Early Prediction of Power Outage Duration Through Hierarchical Spatiotemporal Multiplex Networks

Rafaa Aljurbua[1,2](✉), Jumanah Alshehri[1,3], Shelly Gupta[1], Abdulrahman Alharbi[1,4], and Zoran Obradovic[1]

[1] Center for Data Analytics and Biomedical Informatics, Computer and Information Science Department, Temple University, Philadelphia, PA, USA
{rafaa.aljurbua,shehri.j,shelly.gupta,adalharbi, zoran.obradovic}@temple.edu

[2] Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia

[3] College of Business Administration, Imam Abdulrahman bin Faisal University, Dammam, Saudi Arabia

[4] Department of Computer Science, College of Engineering and Computer Science, Jazan University, Jazan, Saudi Arabia

https://dabi.temple.edu

Abstract. Long power outages caused by weather can have a big impact on the economy, infrastructure, and quality of life in affected areas. It's hard to provide early and accurate warnings for these disruptions because severe weather often leads to missing weather recordings, making it difficult to make learning-based predictions. To address this challenge, we have developed HMN-RTS, a hierarchical multiplex network that classifies disruption severity by temporal learning from integrated weather recordings and social media posts. This new framework's multiplex network layers gather information about power outages, weather, lighting, land cover, transmission lines, and social media comments. Our study shows that this method effectively improves the accuracy of predicting the duration of weather-related outages. The HMN-RTS model improves 3 h ahead outage severity prediction, resulting in a 0.76 macro F1-score vs 0.51 for the best alternative for a five-class problem formulation. The HMN-RTS model provides useful predictions of outage duration 6 h ahead, enabling grid operators to implement outage mitigation strategies promptly. The results highlight the HMN-RTS's ability to offer early, reliable, and efficient risk assessment.

Keywords: power outage · multiplex networks · social media · spatiotemporal learning

1 Introduction

Weather conditions like freezing rain are often critical in predicting power outage risk and estimating the durations of power service disruptions, causing substan-

H. Cherifi et al. (Eds.): COMPLEX NETWORKS 2024, SCI 1189, pp. 320–334, 2025.
https://doi.org/10.1007/978-3-031-82435-7_26

tial economic losses and affecting the quality of human life [13]. Previous research studied and analyzed the power outage problem from a statistical point of view, such as using quantile regression forests and Bayesian additive regression tree models [36]. In the new era of machine learning, research has moved toward predictive models, which use machine learning models to improve predictions of power outages and their causes [7,22,35]. For instance, a Random Forest is recently used to predict the risk of outage disruption [17]. Meanwhile, another disruption risk predictor is developed based on Bayesian deep learning in the KSTAR disruption database [16]. An adaptive ensemble learning approach is also applied for power outage prediction, while a hybrid mechanistic-machine learning outage risk prediction model is proposed to assess the effectiveness of various grid hardening actions [11]. Another study was focused on improving outage prediction risk models caused by a specific weather event such as rainfall [19]. Big data recorded by multiple phasor measurement units (PMUs) is used to improve the detection accuracy of outages in power systems [20]. Previous studies also utilized graph-based models to predict power outages [5,28]. However, none of these studies utilized multiplex networks to enhance predictions; instead, they were restricted to using single-modality input.

Power outage predictions depend crucially on weather data, which often lacks many value during severe weather conditions. Previous studies explored methods to address this challenge, but the potential benefits of combining noisy weather data with information extracted from social media posts should be more adequately studied. In recent years, social media networks have become an integral part of society, allowing critical information to be communicated rapidly during severe weather events [27]. It has been found that weather conditions significantly affect tweeting behavior [15]. Social media data can provide information on weather-related impacts on infrastructure and human behavior and can also provide information back to observers [34]. In addition, a correlation was found between the number of weather-related tweets and the existing weather conditions [33]. Therefore, combining social media data with weather sensor data can provide valuable insights in predicting power outages caused by severe weather events.

While predicting power outages is crucial, it is essential to predict the duration of these outages to implement effective response and mitigation strategies. Several statistical methods were used to predict power outage duration [24]. Others applied machine learning models [6]. However, researchers have yet to fully explore the benefits of multiplex networks that utilize data from multiple modalities to predict power outage duration. In our previous study [1], multiplex graphs successfully predicted the occurrence of power outages three hours in advance at the county level. Still, our previous analysis did not include the duration of the outages. This limitation is addressed in our current study by developing HMN-RTS, an approach that forecasts the likelihood of power outages and their durations while expanding the prediction time intervals at the county level within the entire U.S. Pacific Northwest region. The HMN-RTS is a hierarchical spatiotemporal multiplex network model that leverages structured data, such as weather and transmission line information, and unstructured data from social sensors collected across time and space. The HMN-RTS takes real-time input to update power outage duration prediction three hours in advance based on weather conditions and social media activity changes. This model provides earlier warnings of outage risks, potentially enhancing outage management efficiency and response efforts. The main contributions of this paper are as follows:

1. This study develops a principal framework for improving spatiotemporal classification by integrating multimodal data. It incorporates social media information to address the gaps in weather recording data.
2. A hierarchical county-level, spatiotemporal multiplex network multi-modal HMN-RTS approach is proposed to predict the duration of a power outage three hours ahead.
3. The HMN-RTS model updates the power outage duration prediction in real-time based on changes in weather and social media activity.
4. The HMN-RTS's predictive capabilities for estimating the duration of probable outages are evaluated across multiple time horizons, ranging from 3 to 24 h in advance.

2 Related Work

Three significant components summarize the related work for the proposed approach: 1) power outage duration prediction, 2) handling missing and incomplete weather data, and 3) hierarchical spatiotemporal multiplex network and multi-modal data development for power outage duration prediction.

2.1 Weather Data

The Automated Surface Observing Systems (ASOS) ceilometer measures clouds at or below $12,000$ ft (3.6 km), compromising ASOS data accuracy. Thus, missing data would result from incomplete atmospheric coverage [37]. Additionally, missing data is a common issue when dealing with sensor malfunctions and cloud contamination [10]. Numerous studies have examined various strategies for managing missing data. Mean imputation is the most commonly used method [32]. Common alternatives are disregarding records with missing values and replacing missing data with multiple imputations [30] or regression-based estimation [31]. However, missing data should be handled carefully for accurate analysis since ignoring missing instances can pose significant risks in the context of analysis.

2.2 Power Outage Duration Prediction

For utility companies to plan power restoration more effectively, predicting the duration of power outages early and accurately is essential. Statistical methods to predict power outage durations include accelerated failure time regression, Cox proportional hazards regression, Bayesian additive regression trees, regression trees, and multivariate adaptive regression splines [24]. The researchers also combined statistical and geographic information systems to analyze the performance of a winter storm-affected urban distribution system by using a GIS to plot the data of the repair crews to study the duration of each outage [29]. In another study [18], Accelerated Failure Time (AFT) and Cox Proportional Hazard (CPH) were applied to estimate the duration of storm-caused power outages. The daily Night Time Lights (NTL) imagery data is also used to assess Puerto Rico's outage duration [2]. Previous studies show that socioeconomic factors influence outage duration [12,23]. In recent years, researchers have moved toward machine learning models hoping to predict the duration of power outages. For instance, the duration of hurricane-related power outages is predicted by applying a Random Forests-based forecast mode using input variables such as wind duration and wind speed [25]. The duration during typhoon disasters is predicted by integrating Extra Tree (ET) [9], Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Random Forest (RF), Gradient Boosting Regression (GBR), and Decision Tree (DT).

2.3 Multi-modal Learning in Ahierarchical Spatiotemporal Multiplex Network for Predicting Power Outage Duration

Weather data is often crucial for predicting power outages [14]. However, a high incidence of missing values in weather recordings during severe weather complicates learning. Our research combines weather data with social sensor data to address this challenge. We investigate the influence of social networks (multiplex networks) and social sensor data on predicting the duration of power outages. This study is one of the first to address power outage duration prediction three hours in advance by incorporating social sensors within a multiplex network representation. Precisely, we assess the advantages of learning from a spatiotemporal multiplex network that encompasses data from six sources: Bonneville Power Administration power outages, weather, lightning, Bonneville Power Administration transmission lines, land cover, and social sensor data from two prominent platforms: Reddit and Twitter.

3 Methodology

This study aims to assess whether the proposed HMN-RTS hierarchical model, which integrates multiplex networks and multi-modal data, enhances the prediction of power outage duration three hours ahead. The HMN-RTS model operates in two phases: first, predicting the risk for the occurrence of power outages, and second, estimating the severity of the predicted outage by predicting the duration of the expected outage, as shown in Fig. 1. The research is structured in three phases (1) data collection and spatiotemporal graph construction, (2) the development of the spatiotemporal multiplex network that estimates outage risk, and (3) the development of the model that predicts duration for expected outages.

Each component of this hierarchical framework is detailed in the subsequent subsections.

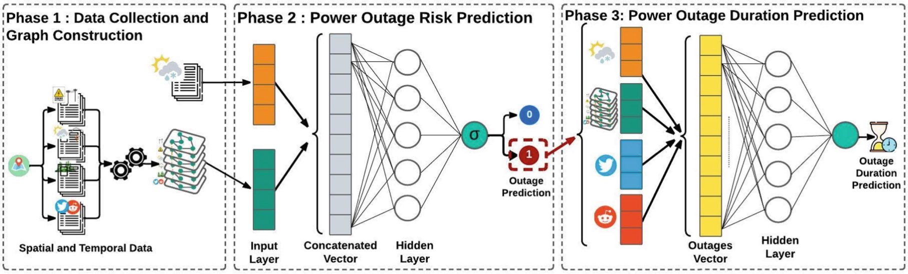

Fig. 1. The architecture of the HMN-RTS hierarchical spatiotemporal multiplex network for multi-modal prediction of power outage duration.

3.1 Data Collection

Previous research has highlighted the benefits of using data from multiple sources to enhance the prediction of power outage severity [1]. As the first step, we identify and collect critical factors to predict outage duration. These factors include power outage data, weather conditions, lightning, land cover, transmission lines, and social sensor data from two prominent platforms: Twitter (now referred to as X) and Reddit. This subsection provides an overview of the data collection process.

1. **Power outages:** We focus on the geographical area of the U.S. Pacific Northwest. We collected transmission services power outage events data covering a territory of more than 15,000 circuit miles over two years, from January 1, 2021, to December 31, 2022. This data is provided by Bonneville Power Administration (BPA), an American federal agency that operates in the U.S. Pacific Northwest. The BPA publicly reports all power outage events, regardless of cause. However, not every outage is relevant to this study; therefore, we only collect weather-related power outages, such as ice and lightning. We identify all weather-related power outages and map them to county and state using our dictionary that links every county to its Federal Information Processing Standards (FIPS) code. This results in 2,411 weather-related outages with a mean duration of 310 minutes[1].
2. **Weather:** We collect historical weather data from Automated Surface Observing Systems (ASOS) stations. These stations are located at airports and include sensors to measure wind, ambient temperature, pressure, obstructions to vision, and sky conditions. We map each station to its county and state using latitude and longitude. As a result, we gather around 39 million weather observations in the five states[2].

[1] https://www.bpa.gov/.
[2] https://www.weather.gov/asos/.

3. **Land cover:** Rapid tree growth or falling trees can damage power lines. To address this issue, we utilize land cover data obtained from the National Historical Geographic Information System (NHGIS). This dataset provides land cover features such as mixed, deciduous, and evergreen forests from the National Land Cover Database (NLCD). We leverage the GISJOIN identifier, which uses the FIPS code to designate counties and states [21].
4. **Lightning:** We obtain lightning information from the National Oceanic and Atmospheric Administration (NOAA) database[3]. As a result, we observe $7,015$ lightning strikes from counties in the five states.
5. **Transmission lines:** This study focuses on transmission line outages rather than distribution line outages. We collect transmission line information in the BPA service region of the Northwest U.S. using the BPA map covering over $15,000$ miles of transmission lines.
6. **Social sensor:** We gather social media activities using weather-related and power-outage-related keywords. **Twitter "currently known as X"**: first, we collect tweets using snscrape. Snscrape is a Python package designed to scrap historical tweets. We scrap tweets within a 10-mile radius of specified geographic coordinates (latitude and longitude) to capture a broad range of posts from the neighborhood. As a result, we collect 8.5 million relevant tweets about weather and power outage events. **Reddit**: we collect Reddit posts from counties' subreddits using Reddit API. We collect all posts and comments from counties' subreddits, followed by filtering the posts and comments using weather and power outage keywords. As a result, we obtained 353,421 posts, of which $95,144K$ were used after filtering and selection based on the keyword selection.

3.2 Modeling of the Spatiotemporal Multiplex Network

We then construct a Hierarchical Spatiotemporal Multiplex Network. Let G denote a spatiotemporal multiplex network defined as $G(V, E, L, T)$, where: $V = \{v_1, v_2, ..., v_n\}$ represents the set of vertices (counties), $E = \{e_1, e_2, ..., e_m\}$ represents the set of edges, $L = \{l_1, l_2, l_3, l_4, l_5, l_6\}$ represents the set of layers, and $T = \{t_1, t_2, ..., t_k\}$ represents a set of time steps. The edges E within each layer L signify a unique type of relationship among the vertices V. These relationships are described as follows:

1. **Transmission lines layer:** In layer l_1, two counties (u_{l_1}, v_{l_1}) are linked if they share the same transmission line. The edge weight represents the number of transmission lines shared between the counties.
2. **Power outage layer:** In layer l_2, two counties (u_{l_2}, v_{l_2}) are linked if both report a power outage on the same date. The edge weight represents the number of shared power outages between the counties.

[3] https://www.ncei.noaa.gov/.

3. **Weather layer:** In some cases, power outages are closely related to weather conditions; therefore, in layer l_3, we connect nodes representing two counties (u_{l_3}, v_{l_3}) that share similar weather properties. In this context, we compute the Euclidean distance between each pair of vertices.
4. **Lightning layer:** In layer l_4, two counties (u_{l_4}, v_{l_4}) are linked if both report a lightning strike on the same day. The edge weight represents the number of shared lightning strikes between the counties.
5. **Land cover layer:** In layer l_5, two counties (u_{l_5}, v_{l_5}) are linked if they share similar land cover properties. Here, we calculate the Euclidean distance between vertices.
6. **Social sensor layers:** In layer l_6, two counties (u_{l_6}, v_{l_6}) are linked if both report social media activities during the power outage. The edge weight represents the number of shared social media activities, such as Tweets or Reddit posts, between the counties.

This graph serves as input for the HMN-RTS model. At the end of each day, a new multiplex graph snapshot, timestamped with that day's data is added to capture the interdependencies between counties. We train the model using the snapshot as input, generating embeddings for the county nodes through the proposed method, which is explained in detail in the following subsection. We aim to predict whether a county will experience a power outage. If an outage is anticipated, we estimate the severity by forecasting its duration. The duration prediction is based on county node embeddings and real-time social media data (from Twitter and Reddit) related to weather conditions and power outages.

3.3 Proposed Model: Hierarchical Spatiotemporal Multiplex Network (HMN-RTS)

This study evaluates whether a hierarchical multiplex network and multi-modal data approach can enhance the prediction of disruption severity three hours in advance. First, we use an annotated dataset to identify weather-related power outages, noting that the frequency of these outages varies across different states and counties. Each data point is labeled as one if a power outage occurs in a county during a specific time frame and 0 otherwise. Second, to forecast outage duration, we classify the outages into five categories, as outlined in Table 1. The model architecture consists of two phases: (1) prediction of the power outage occurrence followed by (2) prediction of the severity of the power outage by estimating the duration of power outages.

Power Outage Risk Prediction. The model uses the multiplex snapshot to generate county node embeddings via a modified version of Node2Vec [8] that incorporates multiple graph layers. These embeddings are then combined with weather data from ASOS and fed into a neural network model consisting of three fully connected layers, with dropout layers in between to mitigate overfitting.

Our dataset is inherently imbalanced, with fewer than two percent of instances classified as outages. We find that the Synthetic Minority Oversampling Technique (SMOTE) [3] is the best technique to address the imbalance issue in the training data. Binary cross entropy is used to train this part of the HMN-RTS architecture. This effectively isolates instances where the model correctly identified outages.

Power Outage Duration Prediction. Once a power outage is predicted for a county, the second part of the model aims to predict the duration of the outage three hours ahead through multiclass classification. First, we gather ASOS weather station recordings within each county at 30-minute intervals. Since there may be multiple stations and recordings per station within a county, we aggregate the data by calculating the mean and standard deviation for each feature, which normalizes the weather data and provides a comprehensive view of local conditions.

Next, we collect all Reddit posts from the U.S. Pacific Northwest region and use BERT [4] to generate 768-dimensional embeddings for each post. We apply max pooling for intervals with multiple posts to create a single representative vector. Similarly, for Twitter data, we gather all tweets posted within each 30-minute interval across the U.S. Pacific Northwest and use BERTweet [26] to generate 768-dimensional vectors for each tweet. Again, we use max pooling to aggregate these vectors into a single representative vector for each interval. This approach effectively captures the most significant patterns from social media, providing the model with a concise yet robust input for each 30-minute segment. Finally, we concatenate the weather data, Reddit vectors, Twitter vectors, and multiplex model embeddings, using this combined dataset as input for the model to predict outage duration. The combined input is fed into a multiclass neural network consisting of three fully connected layers with ReLU activations and interspersed dropout layers to reduce overfitting. The model is trained using multiclass cross-entropy loss to optimize its performance in predicting outage durations across the defined classes. The multiclass cross-entropy loss measures the difference between the predicted probabilities $\hat{y}$ and the actual labels y.

$$\text{Loss}(\hat{y}, y) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{C} y_{ij} \log(\hat{y}_{ij}) \tag{1}$$

Here $\hat{y}_{ij}$ is the predicted probability that the ith sample belongs to the jth class. y_{ij} is the actual label for the ith sample in the jth class, which is one if the ith sample belongs to the jth class and 0 otherwise. N is the total number of samples in the dataset. C is the total number of classes.

4 Experimental Setup

This study aims to determine if hierarchical multiplex network and multi-modal data method can improve early classification of disruption severity to one of five categories (short to very long time). We use two years of data for training and testing. We use data from 2021-01-01 to 2022-06-30 for training and evaluate the model on disjoint data from 2022-07-01 to 2022-12-31. The model is optimized using the *Adam* optimizer and trained for 100 epochs with a batch size 32 and a learning rate of 0.0001. We conduct machine learning experiments to predict the duration of disruptions caused by the occurrence of power outages. Our learning approach is supervised machine learning based. We compare the HMN-RTS model vs Neural Network (NN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Reddit and Twitter Multiplex Network (RTMNO) [1]. We choose metrics suitable for a power system setting to evaluate the model's performance. We select metrics appropriate for a power system context to assess model performance.

Given that our classification involves five classes, we use macro-averaging, which assigns equal weight to each class irrespective of the class frequency. We use macro precision and recall assessing the rates of false positives and false negatives. Moreover, we evaluate the model's performance on the test set using the macro F1 scores for each class. Given that C represents the number of classes, macro F1 can be defined as

$$\text{Macro F1} = \frac{1}{C}\sum_{i=1}^{C} 2 \cdot \frac{\text{precision}_i \cdot \text{recall}_i}{\text{precision}_i + \text{recall}_i} \tag{2}$$

Table 1. Distribution of power outage durations in BPA service territory across five classes of duration in the years 2021-2022.

Class	Duration	Percentage
Class 1	Less than 30 min	63%
Class 2	30 min to 1 h	1%
Class 3	1 to 3 h	8%
Class 4	3 to 6 h	6%
Class 5	Greater than 6 h	22%

5 Results and Discussion

The results for different evaluation metrics of NN, RNN, LSTM, RTMNO, and the proposed Hierarchical Multiplex Network model (HMN-RTS) are shown in Table 2. The performance of the HMN-RTS model outperformed the performance of the traditional models. In addition, the HMN-RTS model obtains a higher macro F1 score in all models considered. As a result, the Hierarchical Multiplex Network model achieves a macro F1 score of 0.76. Further, we can observe that the RTMO model enhances the prediction performance with performance improvement up to 30% versus alternative models. In contrast, despite optimizing the baseline models (NN, RNN, and LSTM), they resulted in macro F1 scores of only 0.16, 0.19, and 0.21, respectively. This can be explained by the limited capabilities of baseline models to capture context compared to our HMN-RTS model.

Table 2. Comparison of macro precision, macro recall, and macro F1 score of Neural Network (NN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Reddit and Twitter Multiplex Network (RTMNO), and the Hierarchical Multiplex Network model (HMN-RTS). Outage duration is classified into five classes defined in Table 1, and the models are evaluated on BPA outage data from 2022-07-01 to 2022-12-31.

Model	Macro precision	Macro recall	Macro F1 Score
NN	0.15	0.21	0.16
RNN	0.17	0.22	0.19
LSTM	0.18	0.24	0.21
RTMNO[1]	0.55	0.50	0.51
HMN-RTS	**0.83**	**0.86**	**0.76**

Further, we evaluate the HMN-RTS model performance using a confusion matrix, which illustrates the number of correct and incorrect predictions for each class, as shown in Fig. 2. Note that Class 1 means duration less than 30 min, Class 2 is for duration between 30 mins to 1 h, Class 3 is for duration between 1 to 3 h, Class 4 is for duration between 3 to 6 h, and Class 5 is for duration greater than 6 h. Due to the ordinal nature of our classes, the model can distinguish between classes with small intervals, while the model struggles to distinguish between classes with high intervals. In the HMN-RTS model, the macro F1 score is lower than macro precision and macro recall because it considers both precision and recall in all classes and is sensitive to lower values in either metric. To investigate this phenomenon, we examine the prediction of each class. Table 3 shows the results of the HMN-RTS model per class. While precision and recall are relatively high for most classes, the lower precision 0.14 and recall 0.45 values cause the macro F1 score to drop significantly, reflecting the harmonic mean's sensitivity to lower values.

The second set of experiments performs an ablation study. We create five different versions of HMN-RTS for this experiment. HMN-R is the model that only utilizes Reddit information. It is modified to process Reddit data only. HMN-T is the model that only uses Twitter information. It is adjusted to handle only Twitter data. HMN-S is the model that only employs multiplex network structure. The model is adjusted to handle multiplex network data. HMN-RT is the model that uses both Reddit and Twitter information, modified to process Reddit and Twitter data. Finally, HMN-RTS is the model that employs Reddit,

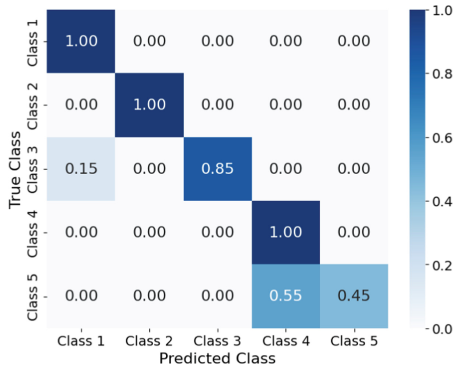

Fig. 2. Normalized confusion matrix of the HMN-RTS model predicting duration of outages for 2022-07-01 to 2022-12-31. Class 1 means duration less than 30 min, Class 2 is for duration between 30 minutes to 1 h, Class 3 is for duration between 1 to 3 h, Class 4 is for duration between 3 to 6 h, and Class 5 is for duration greater than 6 h.

Table 3. Comparison of precision, recall, and F1 score for every class using the HMN-RTS model trained on data from 2021-01-01 to 2022-06-30 and evaluated on 2022-07-01 to 2022-12-31.

Class	Precision	Recall	F1 Score
Class 1	0.99	1	0.99
Class 2	1	1	1
Class 3	1	0.85	0.92
Class 4	0.145	1	0.25
Class 5	1	0.45	0.62

Twitter, and the multiplex network structure. Table 4 outlines the model's performance. It outperforms all other configurations when the HMN-RTS leverages multi-modal data and the network structure.

The third experiment evaluates the HMN-RTS model's ability to predict outcomes at earlier stages. In power systems, predictions made earlier provide more significant benefits. Figure 3 shows the macro F1 score for early prediction. The macro F1 score indicates the HMN-RTS model's performance in early detection. The X-axis shows the number of hours before the power outage when the prediction is made. We can see that the macro F1 score increases as the prediction time approaches the power outage event. This is expected, as impacts are generally more predictable over shorter periods. We can also observe that the proposed HMN-RTS model can predict power outages with a high macro F1 score up to 6 hours before the event. Furthermore, it can make predictions up to 12 hours in advance with slightly reduced performance, which is still adequate for taking appropriate measures to mitigate the side effects of the outage.

Table 4. Ablation study of the HMN-RTS model. Multiplex Network Structure includes layers 1–5 defined in Sect. 3.2 while Reddit and Twitter activities are captured by separate layers.

Settings	Reddit	Twitter	Multiplex Network Structure	Macro precision	Macro recall	Macro F1 Score
HMN-R	✓			0.66	0.73	0.66
HMN-T		✓		0.63	0.79	0.68
HMN-S			✓	0.67	0.80	0.71
HMN-RT	✓	✓		0.83	0.82	0.74
HMN-RTS	✓	✓	✓	**0.83**	**0.86**	**0.76**

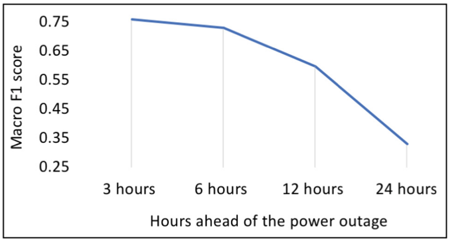

Fig. 3. The HMN-RTS model performance in early detection of outages at the BPA service territory for 2022-07-01 to 2022-12-31 is indicated by the macro F1 score for a five-class problem formulation defined at Table 1. The X-axis shows the number of hours before the power outage when the prediction is made.

6 Conclusion

Power outages can affect daily life, such as transportation and communication. Therefore, predicting the severity of power outages is critical for effective planning. This study proposes a county-level hierarchical multiplex network-based methodology in the U.S. Pacific Northwest. The proposed approach predicts the occurrence of weather-related power outages along with the power outage duration. The HMN-RTS model predicts power outage occurrences and severity across multiple time horizons, including 3, 6, 12, and 24-hour intervals. We use multi-modal data collected over time and space to provide earlier predictions of outage duration. We quantify the benefit of integrating weather-related disruption information with people's behavior.

Consequently, we consider information derived from social sensors. We assess whether the proposed hierarchical multiplex network method can learn better and enhance the prediction performance compared to other machine learning models. Our hierarchical spatiotemporal multiplex network provides a novel way of encoding online social sensor data into a network-based approach that can improve three hours ahead duration prediction accuracy for power outages with a macro F1 score of 0.76, enabling grid operators to implement outage mitigation strategies promptly. The limitation of this work is that the distribution of user ages impacts how they engage on social media platforms, which should be examined in a follow-up study. Future research can explore social sensor data from a broader range of age demographics. In addition, a follow-up study is needed to evaluate our model across different geographic locations and extended prediction horizons (e.g., 36 hours).

Acknowledgment. This research was sponsored by the U. S. Army Engineer Research and Development Center (ERDC) and was accomplished under Cooperative Agreement Number W9132V-23-2-0002. 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 the U.S. Army Research Engineer and Development Center (ERDC) or the U.S. Government.

References

1. Aljurbua, R., Alshehri, J., Alharbi, A., Power, W., Obradovic, Z.: Social media sensors for weather-caused outage prediction based on spatio-temporal multiplex network representation. IEEE Access **11**, 125883–125896 (2023)
2. Azad, S., Ghandehari, M.: A study on the association of socioeconomic and physical cofactors contributing to power restoration after hurricane Maria. IEEE Access **9**, 98654–98664 (2021)
3. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
4. Devlin, J.: BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
5. Dokic, T., Pavlovski, M.: 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 (2019)
6. Doshi, R., Saini, R.D., Kansal, S.: Time duration prediction of electrical power outages. In: Smart Structures in Energy Infrastructure: Proceedings of ICRTE 2021, vol. 2, pp. 31–40. Springer (2022)
7. Eskandarpour, R., Khodaei, A.: Machine learning based power grid outage prediction in response to extreme events. IEEE Trans. Power Syst. **32**(4), 3315–3316 (2016)
8. Grover, A., Leskovec, J.: Node2Vec: scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 855–864 (2016)
9. Hou, H., Liu, C., Wei, R., He, H., Wang, L., Li, W.: Outage duration prediction under typhoon disaster with stacking ensemble learning. Reliab. Eng. Syst. Safety **237**, 109398 (2023)
10. Huang, W., Deng, Y., Hui, S., Wang, J.: Image inpainting with bilateral convolution. Remote Sens. **14**(23), 6140 (2022)
11. Hughes, W., et al.: Assessing grid hardening strategies to improve power system performance during storms using a hybrid mechanistic-machine learning outage prediction model. Reliab. Eng. Syst. Safety **248**, 110169 (2024)
12. Jasiūnas, J., Lund, P.D., Mikkola, J., Koskela, L.: Linking socio-economic aspects to power system disruption models. Energy **222**, 119928 (2021)
13. Jones, K.F.: Freezing fraction in freezing rain. Weather Forecast. **37**(1), 163–178 (2022)

14. Kezunovic, M., Pinson, P., Obradovic, Z., Grijalva, S., Hong, T., Bessa, R.: Big data analytics for future electricity grids. Electr. Power Syst. Res. **189**, 106788 (2020)
15. Kıcıman, E.: OMG, I have to tweet that! a study of factors that influence tweet rates. In: Proceedings Of The International AAAI Conference On Web And Social Media, vol. 6, pp. 170–177 (2012)
16. Kim, J., Lee, J., Seo, J., Ghim, Y.C., Lee, Y., Na, Y.S.: Enhancing disruption prediction through Bayesian neural network in KSTAR. Plasma Phys. Control. Fusion **66**(7), 075001 (2024)
17. Lee, J., et al.: Data-driven disruption prediction using random forest in KSTAR. Fusion Eng. Des. **199**, 114128 (2024)
18. Liu, H., Davidson, R.A., Apanasovich, T.V.: Statistical forecasting of electric power restoration times in hurricanes and ice storms. IEEE Trans. Power Syst. **22**(4), 2270–2279 (2007)
19. Liu, W., Yang, Y., Xu, Q., Xia, Y.: Multi-target prediction model of urban distribution system rainfall-caused outage based on spatiotemporal fusion. Int. J. Electr. Power Energy Syst. **146**, 108640 (2023)
20. Ma, H., Lei, X., Li, Z., Yu, S., Liu, B., Dong, X.: Deep-learning based power system events detection technology using spatio-temporal and frequency information. IEEE J. Emerg. Sel. Topics Circ. Syst. **13**(2), 545–556 (2023)
21. Manson, S.M.: IPUMS national historical geographic information system: version 15.0 (2020)
22. Mensah, A.F., Dueñas-Osorio, L.: Outage predictions of electric power systems under hurricane winds by Bayesian networks. In: 2014 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), pp. 1–6. IEEE (2014)
23. Mitsova, D., Escaleras, M., Sapat, A., Esnard, A.M., Lamadrid, A.J.: The effects of infrastructure service disruptions and socio-economic vulnerability on hurricane recovery. Sustainability **11**(2), 516 (2019)
24. Nateghi, R., Guikema, S.D., Quiring, S.M.: Comparison and validation of statistical methods for predicting power outage durations in the event of hurricanes. Risk Anal. Int. J. **31**(12), 1897–1906 (2011)
25. Nateghi, R., Guikema, S.D., Quiring, S.M.: Forecasting hurricane-induced power outage durations. Nat. Hazards **74**, 1795–1811 (2014)
26. Nguyen, D.Q., Vu, T., Nguyen, A.T.: BERTweet: a pre-trained language model for English Tweet. arXiv preprint arXiv:2005.10200 (2020)
27. Otudi, H., Gupta, S., Albarakati, N., Obradovic, Z.: Classifying severe weather events by utilizing social sensor data and social network analysis. In: Proceedings of the International Conference on Advances in Social Networks Analysis and Mining, pp. 64–71 (2023)
28. Owerko, D., Gama, F., Ribeiro, A.: Predicting power outages using graph neural networks. In: 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp. 743–747. IEEE (2018)
29. Reed, D.A.: Electric utility distribution analysis for extreme winds. J. Wind Eng. Ind. Aerodyn. **96**(1), 123–140 (2008)
30. Robertson, J., Kaptein, M.: Modern statistical methods for HCI, vol. 6. Springer (2016)
31. Streiner, D.V.L.: The case of the missing data: methods of dealing with dropouts and other research vagaries. Canadian J. Psychiatry **47**(1), 68–75 (2002)
32. Tanguma, J.: A review of the literature on missing data (2000)

33. Uddin, M.Y.S., Al Amin, M.T., Le, H., Abdelzaher, T., Szymanski, B., Nguyen, T.: On diversifying source selection in social sensing. In: 2012 Ninth International Conference on Networked Sensing (INSS), pp. 1–8. IEEE (2012)
34. Weyrich, P., Ruin, I., Terti, G., Scolobig, A.: Using serious games to evaluate the potential of social media information in early warning disaster management. Int. J. Disaster Risk Reduction **56**, 102053 (2021)
35. Xu, Z., Liu, J., Fan, W., Wang, Y., Luan, L., Zhou, K.: An power outage prediction method based on XGBoost with improved objective function. In: 2022 37th Youth Academic Annual Conference of Chinese Association of Automation (YAC), pp. 1560–1565. IEEE (2022)
36. Yang, F., Wanik, D.W., Cerrai, D., Bhuiyan, M.A.E., Anagnostou, E.N.: Quantifying uncertainty in machine learning-based power outage prediction model training: a tool for sustainable storm restoration. Sustainability **12**(4), 1525 (2020)
37. Yang, X., You, Z., Hiller, J., Watkins, D.: Updating and augmenting weather data for pavement mechanistic-empirical design using ASOS/AWOS database in Michigan. Int. J. Pavement Eng. **19**(11), 1025–1033 (2018)