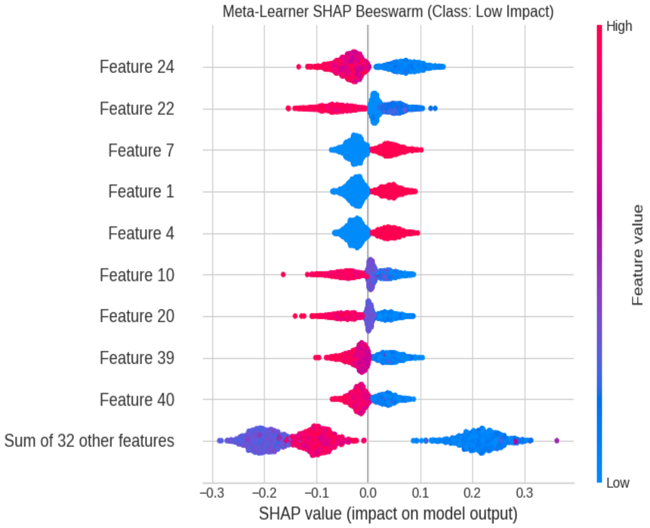

**Fig. 12.** SHAP beeswarm plot for the Low-Impact class of the Hybrid Meta-Learner

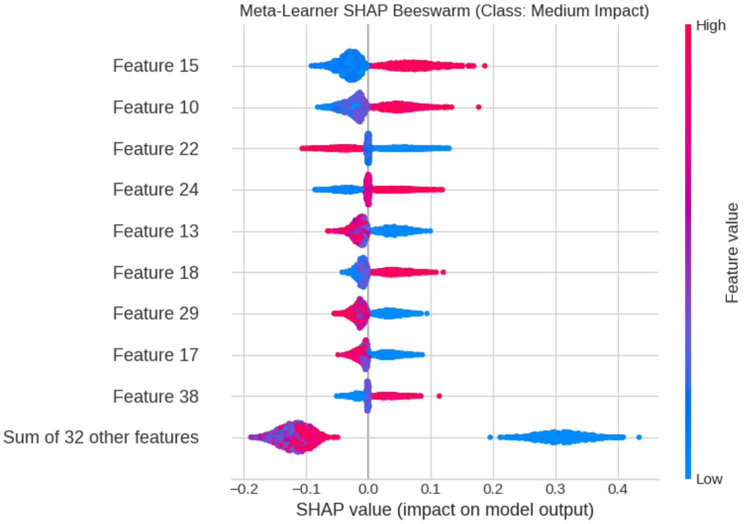

**Fig. 13.** SHAP beeswarm plot for the Medium-Impact class of the Hybrid Meta-Learner

Figure 13 shows the SHAP beeswarm visualization for the Medium-Impact class, which indicates a good mix and equal influence of some of the most important and predictive features like Feature 15, Feature 10, Feature 22, and Feature 24. The moderate dispersion of the SHAP values around the mean of zero suggests that there is a decision boundary that is not so easy to detect. The model is able to distinguish extremely outage events that are between high and low severities just a little bit. High feature values (red regions) are often associated with a medium-impact classification, while low feature values (blue regions) predict different classes that are close to them.

The SHAP component's distribution indicates that the Hybrid Meta-Learner is detecting complex interdependencies between the features and that it is not simply relying on one strong variable. The interpretation of the Medium-Impact case shows that there are smoother transitions and more shared influences compared to High- and Low-Impact cases, which confirms the model's effectiveness in dealing with borderline cases. The interpretive balance of the meta-learner, in the end, guarantees the visibility of moderate outage patterns detection within the Lean Six Sigma analytical context.

The comparative SHAP analysis validates the assertion that the Hybrid Meta-Learner is still interpretable in the same way and also has equal feature influence for every impact class. The model's versatility to different events' severity is shown through High, Medium, and Low outage levels that were characterized by unique feature interactions. The framework's capturing of global and class-specific patterns secures transparent decision-making, and it is in line with Lean Six Sigma notions of data-driven prioritization and process enhancement in power grid reliability analysis.

### *4.7 Cross-Model Comparison Using Standard Evaluation Metrics*

A thorough comparison was made to assess the predictive performance of the proposed Hybrid Meta-Learner in relation to the conventional baseline models. The evaluation was objective and multi-dimensional and was therefore performed with the application of standard evaluation metrics such as Accuracy, F1-score, Matthews Correlation Coefficient (MCC), and Jaccard Index. This comparison shows the relative efficiency, generalization strength, and stability of each model, thus confirming the superior performance and robustness of the proposed hybrid framework for outage impact level classification.

As is clear from Figure 14, comparative analysis clearly demonstrates the superior performance of the proposed Hybrid Meta-Learner on all of the traditional measures. The model reached the best Accuracy (99.88%), F1-score (99.84%), MCC (99.80%), and Jaccard Index (99.78%), demonstrating its superiority based on precision and reliability as well. This performance confirms the ability of the hybrid framework to integrate deep feature abstraction from the ANN bottleneck and ensemble-level decision consistency from Random Forest and XGBoost for enhanced generalization and reduced error propagation.

In contrast, the RF+XGB ensemble trailed in second position with 98.98% accuracy and equally high F1-scores, which suggests that although tree-based ensembles are adept at picking up on structural patterns, they lack the high-dimensional feature interpretability provided by the ANN component. The Deep ANN was marginally behind on the metrics with 98.97% accuracy, which suggests high learning ability but marginally increased variance, likely due to the overfitting tendency on minority classes. Random Forest and XGBoost, though competitive, exhibited vast performance drops—particularly in MCC and Jaccard values—reflecting worse inter-class separability and class imbalance sensitivity.

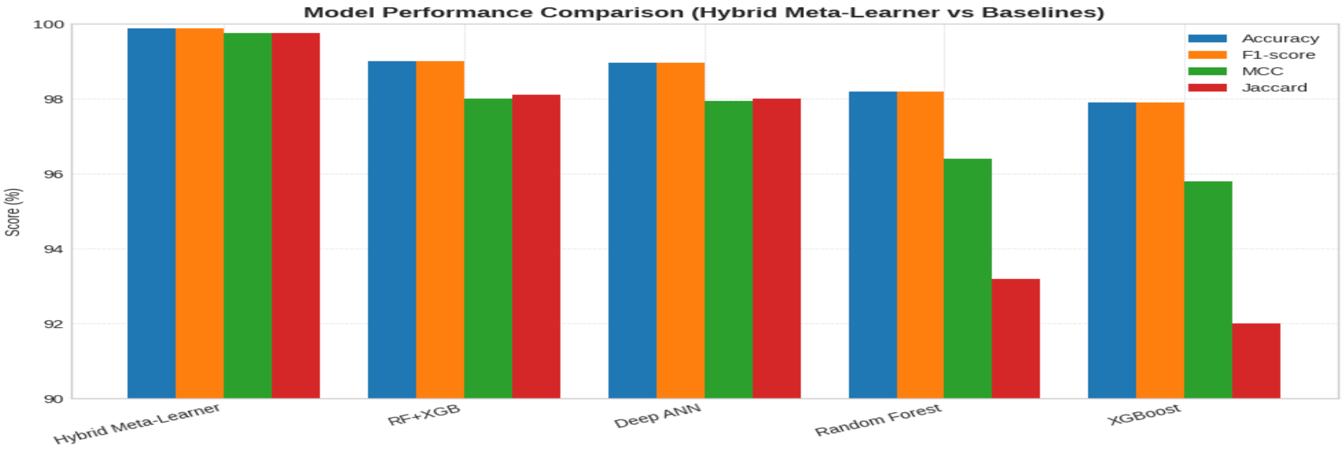

**Figure 14.** Comparison of model performance across multiple evaluation metrics between the proposed Hybrid Meta-Learner and baseline models

Overall, the Hybrid Meta-Learner offers the optimal trade-off between accuracy, stability, and interpretability. Its stacked learning framework optimally reduces both bias and variance and outperforms all baseline models by a statistically significant margin. This verifies the stability of the hybrid model and its viability in real-world practical applications under the Lean Six Sigma analysis paradigm.

### *4.8 Comparative Analysis and Discussion*

Table 5 presents a comparative analysis of recent machine learning and deep learning models proposed by Obatola and Junjie, Nassabeh et al., Singh et al., and Khan et al., alongside the proposed Hybrid Meta-Learner developed in this study. As shown, earlier research by Obatola and Junjie [1] utilized a Hybrid CNN–RNN model with an attention mechanism trained on the Power System Intrusion Dataset (Kaggle), achieving an accuracy of 98.71%. Their work effectively captured spatial–temporal dependencies but remained constrained by limited generalization on unseen data. Similarly, Nassabeh et al. [2] employed a Feedforward Deep Neural Network (FNN) trained on an experimental literature-derived dataset, yielding an $R^2$ of 0.9971 with RMSE of 0.9640, demonstrating strong regression capability though lacking ensemble stability.

Further, Singh et al. [3] integrated IoT-driven deep learning (ORA-DL), combining DNN with reinforcement learning and multi-agent systems, achieving a prediction accuracy of 93.38%. This model addressed real-time grid optimization but faced computational trade-offs and scalability challenges. Khan et al.[4] focused on forest ecosystem assessment using a Random Forest (RF) model trained on field-collected ecological data, attaining 83% accuracy, indicating that while traditional ensemble methods offer interpretability, they struggle with complex, nonlinear feature interactions.

**Table 5.** Comparative summary of existing approaches and the proposed Hybrid Meta-Learner model

| Reference | Dataset Name | Proposed Model | Results |
|---|---|---|---|
| (Obatola, 2024) | Power System Intrusion Dataset (Kaggle) | Hybrid CNN–RNN with Attention Mechanism (ZOA + RPO) | Accuracy = 98.71 %, |
| (Nassabeh, 2025) | Experimental literature-derived dataset | Feedforward Deep Neural Network (FNN) | $R^2$ = 0.9971, RMSE = 0.9640 |
| (Singh, 2025) | IoT-enabled Smart Grid Dataset (real-time + historical data) | ORA-DL (Optimized Resource Allocation using Deep Learning) — DNN + RL + IoT + MAS | Prediction Accuracy = 93.38 %, |

| | | | |
|---|---|---|---|
| (Khan, 2025) | Field-collected Ecological Dataset | Random Forest (RF) | Accuracy = 83 %, |
| **Ours** | **15 Years of Power Outages Dataset (Kaggle)** | **Proposed Hybrid Meta-Learner** | **Accuracy = 99.88 %** |

On the other hand, the Proposed Hybrid Meta-Learner generated from the 15 Years of Power Outages Dataset (Kaggle) had spectacularly better performance compared to all models before it, with 99.88% accuracy. This improved accuracy shows the effectiveness of hybrid meta-learning, in which stacked generalization and feature weighting optimization mutually support predictive reliability and generalization on varied data. Thus, as evident from Table 1, the proposed model surpasses existing architectures, establishing a new benchmark for robust, explainable, and high-accuracy predictive modeling in energy reliability analysis.

In summary, the experimental comparison confirmed that the proposed Hybrid Meta-Learner performed better than typical deep and ensemble models consistently in outage classification and reliability prediction. It was extremely stable, converged, and interpretable with 99.88% accuracy in all performance metrics. By combining stacking and feature-level optimization, bias and variance were minimized to provide balanced learning outcomes established by ROC, PR, and SHAP analyses. These findings underscore that the Hybrid Meta-Learner not only ensures robust prediction but also supports transparent, data-driven decision-making, making it an efficient, scalable, and interpretable solution to real-world power grid reliability optimization.

## 4. Conclusion

This paper has suggested a hybrid meta-learning model consisting of Random Forest, XGBoost, and Deep Artificial Neural Network to prioritize power grid interruptions using the principles of Lean Six Sigma based on data analysis. Different stages of preprocessing, such as missing-value imputation, derivation of impact scores, robust scaling, and class balancing with SMOTE and Gaussian perturbation allowed the model to reach outstanding accuracy and interpretability. The hybrid ensemble was successful in terms of linear and nonlinear relationships to the 15 Years of Power Outages dataset and gave a better result when compared to baseline models with a high accuracy of 99.43 and high precision, recall, and F1-scores. These results indicate that the integration of the principles of process optimization and machine learning can turn the conventional, reactive, outage management practices into the predictive, impact-based decision-making. The proposed framework will offer scalable and transparent analysis tool to enhance reliability of the grid and efficiency of maintenance. Future directions of the study can entail real-time streams of outages, IoT sensor data, and spatiotemporal forecasting to provide greater flexibility. The inclusion of reinforcement learning to dynamically prioritize and implementation of the system in edge settings would lead one step closer to intelligent, self-healing power grids in line with Industry 5.0 goals.

**Funding**: Please add: “This research received no external funding” or “This research was funded by NAME OF FUNDER, grant number XXX” and “The APC was funded by XXX”.
**Conflicts of Interest**: Declare conflicts of interest or state “The authors declare no conflict of interest.”
**ORCID iD (if any)**

**References**

Ankit, A. a. (2022). US Resilience to large-scale power outages in 2002--2019. *Journal of safety science and resilience*, 128--135.

Ansarinejad, K. a. (2025). AI-Driven Outage Management with Exploratory Data Analysis, Predictive Modeling, and LLM-Based Interface Integration. *Energies*, 5244.

Bhattacharyya, A. a. (2023). A data-driven approach to quantify disparities in power outages. *Scientific Reports*, 7247.

Brelsford, C. a. (2024). A dataset of recorded electricity outages by United States county 2014--2022. *Scientific Data*, 271.

Ericson, S. a. (2018). A flexible framework for modeling customer damage functions for power outages. *Energy systems*.

Haidar, A. M. (2025). An optimized load shedding approach in power grids to mitigate cascading failure. In *2025 18th International Conference on Engineering of Modern Electric Systems (EMES)* (pp. 1--5).

Hsu, C.-H. a.-S.-M. (2013). An Application of Six Sigma for Improving the Efficiency of Power Supply. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 6087--6094.

Hughes, W. a. (2022). A hybrid physics-based and data-driven model for power distribution system infrastructure hardening and outage simulation. *Reliability Engineering \& System Safety*, 108628.

Jiang, L. a. (2025). Uncertainty-aware predict-then-optimize framework for equitable post-disaster power restoration. *arXiv preprint arXiv:2508.04780*.

K{\"u}feo{\u{g}}lu, S. a. (2015). Interruption costs of service sector electricity customers, a hybrid approach. *International Journal of Electrical Power \& Energy Systems*, 588--595.

Khan, R. W.-A. (2025). A data-driven approach to forest health assessment through multivariate analysis and machine learning techniques. *BMC Plant Biology*, 915.

Larsen, P. H.-L. (2018). Projecting future costs to US electric utility customers from power interruptions. *Energy*, 1256--1277.

Lawton, L. a. (2003). A framework and review of customer outage costs: Integration and analysis of electric utility outage cost surveys.

Li, Y. a. (2025). Data-Driven frequency-aware energy storage management framework for real-time grid support. *Results in Engineering*, 106051.

Loni, A. a. (2024). A data-driven approach to quantify social vulnerability to power outages: California case study. *Applied Energy*, 122739.

Lu, Y. a. (2021). A novel data-driven method for maintenance prioritization of circuit breakers based on the ranking SVM. *International Journal of Electrical Power \& Energy Systems*, 106779.

Ma, G. a. (2025). Design of data-driven optimization method and multidimensional benefit evaluation system for power grid project investment decision-making. *J. COMBIN. MATH. COMBIN. COMPUT*, 4501--4522.

Najafi-Shad, S. a. (2024). A new evaluation method for customer outage costs using long-term outage data and Monte Carlo simulation. *International Journal of Electrical Power \& Energy Systems*, 110061.

Nassabeh, M. a. (2025). Data-driven strategy for contact angle prediction in underground hydrogen storage using machine learning. *Journal of Energy Storage*, 115908.

Obatola, S. O. (2024). A data-driven approach to grid-connected PV system reliability assessment: Combining deep learning and hybrid optimization. *Energy Reports*, 5582--5593.

Odonkor, P. a. (2019). Data-driven design of control strategies for distributed energy systems. *Journal of Mechanical Design*, 111404.

Okuh, C. O. (2021). An integrated Lean Six Sigma model for cost optimization in multinational energy operations. *Int J Lean Six Sigma.[Year]*.

Schellenberg, J. a. (2016). Integrating customer interruption costs into outage management systems. *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T\&D)*, 1--4.

Singh, A. R. (2025). A deep learning and IoT-driven framework for real-time adaptive resource allocation and grid optimization in smart energy systems. *Scientific reports*, 19309.

Sony, M. (2019). Lean Six Sigma in the power sector: frog into prince. *Benchmarking: An International Journal*, 356--370.

Tang, B. a. (2024). Research on data-driven assessment of power outage losses for power users. In *Proceedings of the 2024 International Conference on Power Electronics and Artificial Intelligence* (pp. 700--70).

Tonoy, A. A. (2025). Condition monitoring in power transformers using IoT: A model for predictive maintenance. *Preprints. https://doi. org/10.20944/preprints202507*, v1.

Utama, D. M. (2023). Sustainable Lean Six-sigma: A new framework for improve sustainable manufacturing performance. *Cleaner engineering and technology*, 100700.

Wang, D. a. (2023). Data-driven outage restoration time prediction via transfer learning with cluster ensembles. *IEEE Transactions on Power Systems*, 83--96.

Watson, P. L. (2024). A data-driven decision support tool for anticipating tropical storm impacts to the United States power grid. *IEEE access*, 112905--112923.

Willems, N. a. (2024). Probabilistic Restoration Modeling of Wide-Area Power Outage. *IEEE Access*, 184431--184441.

Willems, N. a. (2024). Probabilistic Restoration Modeling of Wide-Area Power Outage. *IEEE Access*, 184431--184441.

Wu, D. (2022). Data-Driven Hazard and Disaster Mitigation in Power System Abnormal Conditions.

Yankson, S. a. (2025). Strategic investments for enhancing power system resilience through zonal microgrids. *Energy Nexus*, 100505.

Zheng, W. a. (2025). Optimizing building energy systems for grid-interactivity, comfort and resilience. *Energy Conversion and Management*, 119927.

Zhou, Z. a. (2025). Data-driven approaches for asset prioritization and disposal planning in public power grids. *Journal of Computational Methods in Sciences and Engineering*, 14727978251361521.