Utilizing Multiple Data Sources to Improve Prediction of Severe Weather Events Through Spatio-Temporal Analysis

Hussain Otudi[1,2](✉), Shelly Gupta[1], Ameen Abdel Hai[1], Rafaa Aljurbua[1,3], Abdulrahman Alharbi[1,2], and Zoran Obradovic[1]

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

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

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

Abstract. Predicting rare and disruptive severe weather events presents significant challenges due to class imbalance and data sparsity. Conventional oversampling techniques and unimodal approaches are inadequate for these low-frequency phenomena because they fail to capture the events' intrinsic complexity and spatiotemporal dynamics. Current methods lack the ability to learn modality-specific representations. Herein, we introduce a robust multimodal fusion strategy that directly integrates primary sensor measurements with supplementary modalities including textual descriptions and weather forecasts within a tri-modality framework. Our approach is augmented by advanced spatiotemporal feature engineering, ensuring that both spatial and temporal relationships are preserved and effectively leveraged. Notably, our proposed method, which incorporates Automated Surface Observing System (ASOS) sensor data, textual embeddings, and forecast data, achieves substantial performance improvements, elevating macro F1-scores from 0.04 to 0.89 across a ten-class framework (nine severe event classes and one normal class) for 12-hour forecasting horizons. This integrated approach helps overcome data sparsity, particularly in high-latitude regions. Ultimately, this framework provides an effective early warning system for disaster risk assessment and infrastructure resilience forecasting.

Keywords: Severe Events Classification · Multi-modal Fusion · Spatiotemporal Analysis · Data Sparsity

1 Introduction

Severe weather events are becoming increasingly common and pose significant risks to human life, infrastructure, and the environment. These events including hurricanes, flash floods, severe thunderstorms, and tornado outbreaks cause

L. Iliadis et al. (Eds.): EANN 2025, CCIS 2581, pp. 267–280, 2025.
https://doi.org/10.1007/978-3-031-96196-0_20

widespread damage and require improved predicting and mitigation strategies. Projections indicate significant changes in temperature and precipitation severity throughout the 21st century, with more pronounced increases in extreme minimum temperatures and five-day precipitation amounts compared to mean changes [26]. These climatic shifts are expected to affect various sectors, including infrastructure, food security, and transportation [14].

Alaska is also experiencing an increase in severe climate events, presenting considerable challenges for communities, particularly Alaska Native Villages (ANVs), as illustrated in Fig. 1, posing significant threats to public health and the economy. Although hazard mitigation plans serve as the primary mechanism for addressing climate-related hazards in ANVs, many are overly generic and fail to sufficiently address local concerns [25]. Enhanced planning strategies that incorporate community knowledge, protect subsistence activities, and improve accessibility are necessary. The integration of historical data, climate projections, and localized impact assessments can support more effective climate adaptation planning in Alaska [13]. Predicting severe weather events remains a challenging problem due to the complexity of atmospheric systems. Current predicting methodologies require advancements to improve accuracy and provide actionable insights for emergency response. Traditional models often struggle to account for the dynamic nature of severe weather, necessitating the use of machine learning tools and control systems to enhance prediction capabilities. By integrating heterogeneous environmental data sources including high resolution satellite imagery, radar signals, in situ sensor networks, and numerical weather prediction outputs these systems can offer emergency managers and policymakers the critical lead time necessary for effective risk mitigation. This includes resource allocation, issuing timely evacuation orders, mobilizing emergency response teams, and pre-positioning critical infrastructure.

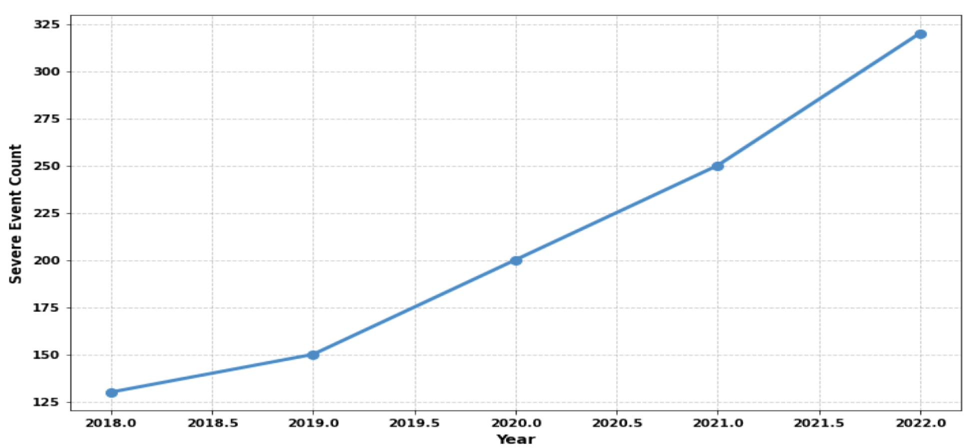

Fig. 1. Observed increasing frequency of severe weather events in Alaska from 2018 to 2022, demonstrating the growing intensity of severe weather events.

This paper presents a variant of a tri-modal fusion framework that integrates high-resolution ASOS sensor data, quantitative NCEI forecasts, and qualitative textual reports from the Storm Events Database to predict severe weather events up to 12 h in advance. Using a late fusion strategy, our model classifies events within a ten-class framework where we have nine severe event classes and one normal class. We employ LSTM and BERT models to extract patterns from different data modalities, enabling early warnings and comprehensive risk assessments. By bridging gaps in Alaskan-specific environmental data, our approach facilitates the robust learning of historical climate patterns associated with severe weather events. This work enhances the reliability of severe weather predictions, ultimately supporting improved decision-making and disaster preparedness strategies. We answer the following research questions in this work-

RQ1: What roles do sensor, forecast, and text modalities play in capturing Alaska's severe weather dynamics?
RQ2: How does our late fusion approach improve prediction accuracy and extend the predicting horizon?
RQ3: How does the model balance recall and precision for effective risk estimation under data limitations?
RQ4: How does a generalized model compare to localized models in capturing severe weather patterns across different regions of Alaska?
RQ5: How does model performance vary across different seasons in severe weather classification?

In this study, we extend a trimodal model to tackle a broader classification challenge in cold, data-scarce regions like Alaska. A key challenge is the presence of certain classes in only one modality, requiring innovative integration of heterogeneous and incomplete data. Our approach enhances severe weather prediction through advanced spatiotemporal analysis, addressing data inconsistencies, sensor failures, and unique environmental conditions. Key innovations include automated 12-hour forecasts, multi-weather adaptability, unified sensor-textual-forecast integration, spatiotemporal pattern recognition, and expert-annotated scalable datasets.

2 Related Works

Event prediction techniques have gained significance across meteorology, pandemic tracking, and financial modeling, extracting spatiotemporal patterns to forecast events [20,29]. Recent AI and big data advances have enhanced prediction accuracy through multimodal analysis of sensor data, forecasts, and social media [22]. Deep learning models excel at capturing non-linear relationships, with studies advancing power outage prediction by combining weather, vegetation, and infrastructure data [2,4]. While GLMs and physics-informed approaches have improved predicting, challenges in severe event prediction and uncertainty quantification persist [3,27].

Random forest (RF) models have proven effective in predicting severe weather across the contiguous United States, often outperforming operational Storm Prediction Center forecasts at longer lead times [11]. They have been successfully applied to predict phenomena such as tornadoes, hail, and severe winds [12], and have demonstrated value in short-term predicting systems and probabilistic precipitation calibration [9,19]. Their capability to capture complex spatiotemporal weather patterns, especially when combined with human expertise, underlines their potential in operational settings [12].

Regression-based methods have also been employed to predict the impacts of severe weather on critical infrastructure. Generalized additive models, for instance, have achieved improved accuracy in predicting hurricane related power outages compared to traditional approaches [10]. Other techniques such as support vector regression, Bayesian additive regression trees, and hybrid data mining regression methods have been used to forecast storm related transmission outages and power outage durations, despite challenges posed by limited data for low probability events [28]. Additionally, big data analytics and spatiotemporal modeling have been leveraged to assess weather impacts on utility assets [5].

Research in weather event prediction has underscored the value of integrating diverse data sources such as weather forecasts, infrastructure data, and historical logs using methods like tree based models, spatially enhanced logistic regression, and two step frameworks. Recent advances fuse numerical weather predictions, satellite imagery, and machine learning, demonstrating the potential of model data fusion to address uncertainties across scales. Studies have shown the benefit of integrating multimodal data to solve forcasting problem [1]. One study proposed multimodal spatiotemporal framework addresses these limitations by integrating diverse data streams for enhanced hurricane scenario predictions [7].

Multiple studies have investigated specialized LLMs to perform urban spatiotemporal prediction and found that they perform better than traditional methods in tasks such as traffic predicting, alignment of time series with natural language and integration of multimodal geospatial data [15,18].

Despite progress, a gap persists between the theoretical capabilities of multimodal approaches and their practical implementation in multiclass classification. In our previous work [23], we addressed a five-class classification problem using dual data modalities with 12- and 24-hour forecast horizons. Tailored for cold, data-scarce regions like Alaska, our approach employs biLSTM architectures with advanced spatiotemporal integration techniques to overcome challenges such as data imbalances, sensor failures, and sparse coverage thereby reducing reliance on expensive equipment and mitigating data gaps.

This work provides a practical framework for enhancing operational prediction in regions where traditional monitoring methods face significant limitations. Integrating multiple data sources, including sensor recordings, weather forecasts, and textual reports, may be effective for predicting severe weather events. The effectiveness of this integration depends on how these heterogeneous data modalities are fused. Three primary fusion strategies early fusion, intermediate fusion, and late fusion offer distinct advantages and challenges in multi-source data modeling.

Early fusion involves concatenating raw data from different modalities at the input stage. This aligns with traditional deep learning architectures where all input features are processed jointly through shared layers. However, this method does not allow for modality-specific processing, making it highly sensitive to missing data and struggling to learn meaningful representations, leading to poor performance. While this approach captures cross-modal correlations and supports standard classifiers, it assumes complete modality availability and synchronization, often unrealistic in real-world scenarios with missing, corrupted, or misaligned data. This issue is evident in speech recognition, where audio-visual synchronization is crucial, but also in clinical settings, where modalities may originate from different time points, such as pre and post interventions, making early fusion less suitable [17].

In contrast, intermediate fusion constructs independent embeddings for each modality before integration, enabling more structured feature learning. This approach is particularly effective when temporal alignments exist between modalities, such as numerical weather predictions and sensor recordings. However, ASOS data often suffer from missing temporal observations, which can hinder the effectiveness of intermediate fusion.

Late fusion, also known as decision-level fusion, processes each modality separately through independent models and combines their outputs at a later stage. This approach allows each modality to be optimized individually, making it more robust to missing or noisy data. However, since late fusion does not leverage cross-modal interactions during feature learning, it may fail to capture deeper correlations between modalities. Despite this limitation, late fusion can be particularly useful in scenarios where different modalities contribute independently to the final prediction. In this study, we explore all three fusion techniques, assessing their impact on model performance in severe event classification.

3 Data

Our study employs a comprehensive data integration approach to underpin robust training and testing of severe weather event prediction models. Three data modalities are selected and sampled hourly over a five-year period (2018–2022). We align them using timestamp mapping to ensure temporal consistency across the data set. These modalities include:

1. **Sensor Data (ASOS - Local):** Collected from 153 Alaskan weather stations at 1- and 5-minute intervals and aggregated into hourly samples, this dataset captures key meteorological variables such as temperature, humidity, wind speed, and precipitation.
2. **Expert Text and Event Logs (Non-Spatial):** Sourced from NOAA's Storm Events Database, this modality comprises detailed narratives of severe weather disruptions. Each event is meticulously labeled and temporally aligned with the sensor and forecast data. The data contains timestamps but does not include spatial coordinates.

3. **Forecast Data (Global):** Derived from the NCEI Archive, this dataset provides hourly predictions of future weather events and climate shifts, serving as an essential input for time-series modeling.

Our labels are sourced from NOAA [21] and consist of 10 distinct weather categories: Normal, Flood, High Wind, Heavy Snow, Winter Storm, Debris Flow, Winter Weather, Coastal Flood, Cold/Wind Chill, and Severe Cold/Wind Chill. We formulate the problem as a multiclass classification task. Although severe weather events can occur simultaneously, suggesting a multilabel approach would be more appropriate, the dataset is structured to assign a single label per instance. When multiple events occur on the same day, they are combined into a new, unique class, enforcing a multiclass classification framework.

4 Methodology

In this study, our methodology is structured into two main components: data preparation and fusion. The data preparation phase involves transforming multimodal inputs to ensure compatibility between different data sources. The fusion phase integrates these processed data streams through a late fusion strategy, allowing the model to leverage complementary information from multiple modalities for improved predictive performance. Each of these components is described in detail in the following subsections. An overview of the entire process is illustrated in Fig. 2, providing a visual representation of how the different data streams are processed and combined within our framework.

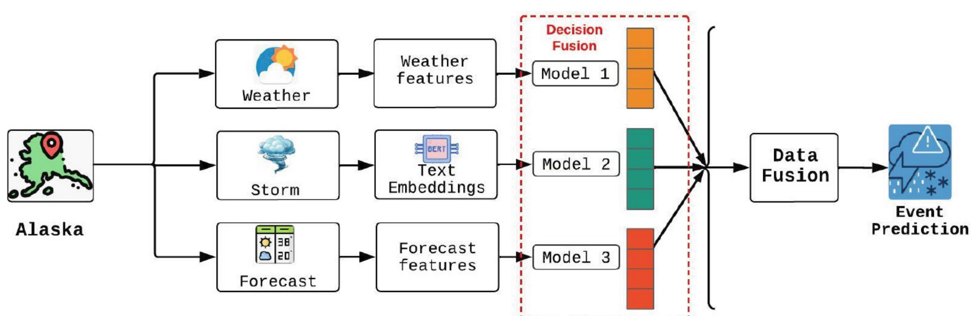

Fig. 2. Overview of the late fusion architecture.

4.1 Multimodal Data Preparation

In this study, we integrate data from ASOS weather sensors, NCEI forecasts, and NOAA storm description texts to develop a tri-modal predictive framework. The weather data is available at an hourly resolution. Our proposed architecture processes data on a per-station basis, with the objective of predicting severe weather events for each ASOS weather station. The framework consists of three specialized processing pipelines with staggered fusion, each designed to extract distinct patterns from different data modalities:

- **Weather Sensor Pipeline (LSTM):** This pipeline processes high-frequency ASOS sensor data from individual stations using a three-layer stacked bidirectional Long Short-Term Memory (LSTM) network with a hidden layer size of 128. It is designed to capture complex temporal dependencies in meteorological measurements.
- **Textual Pipeline (BERT + LSTM):** Since textual storm descriptions are only available for certain hours, we create a uniform temporal sequence by converting the available texts into 768-dimensional BERT embeddings. We choose BERT because it is the state of the art model for creating text embeddings. It captures both semantic and contextual meaning to generate meaningful representations. For hours without textual data, we substitute a zero vector to maintain temporal consistency. The resulting sequence is then passed through an LSTM network to generate a contextual embedding that captures temporal dependencies in the textual data.
- **Forecast Pipeline (LSTM):** Observational forecast data from multiple sources are processed through a parallel LSTM network with a hidden layer size of 64, enabling the capture of global severe weather patterns using a spatial attention mechanism.

By applying this tri-modal approach at the station level, our model predicts severe weather events for each ASOS weather station by leveraging both structured and unstructured data sources to enhance predicting performance.

4.2 Proposed Late Fusion Framework

In our framework, we independently train three models, f_1, f_2, and f_3, corresponding to the three modalities: numeric sensor data, numerical forecast data, and textual descriptions. Each model f_i maps its modality-specific input x_i to a vector of class scores over C classes. Applying the softmax function σ to these scores yields the probability vector $\hat{y}_i = \sigma(f_i(x_i))$ for each modality. To fuse the outputs, we first compute the maximum confidence for each model:

$$m_i = \max_{c \in \{1,\ldots,C\}} \hat{y}_{i,c}.$$

We then determine the modality with the highest confidence:

$$i^* = \arg\max_{i \in \{1,2,3\}} m_i,$$

and select the corresponding prediction as the final output:

$$\hat{c} = \arg\max_{c \in \{1,\ldots,C\}} \hat{y}_{i^*,c}.$$

This late fusion strategy ensures that the final decision is based on the model that is most confident in its prediction.

5 Experimental Setup

We process three primary data modalities (weather data, forecast data, and textual event narratives) each of which is transformed into a compact representation using a dedicated encoder. To rigorously evaluate our tri-modal fusion approach, we implement a temporal hold-out by training on data from 2018 to 2021 and testing on unseen data from 2022. Our model predicts 10 classes, and due to the imbalanced nature of the dataset, we focus our evaluation on the macro F1 score as well as the F1 score for each individual class, ensuring a comprehensive assessment of performance across all classes.

We propose a risk-scoring framework that computes a score $\mathcal{R}_k$ for each event type k as follows:

$$\mathcal{R}_k = \beta_k \,\mathrm{Recall}_k + \big(1 - \beta_k\big)\,\mathrm{Precision}_k,$$

where $\beta_k \in [0.3,\, 3.7]$ is a risk-weighting parameter calibrated using regional economic exposure and population density data. Events with a risk score exceeding 10 are classified as high-impact. In this work, we set $\beta_k = 0.3$ for all event types.

6 Results

We conducted extensive experiments to evaluate the performance of the proposed model and compare it to baselines. In this section, we address the research questions described in Subsects. 6.1 to 6.5 to evaluate the performance of the model.

6.1 What Roles Do Sensor, Forecast, and Text Modalities Play in Capturing Alaska's Severe Weather Dynamics?

To investigate the roles of sensor, forecast, and text modalities in capturing Alaska's severe weather dynamics, we evaluated the 12-hour predictive performance of each modality independently and in an integrated setting (see Table 1). The results indicate that the forecast modality achieves an F1-score of 0.78, outperforming the text-based modality (F1-score of 0.64) and the ASOS sensor modality (F1-score of 0.13). Notably, our proposed trimodal model that fuses all three modalities attains an F1-score of 0.89. This significant performance boost underscores the complementary strengths of each modality and highlights the benefits of multimodal integration for accurately modeling severe weather dynamics in Alaska.

6.2 How Does Our Late Fusion Approach Improve Prediction Accuracy and Extend the Predicting Horizon?

To evaluate how our late fusion approach improves prediction accuracy and extends the predicting horizon, we compared it against two baseline fusion strategies. Table 2 presents the comparative results across these approaches: Early

Table 1. Results for unimodal models vs. proposed trimodal late fusion model.

Modality	F1 Score
Forecast Only	0.78
Text Only	0.64
ASOS Only	0.13
Our Proposed Model	0.89

fusion combines raw data from multiple modalities (e.g., sensors, forecasts, and text embeddings) at the input stage, without modality-specific preprocessing, making it highly sensitive to missing information and resulting in poor performance (F1-score = 0.043). Intermediate fusion integrates independently constructed embeddings for each modality, offering better feature combinations, but its performance remains inconsistent across classes (F1-score = 0.431). In contrast, our late fusion model combines fully processed modality outputs, allowing it to learn modality-specific representations and preserve distinct feature hierarchies while enabling cross-modal interactions. This approach significantly outperforms both early and intermediate fusion methods in terms of overall performance.

Table 2. Comparative Performance of macro F1 scores for ten classes of events obtained by three fusion methods (Early, Intermediate, and Late).

Class	Early	Intermediate	Late
Normal	0.00	0.24	1.00
Flood	0.00	0.40	1.00
High Wind	0.00	0.00	1.00
Heavy Snow	0.00	0.00	1.00
Winter Storm	0.00	0.76	1.00
Debris Flow	0.00	0.51	0.93
Winter Weather	0.43	0.89	0.94
Coastal Flood	0.00	0.38	0.99
Cold/Wind Chill	0.00	0.27	0.88
Severe Cold/Wind Chill	0.00	0.86	0.77
Macro F1	0.04	0.43	**0.89**

These results highlight the advantages of late fusion in mitigating modality-specific limitations and effectively leveraging complementary cross-modal information to improve prediction accuracy and extend predicting capabilities.

6.3 How Does the Model Balance Recall and Precision for Effective Risk Estimation Under Data Limitations?

Our framework builds on established principles in cost-sensitive learning [6,16] and risk-based metrics [8,24], aligning model evaluation with operational risk profiles to support informed decision-making under data limitations. To assess how the model balances recall and precision for effective risk estimation under data limitations, we analyze the computed risk scores ($\mathcal{R}_k$) across different severe weather events, as presented in Table 3. Our approach effectively prioritizes high-impact events, with Flood ($\mathcal{R}_k = 13.15$), High Wind ($\mathcal{R}_k = 13.45$), and Debris Flow ($\mathcal{R}_k = 14.20$) all exceeding the high-risk threshold of 10. In contrast, despite their high frequency of occurrence, Normal conditions and Coastal Flood register below the threshold ($\mathcal{R}_k = 9.95$ and $\mathcal{R}_k = 9.98$, respectively), reflecting their lower operational impact. This demonstrates that the model effectively balances recall and precision, ensuring that high-risk events are prioritized while minimizing false alarms for lower-risk scenarios.

Table 3. Risk-Adjusted Scores and Event Classifications for the Late Fusion Model on Unseen 2022 Data. Risk scores are computed as $\mathcal{R}_k = \beta_k \, \text{Recall}_k + (1 - \beta_k) \, \text{Precision}_k$ (with β_k tuned based on real-world risk factors), and a threshold of 10 is used to differentiate High and Low risk events.

Event	Precision	Risk Score	Risk Class
Normal	1.00	9.95	Low
Flood	0.90	13.15	High
High Wind	0.93	13.45	High
Heavy Snow	0.85	12.25	High
Winter Storm	0.91	12.97	High
Debris Flow	0.98	14.20	High
Winter Weather	0.88	12.52	High
Coastal Flood	1.00	9.98	Low
Cold/Wind Chill	0.90	12.81	High
Severe Cold/Wind Chill	0.77	10.87	High

6.4 How Does a Generalized Model Compare to Localized Models in Capturing Severe Weather Patterns Across Different Regions of Alaska?

To evaluate whether a generalized model outperforms localized models in the Alaskan region, we compare a tri-modal approach with region-specific models trained separately on four distinct subregions: Southeast, Northeast, Southwest, and Northwest Alaska. Our comparative analysis reveals two key limitations in the localized models, despite the Northeast region achieving a seemingly perfect F1-score of 100%.

1. **Data Scarcity and Class Imbalance:** The Northeast experiences fewer severe weather events across our 10-class framework, resulting in a smaller and less imbalanced dataset in that region.
2. **Inadequate Spatial Coverage:** The sparse distribution of ASOS monitoring stations in the Northeast restricts the model's ability to accurately

delineate event boundaries in that region. This limitation increases the likelihood of misclassifying localized anomalies as severe weather patterns, thereby further undermining generalizability.

Regional analysis (Fig. 3) reveals significant performance disparities. While the Northeast model achieves a perfect score, other regions, especially those with frequent severe weather, underperform due to limited spatial coverage and class imbalance. These results demonstrate the shortcomings of geographically isolated training and the failure to capture cross-regional weather dependencies. In contrast, a comprehensive multi-regional model, which leverages shared characteristics and mitigates class imbalance, enhances predictive robustness across all Alaskan subregions, achieving a macro F1 score of 0.89.

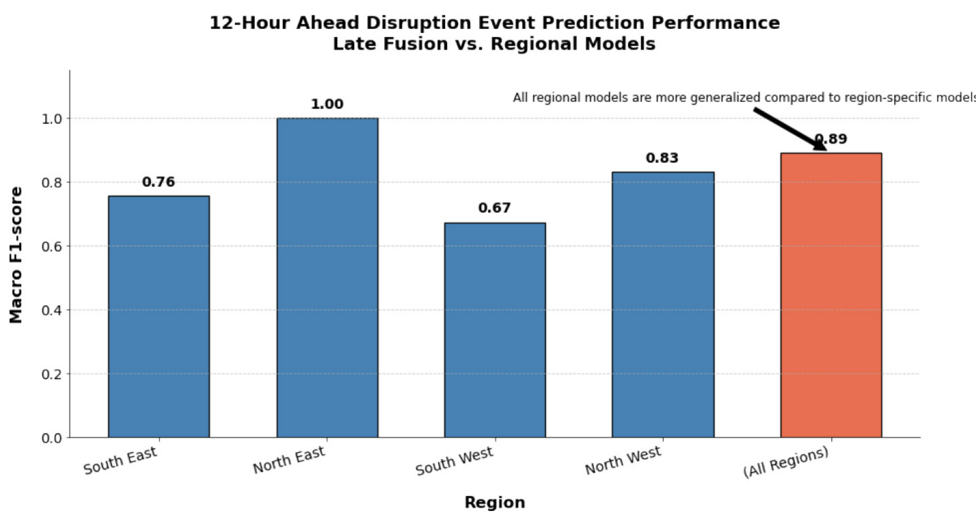

Fig. 3. Geographically isolated macro F1 score of 12-hour ahead disruption predictions by region. Northeast's perfect score is of limited generalizability due to data sparsity.

6.5 How Does Model Performance Vary Across Different Seasons in Severe Weather Classification?

To assess seasonal variability, we conducted an evaluation on unseen year 2022 by reframing the original multi-class problem into a binary classification task normal conditions (class 0) versus severe weather events (classes 1–9). Table 4 reports the season-wise performance metrics.

Our model demonstrates consistently high AU-PRC values (ranging from 0.882 to 0.950) across all seasons, underscoring a robust precision-recall balance. However, AU-ROC results vary considerably. Notably, the winter season exhibits the highest AU-ROC (0.950) along with a high recall (0.950), which aligns with expectations given that most severe weather events tend to occur during winter.

Table 4. Season-Wise Performance Metrics for Severe Weather Classification using a late fusion tri-modal approach 12 h ahead on unseen year 2002.

Season	AU-ROC	AU-PRC	Precision	Recall	F1-Score
Fall	0.85	0.95	0.95	0.94	0.74
Spring	0.55	0.88	0.95	0.95	0.95
Summer	0.89	0.95	0.95	0.95	0.86
Winter	0.95	0.95	0.90	0.95	0.74

In contrast, spring performance is characterized by a markedly lower AU-ROC (0.552) despite high precision and recall scores (both 0.950), suggesting that the model may be less effective at differentiating between normal and severe conditions during this season. Fall and summer yield intermediate AU-ROC values (0.852 and 0.892, respectively), with both seasons maintaining high precision and recall. These findings indicate that the model reliably detects severe weather events throughout 2022.

7 Conclusion

In conclusion, this work presents a novel multimodal learning framework that integrates numeric sensor data (ASOS), textual narratives from the Storm Events Database, and forecast information to predict rare and disruptive weather events in cold, data-scarce regions such as Alaska. By precisely aligning sensor measurements, narrative descriptions, and short-term forecasts, our fusion architecture effectively captures complex spatiotemporal patterns while preserving interpretability through a late fusion strategy. Experimental results demonstrate that the framework accommodates a diverse range of event classes while maintaining transparency in its decision-making process.

Despite these advances, several limitations persist. In particular, the detection of rare events remains challenging due to modest sample sizes in certain regions. Future research will explore alternative temporal encoding strategies (e.g., seasonal or monthly representations), advanced imbalance-handling techniques (such as targeted oversampling and data augmentation), and further incorporation of domain-specific knowledge (including ice thickness and coastal geometry) to enhance predictive accuracy. These improvements are anticipated to bolster the recognition of extreme, infrequent phenomena. Moreover, beyond Alaska, the proposed approach holds promise for application in other cold or remote regions with sparse station coverage and limited specialized equipment. By leveraging widely accessible meteorological networks and publicly available textual logs, our method offers a scalable, cost-effective pathway to more reliable weather predicting, ultimately enabling emergency planners, power grid operators, and local communities to make proactive, informed decisions to mitigate risks. To conclude, our proposed work builds upon our prior study [23], which evaluated models across Alaska, Nevada, and Pennsylvania achieving F1-scores of 0.77, 0.84, and 0.97, respectively thereby demonstrating reproducibility under diverse weather conditions and extreme events. Expanding on this evaluation across climatically distinct regions, we now incorporate sequence models, BERT, and fusion algorithms to learn severe event patterns directly from raw data. This approach enhances computational efficiency and real-time performance through optimized deployment and hardware acceleration, all without the need for more feature engineering.

Acknowledgments. 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 Engineer Research and Development Center (ERDC) or the U.S. Government.

References

1. Aljurbua, R., Alshehri, J., Gupta, S., Alharbi, A., Obradovic, Z.: Early prediction of power outage duration through hierarchical spatiotemporal multiplex networks. In: Proceedings of 13th International Conference on Complex Networks and their Applications. Springer Verlag, Istanbul, Turkey (2024), (in Press)
2. Alqudah, M.: Others: power system measurements for severe weather prediction: a multi-instance learning approach. IEEE Trans. Power Syst. **38**(2), 1234–1245 (2023)
3. Arora, S., Chen, L.: Challenges in extreme event prediction for power systems. Electric Power Syst. Res. **215**, 108789 (2023)
4. Cerrai, D., Anagnostou, M.A., Anagnostou, E.N.: Machine learning models for predicting power outages caused by extreme weather events. IEEE Trans. Power Syst. **35**(4), 3055–3066 (2020)
5. Dokic, K., Texas, A.: Big data analytics and spatiotemporal modeling for assessing weather impacts on utility assets. J. Infrastruct. Syst. **25**(3), 04019015 (2019)
6. Elkan, C.: The foundations of cost-sensitive learning. In: Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI-01), pp. 973–978 (2001)
7. Fatima, Z., Liu, R., Zhang, H.: Multimodal spatiotemporal framework for hurricane-induced outage prediction. Nat. Energy **9**, 178–189 (2024)
8. Fawcett, T.: An introduction to roc analysis. Pattern Recogn. Lett. **27**(8), 861–874 (2006)
9. Flora, M.L., Skinner, P., Stensrud, D.: Short-term severe weather forecasting using the warn-on-forecast system and random forest models. Weather Forecast. **35**(6), 2101–2115 (2020)
10. Han, S.R., Guikema, S.D., Quiring, S.M.: Improving the accuracy of hurricane-related power outage forecasts using generalized additive models. Risk Anal. **29**(10), 1443–1453 (2009)
11. Hill, A.J., Gagne, D.J., McGovern, A.: Calibration and reliability of random forest-based severe weather forecasts. Mon. Weather Rev. **150**(3), 567–582 (2022)

12. Hill, A.J., Gagne, D.J., McGovern, A.: Random forest models for predicting severe weather events across the contiguous united states. Weather Forecast. **35**(5), 1587–1603 (2020)
13. Kettle, N.P., Gladden, L., Martin, J.: Climate adaptation planning in alaska native villages: a systematic review. Arctic **73**(3), 275–291 (2020)
14. Kettle, N.P., Martin, J., Sloan, M.: Building resilience to natural hazards through knowledge networks in alaska native villages. J. Environ. Manage. **215**, 384–394 (2018)
15. Li, E.A.: Urbangpt: integrating spatio-temporal dependencies for urban prediction. J. Urban Intell. (2024)
16. Ling, C.X., Sheng, V.S.: Cost-sensitive learning and the class imbalance problem. Encyclopedia Mach. Learn. 231–235 (2010)
17. Lipkova, J., et al.: Artificial intelligence for multimodal data integration in oncology. Cancer Cell **40**(10), 1095–1110 (2022)
18. Liu, E.A.: St-llm: specialized spatio-temporal prediction in urban settings. IEEE Trans. Intell. Transp. Syst. (2024)
19. Loken, E.D., Clark, A.J., Gagne, D.J.: Calibrating probabilistic precipitation forecasts using random forest models. Mon. Weather Rev. **147**(12), 4321–4336 (2019)
20. McGovern, A.: Others: spatiotemporal relational learning for severe weather prediction. Mach. Learn. **91**(2), 147–180 (2013)
21. National Oceanic and Atmospheric Administration: Storm events database. https://www.ncdc.noaa.gov/stormevents/ (2023)
22. 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, November 2023
23. Otudi, H., Gupta, S., Obradovic, Z.: Leveraging diverse data sources for enhanced prediction of severe weather-related disruptions across different time horizons. In: International Conference on Engineering Applications of Neural Networks, pp. 220–234 (2024)
24. Provost, F., Fawcett, T.: Robust classification for imprecise environments. Mach. Learn. **42**(3), 203–231 (2001)
25. Ristroph, E.: When climate takes a village: legal pathways toward the relocation of alaska native villages. Climate Law **7**(4), 259–289 (2018)
26. Walsh, J.E., Thoman, R.L., Bhatt, U.S.: Alaska climate change assessment: recent changes and future projections. Weather Clim. Extremes **33**, 100456 (2024)
27. Watson, J.P., Guttromson, R.: Uncertainty quantification in power system resilience. IEEE Trans. Power Syst. **37**, 3214–3226 (2022)
28. Yue, X., Guikema, S.D., Nateghi, R.: Problem reformulation to improve regression analysis in granular damage forecasting. Risk Anal. **42**(5), 987–1001 (2022)
29. Zhao, Y.: Advances in event prediction: leveraging artificial intelligence and big data. J. Predictive Anal. **15**(3), 45–60 (2020)