# Classifying Severe Weather Events by Utilizing Social Sensor Data and Social Network Analysis

Hussain Otudi, Shelly Gupta, Nouf Albarakati, Zoran Obradovic
*Center for Data Analytics and Biomedical Informatics*
*Temple University, Philadelphia, PA, USA*
{hussain.otudi,shelly.gupta, nouf, zoran.obradovic}@temple.edu

***Abstract*—Weather-related disruptions have a significant impact on a variety of industries, including agriculture, infrastructure, and public safety. Predicting these unusual weather events remains a significant challenge. The problem is complicated by the lack of high-quality weather data due to the failure of sensors at weather stations during severe weather. In this work, we proposed a novel method to classify rare severe weather-related events by incorporating publicly available tweets with the meteorological conditions readings collected from weather stations across Alaska. The use of multimodal data of varying quality is introduced to compensate for missing meteorological recordings obtained from the weather stations. In our study, we collected geotagged tweets from the region of focus and utilized context-aware BERT embeddings to rigorously analyze and ensure the validity and dependability of the social media texts. Labels for the social sensor data were generated based on weather events associated with the tweet collection. For predicting rare weather events, we proposed a multiclass classification model. This machine learning model was trained and tested using data from the year 2020. The results obtained by learning from the integrated data showed a significant increase in the F1 score when compared to relying on weather data alone. Our findings indicate that a model supplemented with daily weather and social media text data outperforms alternatives enhanced with hourly data. The proposed model achieved an F1-score of 0.83 for multimodal data, compared to 0.30 obtained by the baseline model that relies solely on weather data. Training the proposed model with the combined dataset significantly improved performance, resulting in a 95% accuracy.**

***Index Terms*—rare weather events, social media, multi-modal data, event classification.**

## I. Introduction

Extreme weather events affect millions of people every year around the globe. Hurricanes and extreme storms like Sandy, Katrina, Harvey, and Irma exemplify the significant disruption to human health and community safety caused by these intensified weather patterns. According to the American Public Health Association, climate change has resulted in a higher occurrence and greater intensity of severe weather events, including heavy rainfall, heatwaves, droughts, and storms [1]. Subsequently, numerous studies focus on predicting these phenomena from various perspectives such as numerical methods [2] and statistical and probabilistic methods [3]. In recent years, research has moved toward deep learning algorithms hoping to make a black box model to forecast rare events [4]. However, inherent weather-related challenges remain, including missing values caused by sensor failures at weather stations, as well as the imbalance in data distribution across different classes. Given the rarity of disruptive event occurrences, even weather databases with more than 70 years of high spatial accuracy events are often deemed unreliable and inadequate for accurate machine learning [5]. Several studies have employed simulations and feature construction to address the rare disruptive events by oversampling such data [6]. However, this approach tends to overestimate the algorithm's performance since simulated data cannot fully capture real-world scenarios due to inherent assumptions and fallacies [7].

In a related context, social networks have emerged as the main platforms for individuals to express their reactions to uncommon events, allowing anyone to share their thoughts and experiences through tweets with the hope of saving people from natural disasters. Furthermore, researchers have delved into the study of these exceptional phenomena and identified a substantial correlation between the frequency of tweets that related to the weather and the prevailing weather conditions [8]. This correlation suggests that weather has a discernible impact on people's tweeting behavior [9] [10]. One study successfully established the correlation between tweets and climate data. Furthermore, the use of geotagged tweets allows for a deeper understanding of events in specific locations [8]. Building upon these findings, our study aimed to leverage this information by integrating meteorological data with geotagged social media data to enable multimodal prediction of rare weather events. To access historical data for automated surface observing system (ASOS) by the weather station, we utilized the free API offered by Iowa State University Mesonet that captures weather condition at local airports [11]. By integrating this data source with spatio-temporally collocated tweets, we trained a supervised ML algorithm for our analysis.

Presently, online data is an easy-to-access source of rich, user provided data. While social media represents a wealth of such information, it is still plagued with issues of noise and



*ASONAM '23*, November 6-9, 2023, Kusadasi, Turkey

ACM ISBN 979-8-4007-0409-3/23/11…$15.00
http://dx.doi.org/10.1145/3625007.3627298

bias. To mitigate these concerns, it is imperative to preprocess the data appropriately [12]. In our approach, we focused on aggregating the tweets, aiming to obtain unique and location-specific tweets, particularly during rare events. Our primary area of interest for tweet collection is the Anchorage region of Alaska. Utilizing event-related keywords related to weather, we scraped a substantial volume of social sensor data and integrated them with weather data after ensuring the data quality. A supervised learning model is developed that makes use of multimodal data, the model's performance is assessed on its ability to predict future rare weather events. Additionally, we examined how the model's performance varied when applied to regions exhibiting higher tweet density (such as Anchorage, Alaska) compared to regions with limited social media data (areas away from Anchorage).

**The key contributions of this work are the following:**

1) We describe and address the problem of learning from partially observed weather datasets by employing feature construction techniques to improve the quality and reliability of the weather data.
2) We demonstrates the effectiveness of a multi- modal learning approach that combines social sensor data with weather data outperforming models that rely solely on weather data.
3) We addressed the issue of high data imbalance challenge successfully by changing default loss function with a loss function specifically designed to handle imbalanced class distributions.
4) We Conducted experiments to quantify to the benefits of location-specific tweets and analyze an increased prediction subjectivity when moving away from Anchorage, Alaska's higher tweet density area.

## II. RELATED WORK

Three major components summarize the related work for the proposed approach: 1) handling corrupt and incomplete weather data, 2) integration of social sensors with weather data, and 3) development of a multi-classification model for predicting rare weather events.

- **Weather prediction based on weather data:** While the use of online data is easy nowadays and users can report their observations, despite the richness of information social media provides, the presence of noise and bias in such textual content is high. A weather station's ASOS ceilometer scans the sky overhead and detects clouds below 3.6 km. This makes ASOS data less accurate. As a result, missing data appears, and reporting about the whole atmosphere becomes less accurate than human observations [13]. Consequently, our experiments show that using weather data alone to predict rare weather-related disruptive events is ineffective.
- **Weather Data Quality**: The issue of missing data is an inescapable challenge when working with most of real-world data. Observations with missing data may be the result of sensor failure and cloud pollution [14]. Many studies investigated different approaches to addressing the problem of missing data. Considered approaches include ignoring instances with missing values, replacing missing data with the mean of the remaining data, and estimating missing values using regression techniques. Prior work showed that it is important to handle missing data carefully and contextually, taking it into account for subsequent analysis. However, ignoring missing data events instances could be risky in the context where rare instances are valuable to save lives [15].
- **Methods of Integrating Data Sources (Multi-data Approach):** Multi-data approaches were considered recently to fill the gap of missing data. Researchers investigated missing data in field recordings for power system applications. These real-world observations suffered by missing data and were supplemented by simulation data generated by a real-time digital simulator [16]. However, rare events occur in 1% of the data, and so creating simulations would be expensive and would require large computational resources [17]. Due to such a high cost of existing resource, a method is proposed as an alternative approach that provides additional data with less cost and make them publicly available.
- **Social Sensor Data**: The study of social sensor data, particularly Tweets, aids in the comprehension of how people interact with severe weather incidents. Social sensor data provides an alternative source of textual information to supplement weather data. Our approach generalizes a previous research that explored the influence of area-specific temperature fluctuation on climate change awareness, as evidenced through Spanish tweets [18]. In the he previous study Twitter's data is utilized with machine learning to predict the crime where information extracted form tweets have contributed to learn about incidents in Chicago [19]. Our study utilizes localized, real-time tweets to understand rare weather events while considering data quality issues in social sensor analysis, such as accuracy and bias, with careful evaluation before integrating social and weather data.

  To get the tweets that have meaning related to weather condition, the study compares two NLP approaches to embedding tweets, Word2Vec and Bert. One limitation of Word2Vec's embedding is that it is context-independent and cannot take into account polysemous words [20]. One of many methods for converting text to numerical representation is using TF-ID (Term Frequency-Inverse Document Frequency). However, this approach is ineffective for short text [21], which was addressed in our approach by using BERT (Bidirectional Encoder Representations from Transformers), which can be more effective for short text and work more effectively with words that have multiple meanings.
- **Development of a multi-classification model for classifying rare weather events**: Classifying rare weather events using logistic regression method is not the right option due to the biased that models can assign to the majorityclass [22]. Our work utilized MLP (multilayer

perceptron) on integrated datasets for rare events [23]. As a result. Author applied the MLP algorithm to benchmark datasets for anomaly detection, where the MLP model notably outperformed other models. Additionally, we employed the focal loss function, providing more weight to rare classes. Consequently, a significant improvement over the default setup for MLP models, which typically rely on standard loss functions such as cross-entropy.

## III. Background

Weather data is considered high-quality data, but it has incomplete information due to two reasons: missing data and the low probability of rare events. This section provides more details about these two challenges.

- **Case1: Missing Data.** Detecting rare weather events, such as extreme cold or wind chill, coastal flooding, and winter storms, can be challenging using data gathered from weather observations. This is mainly because many such incidents go unreported. For instance, certain highly disruptive events might only have a few entries in the event log collected over a year. This number is inadequate for training machine learning algorithms.

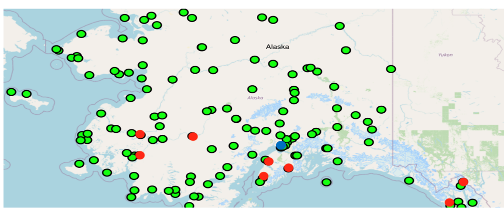

Fig. 1. Red nodes represent the nine weather stations near Anchorage.

Despite the presence of a significant number of stations in Alaska (185 in total), many data coverage gaps still exists, contributing to the need for greater clarity in distinguishing between different weather phenomena. The Anchorage station (blue node in **Figure 1**), for example, the weather dataset lacks comprehensive information, with 64% of its temperature records are missing as shown in **Figure 2.**

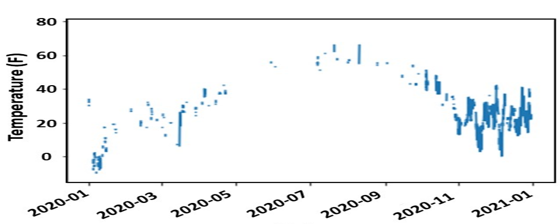


Fig. 2. Daily weather data at a weather station near Anchorage, Alaska.

**Case2: Low frequent event.** The weather dataset contains a few examples of rare but highly disruptive weather events, such as avalanches and astronomical low tides, which occur less frequently than more common incidents like blizzard as shown in **Figure 3**. Consequently, some events may only be represented by a single entry in the event log. This results in an imbalanced dataset, which complicates data analysis and hinders practical model training.

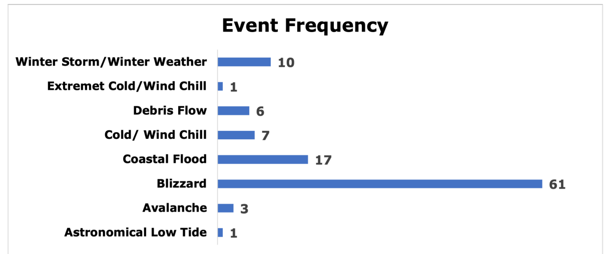


Fig. 3. Distribution of rare events labeled in weather data at one weather stations near Anchorage.

## IV. Proposed Methods

In this section, we discuss our proposed model in eight subsections, starting with data collection, preprocessing, data integration, training, and baseline models. The preprocessing for weather data includes feature construction, time series smoothing, and min-max scaling, while social media data undergoes lowercasing, tokenization, and anonymization of tweets' IDs.

### A. Data Collection, and Preprocessing

We collect weather data from a network of nine stations in Alaska, as depicted with red nodes in **Figure 1**. For the weather data, each station was processed to have 365 daily data per station. Regarding social sensor data, tweets have been collected and analyzed from both local and remote sites based on latitude and longitude, with a radius of 25 km. Most of these tweets originated in Alaska's two major cities, Anchorage and Hoonah. As expected, we found fewer weather-related tweets from Hoonah than from Anchorage due to Hoonah's smaller population. Limited internet access, or fewer active Twitter users. As a result, the Hoonah tweet dataset was relatively small. The event log data used in this study was sourced from the National Climate Data Center (NCDC) [25]. Six categories of disruptive events were gathered as illustrated in **Figure 4**.

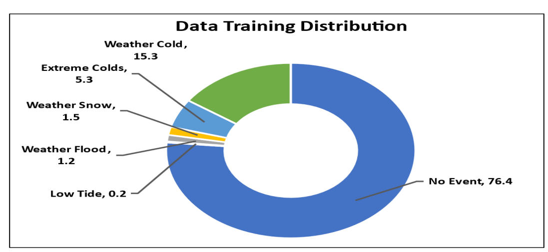


Fig. 4. Training data distribution.

### B. Weather Data

Historical weather data comes from ASOS, which collects data on atmospheric conditions independently. These stations, which are typically positioned near airports, provide critical information for air travel, such as temperature, dew point, wind speed and direction, visibility, atmospheric pressure, and precipitation. Other variables have been collected, such as skyc1, skyc2, skyc3 and 4. Sky Coverage Levels 1-4, Ice accretion (1hr, 3hr, 6hr) Ice accumulation (1hr, 3hr, 6hr).

### C. Social Sensor Data (i.e Twitter Dataset)

We first gathered tweets using snscrape, a python package designed for scraping historical tweets [24], which allowed us to scrape tweets based on specified geographic coordinates (latitude and longitude) for each station location. The time frame for this data extraction was the 2020 calendar year. By using relevant keywords such as avalanches,exterme cold were used to extract the relevant tweets.Keywords were chosen based on an event log provided by National Oceanic and Atmospheric Administration [25]. Tweets were scraped based on keywords related to weather events, such as avalanches, high winds, and other events found using event keywords as shown in **Table I**.

The tweet selection process uses a set of inclusion, relevance, and exclusion criteria. Our inclusion criteria emphasized the relevance of the tweets. We only included tweets that specifically addressed a weather event and were posted within the same day and hour in a specific location. Our exclusion criteria were used to ensure the quality and relevance of our dataset. To avoid repetition, we removed all duplicate tweets, and any tweets that mentioned weather events in Alaska but originated elsewhere.

As a result, the knowledge-based selection contained 1029 relevant tweets about weather events. At the same time, we removed 3338 irrelevant tweets from the study to ensure the accuracy and relevance of our data.

TABLE I
QUERY KEYWORDS USED IN SOCIAL DATA TWEET EXTRACTION

| **Social Sensor Data** | |
|---|---|
| ***Tweet*** | ***Query keywords*** |
| Tonight on Alaska News Nightly two young men have been killed by an avalanche in Haines | Avalanche |
| He entered the gas station a few moments passed and as he opens the door to exit a strong gust of wind blew away several items in his hand | Extreme Wind |

### D. Feature Extraction from Weather Data

Weather-related variables for each weather station have been prepossessed using the following techniques:

- Mean of Variables. The first equation, denoted as (1), computes the variable Mean Range for each station 'i' within a certain period. This is achieved by dividing the difference between the maximum of each variable $T_{max}$ and the minimum of each variable $T_{min}$ by the mean of that variable $T_{mean}$.
- Variance of Variables. The equation (2) computes the variable range variance, which is a measure of the dispersion of variables for each station. This is calculated by dividing the difference between the maximum and minimum variables by their variances.
- The variable range (mean and variance) are computed as described in equation (3) and (4). These values are computed by calculating the standard deviation of the average across all stations to quantify the degree to which the Range Variance differs between stations. A larger value indicates a greater disparity in the range and variability of the variables between stations. Consequently, this statistical values are important to better understand weather related events phenomena.

The equations 1,2,3 and 4 aids in extracting more informative features from raw weather data, that suffered from missing data. As a result, all missing values in the dataset were handled as depicted in **Figure 5** before applying the machine learning model. During the prepossessing, standardization was applied to the entire dataset.

$$\forall period \in \{H, D\}:$$

$$Var_{range\ mean\ (i)} = \frac{T_{max\ (i)} - T_{min\ (i)}}{T_{mean\ (i)}} \tag{1}$$

$$Var_{range\ var\ (i)} = \frac{T_{max\ (i)} - T_{min\ (i)}}{T_{var\ (i)}} \tag{2}$$

$$Var_{range\ mean} = (\frac{1}{N}\sum_{i=1}^{N} Var_{range\ mean\ (i)}) \tag{3}$$

$$Var_{range\ var} = (\frac{1}{N}\sum_{i=1}^{N} Var_{range\ var\ (i)}) \tag{4}$$

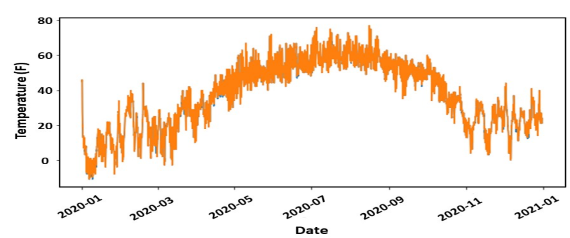


Fig. 5. Weather data after imputation.

### E. Feature Extraction from Social Sensors Data

Tweets were collected, cleaned, aggregated, and labeled as shown in **Figure 6**.

BERT, a well-known deep learning model for natural language processing tasks, was used. Our aim was to utilize the BERT model as a feature extractor. The BERT model analyzes the encoded input and generates outputs that include hidden states. These hidden states, essentially internal representations of the input text encapsulating the crucial information within the tweet, are transformed into a numerical form and stored as

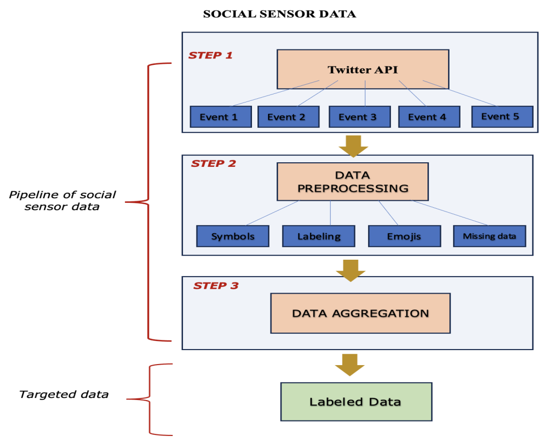


Fig. 6. Social sensors processing pipeline.

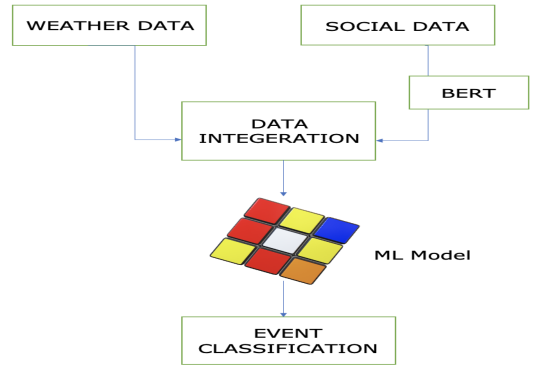


Fig. 7. Overall architecture of the proposed multi-modal classifier.

a list of features specific to the tweet. The extracted embedding from social sensor data for our downstream task of rare event classification tasks [26].

### F. Data Integration

Data integration involves combining weather data with data extracted from Twitter. Tweets are fed to BERT to generate embedding that could improve the detection and classification of rare weather events.We leverage the power of social media by reading tweets as virtual sensors that offer information about weather-related topics. By adapting context-aware BERT embeddings, we can reduce the noise associated with tweets. Weather and tweets datasets are aggregated hourly and daily before model training. Data integration strategy enhances weather data with insights from unstructured social media data for a more detailed under- standing of rare weather events. The architecture of the proposed model is illustrated in **Figure 7**. The data integration process aligns weather station observations and tweets based on timestamp. Regarding social data, we filter out duplicate and irrelevant tweets, and then aggregate this multimodal data into train and test sets, ensuring temporal coherence between both sources. Multimodal data has the challenge of each data type having its own characteristics.

### G. Training Dataset

Aside from the tasks involved in feature extraction, labeling integrated weather and social data is an important aspect of our proposed machine learning workflow. Specific weather-related events were used to categorize Twitter and weather data. For example, data was labeled as 0 if it follows normal or no event conditions. Wind Chill and Storm events were assigned label number one. Label two was assigned to Coastal Flood and Debris Flow. Extreme Wind and Extreme Cold events obtained a label of three. Avalanche and snowstorm incidents were assigned the number four. Finally, Astronomical Low Tide events were designated as label five as shown in **Table II**. This labeling technique aided in the simple identification and analysis of weather events. Therefore, four datasets were created and integrated into the ML model. These datasets were partitioned temporally, with 70% designated for training and 30% for testing. This partitioning strategy was consistently applied to both the hourly and daily data. Both datasets were inspected before passing them to the ML models to ensure there were no data quality issues.

TABLE II
WEATHER-RELATED EVENTS AND THEIR DESCRIPTORS

| **Weather-Related Events** | |
|---|---|
| ***Event Descriptor (Original event)*** | ***Category of events*** |
| Normal (i.e no event) | 0 |
| Wind Chill AND Storm | 1 |
| Coastal Flood AND Debris Flow | 2 |
| Extreme Wind AND Extreme Cold | 3 |
| Avalanche AND Snowstorm | 4 |
| Astronomical Low Tide | 5 |

### H. Baseline Traditional Methods vs Multi Layer Perceptron

In this study, the experimental strategy began by applying classic ML methods to our dataset. We utilized Random Forest (RF) and Logistic Regression (LR) models as our baseline approaches. These models are well-established and widely utilized as they are robust, interpretable, and efficient in a variety of classification tasks. These baselines are compared to using Multi Layer Perceptrons (MLPs), which are more complex models that take into account the nonlinearities of the rare event patterns.

## V. EXPERIMENTS

Several factors influence the performance of the proposed pipeline, including the classification methods utilized, the type of integrated dataset (weather or tweets), and the specified parameters. To encode class labels as integers, the Label Encoder is used, and stratified 5-fold cross-validation is used to ensure that each fold is a representative subset of the entire dataset.

The ML model includes several hyperparameters. The number of epochs (30) specifies how many times the learning algorithm will traverse the entire dataset, and the batch size is (16), and the learning rate which is a hyper-parameter, was set to 1e-4 (i.e., 0.0001) as an optimization parameter. The model's architecture consists of dense layers with a fixed number of neurons (32 and 16). In MLP these layers' activation functions are the relu and softmax functions.

For model's loss function, Focal Loss function was adapted and it has two hyperparameters, Gamma and Alpha aimed to reduce the importance of examples that are easy to classify and to account for difficulties of samples to be classified [27].The optimization strategy is defined by the optimizer, which is set to Adam in this case. During each cross-validation fold, the training and test labels are encoded, and samples are calculated to address class imbalance. We report the macro-averaged of the F1 scores for each class as shown in equation (5) where $N$ is the number of classes, $n_i$ is the number of instances in class $i$, and $F1_i$ is the F1 score for class $i$.

The macro-averaged F1 score is calculated as:

$$F1_{\text{macro}} = \frac{1}{N} \sum_{i=1}^{N} 2 \cdot \frac{\text{precision}_i \cdot \text{recall}_i}{\text{precision}_i + \text{recall}_i} \tag{5}$$

## VI. Results and Discussion

**Table III** and **Table IV** present the results of the proposed model's classification of rare weather events using weather data alone and integrated data (weather and tweets).

TABLE III
TEST RESULTS WHEN LEARNING FROM INTEGRATED WEATHER AND SOCIAL DATA

| Classification Report for Each Class | | | |
|---|---|---|---|
| ***Class*** | ***Precision*** | ***Recall*** | ***F1-score*** |
| No event | 1.00 | 1.00 | 1.00 |
| Wind Chill AND Storm | 0.85 | 0.93 | 0.89 |
| Coastal Flood AND Debris Flow | 1.00 | 0.38 | 0.56 |
| Extreme Wind AND Extreme Cold | 0.75 | 0.27 | 0.40 |
| Avalanche AND Snowstorm | 0.61 | 0.65 | 0.63 |
| Astronomical Low Tide | 1.00 | 1.00 | 1.00 |

TABLE IV
TEST RESULTS WHEN LEARNING FROM WEATHER DATA

| Classification Report for Each Class | | | |
|---|---|---|---|
| ***Class*** | ***Precision*** | ***Recall*** | ***F1-score*** |
| No event | 0.91 | 0.95 | 0.93 |
| Wind Chill AND Storm | 0.85 | 1.00 | 0.92 |
| Coastal Flood AND Debris Flow | 0.00 | 0.00 | 0.00 |
| Extreme Wind AND Extreme Cold | 0.00 | 0.00 | 0.00 |
| Avalanche AND Snowstorm | 0.00 | 0.00 | 0.00 |
| Astronomical Low Tide | 0.00 | 0.00 | 0.00 |

TABLE V
TEST RESULTS WHEN LEARNING FROM INTEGRATED WEATHER AND SOCIAL DATA

| | **Weather Data** | **Integrated Data** |
|---|---|---|
| **Average Total False Positives** | 72/6=12 | **40/6=6.67** |
| Average F1 Score | 0.30 | **0.83** |
| Average Accuracy | 88% | **95%** |

Developed model trained on daily social data integrated with weather data from 9 stations performs with an average F1-score of 0.83 comparre to the model, that was trained only on weather data, obtained a lower F1-score of 0.30. Moreover, the average false positives per class were calculated, providing insights into the model's errors. For instance, the weather dataset shows an average of 12 false positives, compared to only 6.67 in the integrated dataset as shown in **Table V**.

Based on results reported in Table III, analyses of social sensor data events and text were visualized through Word-Clouds. Social sensors add more value to rare weather data events, even though these events remain as rare as illustrated in the word cloud as shown in **Figure 8**.


Fig. 8. Social sensor word cloud.

Nevertheless, this is extremely beneficial for amplifying the weak signals of weather events. Additionally, model's performance were examine by utilizing a confusion matrix, which shows the percentage of correctly classified examples per class, as depicted in **Figure 9**. Note, Labels: 0 for Normal (no event), 1 for Wind Chill and Storm, 2 for Coastal Flood and Debris Flow, 3 for Extreme Wind and Cold, 4 for Avalanche and Snowstorm, and 5 for Astronomical Low Tide.

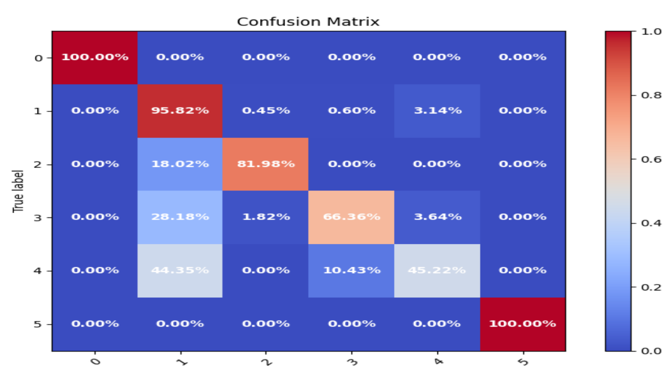


Fig. 9. Confusion matrix for the model trained on integrated data.

The proposed model exhibits excellent performance when using weather integrated with social sensor data, as opposed to using only weather data, as depicted in **Figure 10**. High-volume tweets from densely populated cities improve low-frequency event detection, with diverse reactions aiding in distinguishing rare event classes.

Utilizing social sensor data, such as tweets, is effective, however, omitting or ignoring important tweets related to severe weather can result in a slight degradation of the model's performance. **Figure 11** shows the results after excluding

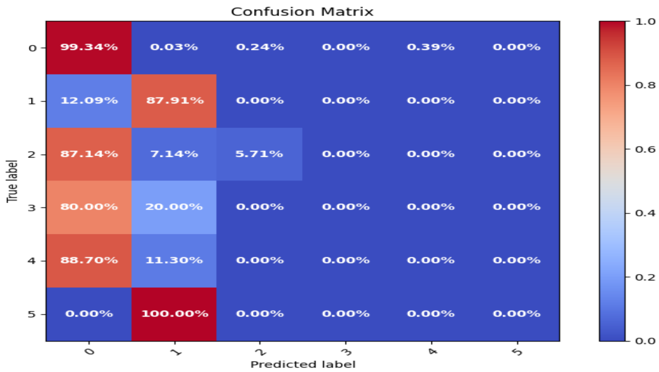


Fig. 10. Confusion matrix for the Weather-based model.

tweets related to avalanches. Consequently, we observe a decline in the model's performance.

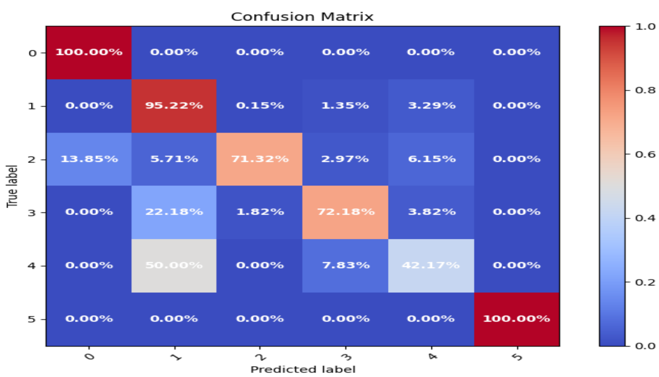


Fig. 11. Confusion matrix for the model trained on a subset of tweets from the integrated dataset.

In our study, we conducted a comparison to assess the performance of several machine learning algorithms. We specifically examined the performance of four methods while learning from combined weather and social media data: logistic regression, random forest, multilayer perceptron, and the proposed method. To conduct this comparative analysis, we used integrated dataset (weather and tweets) and the rare event were consistent across all model training and testing settings. The area under the receiver operating characteristic precision-recall Curve (PR AUC) was utilized as the critical assessment metric in this experiment, as result, **Table VI** demonstrate our findings, the logistic regression model achieved a PR AUC score of 0.89, demonstrating strong predictive performance. Random Forest and Multilayer Perceptron models both achieved PR AUC scores of 0.88, indicating a similarly high performance.The proposed method proposed resulted in a PR AUC score of 0.91, outperformed the other models. This suggest that our method provides the best balance between precision and recall among the four models tested.

In this study integrated a dataset composed of tweets about weather events. Irrelevant tweets were filtered out, resulting in a dataset of 1,029 tweets pertinent to rare weather events.

TABLE VI
COMPARISON OF THE PROPOSED WITH FOUR ALTERNATIVES WHEN LEARNING FROM INTEGRATED WEATHER AND SOCIAL MEDIAL DATA.

| Result | |
|---|---|
| ***Algorithm*** | ***PR AUC Score over all folds*** |
| Proposed method | **0.91** |
| Logistic regression | 0.89 |
| Random Forest | 0.88 |
| Multilayer Perceptron | 0.88 |

Four distinct datasets were created for this experiment.

1) Dataset A combined both weather and social sensor data. Spatial proximity was taken into account while collecting social sensor data (i.e tweets).
2) Dataset B is using social information posted in the proximity of the locations of interest and is not using weather data.
3) Dataset C combines both weather and social sensor data without taking spatial proximity while collecting social sensor data (i.e tweets).
4) Dataset D combines only a random sample of weather data from nearby weather stations.

The results outlined in **Table VII** demonstrates the model's performance achievable when relying solely on social data to make predictions, and an additional column that shows standard deviation (SD) of F1 score on each dataset in 5-fold cross validation. Models trained using Datasets A a using social media posts had much higher precision and recall and much larger confidence than a baseline relaying on social sensor data alone Dataset B or weather data Dataset D.

TABLE VII
INFLUENCE OF SOCIAL INFORMATION AND PROXIMITY TO F1 SCORE. AN INTERVAL OF F1 SCORE IS PRESENTED FOR 5-FOLD CROSS VALIDATION EXPERIMENTS AND SD IS STANDARD DEVIATION OF F1 SCORE

| Result | | | | | |
|---|---|---|---|---|---|
| ***Dataset*** | ***weather*** | ***social*** | ***proximity*** | ***F1 range in 5 folds*** | ***SD*** |
| A | ✓ | ✓ | ✓ | **[0.78, 0.83]** | **0.01** |
| B | X | ✓ | ✓ | [0.72, 0.78] | 0.06 |
| C | ✓ | ✓ | X | [0.71, 0.77] | 0.02 |
| D | ✓ | X | ✓ | [0.26, 0.33] | 0.03 |

The proposed model demonstrates high predictive confidence, with performance improving as more social sensor data is included. Note,the computational complexity of the model is driven by BERT feature extractor and the classifier. Our method enhances weather monitoring by integrating Twitter data with weather station data, enabling the prediction of rare and extreme events, and enhancing weather data quality.

## VII. CONCLUSIONS

Our study proposed a novel method for rare event classification using a multimodal data technique. The proposed model is trained on an integrated dataset comprising weather measurements and social sensor data. Our strategy integrates twitter and weather data to classify rare events in the Northwest region of the USA, specifically in Alaska. It is evident that training classification models on integrated weather data and tweets is more effective than using only weather data. Our proposed feature construction method and classifier algorithms, which include an adaptation loss function, can automatically classify rare events from weather datasets. We extracted and label tweets to aid in learning patterns for six rare events, enabling the ML model to generalize to unseen events. Our study shows limitations in rural areas due to sparse social data. Future work aims to improve early risk prediction in these regions by leveraging spatial correlations between weather events and social sensor data from nearby populated areas.

## Acknowledgment

The research was sponsored by the United States Army Corps of Engineers (USACE) Engineer Research and Development Center (ERDC) Geospatial Research Laboratory (GRL) 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 ERDC 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. During this work, Mr. H. Otudi was funded by the College of Computer Science and Information Technology at Jazan University in Saudi Arabia.

## References

[1] Ebi, K., Vanos, J., Baldwin, J., Bell, J., Hondula, D., Errett, N., Hayes, K., Reid, C., Saha, S., Spector, J. & Others Extreme weather and climate change: population health and health system implications. *Annual Review Of Public Health*. **42**, 293-315 (2021).

[2] Vitart, F. & Robertson, A. The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *Npj Climate And Atmospheric Science*. **1**, 3 (2018).

[3] Naveau, P., Hannart, A. & Ribes, A. Statistical methods for extreme event attribution in climate science. *Annual Review Of Statistics And Its Application*. **7** pp. 89-110 (2020).

[4] Jacques-Dumas, V., Ragone, F., Borgnat, P., Abry, P. & Bouchet, F. Deep learning-based extreme heatwave forecast. *Frontiers In Climate*. **4** (2022).

[5] Salcedo-Sanz, S., Pérez-Aracil, J., Ascenso, G., Del Ser, J., Casillas-Perez, D., Kadow, C., Fister, D., Barriopedro, D., Garcıa-Herrera, R., Restelli, M. & Others Analysis, characterization, prediction and attribution of extreme atmospheric events with machine learning: a review. *ArXiv Preprint ArXiv:2207.07580*. (2022).

[6] Asch, A., J Brady, E., Gallardo, H., Hood, J., Chu, B. & Farazmand, M. Model-assisted deep learning of rare extreme events from partial observations. *Chaos: An Interdisciplinary Journal Of Nonlinear Science*. **32** (2022).

[7] Matsuoka, D. & Carley, J. Can machine learning models trained using atmospheric simulation data be applied to observation data?. *Experimental Results*. **3** pp. e7 (2022).

[8] Uddin, M., Al Amin, M., Le, H., Abdelzaher, T., Szymanski, B. & Nguyen, T. On diversifying source selection in social sensing. *2012 Ninth International Conference On Networked Sensing (INSS)*. pp. 1-8 (2012).

[9] Kıcıman, E. OMG, i have to tweet that! a study of factors that influence tweet rates. *Proceedings Of The International AAAI Conference On Web And Social Media*. **6**, 170-177 (2012).

[10] Lu, X., Zhou, M. & Qi, L. Analyzing temporal-spatial evolution of rare events by using social media data. *2017 IEEE International Conference On Systems, Man, And Cybernetics (SMC)*. pp. 2684-2689 (2017).

[11] Brendel, C., Dymond, R.Aguilar, M. An interactive web app for retrieval, visualization, and analysis of hydrologic and meteorological time series data. *Environmental Modelling Software*. **117** pp. 14-28 (2019).

[12] Stanojevic, M., Alshehri, J. & Obradovic, Z. Surveying public opinion using label prediction on social media data. *Proceedings Of The 2019 IEEE/ACM International Conference On Advances In Social Networks Analysis And Mining*. pp. 188-195 (2019).

[13] Dai, A., Karl, T., Sun, B. & Trenberth, K. Recent trends in cloudiness over the United States: A tale of monitoring inadequacies. *Bulletin Of The American*.

[14] Huang, W., Deng, Y., Hui, S. & Wang, J. Image Inpainting with Bilateral Convolution. *Remote Sensing*. **14**, 6140 (2022)

[15] Kotsiantis, S., Kostoulas, A., Lykoudis, S., Argiriou, A. & Menagias, K. Filling missing temperature values in weather data banks. *2006 2nd IET International Conference On Intelligent Environments-IE 06*. **1** pp. 327-334 (2006).

[16] Otudi, H., Dokic, T., Mohamed, T., Kezunovic, M., Hu, Y. & Obradovic, Z. Line Faults Classification Using Machine Learning on Three Phase Voltages Extracted from Large Dataset of PMU Measurements. *Proceedings Of The Annual Hawaii International Conference On System Sciences Proceedings Of The 55th Hawaii International Conference On System Sciences*. (2022).

[17] Hahner, M., Sakaridis, C., Dai, D. & Van Gool, L. Fog simulation on real LiDAR point clouds for 3D object detection in adverse weather. *Proceedings Of The IEEE/CVF International Conference On Computer Vision*. pp. 15283-15292 (2021).

[18] Mumenthaler, C., Renaud, O., Gava, R. & Brosch, T. The impact of local temperature volatility on attention to climate change: Evidence from Spanish tweets. *Global Environmental Change*. **69** pp. 102286 (2021).

[19] Chen, X., Cho, Y. & Jang, S. Crime prediction using Twitter sentiment and weather. *2015 Systems And Information Engineering Design Symposium*. pp. 63-68 (2015).

[20] Meijer, H., Truong, J. & Karimi, R. Document embedding for scientific articles: Efficacy of word embeddings vs TFIDF. *ArXiv Preprint ArXiv:2107.05151*. (2021).

[21] Purwandari, K., Sigalingging, J., Cenggoro, T. & Pardamean, B. Multi-class weather forecasting from twitter using machine learning aprroaches. *Procedia Computer Science*. **179** pp. 47-54 (2021).

[22] Triasmoro, S., Ratnasari, V. & Rumiati, A. Comparison performance between rare event weighted logistic regression and truncated regularized prior correction on modelling imbalanced welfare classification in Bali. *2018 International Conference On Information And Communications Technology (ICOIACT)*. pp. 108-113 (2018)

[23] Abbasi, A., Javed, A., Yasin, A., Jalil, Z., Kryvinska, N. & Tariq, U. A large-scale benchmark dataset for anomaly detection and rare event classification for audio forensics. *IEEE Access*. **10** pp. 38885-38894 (2022).

[24] Turner, J., McDonald, M. & Hu, H. An Interdisciplinary Approach to Misinformation and Concept Drift in Historical Cannabis Tweets. *2023 IEEE 17th International Conference On Semantic Computing (ICSC)*. pp. 317-322 (2023).

[25] NCEI Storm Events Database — National Centers for Environmental Information — ncdc.noaa.gov. (https://www.ncdc.noaa.gov/stormevents/,0).Accessed 17-Jun-2023.

[26] Müller, M., Salathé, M. & Kummervold, P. Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter. arXiv 2020. *ArXiv Preprint ArXiv:2005.07503*.

[27] Nemoto, K., Hamaguchi, R., Imaizumi, T. & Hikosaka, S. Classification of rare building change using cnn with multi-class focal loss. *IGARSS 2018-2018 IEEE International Geoscience And Remote Sensing Symposium*. pp. 4663-4666 (2018).

[28] Mahmood, M., Patra, R., Raja, R. & Sinha, G. A novel approach for weather prediction using forecasting analysis and data mining techniques. *Innovations In Electronics And Communication Engineering: Proceedings Of The 7th ICIECE 2018*. pp. 479-489 (2019).