EAI Endorsed Transactions
on Energy Web

Research Article EAI.EU

Technology for Power Outage Research and Judgment-dependent Data Feature Noise Analysis

Xiang Li[1,*]

Yunnan Power Grid Co.,Ltd, Kunming 650000, Yunnan, China

Abstract

INTRODUCTION: Power grid blackouts occur frequently, which significantly impacts social impact. Because these accidents are dynamic and random, predicting and evaluating them is challenging.
OBJECTIVES: To explore the complexity of the power grid itself, analyzes the critical changes of the self-organizing model during power grid fault, extracts the data characteristics related to the steady-state maintenance of abnormal systems, and puts forward an effective outage prediction model.
METHODS: Starting with cluster analysis, The authors can reduce data fluctuation and eliminate noise interference to optimize data. The evaluation indexes of initial fault occurrence possibility and fault propagation speed in the power grid are constructed.
RESULTS: The validation of the outage forecasting model has produced promising results, achieving 96.4% forecasting accuracy and a meager error rate. In addition, the evaluation index developed in this study accurately reflects the possibility and spread speed of power outage accidents.
CONCLUSION: The research proves the feasibility of establishing an outage prediction model based on the power grid system data characteristics. The model has high accuracy and reliability and is a valuable tool for power outage research and judgment.

Keywords: power outage research and judgment, data characteristics, noise analysis

Received on 15 January 2023, accepted on 16 September 2023, published on 22 September 2023

doi: 10.4108/ew.3949

[*]Corresponding Author. Email: 18208891009@163.com

1. Introduction

As the power grid continues to expand and evolve, its security and reliability demands are steadily rising. Power outages represent one of the most prevalent issues in the power system, significantly impacting grid operations. As a result, the exploration, analysis, prediction, and control of power outage incidents are crucial research areas in the power network domain. These endeavors play a pivotal role in ensuring the consistent and stable operation of the power grid while enhancing user satisfaction. Among these endeavors, accurately and swiftly identifying the causes and extent of power outages holds immense practical significance.

2. Model analysis

2.1 Literature review

Introduction:
Power grid faults and outages have been subjects of extensive research, with various studies exploring predictive and analytical approaches. Fan Min et al. (2023) investigated distribution network fault-related power outages, identifying root causes based on outage fault data to enhance predictive models. They developed a random cost-sensitive Convolutional Neural Network

(CNN) using Mo Xingguo's approach and applied random sampling with replacement to mitigate false positives in abundant average data [1]. Li Guoqing et al. (2022) devised a prediction model for node outage risk in the new energy grid. They assessed power outage risks in the power grid system through real-time grid system pressure prediction and conducted case studies to validate model effectiveness [2]. Nan Dongliang et al. (2021) constructed a data-driven prediction model for distribution network power outages. This model balanced original data via k-means clustering and employed the Adaboost algorithm for classification, significantly enhancing predictive accuracy [3]. Yu Qun et al. (2018) delved into the autocorrelation of power grid blackout incidents, analyzing decades of power outage data and identifying long-range autocorrelations within the loss load sequence. Their research facilitated time-series predictions of blackout incidents and associated loss loads [4].

These studies have tackled the issue of power outages in distribution networks from diverse angles and dimensions[5]. They have explored fault analysis and modeling using outage data, risk prediction based on new energy grid nodes, and data-driven analysis for distribution networks. By deconstructing power grid fault outage prediction models, they have offered effective solutions through various methodologies and technical approaches. While these studies have made significant theoretical and practical contributions, many models still grapple with false optimistic predictions stemming from inadequate sampling of power failure occurrences and an overabundance of standard samples[6].

In light of these challenges, this paper aims to explore the cascade effect of widespread faults in predictions from the perspective of self-organized criticality theory.

2.2 Self-organized criticality theory

The theory of self-organized criticality posits that many complex systems naturally exist in a critical state, where the system's behavior undergoes abrupt changes or phase transitions, exhibiting features such as instability, diffusion, and long-range correlation. These characteristics serve as essential foundations for the self-organization and fractal patterns observed in the system. Power outage accidents, shaped by the interplay of multiple factors within the power grid, exemplify the typical traits of self-organized critical state transformations[7]. This article builds upon this analysis to mine the data characteristics within the grid system.

Self-organization manifests in the similarity and scale invariance of a system's properties across different time scales. Leveraging these phenomena, researchers have equated mathematical models and computational methods to simulate and predict a wide array of natural and societal occurrences. Some scholars have examined the self-organized criticality of power grid fault outages in prior research. For instance, Yu Qun and Guo Jianbo (2006) systematically organized data on significant power outages in China's power grid, delving into their self-organized criticality traits[8]. Xinyao L et al. (2014) introduced a novel cascading fault prediction methodology rooted in self-organized criticality principles, dissecting the typical progression of cascading faults[9]. Drawing from the non-uniform self-organized criticality characteristics, they proposed an assessment index for estimating the overall self-organized critical line. This index incorporates state and structural transformations following power failures, enabling the prediction of cascading failures within clusters of self-organizing essential lines based on a composite margin index.

This paper endeavors to uncover and analyze the inherent self-organized criticality within power grid outages, shedding light on data characteristics that can enhance our understanding and predictive capabilities in managing grid system disruptions.

2.3 Distribution of blackout accidents

To analyze power grid fault outage incidents through the lens of self-organized criticality theory, the initial step involves organizing the historical time series data about these outages. In this context, the study conducted by Yu Qun and Guo Jianbo (2006) presented critical information concerning major power outages, including factors such as load loss and frequency. Upon meticulous organization and categorization of this raw data, the following results emerged:

Table 1. Statistics on the scale of power outage loss load in China's power grid

power outage loss load (MW)	quantity	frequency
100＜Q＜200	96	32.0%
200＜Q＜300	74	24.7%
300＜Q＜400	40	13.3%
400＜Q＜500	30	10.0%
500＜Q＜600	21	7.0%
600＜Q＜700	15	5.0%
700＜Q＜800	13	4.3%
800＜Q	11	3.7%

Among these findings, it is worth highlighting that R-squared (R2) equals 0.9696, indicating a notably improved predictive performance of the model. This result substantiates the observed patterns in accident frequency and underscores their regularity.

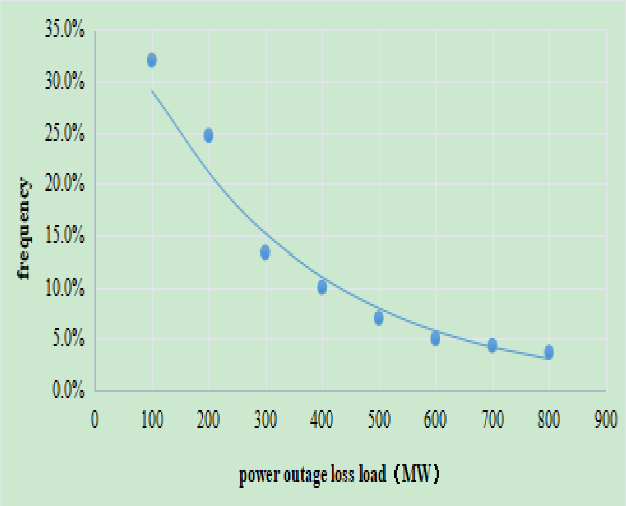

Figure 1. The fitting results of the scale statistics of power outage loss load in ChinaThe appropriate results of the scale statistics of power outage loss load in China

This data can be further corroborated by examining blackout occurrences at the scale of power system blackout events[10]. When assessed on this broader observational scale, the frequency of blackout phenomena exhibits notable stability[11]. This validation reinforces the efficacy of the analysis conducted using the overall data characteristics of the system.

3. Model Construction

3.1 Introduction of Initial Failure

Power grid fault blackouts often originate from specific initial fault causes. Let's assume there are multiple types of responsibility causes and the probability of the i-th fault cause occurring is denoted as P_i^a . If there are b lines, the probability of the j-th line failing is represented as P_j^b , and the occurrence probability of the *i* type of failure is P_{ij}^b . If the line j is further divided into c sections, then the probability P_k^c Of faults occurring in section k is:

$$P_k^c = \frac{f_k}{\sum_{k=1}^{c} f_k} \tag{1}$$

Now, assuming the existence of an incompatible fault type d, where the probability of occurrence of the s type of fault is P_s^d, there are:

$$\sum_s^d P_s^d = 1 \tag{2}$$

Consequently, the probability of fault s occurring in segment k of line j, caused by the i-th fault cause, is computed as:

$$\mathrm{P}(i,j,k,s) = P_i^a P_j^b P_k^c P_s^d \tag{3}$$

This model establishes the essential simulation state based on the initial failures' probabilities. It can analyze the likelihood of failure at each key stage resulting from various shortcomings and their chain reactions. Cascading failures, induced by multiple random factors, lead to overall failure and power outage incidents.

3.2 Fault creep assessment

The model constructs initial faults using the aforementioned random process and designs multiple fault propagation paths based on it to analyze propagation speeds on specific lines. These propagation paths are defined as:

$$\mathrm{T} = \frac{1}{\mathrm{N}(\mathrm{E_{CAS}})} \tag{4}$$

Here$\mathrm{N}(\mathrm{E_{CAS}})$, represents the number of sequential circuit breaks in the specific fault-cascading process. Simultaneously, the model also considers the extent of fault spread, i.e., the proportion of skipped lines in the total power grid, as follows:

$$\mathrm{S} = \frac{\mathrm{NUM}(\mathrm{E_{CAS}})}{\mathrm{N}} \tag{5}$$

Here $\mathrm{NUM}(\mathrm{E_{CAS}})$, is the number of line jumps in a specific fault cascading process, and N is the number of lines in the power grid.

The overall characteristics of fault development can be established by assessing the original fault and propagation scenario, which allows for distinguishing different fault propagation processes and analyzing the fault propagation path leading to power outage incidents[12].

However, during this analysis, the model encounters substantial environmental noise points caused by initial faults that do not trigger further diffusion and defects that do not propagate. Similarly, fault outage predictions based on many average values in actual sample data often result in false positives, which can be attributed to the probability impact of such non-diffusing fault propagation and fault outage incidents[13]. Therefore, additional noise processing is required for data generated based on monitoring and simulation.

3.3 Noise processing

Many existing models used by various research institutions for power outage research and assessment encounter the challenge of data characteristic noise interference, resulting in issues like low accuracy and high time consumption. Researchers use various methods to filter data and enhance models to address this. This paper proposes a data feature noise analysis technique for power outage research and assessment. By identifying and analyzing characteristic noise points in the data, people perform data clustering and subsequently conduct model predictions based on the clustered data, thus improving the accuracy and efficiency of power outage research and assessment[14]. People employ DBSCAN (Density-Based Spatial Clustering of Applications with Noise) for noise analysis.

Given the original dataset:

$$D = \{x_1, x_2, \cdots, x_m\} \tag{6}$$

Where element x_j in D contains all the points in D whose distance from x_j is not greater than ε is ε . i.e.,

$$N_\varepsilon(x_j) = \{x_j \in D \mid dist(x_i, x_j) \le \varepsilon\} \tag{7}$$

Among these, core points are defined as:

$$\left|N_{\varepsilon}\left(x_{j}\right)\right| \geq \min \ Pts \tag{8}$$

In the above equation, core points are defined as x_j .
If x_j It is located in ε, whose core object is x_i, then x_j is said to be directly reachable by the density of x_i.
Set up a sample sequence:

$$p_1, p_2, \cdots, p_n \tag{9}$$

Here, where p_1 is x_i, p_n, is x_j, and the sample sequence meets the requirement that any p_i can meet the requirement that p_{i+1} F can be directly reached by p_i density, then x_j and x_i They are said to be density-reachable. When both x_j and $x_i$can reach the density of x_k , it is said that x_j and x_i Are connected in density.
The DBSCAN clustering algorithm requires an input point set, given a radius. (eps) and a minimum number of points min(Pts) .
The algorithm first arbitrarily selects a point if the number of points is less than (eps) is greater than or equal to min(Pts), the specific point arbitrarily selected by the algorithm is first marked as a cluster, and then all topics are recursively processed. Therefore, the DBSCAN clustering algorithm starts from the core point and expands to all density-reachable εfields. In the process of the ε field, the obtained area is maximized, and any two regional issues maintain a density-connected relationship. If the given number of points is less than (eps) is less than the min(Pts) limit, the resulting points that cannot be included in the cluster are classified as noise points.

4. Model Fitting and Data Analysis

4.1 Noise Analysis Results

Drawing from the preceding analysis, the author constructed a model and utilized power grid monitoring data for noise analysis. As a result, the author has organized and generated 14 indicators to participate in the model construction. The DBSCAN clustering results are as follows:

Table 2. Classification results of bicluster centers

variable	Cluster species 1	Cluster species 2
x1	0.05787037	0.168595041
x2	0.08912037	0.181818182
x3	0.540509259	0.228099174
x4	0.142361111	0.256198347
x5	0.148148148	0.000000000
x6	0.278935185	0.000000000
x7	0.000000000	1
x8	0.399305556	0.000000000
x9	0.091435185	0.117355372
x10	0.22337963	0.244628099
x11	0.31712963	0.262809917
x12	0.243055556	0.252892562
x13	0.116898148	0.137190083
x14	0.128472222	0.14214876

As the table above shows, the model's classification results exhibit minimal variations across most variables. The author can organize these results into a scatter plot for a more precise visual representation. The scatter plot reveals the following patterns:

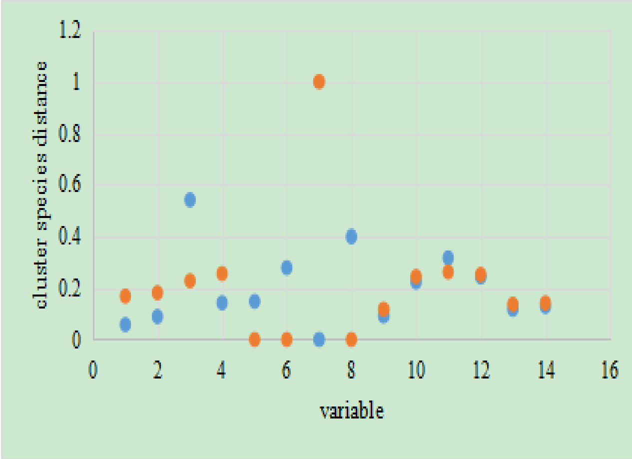

Figure 2. Classification results of bicluster centers

As illustrated in the figure above, noise analysis results reveal that certain variables, including x3, x6, x7, x8, and others, exhibit notable data discrimination. This observation suggests that the sample data has been partitioned into distinct groups based on two cluster centers, allowing for the identification of noise data. This methodology effectively removes noise while preserving information unrelated to the noise within the original dataset. Consequently, it enhances the validity and reliability of the data for predictive analysis, ensuring that the information essential for research remains intact.

4.2 Model prediction results

Feature extraction is carried out using the preprocessed data mentioned above, focusing on extracting crucial characteristic parameters that reflect the operational state of the power system, which will be used for model predictions[15]. In this process, 70% of the sample data is allocated for training to optimize the model's performance. Another 15% is earmarked for validation, assessing the model's convergence and termination during training. The final 15% is reserved for testing, crucial for evaluating the model's predictive performance.
A noteworthy trend is observed in the gradient descent process during model training and optimization. This trend illustrates the iterative progression towards enhancing model performance.

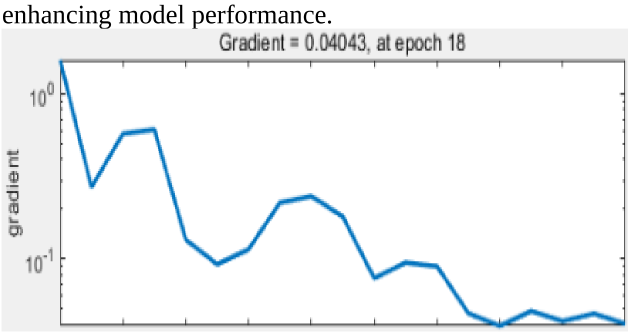

Figure 3. Gradient downward trend curve in iteration

The figure above vividly illustrates a critical trend. As the number of iterations increases, the curve within the figure showcases a gradual reduction in fluctuations, which signifies that when the model parameters are fine-tuned using the gradient descent method, the model's deviation from the optimal point diminishes[16]. Consequently, the model's performance undergoes gradual optimization, reducing the disparity between the model's predictions and the actual values. With increasing iterations, the model parameters steadily converge toward the optimal point, leading to a smoother and more refined optimization process[17].

Simultaneously, during this iterative process, changes in the validation set's performance (val fal) are observed:

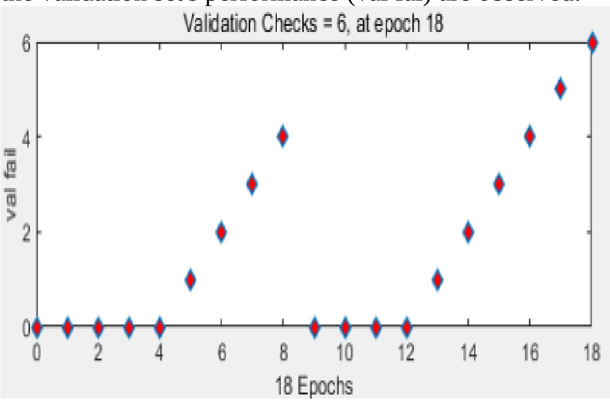

Figure 4. Validation checks result in iteration

4.3 Model prediction performance

Next, let's delve into the organization of the model's prediction performance, particularly concerning the ROC curve for the training samples:

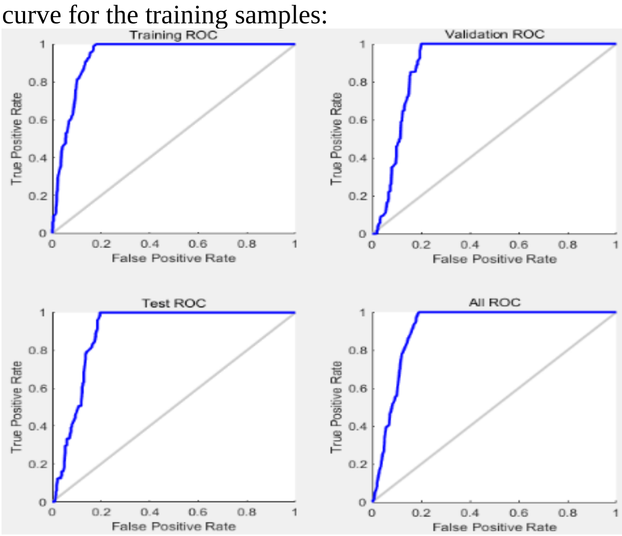

Figure 5. ROC curve

As depicted in the figure above, the ROC curve for the training samples showcases several noteworthy characteristics. The curve reveals a strong hit rate performance, with a limited increase in the false positive rate as the actual rate increases. This implies that the model maintains relatively tight control over the false positive errors resulting from its predictive accuracy. Even when managing the false positive rate effectively, the model retains a commendable hit rate. Furthermore, when striving for a high hit rate of 100%, the overall false positive rate generated by the model's predictions on the sample data remains below 20%, demonstrating significant practical applicability[18]. The model's prediction performance on the validation samples exhibits slightly lower results than the training samples. However, it's important to note that the model's predictive ability remains valid and reliable, demonstrating its capacity to generalize beyond the training data. The model's performance on the test samples slightly improved compared to the validation samples[19]. This outcome reaffirms the model's optimization results and indicates that it has successfully avoided the negative impact of overfitting, which can sometimes occur when training on the training samples. The ROC curve of the total sample data is smoothed, and based on this, the balance between the false positive rate and the false negative rate can be precisely analyzed.

4.4 Model Performance Comparison

To assess the model's predictive prowess, the author compares it with a traditional model's performance. Organize it into a visualization result, including:

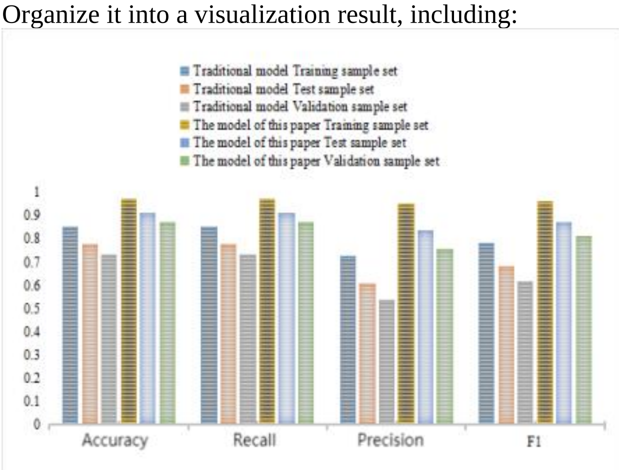

Figure 6. The prediction performance of the model used in this paper is compared with the traditional model

As illustrated in the figure above, the model prediction results indicate that the traditional model's predictive efficacy falls significantly short. Notably, the conventional power outage prediction model exhibits a substantially lower accuracy rate, resulting in a notably high false positive rate and undermining the overall performance of the model's prediction results[20]. This disparity in accuracy rates underscores the limitations of the existing analysis methods, highlighting the potential for the noise analysis technique employed in this study, as introduced by Benne, to be refined and optimized for superior performance, particularly in terms of F1 score.

The model developed in this paper displays distinct characteristics in comparing the three model types[21]. It excels in predicting training samples but performs poorly when applied to verification and test samples. This observation suggests that there is significant room for further optimization in existing models, offering promising opportunities for the refinement and expansion of this model in practical applications.

5. Conclusion

The conclusions drawn from this comprehensive research can be summarized as follows: The outage prediction model developed in this study achieves an impressive accuracy rate of 96.4%, underscoring its practical utility and potential to provide a more dependable safeguard for power system safety. This research introduces a novel approach to power outage analysis and assessment, leveraging data characteristic noise analysis technology. The application of this technology substantially enhances the accuracy and efficiency of power outage research and evaluation. Effective data preprocessing and feature extraction techniques are essential for successfully applying this methodology. Additionally, practical considerations such as data timeliness and accuracy must be addressed when implementing this approach in real-world applications. Building upon this technological foundation, future work could focus on optimizing data preprocessing and feature extraction techniques to enhance algorithm performance and accuracy further. Model optimization remains an area of consideration as well. Addressing practical challenges such as data timeliness and accuracy is paramount, especially for real-time power system monitoring. To meet the demands of real-time power system monitoring, this technology can be integrated with existing power system safety monitoring technologies to establish a comprehensive power system safety monitoring system.

References

[1] Fan, M., Yang, Q., Guo, X.F., Liu, H., Xia J.L., Peng Y.W.. (2023). Prediction method of power outage in a distribution network for unbalanced data, Power System Protection and Control, 51(8): 96-106

[2] Li, G.Q., Liu, D.G., Xiao, G.L., Zhang, B., Wang, G.W., Ren, H., Zhen, Z.. (2022).Risk Prediction of Node Outage in High Proportion New Energy Grid, Power System and Clean Energy, 38(10), 106-115.

[3] Nan, D.L., Feng, C.Y., Cao, H., Wang, X., Li, Y.D.. (2021). Data-driven predictive model of distribution system blackout， Advanced Technology of Electrical Engineering and Energy， 40(12), 56-63.

[4] Yu, Q., Qu, Y.Q., Shi, L.. (2018). Self-correlation Analysis of Power Grid Blackouts Based on Relative Value Method and Hurst Exponent." Automation of Electric Power Systems, 42(01), 55-60,124.

[5] Hassani, H., Razavi-Far, R., & Saif, M.. (2022). Real-time out-of-step prediction control to prevent emerging blackouts in power systems: A reinforcement learning approach.Applied Energy,314, 118861.

[6] Dong, M.. (2019). Combining unsupervised and supervised learning for asset class failure prediction in power systems. IEEE Transactions on Power Systems, 34(6), 5033-5043.

[7] Corcoran, J., Davies, C. M.. (2018). Monitoring power-law creep using the failure forecast method. International Journal of Mechanical Sciences, 140, 179-188.

[8] Dong, M., & Nassif, A. B.. (2018). I am combining modified Weibull distribution models for power system reliability forecast. IEEE Transactions on Power Systems, 34(2), 1610-1619.

[9] Tang, Y., Li, F., Wang, Q., Xu, Y.. (2018). Hybrid method for power system transient stability prediction based on two-stage computing resources. IET Generation, Transmission & Distribution, 12(8), 1697-1703.

[10] Kamali, S., Amraee, T.. (2017). Blackout prediction in interconnected electric energy systems considering generation re-dispatch and energy curtailment. Applied Energy, 187, 50-61.

[11] Salimian, M. R., Aghamohammadi, M. R.. (2017). A three-stage decision tree-based intelligent blackout predictor for power systems using brittleness indices.IEEE Transactions on Smart Grid, 9(5), 5123-5131.

[12] Wetherill, R. R., Fromme, K.. (2016). Alcohol-induced blackouts: A review of recent clinical research with practical implications and recommendations for future studies. Alcoholism: clinical and experimental research, 40(5), 922-935.

[13] Gupta, S., Waghmare, S., Kazi, F., Wagh, S., Singh, N. (2016, March). Blackout risk analysis in intelligent grid WAMPAC system using KL divergence approach. In 2016 IEEE 6th International Conference on Power Systems (ICPS) (pp. 1-6). IEEE.

[14] Wang, C., Grebogi, C., & Baptista, M. S.. (2016). Control and prediction for blackouts caused by frequency collapse in intelligent grids—chaos: An Interdisciplinary Journal of Nonlinear Science, 26(9), 093119.

[15] Vasseur, J., Wadsworth, F. B., Lavallée, Y., Bell, A. F., Main, I. G., & Dingwell, D. B.. (2015). Heterogeneity: The key to failure forecasting. Scientific reports, 5(1), 13259..

[16] Gupta, S., Kazi, F., Wagh, S., & Kambli, R. (2015). Neural network-based early warning system for an emerging blackout in smart grid power networks. In Intelligent distributed computing(pp.173-183).SpringerInternationalPublishing.

[17] Xinyao,L.,Huaqiang,L.,Yimiao,W.,Peiqing,L.,&Qiang,H.(2014,October).Cascadingfailuresforecastingresearchtopowergridbasedonself-organizedcriticality.In2014InternationalConferenceonPowerSystemTechnology(pp.820-825).IEEE..

[18] Shiroei,M.,&Ranjbar,A.M.(2014).SupervisorypredictivecontrolofpowersystemloadfrequencycontrolInternationalJournalofElectricalPower&EnergySystems,61,70-80.

[19] Mcdermit,D.,Shipp,D.D.,Dionise,T.J.,Lorch,V..(2012).Mediumvoltageswitchingtransientinducedpotentialtransformerfailures;prediction,measurement,andpracticalsolutions.48thIEEEIndustrial&CommercialPowerSyst

emsConference.IEEE.Agarwal,Mridul,etal."Circuitfailurepredictionanditsapplicationtotransistoraging."25thIEEEVLSITestSymposium(VTS'07).IEEE,2007.

[20] Agarwal,M.,Paul,B.C.,Ming,Z.,Mitra,S..(2007).CircuitFailurePredictionandItsApplicationtoTransistorAging.IEEEVlsiTestSymposium.IEEE.

[21] Yu,Q.,Guo,J.B..(2006).StatisticsandSelf-organizedCriticalityCharactersofBlackoutsinChinaElectricPowerSystems,AutomationofElectricPowerSystems,30(2),16-21.