IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2022.Doi Number

Enhancing Weather-Related Outage Prediction and Precursor Discovery through Attention-Based Multi-Level Modeling

Mohammad Alqudah[1], Student Member, IEEE and Zoran Obradovic[1], Senior Member, IEEE

[1]Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122, USA

Corresponding author: Mohammad Alqudah (e-mail: mohammad.alqudah@temple.edu).

This work was supported in part by the United States Army Core of Engineers (USACE) Engineer Research and Development Center (ERDC) Geospatial Research Laboratory (GRL). It was accomplished under Cooperative Agreement Federal Award Identification Number (FAIN) W9132V-22-2-0001.

ABSTRACT Electric grid continually monitors spatiotemporal data from sparse service areas. As power systems grow and get more complex, and with the deployment of more sensors and data collection capabilities, monitoring and analyzing data streams for outage prediction will get more complicated. In addition, the burden on human operators to analyze such data is getting challenging. Furthermore, climate change introduces new challenges to power grid reliability and makes the human grid operators' task more critical. To address some of these challenges, this research proposes a novel model to jointly predict power grid outages and discover precursors from spatiotemporal data using multi-level data. The new method utilizes multi-task learning (MTL) and multi-instance learning (MIL) to jointly predict outages and learn event precursors. This is achieved by introducing distance-aware self-attention to capture relationships between locations and improve event detection and precursor discovery while utilizing multi-level data (local weather data, global demand, and forecast data) in a sparse setting. Experiments are conducted using five years of data collected in the U.S. Pacific Northwest. The proposed methodology achieves an Area Under the Precision-Recall Curve (AU-PRC) of 0.97 using 12 hours of data before the event. Experiments showed that the proposed model could predict events several hours ahead with high accuracy, where such early predictions allow grid operators to deploy outage mitigation plans. In addition, the new framework effectively discovers spatiotemporal precursors for power outages. Grid operators can use such event precursors to help mitigate outages and improve grid reliability.

INDEX TERMS Weather, event detection, event precursors, machine learning, power system faults, smart grids, time series analysis, big data, climate change.

NOMENCLATURE

N	Set of events.
n	Event $n \in N$.
$\mathcal{L}$	Set if locations, $\lvert\mathcal{L}\rvert = Q$.
$\mathbb{L}_q$	Location in set of locations $\mathcal{L}$, where $\lvert\mathcal{L}\rvert = Q$.
$\mathbb{Y}_n$	Label for event n, $\mathbb{Y}_n \in \{0,1\}$.
$\mathcal{Y}$	Set of labels for all events .
$\mathcal{B}$	Set of data bags.
$\mathbb{B}_n$	One data bag representing event n.
t_n	Time for event n.
θ^q	Model for location q.
N_q	Events for location q.
γ^q	Neighbors of location q where $\gamma^q \in \mathcal{L}$.
$\hat{\theta}$	Global model across all locations.
$\mathcal{F}$	Generic function.
λ, β	Regularizartion hyper pramater represented by $\lambda_1, \lambda_2, \lambda_3, \beta$.
σ	Scaled dot product between two vectors.
x	Data vector.
α	Calculated probabilities from σ.
$\acute{\gamma}$	One location where $\acute{\gamma} \in \gamma$.

$d_{q,l}$	Normalized inverse distance between locations q and l.
p_j	Probability elation of x_j to the original labels $\mathbb{Y}_n$.
$\hat{y}$	Predicted label for x.
$\widehat{\mathbb{Y}}_j$	Predicted label for $\mathbb{B}_j$.
$\mathcal{A}$	Aggregate function.
$\mathbb{P}_n$	Aggregated probabilitt over $\widehat{\mathbb{Y}}_j$.
$\hbar(\theta)$	Loss function.
Δ_1	Log-loss, which equals: $-(\mathbb{Y}_n \log(\mathbb{P}_n) + (1 - \mathbb{Y}_n) \log(1 - \mathbb{P}_n))$.
$\mathcal{X}_i$	nMIL temporal group. $\mathcal{X}_i = \{x_{ij}\}$.
x_{ij}	nMIL data vectors for time i and the j-th data source.
P_i	Probabiliy of $\mathcal{X}_i$.
p_{ij}	Probabiliy of x_{ij}.
g	nMIL temporal penalization.
$\mathcal{K}$	Similarity metric, such as cosine.
Δ_2	Sequare loss, $\Delta_2(a,b) = (\mathrm{a} - \mathrm{b})^2$.
h	Hinge loss.
$\mathrm{m}_0, p_0, \varepsilon$	Hing loss hyper parameters.
τ	Precursor thresholds.
s	String.
l_p	The length of a common prefix.
m	The number of matching characters.
u	The number of transpositions for s.
sim_j	Jaro-Winkler similarity.
$L_{\text{Similarities}}$	List if string similarities.
$\bar{s}$	Similarity between two strings.

I. INTRODUCTION

Electric grids are a vital part of any modern economy and are deemed a critical infrastructure that serves society in many ways. Climate change presents significant challenges to power grid reliability. The increasing frequency and intensity of extreme weather events, such as storms and wildfires, can damage infrastructure and disrupt power transmission. Rising temperatures and cooling demands during heat waves strain the grid, potentially leading to failures and blackouts. Changing precipitation patterns and adding renewable energy sources introduce grid stability challenges. Addressing these issues requires investments in grid resilience, infrastructure upgrades, advanced forecasting, renewable energy, and collaboration among various governing entities to ensure a reliable power grid in the face of climate change [1, 2, 3]. Given the importance of power grids, power outages can cost countries billions of dollars and disrupt many people's lives [1]. In the last two decades, weather-related events have been the leading cause of power outages in the United States [1, 2]. As such, they lead to significant economic and social costs and implications. Weather-related power outages can cost countries' economies billions of dollars in the form of lost wages, spoiled inventory, impeded emergency services, and damage to infrastructure.

Between 2018 and 2020, more than 231,000 power outages occurred in the United States that lasted more than one hour, out of which 17,484 lasted at least eight hours, where power outages lasting eight hours or more can be deemed medically relevant [2]. In the same period, the power outages resulted in an annual loss of 520 million customer hours across 2,447 US counties [2]. Between 2000 and 2021, approximately 83% of significant power outages impacting a minimum of 50,000 customers in the United States were attributed to severe weather conditions [4]. As the frequency and intensity of extreme weather events increase and the power grid infrastructure ages, the occurrence and severity of power outages are expected to increase. The power grid continues to be vulnerable to outages due to weather events, where 70% of the U.S. power grid is over 25 years old. The average number of weather-related power outages increased by approximately 78% between 2011 and 2021 [4]. Fig. 1 shows the increase in the number of major weather-related outages in the years 1992-2012. Achieving grid resilience against weather events can be done on multiple fronts, such as strengthening the aging infrastructure, increasing systems flexibility and robustness, and introducing situational awareness using advanced systems such as Phasor Measurement Units.

This paper introduces a novel machine-learning methodology to predict power outages and provide leading indicators (event precursors) to assist grid operators in power outage mitigation. This paper presents a novel method that learns the system's state by examining data spatially and temporally using both local and global system data, then simultaneously providing an event prediction and prediction explanation using event precursors.

A. PROBLEM STATEMENT AND OBJECTIVES

Using advanced machine learning tools and control systems is important to combat power outages. When applied to critical grid components, these advanced systems provide grid operators with alerts and actionable information that provides adequate time in advance for planning and risk mitigation (e.g., taking equipment offline, load transfers, and crew deployments). The problem of designing effective event prediction and precursor discovery of weather-related power outages relies on understanding *each grid component's state* and the relations between each grid component and *other components within the same vicinity*.

Furthermore, the state of each component can be understood by examining its *local data* and the global data that represents the system's status. The hypothesis is twofold: first, power outage event prediction and precursor discovery can be improved by examining locations' local weather and global system data. This data modeling scheme can be referred to as multi-level data. Secondly, this examination is performed spatially and temporally by studying each location's data and considering data from neighboring locations where such locations can affect and interact.

Proposing solutions for event prediction and precursor discovery in power grids requires a multi-faceted approach. Such approaches are needed due to the complexity of the involved systems. This research proposes models to answer the following research questions:

1. *how to predict weather-related power outages* using *integrated spatiotemporal and multi-level data* representing local data at spatial locations and global system state conditions data.
2. *how to provide explainable spatiotemporal predictions* to help grid operators mitigate outages ahead of time and plan for effective action plans.

This study suggests a principal approach to answer the following questions:

1. *when* an event might happen.
2. *why* an event may occur, providing explainable insights that can assist power grid operators in defining timely plans.

This research assumes that weather conditions directly cause the studied outages. While the event logs used (described in Section III.B.2) are selected based on the main reason, the weather, the proposed model doesn't consider other outage factors, such as equipment status or age as such data is not available for our current study.

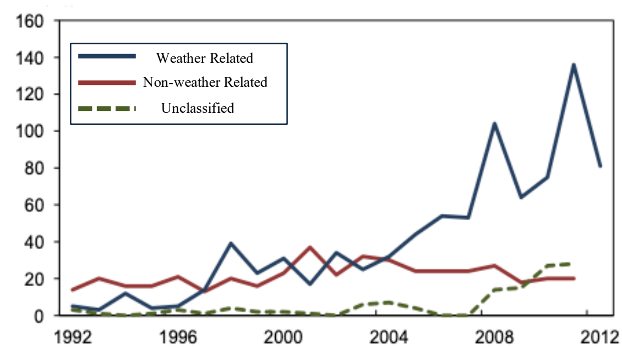

FIGURE 1. Observed major bulk outages in the United States electric systems [1]. The figure shows the increase in weather-related outages from 1992 to 2012. The *y*-axis represents the number of major outages in the United States.

B. RELATED WORK

Event prediction has many applications and is studied in domains such as pandemic and disease outbreaks, finance and the stock market, and crime prevention [5]. Machine Learning (ML) based power outage prediction models have recently gained more popularity with machine learning techniques advancement. Several models were used to predict power outages. A tree-based model (random forest) was utilized in [6] to predict power system outages caused by tropical cyclones. Similarly, tree-based models were used in [7] to study parts of the U.S. in New England to predict power outages. A tree-based outage prediction model is introduced in [8] for the northeastern united states using multi-modal optimizations, where data presentations for weather parameters and vegetation are created. Logistic regression was used in [9] to define a decision boundary to predict component outages efficiently. Quantile regression forests (QRF) and Bayesian additive regression trees (BART) were used in [10] for predicting storm outages on electric distribution networks. A combination of a data preprocessing heuristic and standard methods was used in [11]. In this work, the data preprocessing heuristic divides the training dataset into subsets of events representative of the predicted power grid event's severity by calculating the quantile weight distance (QWD) between severe weather-related events in the dataset and the predicted event. After that, standard methods such as tree-based approaches are used for prediction. An extension of logistic regression was used in [12], where the model embeds the spatial configuration of the network used for outage prediction in the western United States. Similarly, a two-step model was used in [13] to predict damages from weather-related events. A model training combined clustering to group spatial coherent data, followed by training linear and tree models to predict outages and using a scoring model to assess the severity on the grid. Predicting power outages due to equipment failure is a widely studied problem [14, 15, 16]. Such models rely on understanding the equipment status (such as transformers) and use their physical properties to predict outages. Such methods require access to data representing the current physical status of the equipment.

Model interpretability can be classified based on the interpretation stage: pre-model, in-model, and post-model [14]. Pre-model interpretation is performed before the machine learning stage using exploratory data analysis and feature engineering methods. Pre-model interpretation aims to gain a deeper understanding of the data, leading to a better understanding of the modeling results. Pre-model interpretation can be labor intensive and requires domain experts to assess the analysis. Post-model interpretation attempts to explain the trained ML model and perform post-moretem analysis of the model. Due to the increasing complexity of ML models, post-model interpretability has become the main direction of current interpretable ML research. It is mainly focused on the field of deep learning. In-model aims to train machine learning models that are intrinsically interpretable [17]. Such interpretations are contained within the model, allowing the decisions to be understood without additional preprocessing.

Our paper introduces a model that falls under the in-model class of explainable models. Simple ML models (such as linear regression and generalized additive models) [11, 12, 18] can achieve in-model interpretability since these models are inherently transparent. In addition, rule-based models (such as decision trees) [6, 7, 8, 10] can be regarded as interpretable models. In the power grid domain, there is still a need for interpretable and generalizable models. The gap can be described as follows: new models are needed to generate human-centered interpretations, where interpretations are produced according to the audience's operational needs. At the same time, this interpretation should expose the logical reasoning behind the model while giving the desired decisions. Furthermore, interpretable models are needed to address spatiotemporal data [17]. Also, interpretable models are required to assess more grid operational areas, such as dispatch, control, and power safety. Previously discussed literature still needs to include these aspects, and additional models have been recently developed to address these needs. For example, several recent works are focused on designing in-model interpretable ML techniques. Multi-instance learning is introduced in [19] that jointly predicts power outages and discovers precursors; this method focuses on discovering temporal precursors. In other applications, such as societal events, multi-instance learning was utilized [20] to develop a methodology to propagate information from bag labels to individual instances and allow the distribution of labels to group features. In [21, 22], multi-instance learning was expanded by introducing nested multi-instance learning and applying it to spatial correlations. A spatial correlation approach was introduced in [22] that utilizes shared labels between locations rather than a data-driven approach where the similarity of the underlying data is utilized.

C. PURPOSE AND NOVELTY

This paper aims to introduce a model to aid in analyzing complicated and correlated data streams. Where such a task can be challenging and time-consuming, and the use of automated systems is preferred. Event precursors add challenges to grid operators since event precursors aim to find useful indicators within large amounts of data streams. Effective and fast precursor discovery allows grid operators to design better power outage mitigation methods.

The novelty of this paper is in using multi-level data in a spatiotemporal setting to predict outages and discover precursors jointly. Modeling prediction and precursor discovery jointly enable leveraging event labels for precursor discovery, where labels for individual precursors do not exist. In addition, joint models can alleviate the need of designing and training two separate models. Using multi-level data provides a methodology to incorporate different data streams from disjointed monitoring systems where it will be beneficial to be used within power grid monitoring systems. Furthermore, this paper introduces a novel methodology that learns spatially and temporally, which to the best of our knowledge, has not been reported before. This methodology allows for predicting events accurately and discovering spatial and temporal precursors. This is achieved by modeling the data in a multi-level fashion, then utilizing multi-task learning (MTL) and multi-instance learning (MIL) to learn spatially and temporally. A local attention mechanism is also introduced to allow MTL models to learn complex relations within spatial locations. This attention mechanism allows for better event prediction and the discovery of event precursors.

D. CONTRIBUTION AND ADVANTAGES

Event prediction and precursor discovery present a multitude of challenges. These challenges can be described under two main categories: modeling and data labeling. There are no standard models for event prediction and precursor discovery for modeling challenges, and when it comes to spatiotemporal modeling of this problem, this becomes even more challenging. Regarding data labels, event precursors are parts of the data that indicate the probability of an event occurring in the future. In most cases, there are no predefined labels for event precursors. This adds further challenges to modeling event prediction and precursor discovery. Modeling event prediction and precursor discovery for power grids present additional challenges. Power grid data is noisy, with many details hidden for confidentiality reasons. The data collected from the electric grid has missing values, measurement outliers, and data labels which are inaccurate and lack information. Furthermore, power grid data, when available, is generally sparse and suffers from the big data paradox (large amounts of data with a small number of events of interest). Modeling such datasets to make informed decisions can be complex and resource intensive; however, such datasets can be particularly useful if paired with models designed to inform operators and facilitate data-driven decisions. The advantages of the approach proposed in this paper can be summarized as the following:

1. introducing automated models to assist power grid operators in predicting events in an informed fashion. Where with each prediction, event precursors are identified. This allows grid operators to know when an event might happen and to have enough information to assist in outage mitigation.
2. introduce a methodology to predict events and discover precursors in a spatiotemporal fashion jointly. This methodology utilizes attention to capture relations between locations.
3. utilize local and global data in a multi-level fashion; this allows for capturing the system's status from different angles.
4. utilize multi-task learning with multi-instance learning to extract spatial relations and discover precursors. This enables the models to extract precursors without predefined labels.
5. utilize large datasets without resorting to extensive data studies and feature engineering. This approach allows extending the model with more datasets if needed.

E. PAPER ORGANIZATION

The remainder of this paper is organized as follows. Section II describes the proposed methodology for spatiotemporal outage prediction and precursor discovery. Section III described the used data and the preprocessing steps. Section IV introduces the experimental setup and the results. Section

V concludes the paper, and Section VI discusses future work. The references are provided at the end.

II. SPATIO-TEMPORAL EVENT PREDICTION AND PRECURSOR DISCOVERY

Formulating spatiotemporal event prediction and precursor discovery relies on multi-task learning with multi-instance learning. This section will start by formally defining the problem, then describe the proposed methodology.

The power grid consists of multiple locations within a geographical area of interest. For a location $\mathbb{L}_q \in \mathcal{L}$, where $\mathcal{L}$ is a group of locations in a geographical area, the goal of the proposed methodology is to predict the probability of an event $\mathbb{Y}_n \in \mathcal{Y}$ at the location $\mathbb{L}_q$, where $\mathbb{Y}_n$ is an event that occurs at a time t_n at the location $\mathbb{L}_q$.

To predict the probability of an event, and for every event $\mathbb{Y}_n \in \mathcal{Y}$ there exists a bag of data $\mathbb{B}_n \in \mathcal{B}$. Each bag of data $\mathbb{B}_n$ represents all the data related to the period before the event. Bags ($\mathbb{B}_n$) are a collection of data instances $\{x_j\}$ in an unordered fashion.

One commonly studied task within the power grid domain is *event detection* [9], which vastly differs from the work presented in this paper. Event prediction and precursor discovery differ from *event detection* in the following aspects:

1. event detection focuses on studying the event signatures at the time t_n. In event detection and precursor discovery, *event data at the time t_n is not used,* and only the data before the event is utilized. For each event at the time t_n, the period used is formally defined as $[\mathrm{t_n} - \mathrm{k}, t_n)$ to indicate the omission of time t_n. k represents the period length (in hours) examined before the event at the time t_n. All event signatures at the time t_n are discarded, and event detection and precursor discovery are performed without the data at the time t_n.
2. *there are no labels for data instances x_j* within the bag $\mathbb{B}_n$. Labels $\mathbb{Y}_n$ are only assigned at the bag level, and there are no labels for any individual instance within $\mathbb{B}_n$.

Every event $\mathbb{Y}_n \in \mathcal{Y}$ is mapped to a label $\mathbb{Y}_n \in \{0,1\}$ where the value of 1 represents the occurrence of an event at the time t_n.

A. *MULTI-TASK LEARNING FOR EVENT PREDICTION AND PRECURSOR DISCOVERY*

Multi-task learning (MTL) is a paradigm that leverages shared and useful information in multiple tasks. Leveraging MTL helps improve task performance and helps achieve greater generalization of learning.

1) MULT-TASK LEARNING

MTL has gained fame over the past decade with many applications in machine learning. MTL has been applied to many tasks, such as computer vision and speech analysis, especially in domains with sparse and high-dimensional spaces [23]. MTL utilizes data from different tasks simultaneously. Compared to single-task learning, MTL can learn from sparse data and produce more robust models. The simultaneous learning paradigm leads to better knowledge sharing between tasks and lowers the risk of model overfitting in single tasks [23].

As discussed in Section I, power grid data suffers from data sparsity challenges. The MTL paradigm helps alleviate the data sparsity problem by utilizing knowledge sharing. More than the labeled data is needed in a sparse setting to train individual tasks effectively. Using MTL, the simultaneous learning aggerates all labeled data and acts as data augmentation for single tasks. Generally, in MTL, all tasks are treated equally, and there is no distinction between individual tasks. MTL can be formulated in several ways; one common methodology to formulate MTL is regularized multi-task learning (R-MTL) [24]. R-MTL assumes that tasks are close to each other and should inhabit similar behaviors. This allows sharing of information between tasks and helps regularize individual tasks.

2) ATTENTION-BASED MULTI-TASK LEARNING FOR EVENT PREDICTION AND PRECURSOR DISCOVERY

This section describes the novel methodology that aims to predict events and discover precursors using attention mechanisms.

Task: Assuming multiple locations $\mathcal{L}$, where $|\mathcal{L}| = Q$, within a geographical area of interest in the power grid for a location $\mathbb{L}_q \in \mathcal{L}, \mathrm{q} \in \mathrm{Q}$, the goal is to learn a model θ^q for each $\mathbb{L}_q$, where:

1. θ^q predicts the probability of weather-related events and discovers event precursors.
2. θ^q predicts events only for one location $\mathbb{L}_q$, but in its learning, θ^q utilized shared knowledge between different locations.

R-MTL assumes that there are inherent relationships between tasks. Formally, the loss function to learn R-MTL is defined as:

$$\min_\theta \sum_{q \in Q} \left(\frac{N_q}{N} \mathcal{F}(\theta^q) + \frac{\lambda_1}{2} \left\| \hat{\theta} - \theta^q \right\|_2^2 + \frac{\lambda_2}{2} \|\theta^q\|_2^2 \right) \quad (1)$$

where $\lambda_1$and λ_2 are hyperparameters for the regularization, and $\hat{\theta}$ is the average model across all locations. $\mathcal{F}$ represents a general loss function to learn event detection. N is the total number of events for all locations, and N_q is the number of events at location q.

Power systems can span large geographical areas, and the weather is a local factor that affects the power grid. A novel self-attention approach is introduced to capture the spatial relationships between the locations.

Hypothesis: for each location $q \in Q$, the spatial information in R-MTL can be captured through self-attention α calculated on γ^q, where γ^q represents the nearest neighbors of q. The hypothesis is to utilize self-attention to parametrize spatial coherence between locations and direct model learning to utilize nearby locations' data. The rationale behind this hypothesis is as follows:

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3303110

1. effects of weather tend to degrade as distances grow. Using the nearest locations helps narrow the model search for the best parameters.
2. self-attention helps to direct models with similar conditions to learn similar parameters.

To formally define the attention-based multi-task learning model (A-MTL), equation (1) is redefined as follows:

$$\min_{\theta} \sum_{q\in Q} \left(\begin{array}{c} \frac{N_q}{N}\mathcal{F}(\theta^q) + \frac{\lambda_1}{2}\left\|\hat{\theta} - \theta^q\right\|_2^2 + \frac{\lambda_2}{2}\|\theta^q\|_2^2 \\ + \frac{\lambda_3}{2}\sum_{i}^{N_q}\sum\nolimits_{l\in\gamma}\alpha_{q,l}(\theta^q - \theta^l)^2 \end{array} \right) \quad (2)$$

where λ_3 is a hyperparameter for balancing. To calculate α, the scaled dot product (σ) is calculated between the respective data vectors in the data bags of q, l, which is formally defined as:

$$\sigma = \sum \frac{x_q^T x_l}{\sqrt{|x_q|}} \;\forall\, x_q, x_l \in \mathbb{B}_q, \mathbb{B}_l \quad (3)$$

Then, the next step is to map σ to respective probabilities by calculating the data vector α, where each value $\alpha_z \in \alpha$ is calculated as:

$$\alpha_z = \frac{e^{\sigma_z}}{\sum_{i=1}^{|\gamma|} e^{\sigma_i}} \quad (4)$$

where $|\alpha| = |\gamma|$.

3) DISTANCE AWARE ATTENTION-BASED MULTI-TASK LEARNING FOR EVENT PREDICTION AND PRECURSOR DISCOVERY

Section II.A.2 defines the attention-based multi-task learning model (A-MTL). This model utilizes the nearest γ locations without considering the spatial distance. In this variation of the proposed methodology, a distance-aware method in addition to attention is introduced.

Hypothesis: for each location $\acute{\gamma} \in \gamma$, the self-attention α calculations can be further improved by relying on the absolute distances. The distances are converted to a ratio between 0 and 1 using a combination of distance inverse and range normalization. Equation (3) defines the distance aware A-MTL (DA-MTL) as:

$$\min_{\theta} \sum_{q\in Q} \left(\begin{array}{c} \frac{N_q}{N}\mathcal{F}(\theta^q) + \frac{\lambda_1}{2}\left\|\hat{\theta} - \theta^q\right\|_2^2 + \frac{\lambda_2}{2}\|\theta^q\|_2^2 \\ + \frac{\lambda_3}{2}\sum_{i}^{N_q}\sum\nolimits_{l\in\gamma} d_{q,l}\alpha_{q,l}(\theta^q - \theta^l)^2 \end{array} \right) \quad (5)$$

where $d_{q,l}$ is the normalized inverse distance between locations q and l.

4) TEMPORAL INFORMATION PROPAGATION MULTI-INSTANCE LEARNING

Each task θ^q is learned through a function $\mathcal{F}$ as part of the equations (2-5). Labels for data instances (x_j) are needed to learn individual models. As discussed earlier, obtained labels ($\mathbb{Y}_n$) are assigned to only the bag level where data instances (x_j) don't have any associated labels. Information is propagated from bag labels to data instances to achieve model learning and utilize the bag labels. This information propagation process is achieved through multi-instance learning. This section discusses how this process is utilized. Assuming an unknown function $\mathcal{F}$, the goal is to train $\mathcal{F}$ where:

$$\mathcal{F}(\mathbb{B}_n|\theta) \rightarrow \mathbb{Y}_n \;\forall\, n \in [1, N] \quad (6)$$

and $N = |\mathcal{Y}|$. $\mathcal{F}$ is modeled as a logistic binary function on the data instances. This modeling allows to learn probabilistic values (p_j) that represent the relation of x_j to the original labels $\mathbb{Y}_n$. Formally, p_j is defined as:

$$p_j = \mathrm{sigmoid}(\theta^T x_j) = \frac{1}{1 + e^{-\theta^T x_j}} \quad (7)$$

Each individual instance x_j is assigned a label $\hat{y}_j$ using the probability p_j. To extract the predicted labels for the bags $\hat{\mathbb{Y}}_j$, an aggregate function on the individual instance probabilities. This process is formally defined as follows:

$$\mathbb{P}_n = \mathcal{A}(\mathcal{F}(\mathbb{B}_n|\theta)) \quad (8)$$

By modeling $\mathcal{F}$ on the data instances using a sigmoid function, it allows *propagating information* from bag labels and learning labels for individual data instances. The information propagation is controlled through the loss function $\hbar(\theta)$. One effective loss function is the Nested Multi-Instance Learning (nMIL) loss function, which performed well on several applications [21]. nMIL defines the loss function $\hbar_{nMIL}(\theta)$ as a combination of bag-level loss $\hbar(\theta)_{Bag}$, instance-level loss $\hbar(\theta)_{Instance}$, and regularization $\hbar(\theta)_{reg}$ [21]. Formally nMIL loss is computed as:

$$\hbar_{nMIL}(\theta) = \hbar(\theta)_{Bag} + \hbar(\theta)_{Instance} + \hbar(\theta)_{reg} \quad (9)$$

The intuition behind the bag level loss is to control the learning process and ensure that the aggregated bag level labels calculated by (8) are similar to the true bag label. This can be defined as:

$$\hbar(\theta)_{Bag} = \frac{\beta}{N}\sum_{n=1}^{N}\Delta_1(\mathbb{Y}_n, \hat{\mathbb{Y}}_n) \quad (10)$$

where Δ_1 is the log-loss of the bag level prediction, defined as:

$$\Delta_1(\mathbb{Y}_n, \widehat{\mathbb{Y}}_n) = -(\mathbb{Y}_n \log(\mathbb{P}_n) + (1 - \mathbb{Y}_n) \log(1 - \mathbb{P}_n)) \quad (11)$$

where β is a hyperparameter. To calculate the instance cost, nMIL introduced a nested data approach [21], which allowed it to account for the temporal information within the bag. To achieve this, nMIL redefined the bags from an unordered set of data instances $\{x_j\}$ to an ordered set of temporal groups $\mathbb{B} = [\mathcal{X}_i]$. Each temporal group is an unordered set of data instances $\mathcal{X}_i = \{x_{ij}\}$ where x_{ij} represents data from grouping time i for the j-th data source.

Temporal dependencies in the instant cost rely on the assumption that temporal groups $\mathcal{X}_t$ that are close in time have similar data properties and thus should have similar probabilities P_i. This relationship was captured through a specific function g, which uses temporal order in combination with a similarity metric $\mathcal{K}$ to account for temporal relationships [21]. This is defined as

$$g = \mathcal{K}(\mathcal{X}_i, \mathcal{X}_{i-1})\, \Delta_2(P_i, P_{i-1}) \quad (12)$$

and P_i is defined as the aggregation of individual instance probabilities, and Δ_2 is defined as the square loss function. More formally,

$$P_i = \mathcal{A}\left(\mathcal{F}\left(x_{ij} \in \mathcal{X}_i \middle| \theta\right)\right) = \mathcal{A}(p_i) \quad (13)$$

$$\Delta_2(a, b) = (\mathrm{a} - \mathrm{b})^2$$

Finally, $\hbar(\theta)_{Instance}$ is defined as:

$$\hbar(\theta)_{Instance} = \frac{1}{\mathrm{N}} \sum_{i=1}^{B} \frac{1}{t} \sum_{t=1}^{T} g(\mathcal{X}_{i,t}, \mathcal{X}_{i,t-1}) \quad (14)$$

Regularization is added to prevent the model from overfitting the data and help produce generalizable models. Regularization is defined as:

$$\mathcal{J}(\theta)_{Reg} = \sum_{\substack{\mathbb{B} \in \mathcal{B} \\ \mathcal{X}_i \in \mathbb{B} \\ x_{ij} \in \mathcal{X}_i}} \frac{1}{T} \sum_{i=1}^{T} \frac{1}{B} \sum_{j=1}^{B} h(x_{ij}, \theta) + \varepsilon R(\theta) \quad (15)$$

where h is unsupervised hinge loss. Equation (16) defines h formally. In (16), m_0, p_0 are hyperparameters and sgn is the sign function.

$$h(x_{ij}, \theta) = \max\left(0, \mathrm{m}_0 - sgn(p_{ij} - p_0)\theta^T x_{ij}\right) \quad (16)$$

5) PRECURSOR DISCOVERY

Precursor discovery is performed after event detection. The precursors can be determined on a data instance or temporal group level. Once the predicted label is $\widehat{\mathbb{Y}}_j = 1$, p_{ij} and P_i are examined using a predefined threshold τ. A precursor is identified if $p_{ij} \geq \tau$ or $P_i \geq \tau$.

III. DATA MANAGEMENT

This section discusses the data sets utilized in the study and the preprocessing steps performed. One of the goals of this study is to use datasets without the need for manual feature engineering, which can be labor extensive. Firstly, this section introduces the datasets. After that, the preprocessing steps performed on the datasets are discussed.

A. DATA DESCRIPTION

This study utilizes multi-level data, where data exists on the local level and global levels. Here, weather data will represent local data, while power systems demand and forecast data will present global data.

1) WEATHER DATA

Comprehensive and accurate weather data is needed to predict weather-related power outages effectively. One public and reliable data source for weather data is the Automated Surface Observing System (ASOS) network data [25]. ASOS is a network of weather stations deployed across various locations in the United States. ASOS stations are automated weather observation systems that provide real-time weather data and observations. The network is operated by the National Weather Service (NWS). It is a joint effort of the National Weather Service (NWS), the Federal Aviation Administration (FAA), and the Department of Defense (DOD). The ASOS data is collected regularly, typically reported at intervals such as every 1-minute and 5 minutes. ASOS contains weather data from more than 900 sites in the United States. The exact location of each weather station is known.

ASOS is a widely studied data set in many applications, including power outage prediction. ASOS data reports many variables describing the current conditions of data. In this study, the following measurements are utilized: air temperature in Fahrenheit, dew point temperature in Fahrenheit, wind direction in degrees from north, wind gust in knots, wind gust direction, visibility in miles, atmospheric pressure, precipitation total, and wind speed in knots.

2) POWER SYSTEMS DEMAND AND FORECAST DATA

The power systems demand and forecast data is obtained from the U.S. Energy Information EIA-930 data [26]. The EIA-930 dataset comprehensively aggregates hourly demand and demand forecast data. The data is collected for the lower 48 states in the United States. The dataset covers various aspects of energy production, consumption, and infrastructure in the United States. EIA-930 is collected from balancing authorities responsible for balancing the power system demand and supply in real time. This dataset reports on an hourly basis. The reported measurements include demand, demand forecast, and net generation. The following measurements are used from EIA-930: demand forecast (MW), demand (MW), net generation (MW), the total interchange (MW), demand

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3303110

(MW) adjusted, net generation (MW) adjusted, Direct Interchange with directly interconnected Balancing Authorities (DIBAs) in (MW). The adjustment means incorporating dynamic scheduling arrangements and interchanges, balancing authorities adjusting metered physical flow values to produce an alternative view of power system operations [26].

3) LOCAL DATA VS. GLOBAL DATA

Weather data is local by nature, and weather conditions are measured in small geographical regions compared to the power grids. The power grid can span vast areas with multiple weather regions, where each region can have different weather conditions. In contrast, the EIA-930 data is considered global data since it represents the state of the whole examined grid, where for example, demand forecast (MW) represents the electric demand on the entire grid. While temperature measured from a weather station only represents the area near that station.

B. DATA PREPROCESSING

This section describes the preprocessing steps performed on the two utilized datasets. The explained steps don't include any feature engineering.

1) GEOGRAPHICAL AREA SELECTION

The Western Interconnection contains 136,000 miles of transmission lines, and it spans 1.8 million square miles across 14 western U.S. states, in addition to parts of Canada and Mexico [27]. This paper focuses on studying the Pacific Northwest (PNW), specifically the events in the geographical area of the Bonneville Power Administration (BPA) [28]. Using this area gives two important aspects: (1) the PNW area has diverse climatological topography, and (2) weather plays a role in affecting power system stability. BPA operates in the following states: Washington, Oregon, and Idaho, in addition to eastern Montana, a small part of eastern Wyoming, northern Utah, northern Nevada, and a small part of northern California. Secondly, BPA provides public historical logs of outages caused by weather conditions, and this is crucial to this study since publicly available data for power systems are rare. The overlapping weather stations with the BPA maps were selected [28], which resulted in 81 weather stations. After that, the power systems demand and forecast data for BPA is filtered by using identifier columns part of the EIA-930 data.

2) FILTERING POWER OUTAGE EVENTS

After selecting the geographical area in section III.B.1, the next step is to select the power outage events. BPA publicly reports all power outage events for all reasons. Since not every power outage is of interest to this study, the following criteria were used:

1. *selecting power outages for five years (2018-2022).* This period allows for enough data to train and test the prediction models.
2. *selecting weather-related events only*. BPA reports the general causes of the outages. Using the outage codes, maintenance, and planned outages were excluded from the event log
3. *excluding overlapping events.* Since the model utilizes the period before the event at t_i, this event will be excluded if any other events are in the period $[t_{i-k}, t_i)$.
4. *excluding non-transmission line events.* Transmission line events affect more customers and can span wide geographical areas.

3) MAPPING POWER OUTAGE EVENTS TO LOCATIONS

BPA publicly publishes Power outage logs. After filtering out unneeded outages with the process described in section III.B.2. The next step is to map the outages to locations on the map. BPA releases information about the grid's topology and some grid assets. Despite this, mapping outages to transmission lines are only partially achievable. BPA outage logs provide a textual description of the location of the outage, which doesn't necessarily have a one-to-one mapping to one of the transmission lines. To overcome this, an approach was adopted to search for matching transmission lines. Table 2 describes the Algorithm. To calculate the similarity between the outage description and the transmission line name, the Jaro-Winkler similarity metric is utilized [29]. Equations (17, 18) define the Jaro-Winkler similarity between two strings s_1, s_2. Where m is the number of matching characters, and u is the number of transpositions. A transposition is the number of matching characters not in the right order divided by two. l_p is the length of a common prefix at the start of the string up to a maximum of 4 characters, and v is a constant scaling factor for how much the score is adjusted upwards for having common prefixes.

4) DATA SPATIAL CORRELATION

Mapping outages to transmission lines helped narrow down the events' locations. A more granular correlation is needed to map weather data to outage locations. To achieve this, each outage was mapped to the nearest power substations. This allowed outages cross-correlated with weather data, power demand, and forecast data. This step was achieved using the transmission assets data published by BPA in addition to public map data available from OpenStreetMaps. This step is needed since BPA doesn't explicitly disclose the locations of substations.

$$sim_j = \begin{cases} 0 & if\ m = 0 \\ \frac{1}{3}\left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-u}{m}\right) & otherwise \end{cases} \quad (17)$$

$$sim_w = sim_j + v \times l_p \times (1 - sim_j) \quad (18)$$

TABLE I. **Outage description to transmission line matching algorithm.**

Algorithm 1 Outage Location Matching

1: **procedure** Outage-Location-Match
2: **Input**:
BPA outage logs ($logs$)
$BPA\ transmission\ line$ ($names$)
3: **Output**: $logs$ with mapped locations
4: **for** $outage\_description \in logs$ **do**
5: $L_{\text{Similarities}} = []$
6: **for** $line\_name \in names$ **do**
7: $\bar{s} = sim_w(line\_name, outage\_description)$
8: $L_{\text{Similarities}} \leftarrow \bar{s}$
9: sort $L_{\text{Similarities}}$
10: assign the most similar $line\_name$ to $outage\_description$
11: return $logs$

IV. EXPERIMENTAL EVALUATION AND RESULTS

A. EVENT PREDICTION AND PRECURSOR DISCOVERY MODELS

The attention-based model (A-MTL) introduced in section II.A.2 and the DA-MTL introduced in section II.A.3 are utilized in the next experiments. There have been some advances in event prediction in the last few years. To compare, models must be able to predict events and discover precursors jointly with in-model interpretability capabilities. This task is different from traditional event prediction, and models to be compared with must have these properties. Hence baseline experiments are performed using R-MTL.

B. UTILIZED DATA AND EXPERIMENTAL SETUP

As discussed in Section III. five years of data were used for training and testing. When testing data containing a temporal aspect, it's crucial to split it temporally to gain accurate results. The years 2018 to 2020 are used for training, and 2021 and 2022 are used as testing datasets. This ensures there is no information leakage between training and testing data. The models above use the full year for training and testing. These models are named *global models*.

Weather data contains a great amount of seasonality. Seasonal models were trained and tested to study the effects of seasonality on the modeling results. In this setup, seasons were used to split the data further. For example, the summer data for 2018 to 2020 are used for training, and summer data for 2021 and 2022 are used for testing.

The number of outages in the training dataset is 796, and the number of outages in the testing dataset is 488. To get data with no events, random time stamps across all years and all seasons were chosen as negative events; these timestamps don't overlap with any event or the k hours before the event. The total number of events in the training data is 1,624 and the total in the test data is 1,074. A total of 482 substations exist in the dataset. Weather data was pulled using 5 min intervals. k is set to 12 hours of data before each event. Using hyperparameter tuning, the mini-batch size is set to 5, $\lambda_1 = 0.25$, $\lambda_2 = 0.4$, $\lambda_3 = 0.4$, $\beta = 0.25$, learning rate = 0.01. m_0 and p_0 are kept at the default value of 0.5. The number of neighbors $\gamma = 5$.

The metrics to evaluate the performance of the models were chosen to be suitable in a power system setting. Precision and recall were used to measure the false positives and false negatives, which can give a detailed picture of the model's performance. Furthermore, the Area Under the Receiver Operator Curve (AU-ROC) is reported to measure the effects of different decision boundary limits. Finally, the Area Under the Precision-Recall Curve (AU-PRC) is reported as an indicator of the tunability and flexibility of the model. All metrics (AU-ROC, AU-PRC, Precision, Recall) have values between 0 and 1, with a higher value indicating a better-performing model.

C. EXPERIMENTAL RESULTS

The first set of experiments aims to evaluate the performance of the models DA-MTL, A-MTL, and R-MTL. All these models were evaluated using the global models. Table II shows the results of detecting the events using 12 hours of data, and the decision is made 1 hour ahead. One can notice that the proposed models (A-MTL and DA-MTL) achieved a lift in recall equal to 8.1% and 15% compared to the baseline recall. Recall, also known as the true positive rate (TPR), is the percentage of events that the trained model correctly identifies as weather-related outages out of the total studied weather-related outages. The recall is specifically useful in scenarios where identifying all positive instances is crucial, and the cost of false negatives (missed detection of actual positive instances) is high. In power grids, where the cost of missing the prediction of an actual outage is high, it is crucial to prioritize event recall over precision and general detection metrics. Precision measures the model's ability to correctly identify power outages out of the total data points predicted as outages. Table II shows that DA-MTL has the highest Precision and Recall metric among all models.

The second set of experiments is to train and test seasonal models. This setup uses the best-performing model from Table II (DA-MTL) for four experiments. The model was retrained and tested in each experiment on a year's season. Table III represents the results of the seasonal experiments.

TABLE II. **Global model experiment results.**

Model	AU-ROC	AU-PRC	Precision	Recall
R-MTL	0.75	**0.89**	0.75	0.64
A-MTL	**0.76**	**0.89**	0.77	0.71
DA-MTL	0.74	**0.89**	**0.80**	**0.75**

Bold values represent the best performance for each model.

TABLE III. Seasonal model experiment results using DA-MTL

Model	AU-ROC	AU-PRC	Precision	Recall
Winter	**0.88**	**0.97**	**0.86**	**0.93**
Spring	0.56	0.71	0.63	0.48
Summer	0.68	0.85	0.85	0.77
Fall	0.64	0.83	0.73	0.48

Bold values represent the best performing for each model.

From Table III, the best-performing model is the Winter model. This is expected since weather-related outages tend to increase in the wintertime. Furthermore, since this model studies one season, the data tend to be more homogeneous. Several factors can explain the degradation of the performance for Spring and Fall. The Pacific Northwest has a microclimate where weather is less predictable. In addition, the outage labels need to be more accurate.

The next study is to assess how early the model can predict. In power systems, the earlier the model predicts, the more beneficial the prediction is. An early prediction can be crucial to allow the grid operators enough time to deploy a mitigation plan. Fig. 2 shows the AU-PRC for early prediction. The x-axis represents how many hours ahead of the event used to predict the event. Fig. 2 shows that the accuracy increases as the time gets closer to the event. This behavior is expected since weather effects tend to be more predictable within short time frames. On the other hand, Fig. 2 shows that the event detection is still accurate far ahead of the event. This behavior is specifically beneficial to allow grid operators to act early.

D. PRECURSOR DISCOVERY

As discussed in section II, precursor discovery uses the threshold τ to determine the significance of data towards the predicted label $\hat{\mathbb{Y}}_n$. A threshold of $\tau \geq 0.7$ is used to visualize the events. Fig. 3-6 shows an example of an event and how the model shows precursors. Each figure shows a heat map of how the probabilities are captured, where the event location is marked by the power tower symbol and is marked as the number 5 site in Fig. 3. Each figure represents a heat map, where the heat map represents the probabilities learned by each model $\theta^{\acute{\gamma}}$ for each of the neighbors' $\acute{\gamma}$ where $\acute{\gamma} \in \gamma$ and the event location. By examining the probabilities P_i across locations $\acute{\gamma} \in \gamma$ and the event location, the model can visualize how each location (through its model) is analyzing its local data at a specific time. In Figs 3-6, the areas with higher outage probability are colored darker. This visualization aims to demonstrate the discovery process of event precursors over time. This illustrates how grid operators can monitor the locations and observe how the models discover and highlight the patterns. Each figure represents the probabilities for all locations at a certain time. This can be considered a precursor, where multiple locations determined that the probabilities at this time (12 hours ahead) are significant enough. As the time progresses, 9 hours before the event and as shown in Fig. 4, the larger outage probabilities shift towards the event location.

As time progresses, Fig. 4 shows that location 1 has a lighter color than Fig. 3 (representing lower probability), but location 3 is still darker. This indicates that the probabilities are shifting south from locations 1 to 3. Fig. 5, which represents the probabilities 3 hours before the event, shows probabilities shifting even closer to the event location and getting darker in color (indicating higher outage risk). Location 4 shows a dark color now, while location 1 has an insignificant probability.

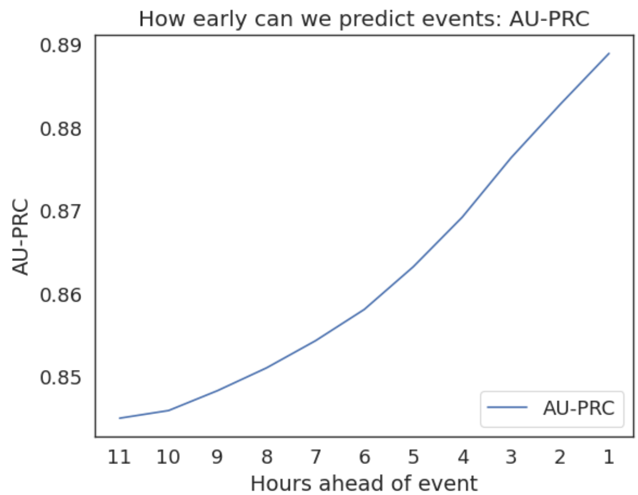

FIGURE 2. Model early detection represented by AU-PRC. X-axis represents the hours ahead of the event where the prediction is made.

Finally, Fig. 6 shows the outage probabilities 1 hour before the event, with the highest probabilities at the event location (location 5). At this time, one can notice the darker and more widespread probabilities. The progress of change in outage probabilities is considered as the spatiotemporal precursors of the outage, where hours before the event, one can see how the probabilities are shifting with a specific pattern. As it gets closer, the probabilities concentrate in one location and get darker as time progresses. Fig. 3-6 visualize how the model captured the spatiotemporal relationships between the locations. The model reliably captured how the probabilities' progress started early (and at a relatively far location), then shifted toward the event location. One could see such information useful to a human operator, where large amounts of data are streamed to control rooms.

The proposed model can analyze, combine and summarize multiple data streams using simple heat maps. Such information can be part of many data streams shown to the grid operators. On the other hand, other data streams can be used as inputs to the proposed methodology. When such information is utilized, it can provide a more complete picture of the status of the power grid. This research only utilized the datasets that could be reliably accessed in a research setting.

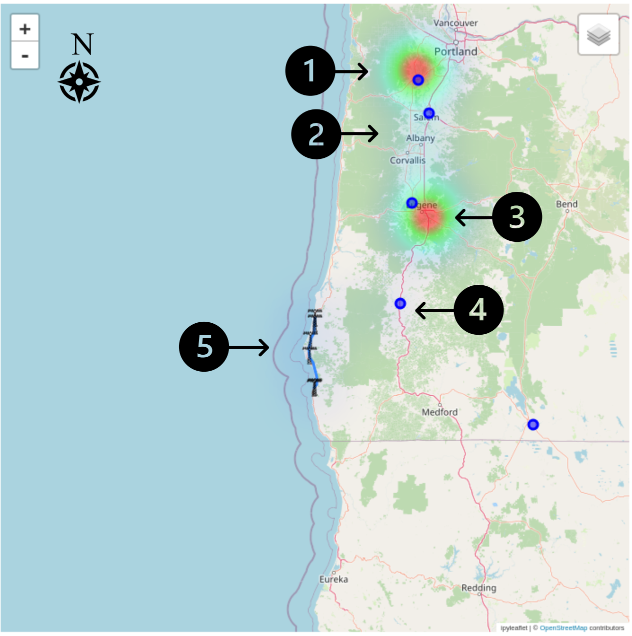

FIGURE 3. Heatmap of prediction probabilities 12 hours before the event. The event location is marked with the tower symbols. The blue dots are weather stations.

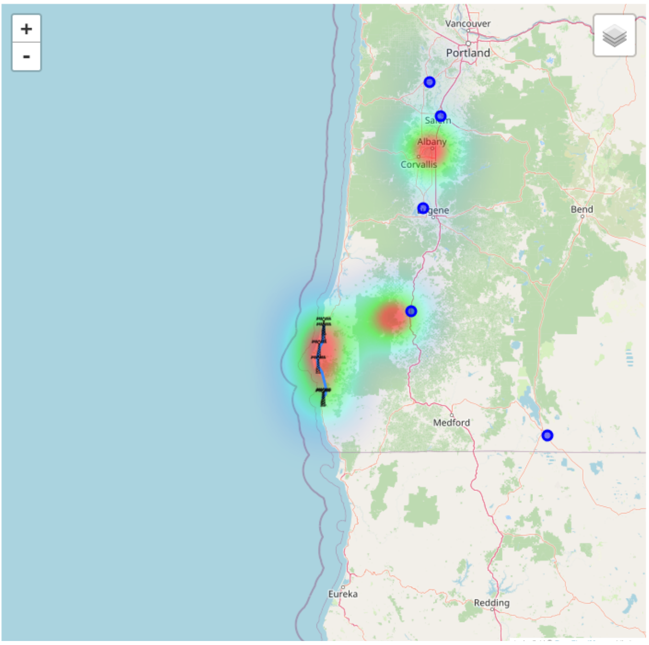

FIGURE 5. Heatmap of prediction probabilities 3 hours before the event. The event location is marked with the tower symbols. The blue dots are weather stations.

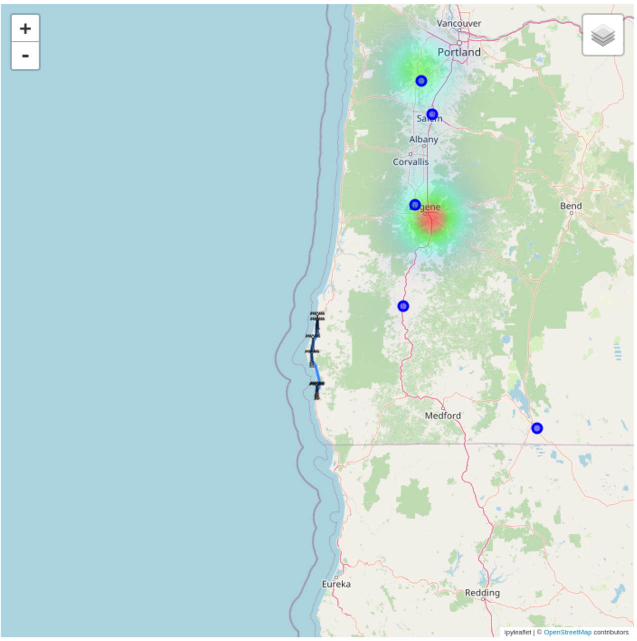

FIGURE 4. Heatmap of prediction probabilities 9 hours before the event. The event location is marked with the tower symbols. The blue dots are weather stations.

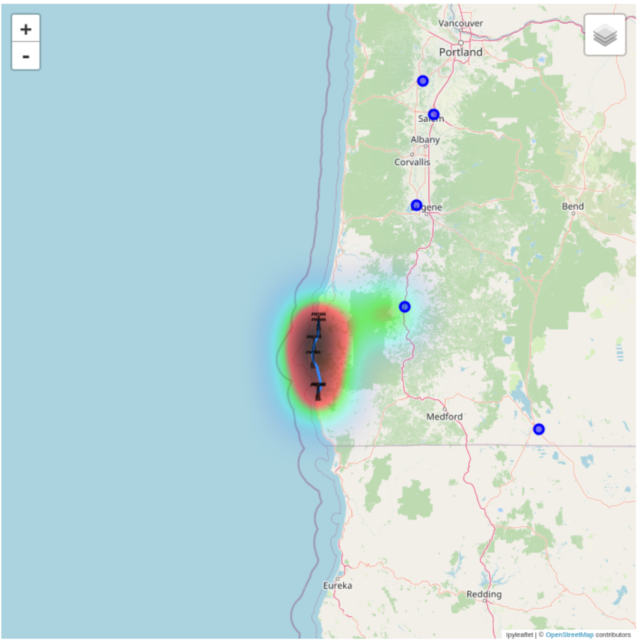

FIGURE 6. Heatmap of prediction probabilities 1 hour before the event. The event location is marked with the tower symbols. The blue dots are weather stations.

V. CONCLUSION

To conclude, this study:

- proposed a novel methodology for spatiotemporal event detection and precursor discovery designed to help power grid operators analyze large amounts of data and utilize it for power outage mitigation.
- showed how to use the proposed methodology with minimal preprocessing and without the need for feature engineering.
- showed that the proposed method can predict events with good performance, and seasonal models can improve detection accuracy.
- showed how the proposed model can capture spatiotemporal relations and expose event precursors.
- showed how to utilize multi-level data for predicting events and how to achieve event detection and precursor discovery at the same time.
- demonstrated how early detection could be achieved with good performance.

The introduced models are designed to be generalizable to other components of the power system, such as generation and load centers. The proposed models can be used to predict component failures and power generation issues beyond the scope of this study. By leveraging the proposed models, it becomes feasible to predict potential component failures within the power system infrastructure and anticipate power generation issues that may arise. However, it should be noted that the experiments necessary to validate the effectiveness of these models in those specific applications were conducted outside of this particular study. The broader applicability and performance of the models in predicting component failures and power generation issues would require further basic research and empirical evaluation.

The data used in this study reports outages on a transmission line granularity. In the preprocessing steps, each outage was mapped to the nearest substations. This can be a limitation since some transmission lines cross long distances. To our knowledge, no public data is available that reports outages on a more granular scale. With such data, better models could be trained. Furthermore, the proposed model doesn't have access to internal data about the electric grid, such as PMU data. If such data is available, it can give a better picture of the internal state of the grid at a local level.

VI. FUTURE WORK

This paper utilized non-parametric attention to capture relations within the power grid. One future step is to learn a joint attention model that can, in parallel, learn relationships within the studied space. On the data side, one dataset that could be helpful is temporal weather forecasts. Such data can be utilized within the same model, further improving the detection. Another area of future research is to explore prediction for longer time horizons (24, 36 hours), which allows for better planning by utility companies.

ACKNOWLEDGMENT

The research was sponsored in part by the United States Army Core of Engineers (USACE) Engineer Research and Development Center (ERDC) Geospatial Research Laboratory (GRL). It was accomplished under Cooperative Agreement Federal Award Identification Number (FAIN) W9132V-22-2-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of USACE EDRC GRL or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. Partial support by Temple University office of the Vice President for Research 2022 Catalytic Collaborative Research Initiative Program AI/ML Focus Area is acknowledged.

REFERENCES

[1] Executive office of the United States President, The White House "Economic Benefits of Increasing Electric Grid Resilience to Weather Outages" August 2013, accessed May 21, 2023. [Online]. Available: https://www.energy.gov/articles/economic-benefits-increasing-electric-grid-resilience-weather-outages

[2] Do, Vivian, Heather McBrien, Nina M. Flores, Alexander J. Northrop, Jeffrey Schlegelmilch, Mathew V. Kiang, and Joan A. Casey. "Spatiotemporal distribution of power outages with climate events and social vulnerability in the USA." Nature communications 14, no. 1 (2023): 2470.

[3] The United States Federal Emergency Management Agency (FEMA). "Community Lifelines." Accessed May 21, 2023. [Online]. https://www.fema.gov/emergency-managers/practitioners/lifelines

[4] Larson, Eric. "Surging Weather-related Power Outages". Climate Central. [Online]. Accessed May 21, 2023. https://www.climatecentral.org/climate-matters/surging-weather-related-power-outages

[5] Zhao, Liang. "Event prediction in the big data era: A systematic survey." ACM Computing Surveys (CSUR) 54, no. 5 (2021): 1-37.

[6] R. Nateghi, S. Guikema, and S.M. Quiring, "Power outage estimation for tropical cyclones: Improved accuracy with simpler models," Risk Anal., vol. 34, no. 6, pp. 1069–1078, Jun. 2014.

[7] D. W. Wanik, E. N. Anagnostou, B. M. Hartman, M. E. B. Frediani, and M. Astitha, "Storm outage modeling for an electric distribution network in northeastern USA," Natural Hazards, vol. 79, no. 2, pp. 1359–1384, Nov. 2015.

[8] D. Cerrai et al., "Predicting Storm Outages Through New Representations of Weather and Vegetation," in IEEE Access, vol. 7, pp. 29639-29654, 2019, doi: 10.1109/ACCESS.2019.2902558.

[9] Eskandarpour, Rozhin, and Amin Khodaei. "Machine learning based power grid outage prediction in response to extreme events." IEEE Transactions on Power Systems 32, no. 4 (2016): 3315-3316.

[10] He, Jichao, David W. Wanik, Brian M. Hartman, Emmanouil N. Anagnostou, Marina Astitha, and Maria EB Frediani. "Nonparametric tree-based predictive modeling of storm outages on an electric distribution network." Risk Analysis 37, no. 3 (2017): 441-458.

[11] F. Yang, P. Watson, M. Koukoula and E. N. Anagnostou, "Enhancing Weather-Related Power Outage Prediction by Event Severity Classification," in IEEE Access, vol. 8, pp. 60029-60042, 2020, doi: 10.1109/ACCESS.2020.2983159.

[12] Dokic, Tatjana, and Martin Pavlovski. "Spatially aware ensemble-based learning to predict weather-related outages in transmission." In The Hawaii International Conference on System Sciences–HICSS, Maui, Hawaii, January 2019.

[13] Li, Zhiguo, Amith Singhee, Haijing Wang, Abhishek Raman, Stuart Siegel, Fook-Luen Heng, Richard Mueller, and Gerard Labut. "Spatio-temporal forecasting of weather-driven damage in a distribution system." In 2015 IEEE Power & Energy Society General Meeting, pp. 1-5. IEEE, 2015.

[14] Darwish, Mohamed MF, Mohamed HA Hassan, Nagat MK Abdel-Gawad, and Diaa-Eldin A. Mansour. "A new method for estimating transformer health index based on ultraviolet-visible spectroscopy." In 2022 23rd International Middle East Power Systems Conference (MEPCON), pp. 1-5. IEEE, 2022.
[15] Darwish, Mohamed MF, Mohamed HA Hassan, Nagat MK Abdel-Gawad, and Diaa-Eldin A. Mansour. "Application of infrared spectroscopy for discrimination between electrical and thermal faults in transformer oil." In 2022 9th International Conference on Condition Monitoring and Diagnosis (CMD), pp. 255-258. IEEE, 2022.
[16] Ghoneim, Sherif SM, Sobhy S. Dessouky, Ahmed Boubakeur, Adel A. Elfaraskoury, Ahmed B. Abou Sharaf, Karar Mahmoud, Matti Lehtonen, and Mohamed MF Darwish. "Accurate insulating oil breakdown voltage model associated with different barrier effects." Processes 9, no. 4 (2021): 657.
[17] Xu C, Liao Z, Li C, Zhou X, Xie R. Review on Interpretable Machine Learning in Smart Grid. Energies. 2022; 15(12):4427. https://doi.org/10.3390/en15124427
[18] O. F. Eikeland, I. S. Holmstrand, S. Bakkejord, M. Chiesa and F. M. Bianchi, "Detecting and Interpreting Faults in Vulnerable Power Grids With Machine Learning," in IEEE Access, vol. 9, pp. 150686-150699, 2021, doi: 10.1109/ACCESS.2021.3127042.
[19] M. Alqudah, M. Kezunovic, and Z. Obradovic, “Automated Power System Fault Prediction and Precursor Discovery Using Multi-Modal Data,” in IEEE Access, vol. 11, pp. 7283-7296, 2023, doi: 10.1109/ACCESS.2022.3233219.
[20] Kotzias, Dimitrios, Misha Denil, Nando De Freitas, and Padhraic Smyth. “From group to individual labels using deep features.” In Proceedings of the 21st ACM SIGKDD international conference on knowledge discovery and data mining, pp. 597-606. 2015.
[21] Ning, Yue, Sathappan Muthiah, Huzefa Rangwala, and Naren Ramakrishnan. “Modeling precursors for event forecasting via nested multi-instance learning.” In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1095-1104. 2016.
[22] Ning, Yue, Rongrong Tao, Chandan K. Reddy, Huzefa Rangwala, James C. Starz, and Naren Ramakrishnan. “STAPLE: Spatio-temporal precursor learning for event forecasting.” In Proceedings of the 2018 SIAM International Conference on Data Mining, pp. 99-107. Society for Industrial and Applied Mathematics, 2018.
[23] Zhang, Yu, and Qiang Yang. "A survey on multi-task learning." IEEE Transactions on Knowledge and Data Engineering 34, no. 12 (2021): 5586-5609.
[24] Evgeniou, Theodoros, and Massimiliano Pontil. "Regularized multi-task learning." In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 109-117. 2004.
[25] ASOS Network. Iowa Environmental Mesonet, Iowa State University. https://mesonet.agron.iastate.edu/ASOS/. Retrieved May 31, 2022.
[26] Hourly Electric Grid Monitor, U.S. Energy Information Administration. https://www.eia.gov/electricity/gridmonitor/dashboard/electric_overview/US48. Retrieved May 28, 2023.
[27] Western Interconnection, Western Electricity Coordinating Council. https://www.wecc.org/epubs/StateOfTheInterconnection/Pages/Western-Interconnection.aspx. Retrieved May 28, 2023.
[28] BPA Geospatial Open Data. Bonneville Power Administration. https://data-bpagis.hub.arcgis.com/. Retrieved May 28, 2023.
[29] Wang, Yaoshu, Jianbin Qin, and Wei Wang. “Efficient approximate entity matching using Jaro-Winkler distance.” In Web Information Systems Engineering–WISE 2017: 18th International Conference, Puschino, Russia, October 7-11, 2017, Proceedings, Part I, pp. 231-239. Cham: Springer International Publishing, 2017.

Mohammad Alqudah (S’20) received B.Sc. from Jordan University of Science and Technology, Irbid, Jordan, in 2012 and an M.S. degree from Binghamton University, NY, in 2015. Currently, he is a Ph.D. candidate in Computer and Information Sciences at the Center for Data Analytics and Biomedical Informatics at Temple University, PA. His main research interests are Machine Learning for spatiotemporal and graph data and the application of machine learning in smart grids.

Zoran Obradovic (S’85-M’87-SM’91) is a Distinguished Professor and a Center director at Temple University, an Academician at the Academia Europaea (the Academy of Europe), and a Foreign Academician at the Serbian Academy of Sciences and Arts. He mentored 45 postdoctoral fellows and Ph.D. students, many of whom have independent research careers at academic institutions and industrial research labs. Zoran is the editor-in-chief of the Big Data journal and the steering committee chair for the SIAM Data Mining conference. He is also an editorial board member of 13 journals and was the general chair, program chair, or track chair for 11 international conferences. His research results were published in about 400 data science and complex networks articles addressing challenges related to big, heterogeneous, spatial, and temporal data analytics motivated by applications in healthcare management, power systems, earth, and social sciences.