CWE-8
June 7-11, 2026
London, ON, Canada

Automated classification of thunderstorm wind events using deep learning: Implications for United States wind climatology

David T. Roegner[a], Franklin T. Lombardo[b], Robert J. Trapp[c], Evan McDonald[d], Faith Shepherd[e]

[a]University of Illinois Urbana-Champaign, Urbana, IL, USA, dtr2@illinois.edu
[b]University of Illinois Urbana-Champaign, Urbana, IL, USA, lombaf@illinois.edu
[c]University of Illinois Urbana-Champaign, Urbana, IL, USA, jtrapp@illinois.edu
[d]University of Illinois Urbana-Champaign, Urbana, IL, USA, evanjm4@illinois.edu
[e]University of Illinois Urbana-Champaign, Urbana, IL, USA, faithms4@illinois.edu

Summary

This study uses a CNN with a ResNet-34 backbone and ArcFace loss in order to accurately classify thunderstorm wind events captured by the ASOS network in the United States into respective thunderstorm types. A testing accuracy of nearly 89% was achieved on an independent dataset. Other work has shown that for Mesoscale Convective Systems (MCS), the rear-inflow jet mechanism is primarily responsible for generating extreme winds and not the downburst mechanisms. Analysis of events from 2000–2005 reveals that MCSs were responsible for 67.1% of extreme thunderstorm wind events ($>$ 50 kts), whereas Single-Cell thunderstorms produced only 5.6% of events.

Keywords: *Thunderstorm, Machine Learning, Convolutional Neural Network, Mesoscale Convective System, Rear-inflow jet, Downburst*

1 INTRODUCTION

Thunderstorm winds dominate the extreme wind climatology across the United States (Kelly et al., 1985). However, current engineering models rely heavily on simulations of isolated downbursts, despite evidence that different thunderstorm types possess distinct wind-producing mechanisms resulting in different wind characteristics (Gunter and Schroeder, 2015). Recent studies indicate that the rear-inflow jet (RIJ) in Mesoscale Convective Systems (MCS) may be a more significant driver of extreme winds than the downburst mechanism (Killion, 2024). To evaluate the validity of the downburst assumption, extreme wind events must be categorized by their parent thunderstorm type.

The thunderstorm types of interest—MCS, Single-Cell, and Unorganized Convection— can be distinguished in radar reflectivity data. An MCS is defined by a continuous convective area exceeding 75 km, while a Single-Cell thunderstorm consists of a small, isolated convective core representative of the classic downburst mechanism. Unorganized Convection represents multiple scattered areas of convection in no organized fashion. Since manual classification of decades worth of thunderstorm events is infeasible due to the volume and subjectivity of the classes, this study utilizes transfer learning and ArcFace loss to train a Convolutional Neural Network (CNN) to automate the classification of radar reflectivity images.

2 METHODOLOGY

2.1 Data Acquisition and Pre-processing

Thunderstorm wind events with gusts greater than 25 kts used for training were identified by the Automated Surface Observing Systems (ASOS) network using the methodology outlined in Lombardo et al. (2009). Independence between thunderstorm events was ensured by keeping only the maximum wind gust recorded within any 3-hour time window and 250 km spatial radius around the ASOS station. The dataset was split into an 80/20 ratio for training and validation.

For each independent thunderstorm event, a 500 km × 500 km radar reflectivity plot centered on the ASOS station of interest was generated using the Iowa State Environmental Mesonet dataset, which provides reflectivity mosaics at a 5-minute and 1 km resolution. Thunderstorm candidates were identified by searching for pixels exceeding the 35 dBZ reflectivity threshold (yellow pixels). A flood-fill algorithm was then used on all connected pixels representing reflectivity values in the 20 dBZ bin or higher (green, yellow, orange and red). This process fills internal voids within the convective region and generates a final mask for the thunderstorm candidate.

Using the resulting segmentation, a thunderstorm candidate image was created by masking areas outside the polygon and resizing the bounding box to 224 × 224 pixels while maintaining the aspect ratio, resulting in images like the ones in Figure 1. This process was repeated until all convective pixels in the original domain were assigned to a thunderstorm candidate. An area threshold of 30 km^2 was applied to reduce the dataset size and filter out noise.

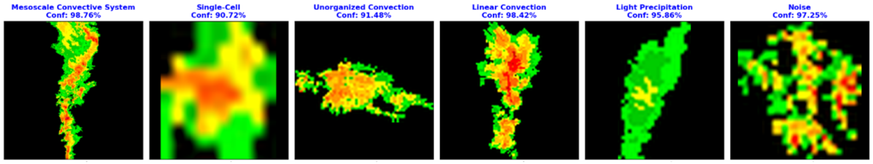

Figure 1: Example radar reflectivity images used in the CNN for all the classification categories. The prediction confidence is included with the classification for each image.

2.2 Labeling Strategy

Each thunderstorm candidate was labeled by three experts into six categories: Mesoscale Convective System (MCS), Single-Cell thunderstorm, Unorganized Convection, Linear Convection (elongated convective regions not meeting the 75 km MCS threshold), Light Precipitation (areas with only small clusters of 35 dBZ pixels), and Noise (non-meteorological targets).

To combat the subjectivity inherent in thunderstorm type classification, random subsets of the training data were labeled twice by two experts to create a ”high-confidence” dataset where labeler’s agreed. This high-confidence dataset was used to train an initial model reaching 80% accuracy, which was used to label the remaining thunderstorm candidates of the training dataset. A pseudo-labeling approach was used, where if both the model prediction and the human labeler agreed on a classification, the event was added to the final training dataset. A separate testing dataset was created with thunderstorm wind events from 2001 that exceeded 40 kt using expert labelers.

2.3 Classification Model

Thunderstorm events were classified using a deep learning framework consisting of a ResNet-34 backbone and a secondary Multi-Layer Perceptron (MLP) for geometric attributes. The CNN extracted visual data from the radar reflectivity images, while the MLP simultaneously processed the geometric metrics, log-precipitation area, log-bounding box aspect ratio, precipitation density (ratio of polygon area to bounding box area), precipitation linearity, and edge smoothness. These inputs were combined and trained using ArcFace loss to maximize the separation between categories.

The training process involved fine-tuning the ResNet-34 parameters with a learning rate of

5×10^{-5}. In addition, heavy class imbalance was addressed by using class-weighted Focal Loss in order to heavily penalize wrong predictions for minority classes. The model was optimized using AdamW with a Cosine Annealing scheduler to escape local minima. The final model was selected based on the highest macro F1-score to ensure balanced performance across all thunderstorm categories.

3 RESULTS

3.1 Model Performance

The model achieved a training accuracy of 93.8% and a validation accuracy of 91.8% after 36 epochs. The final validation macro F1-score was 0.879. On the independent testing dataset, the model achieved an overall accuracy of 88.8% and a macro F1-score of 0.83. Performance was strong for the Single-Cell Thunderstorms (98%), Light Precipitation (94%), and Unorganized Convection (90%). MCS thunderstorms achieved 83% accuracy where 14% of MCS cases were mislabeled as Linear Convection. Linear Convection achieved an F1-score of 0.68, with the confusion matrix indicating that 33% of Linear events were misclassified as Unorganized Convection.

3.2 Thunderstorm Wind Climatology

The trained model was then used to classify thunderstorm events from 2000 to 2005 for independent thunderstorm events greater than 25 kt. A determination of which thunderstorm type produced the wind gust at the ASOS station was determined by using the closest thunderstorm to the actual station. Figure 2 shows the breakdown of the thunderstorm types responsible for producing the winds seen at the ASOS station where the blue bar shows events with a threshold of 25 kt, and the orange bar shows the density of events with a threshold of 50 kt. It can clearly be seen that MCS thunderstorms are responsible for the majority of the thunderstorm winds in the United States, and for the extreme winds the proportion compared to other categories increases.

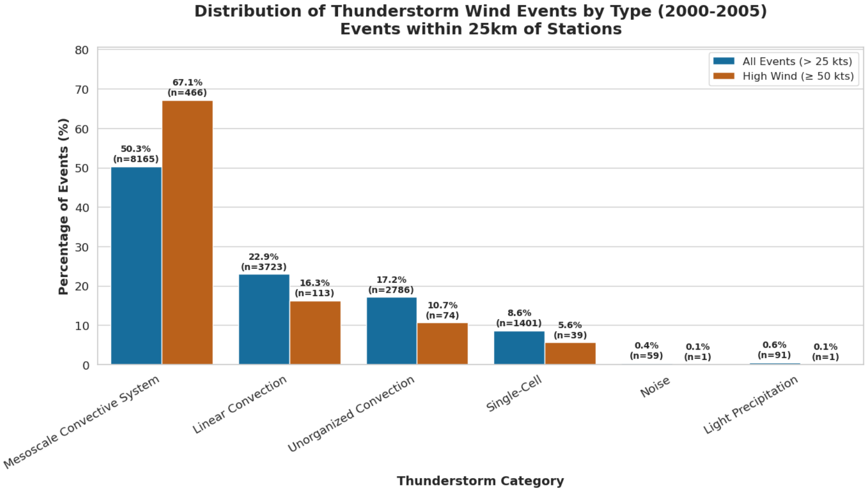

Figure 2: Preliminary analysis showing the distribution of wind events by the respective thunderstorm type from 2000 to 2005. Only wind events with an identifiable thunderstorm within 25 km of the station is included. The blue bar is all wind events above 25 kt and the orange bar is events greater than 50 kt.

4 DISCUSSION

The classification errors between Linear Convection and Unorganized Convection likely is a result of the similarity in appearance between the types. In addition, the Linear category serves as a hybrid class between Unorganized Convection and MCS. Similarly, the confusion between Linear Convection and MCSs suggests that the model fails to recognize the spatial thresholds (e.g., the 75 km criterion) which is not unexpected. However, the distinction between these categories may not be critical, as both thunderstorm types likely share similar wind-producing mechanisms.

Despite these classification challenges, the results show a clear trend in how extreme winds are produced. For all thunderstorm winds in the United States, a MCS thunderstorm is likely the source. As the wind speed increases, it becomes even more likely that an MCS produced the wind. This finding challenges the current reliance on the isolated downburst model for simulating thunderstorm winds. Since MCS winds are often driven by the rear-inflow jet rather than downbursts (Killion, 2024), current design standards may not accurately represent the duration and characteristics of the thunderstorm winds that cause the most damage.

5 CONCLUSIONS

This study successfully developed a CNN to automate the classification of thunderstorm types, enabling the generation of a comprehensive extreme wind climatology. The results demonstrate that MCS thunderstorms are responsible for the majority of extreme thunderstorm wind events ($>$ 50 kt) in the United States.

These findings suggest that the isolated downburst model, which is primarily represented by Single-Cell thunderstorms, may not accurately reflect the dominant wind threat to structures and crops. Since MCS winds are likely driven by the rear-inflow jet (RIJ) rather than isolated downdrafts, current simulations may not accurately represent the most damaging events. Consequently, future work must focus on statistically quantifying the distinct wind characteristics and production mechanisms associated with each thunderstorm type to improve wind load modeling.

ACKNOWLEDGEMENTS

Funding is gratefully acknowledged from the National Science Foundation (NSF). This research was supported by StEER CMMI-1841667, NSF RAPID AGS-2054706, NSF CAREER CMMI-2144760, and NSF Disaster Resilience Research Grant CMMI-2242578.

REFERENCES

Gunter, W.S., Schroeder, J.L., 2015. High-resolution full-scale measurements of thunderstorm outflow winds. J. Wind Eng. Ind. Aerodyn. 138, 13–26. https://doi.org/10.1016/j.jweia.2014.12.005

Kelly, D.L., Schaefer, J.T., Doswell III, C.A., 1985. Climatology of nontornadic severe thunderstorm events in the United States. Mon. Wea. Rev. 113, 1997–2014.

Killion, K.J., 2024. Quantification of mechanisms for damaging wind generation in simulated derecho events (Master's thesis). University of Illinois Urbana-Champaign, Urbana, IL. https://hdl.handle.net/2142/124382

Lombardo, F.T., Main, J.A., Simiu, E., 2009. Automated extraction and classification of thunderstorm and non-thunderstorm wind data for extreme-value analysis. J. Wind Eng. Ind. Aerodyn. 97, 120–131. https://doi.org/10.1016/j.jweia.2009.03.001