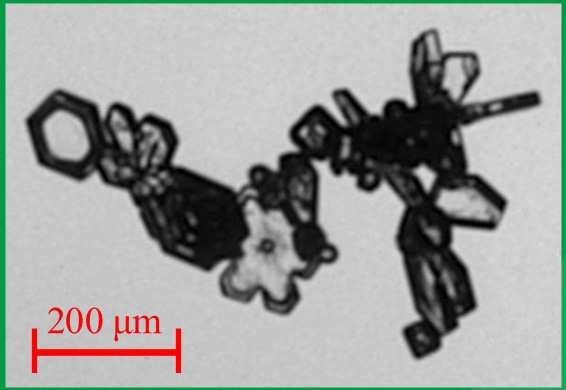
Identifying Ice Crystal Chain Aggregates in Cold-Season Storms: Leveraging Machine Learning to Map Occurrence and Distribution

Christian M. Nairy[1] (christian.nairy@und.edu), David J. Delene[1], Shawn W. Wagner[1], Joseph A. Finlon[2, 3], John E. Yorks[2]

[1]University of North Dakota, Department of Atmospheric Science

[2]National Aeronautics and Space Administration (NASA) - Goddard Space Flight Center

[3]Earth System Science Interdisciplinary Center (ESSIC), University of Maryland

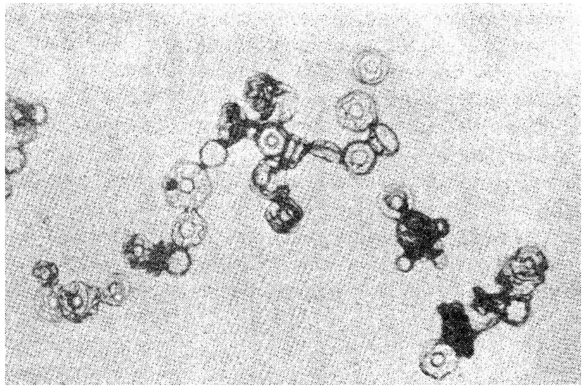


5th International Summer Snowfall Workshop - 15 - 17 September 2025, Department of Meteorology, University of Reading, UK

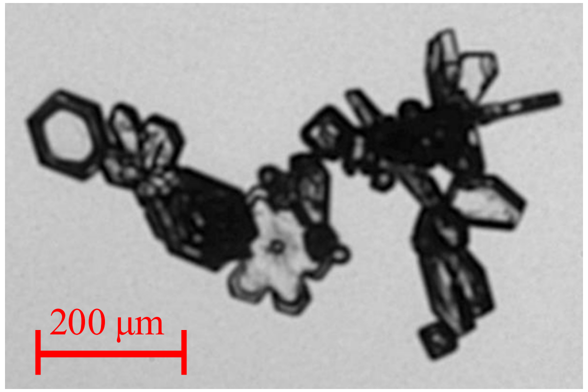
Background

- Chain Aggregate: Unusually elongated, quasi-linear ice crystal aggregate consisting of three or more discernable ice monomers joined together end-to-end or by small joints (Nairy, 2022).

- Laboratory experiments show that chain aggregation occurs in the presence of strong electric fields (>60 kV m^{-1}) between -5 and -37 °C (Maximum efficiency at -8 °C) [Saunders & Wahab, 1975].
    - Chain aggregation efficiency found to be temperature depended.



- Chains are observed in mid- to upper-level tropical, subtropical, and midlatitude summertime storms.
    - Cirrus anvils during CapeEx19 (Florida) *where in-situ E-Fields < 60 kV* m^{-1}.
    - Deep stratiform cirrus anvils during ABFM-II (Florida) *where in-situ E-Fields < 60 kV* m^{-1}.
    - Cirrus outflow during EMERALD-II (Darwin)
    - Convective cloud tops during CIRCLE-II (western Europe)
    - Overshooting tops during DC3 (eastern Colorado)



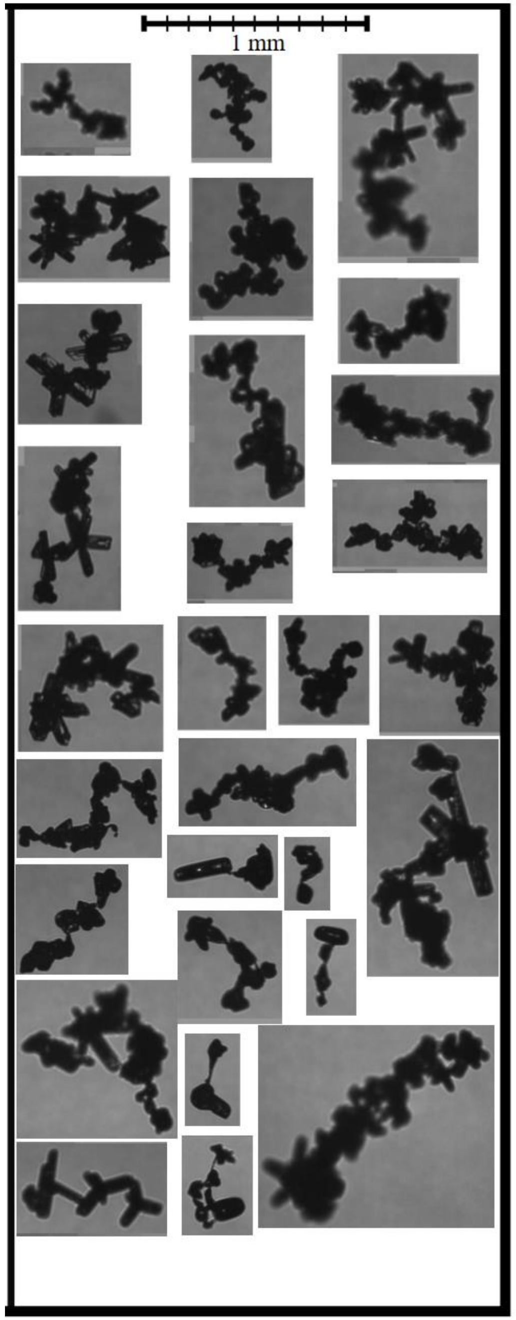
Motivation for Research

- The chain aggregation process is still not well understood.
  - Where and how do these aggregates form?
  - Laboratory results and aircraft observations are not in agreement.
  - Chain aggregates are not represented in current cloud models.

- Recent observations of chain aggregates in cold-season storms (NASA IMPACTS) challenge existing theories and add uncertainty to our understanding.

  Nairy, Christian M., David J. Delene, Joseph A. Finlon, John E. Yorks, Emma Järvinen, Martin Schnaiter, Andrew J. Heymsfield, Andrew G. Detwiler, Lynn A. McMurdie, Unveiling In-situ Observations of Ice Crystal Chain Aggregates in Winter Storms. Journal of Geophysical Research Letters, In Review, 2025. Preprint: https://d197for5662m48.cloudfront.net/documents/publicationstatus/271990/preprint_pdf/0e470a6a2eee19f7dad1defe572d8aea.pdf

Hawkeye-CPI Images

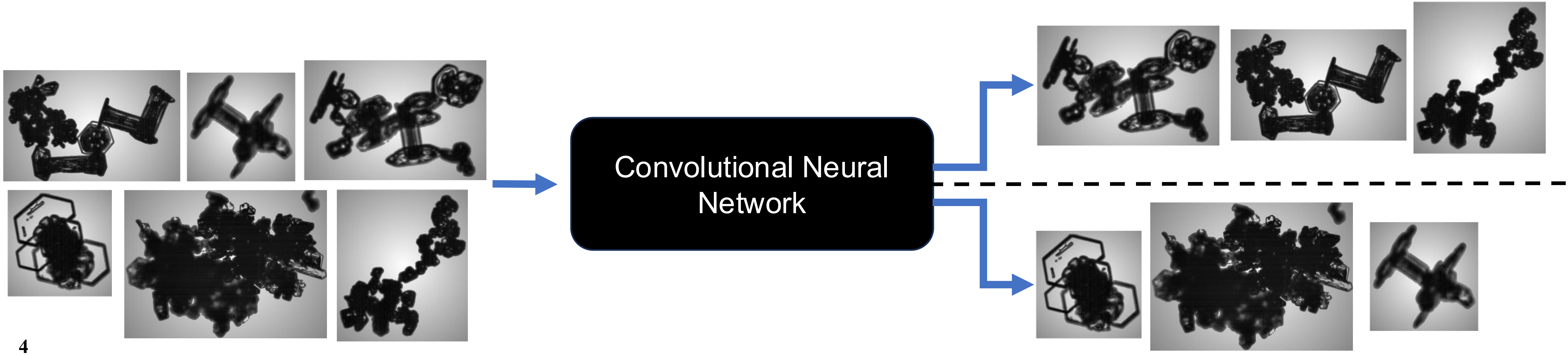


IMPACTS

Objective

Understanding chain aggregate formation within cloud systems requires contextual analysis, but manual classification of cloud particles is extremely time-consuming.

Presentation Objective: Employ a supervised Convolutional Neural Network (CNN) to differentiate ice crystal chain aggregates from non-chain aggregates.



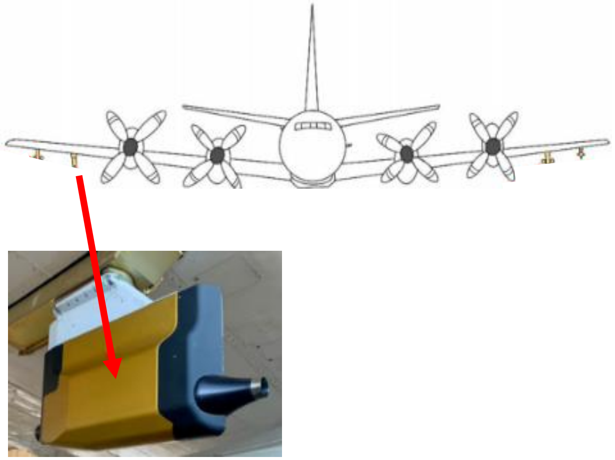
Methodology

1. Utilize 56,193 manually classified ice crystals ($D_{max} > 150$ μm) obtained from Cloud Particle Imager (CPI) observations during NASA's IMPACTS field campaign.

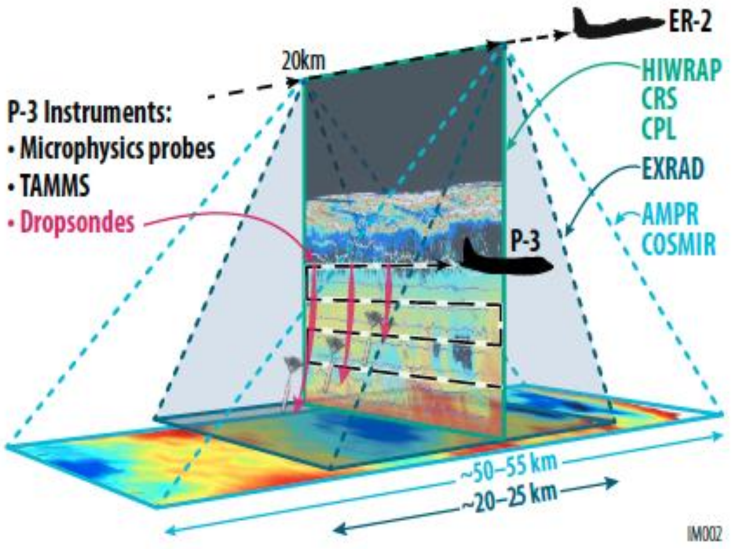
2. Train, validate, and evaluate multiple CNN architectures.

3. Identify the most effective CNN for distinguishing chain aggregates.

4. Apply the best CNN model to classify CPI data from other IMPACTS flights.



Hawkeye-CPI



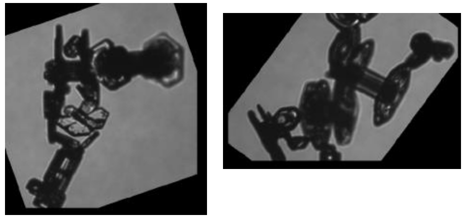
End-to-End CNN Pipeline for Chain Aggregate Classification

Classified Images

Non-chains: 55,162
Chains: 1,031

Class Balancing

Chain Agg. Image Augmentation



Input

Non-chains: **55,162**
Chains: **55,162**

80/20 Train/Validation Split

Training Dataset

Non-chains: 44,130 Chains: 44,130

Validation Dataset

Non-chains: 11,032 Chains: 11,032

CNN Architecture

ResNet18/34/50/100
VGG16/19

Training

- *Epochs:* 30
- *Learning Rate:* 0.0001
- *Batch Size:* 64
- *Optimizer:* Adam
- *Criterion:* Binary Cross Entropy Loss

Epoch Validation

- Model Calibration Analysis
- F1 Threshold Tuning
- Save Best Model Epoch

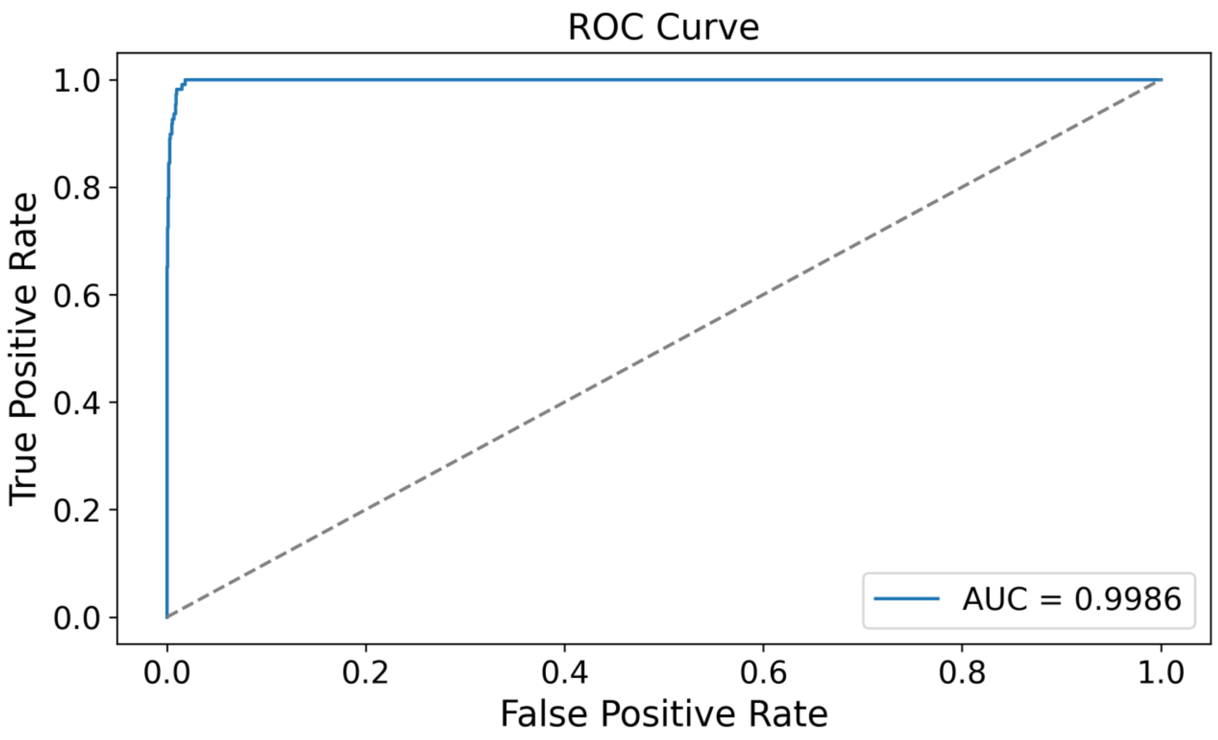
Evaluation on Unseen, Labeled Data

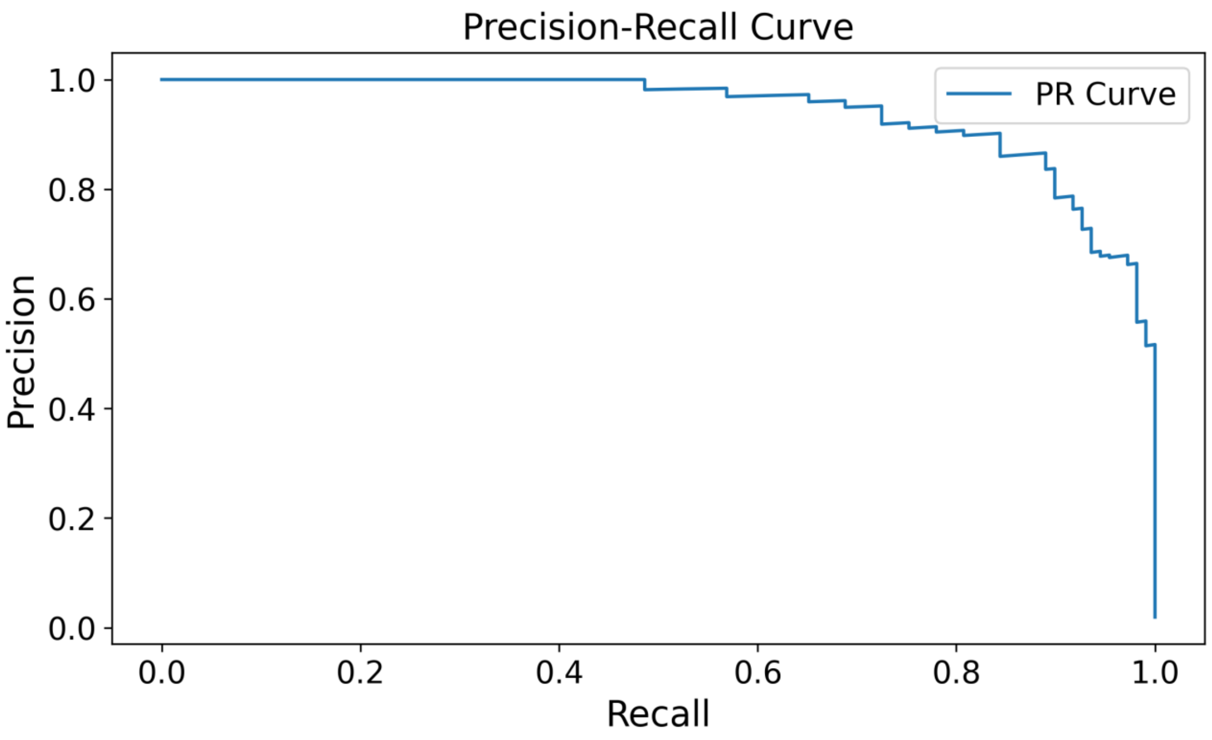
Non-chains: 5,520; Chains: 110

- Accuracy, Precision, Recall, F1 Scores
- ROC-AUC
- Error & Statistical Analysis
- Prediction Metrics

UNIVERSITY OF NORTH DAKOTA

CNN Validation on Unseen Labeled Dataset





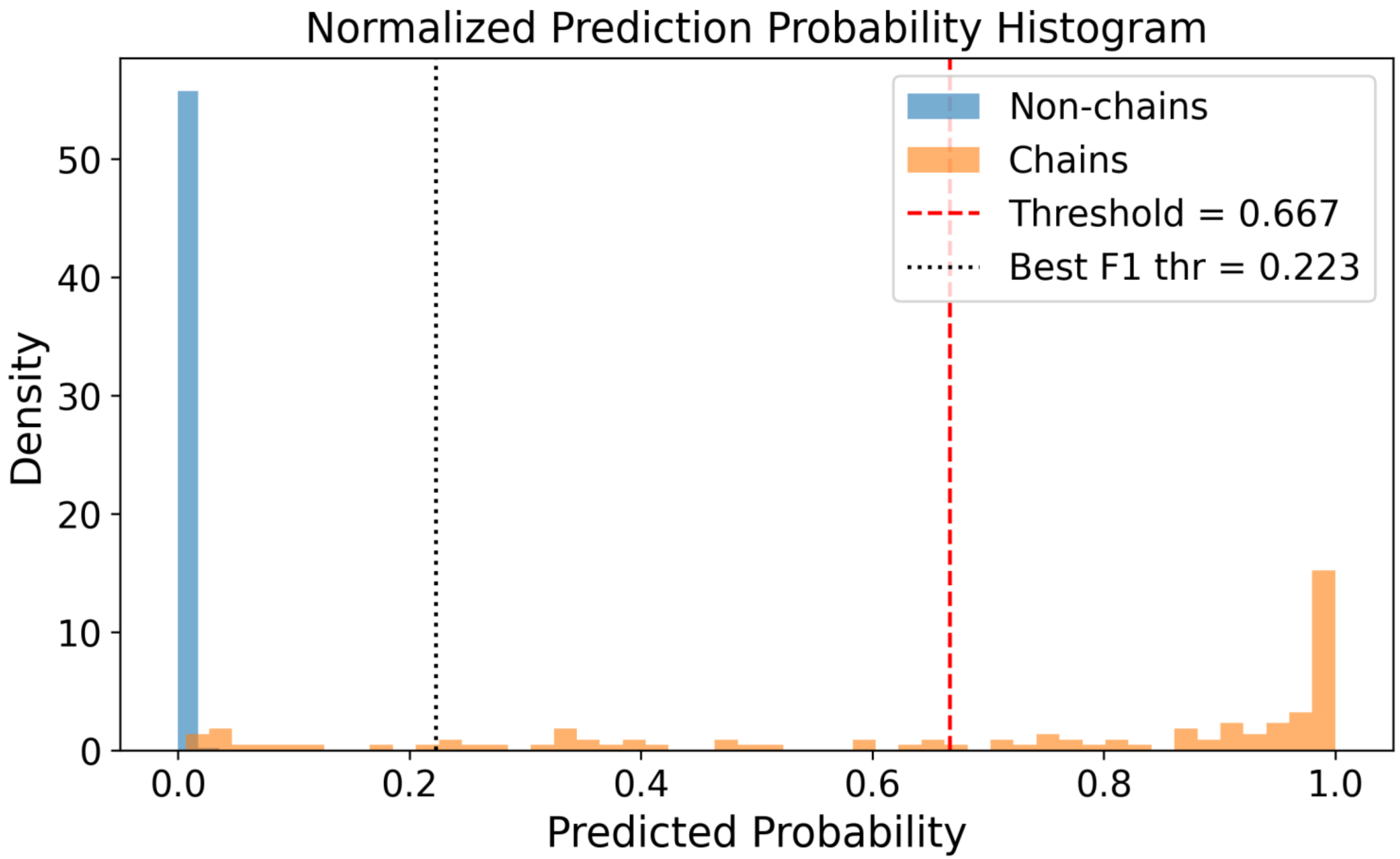
Key Takeaways (Receiver Operating Characteristic-Area Under the Curve [ROC-AUC] plot):

- **AUC = 0.9986**, indicating *near-perfect discrimination* between chain and non-chain classes.
- Very low false positive rate at high true positive rates.

Key Takeaways (P-R Curve):

- Precision remains >90% until very high recall values.
- Confirms strong detection of chains even in a highly imbalanced dataset.
- **Do we maximize precision or recall?**

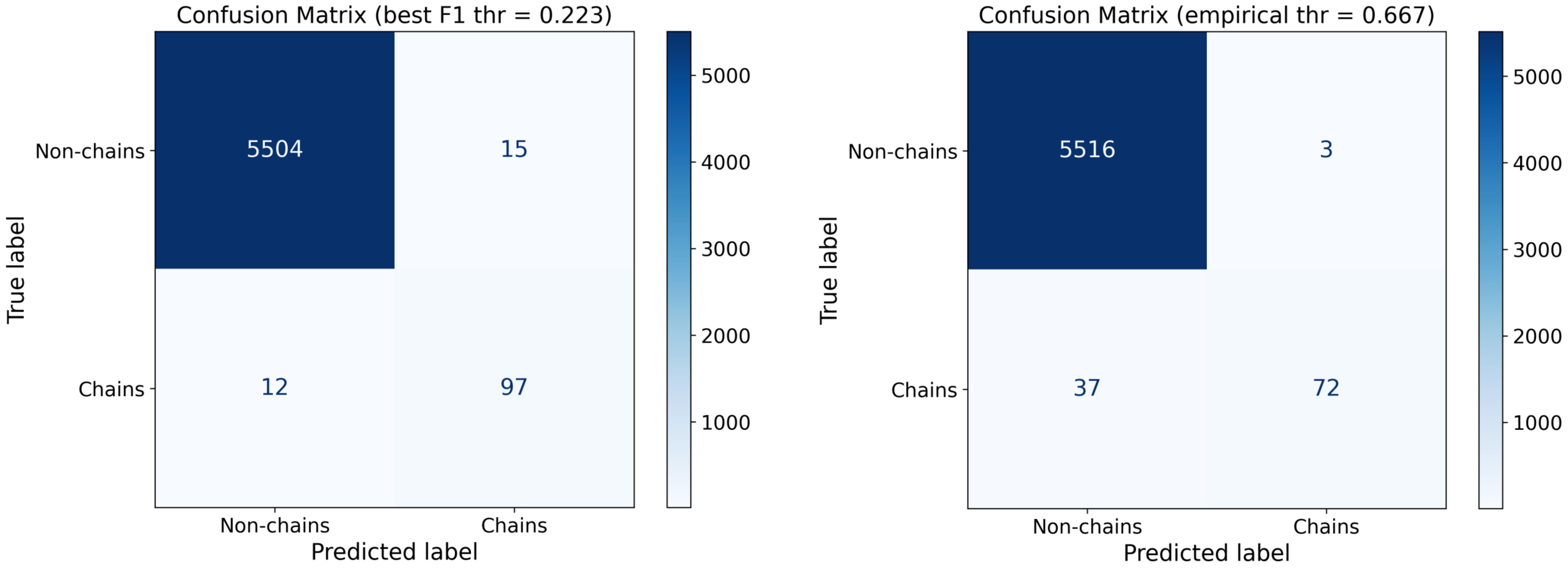
CNN Validation on Unseen Labeled Dataset



Key Takeaways:

- Model outputs are sharply separated
- Threshold choice impacts detection
  - **Optimized F1 threshold (0.223)** → more chains detected (higher recall) but susceptible to more FP.
  - **Empirical threshold (0.667)** → fewer false positives (higher precision and more conservative).

CNN Validation on Unseen Labeled Dataset



Key Takeaways:

- Lower threshold finds more chains; higher threshold ensures precision.

CNN Validation on Unseen Labeled Dataset

Class	Classification Report (Pred Thresh: 0.223)			Classification Report (Pred Thresh: 0.667)		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Non-Chain Aggregate	99%	99%	99%	99%	99%	99%
Chain Aggregate	**87%**	89%	88%	**96%**	66%	78%
Macro Accuracy	**93%**	94%	0.94%	**98%**	83%	89%
Correlation Coef.	**0.88**			**0.80**		
Brier Score	**0.004**					

Key Takeaway:
- Lower threshold boosts recall, higher threshold boosts precision.
- Low Brier score (0.004) indicates our CNN's probabilities are reliable, not just the hard classifications.

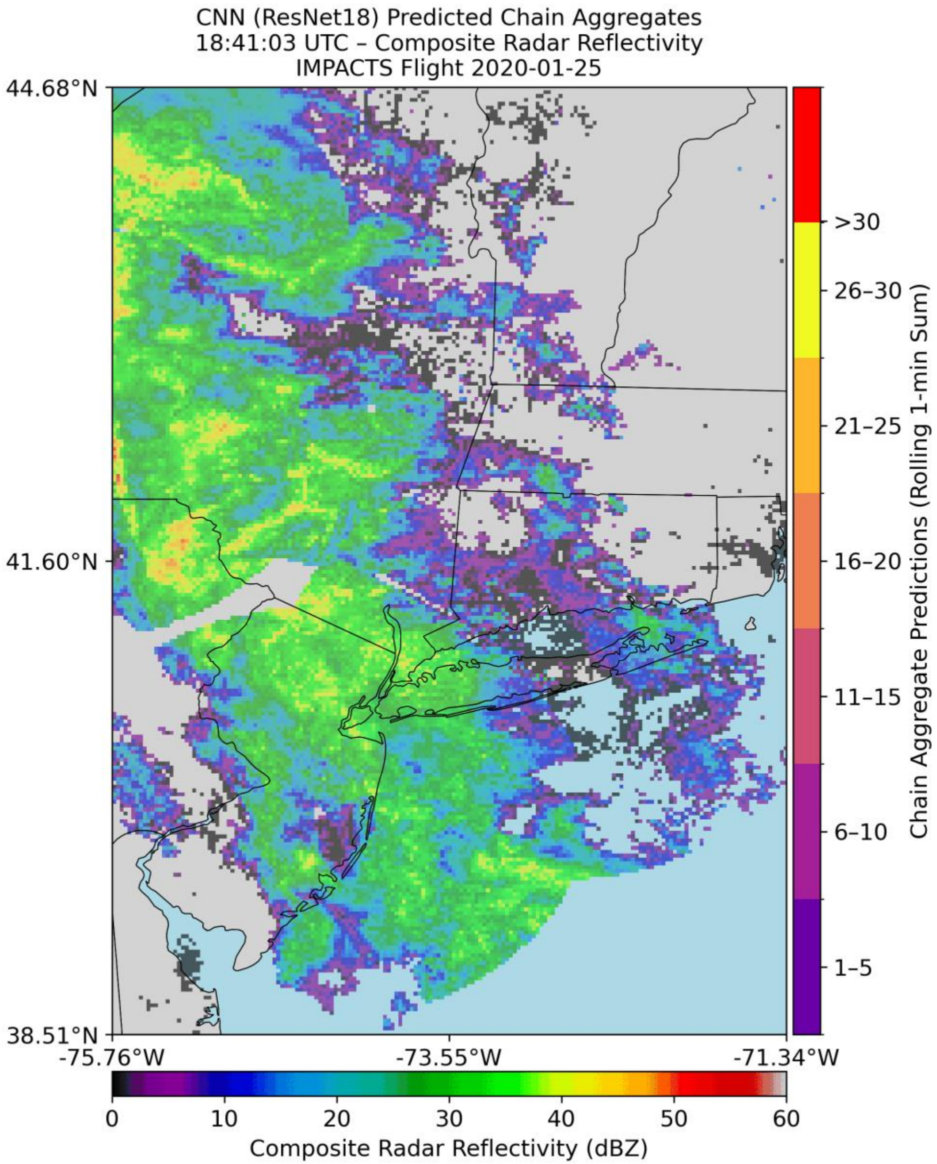
CNN Validation on Unseen Labeled Dataset

Class	Classification Report (Pred Thresh: 0.223)			Classification Report (Pred Thresh: 0.667)		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Non-Chain Aggregate	99%	99%	99%	99%	99%	99%
Chain Aggregate	**87%**	89%	88%	**96%**	66%	78%
Macro Accuracy	**93%**	94%	0.94%	**98%**	83%	89%
Correlation Coef.	**0.88**			**0.80**		
Brier Score	**0.004**					

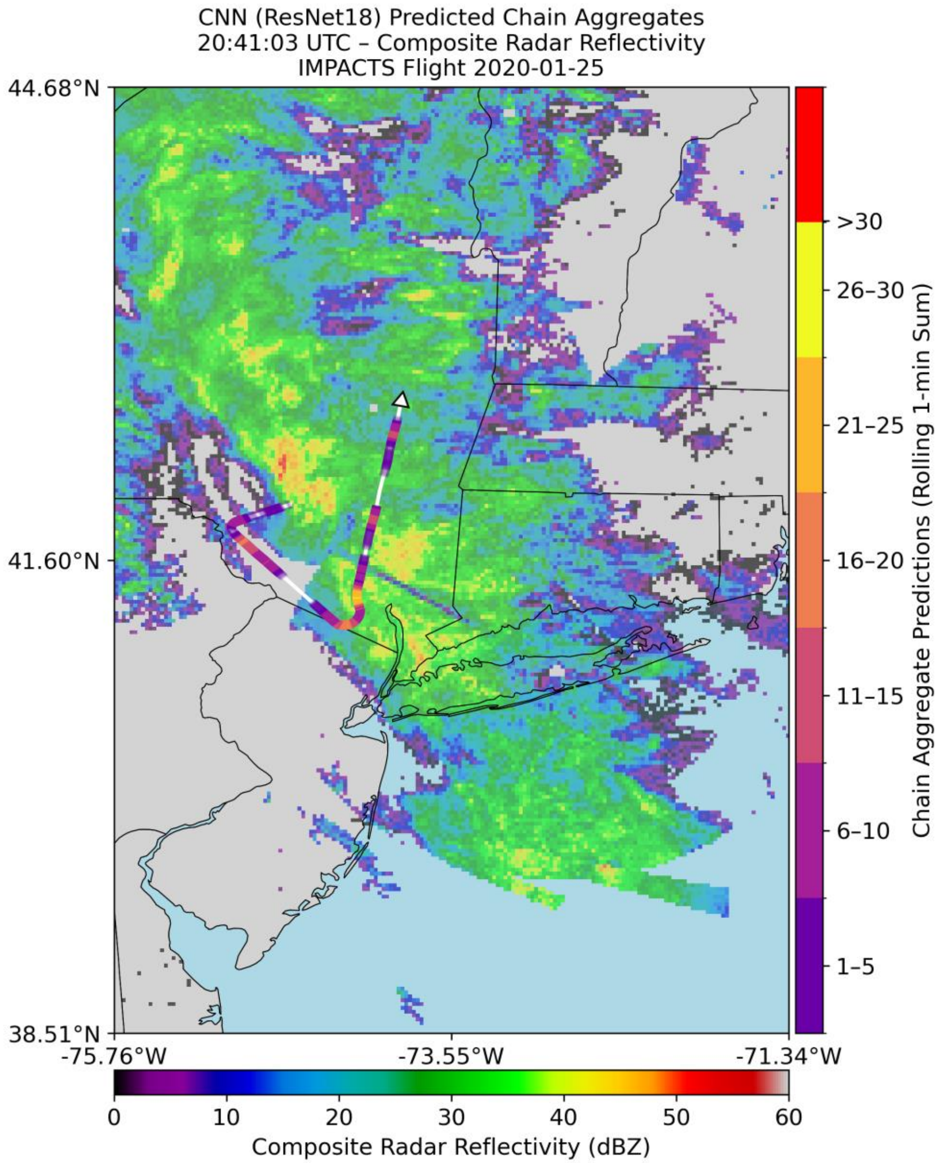
Key Takeaway:
- Lower threshold boosts recall, higher threshold boosts precision.
- Low Brier score (0.004) indicates our CNN's probabilities are reliable, not just the hard classifications.

With strong validation and reliable probability calibration, our CNN can now be applied to map the occurrence and distribution of chain aggregates throughout the IMPACTS campaign.

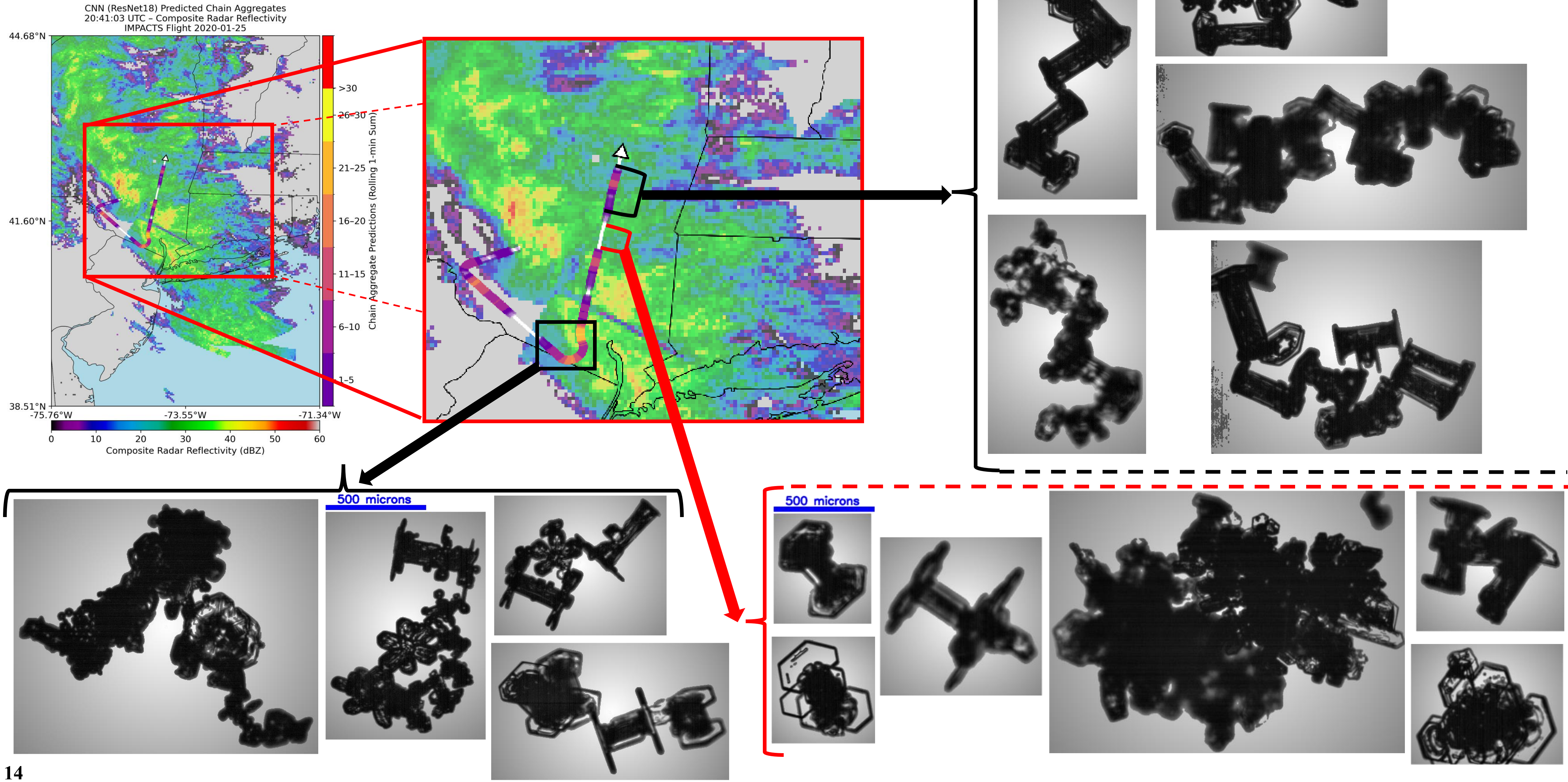
Applying CNN on IMPACTS Flight



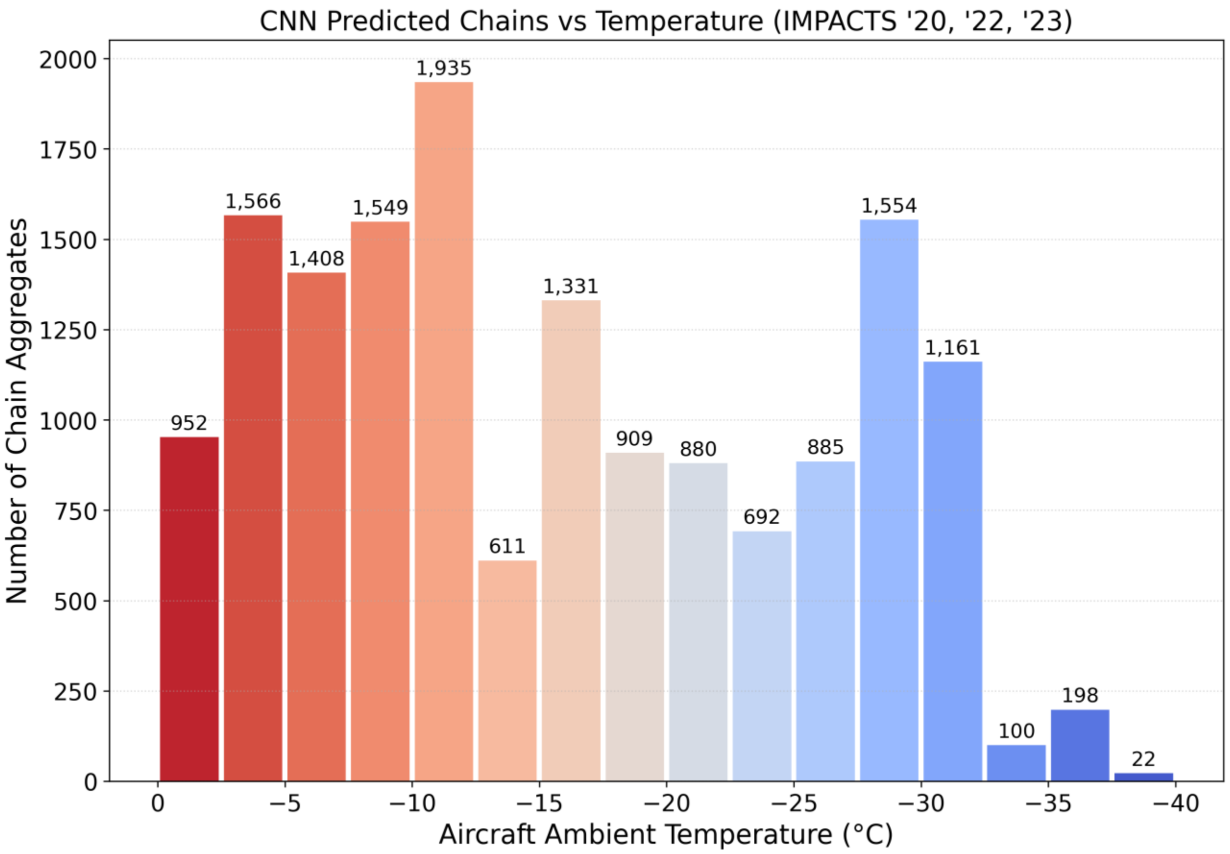
Applying CNN on IMPACTS Flight

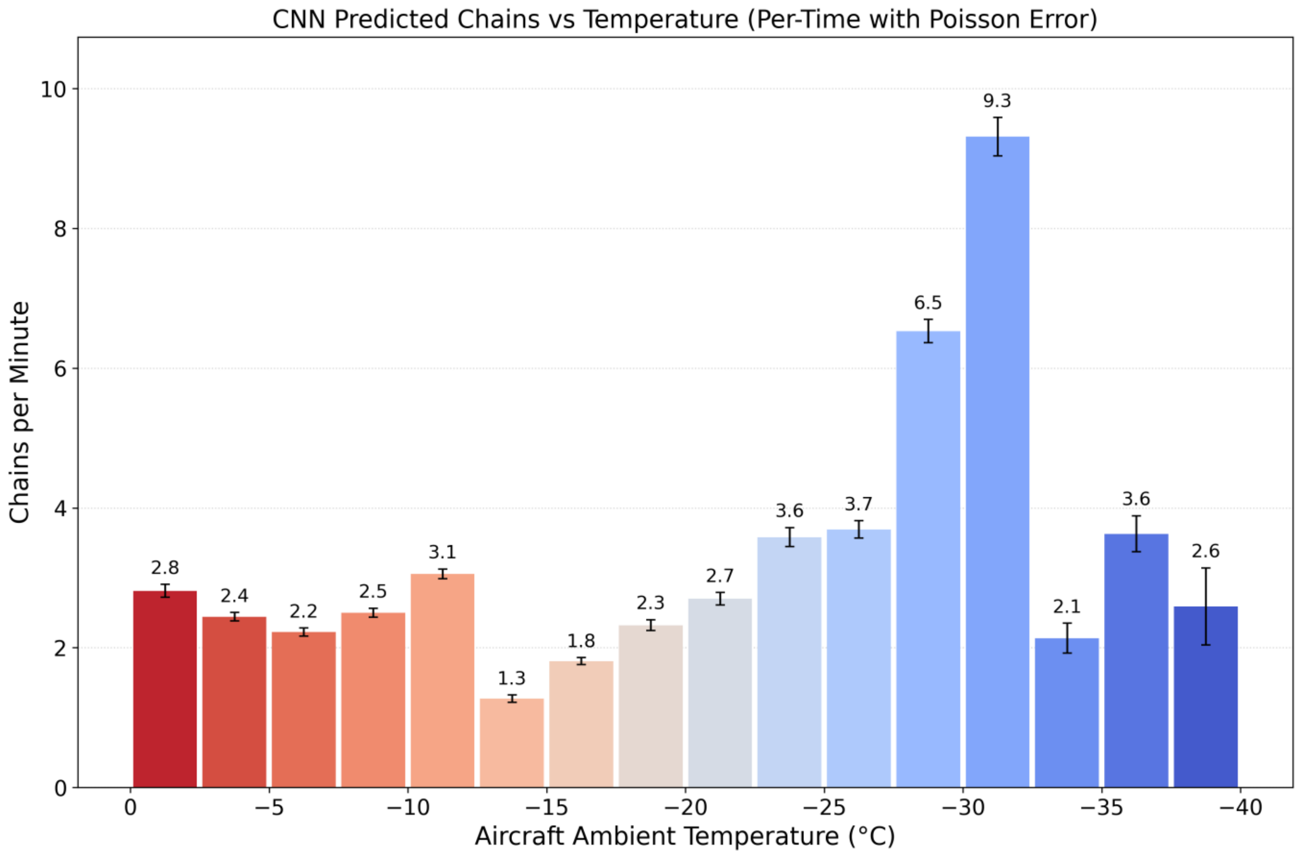


Applying CNN on IMPACTS Flight



All IMPACTS Flights (34 Total)





UND UNIVERSITY OF NORTH DAKOTA

Summary

- **Targeted CNN pipeline** identifies chain aggregates in CPI imagery and **generalizes** to unseen flights.
- **ResNet18 is the top performer**; with a conservative probability threshold (**$P > 0.66$**) it achieves high precision and **minimizes false positives**.
- The approach enables flight-wide maps of chain-aggregate occurrence/concentration at second-level resolution.

Future Work:

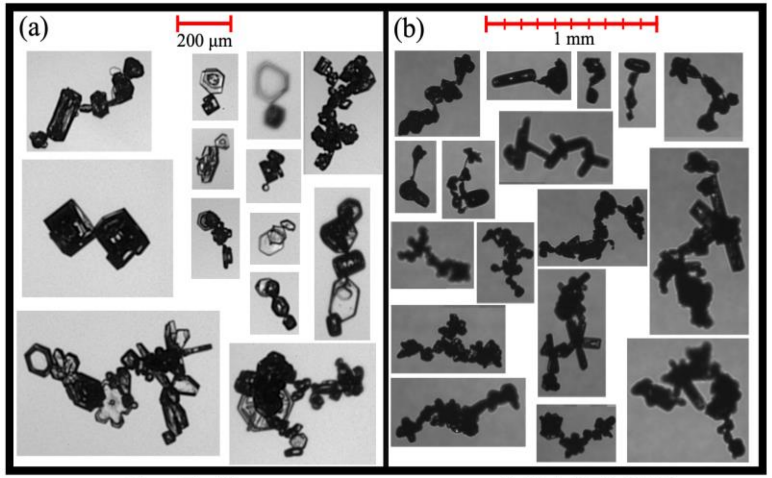
- **Contextualize detections:** identify high/low concentration regions and relate them to storm phase and environment (from coarse to fine scale).
- **Fuse measurements:** improve along-track and 3-D localization by integrating ER-2 radar/lidar and electric-field data.
- **Broaden scope:** apply across additional field campaigns and other habits; release code and catalogs.

UND UNIVERSITY OF NORTH DAKOTA

References & Acknowledgements

- Baran, A. J. (2009). A review of the light scattering properties of cirrus. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *110*(14), 1239–1260. https://doi.org/10.1016/j.jqsrt.2009.02.026
- Connolly, P. J., Saunders, C. P. R., Gallagher, M. W., Bower, K. N., Flynn, M. J., Choularton, T. W., et al. (2005), Aircraft observations of the influence of electric fields on the aggregation of ice crystals. Quarterly Journal of the Royal Meteorological Society , 131 (608), 1695–1712. https://doi.org/10.1256/qj.03.217
- Liou, K. N. (1973). Transfer of Solar Irradiance through Cirrus Cloud Layers., *J. Geophys. Res.*, *78*, 1409–1418.
- Nairy, C. M. (2022). Observations of Chain Aggregates in Florida Cirrus Cloud Anvils on 3 August 2019 during CAPEEX19 (Master's thesis), Dept. of Atmospheric Sciences, University of North Dakota, Grand Forks, North Dakota. Retrieved from https://commons.und.edu/theses/4363/
- Saunders, C. P. R., & Wahab, N. M. A. (1975). The Influence of Electric Fields on the Aggregation of Ice Crystals. *Journal of the Meteorological Society of Japan. Ser. II*, *53*(2), 121–126. https://doi.org/10.2151/jmsj1965.53.2_121
- Schmitt, C. G., and A. J. Heymsfield (2014), Observational quantification of the separation of simple and complex atmospheric ice particles, Geophys. Res. Lett., 41, 1301–1307, doi:10.1002/2013GL058781.
- Stephens, G. L., Tsay, S.-C., Stackhouse, P. W., & Flatau, P. J. (1990). The Relevance of the Microphysical and Radiative Properties of Cirrus Clouds to Climate and Climatic Feedback. *Journal of the Atmospheric Sciences*, *47*(14), 1742–1754. https://doi.org/10.1175/1520-0469(1990)047<1742:TROTMA>2.0.CO;2
- Stith, J. L., Avallone, L. M., Bansemer, A., Basarab, B., Dorsi, S. W., Fuchs, B., et al. (2014). Ice particles in the upper anvil regions of midlatitude continental thunderstorms: the case for frozen-drop aggregates. *Atmospheric Chemistry and Physics*, *14*(4), 1973–1985. https://doi.org/10.5194/acp-14-1973-2014

This research was supported by a NASA MOSAICS research grant to the University of North Dakota. Grant #: 435002210UND0029914. The IMPACTS dataset is publicly available at the NASA GHRC http://dx.doi.org/10.5067/IMPACTS/DATA101. We would also like to thank Andrew Heymsfield and Stephen Nicholls for their added expertise in this analysis.



CapeEx19 IMPACTS 2022

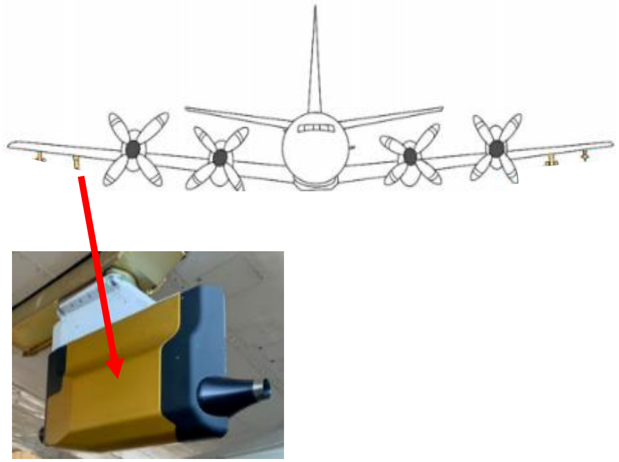
Thank you.

UND UNIVERSITY OF NORTH DAKOTA

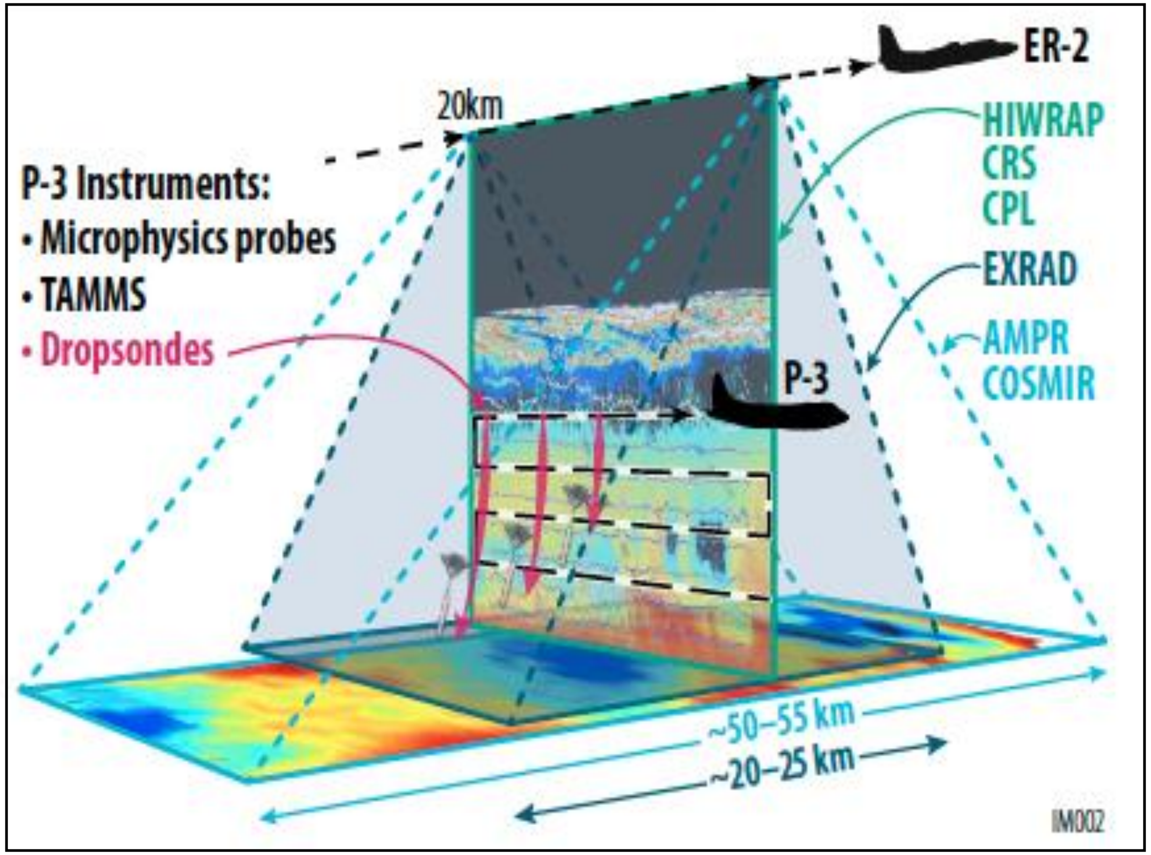
Extra Slides

Dataset/Instrumentation

- **I**nvestigation of **M**icrophysics and **P**recipitation for **A**tlantic **C**oast-**T**hreatening **S**nowstorms (IMPACTS).
- NASA P-3b Orion Research Aircraft
  - Hawkeye-Cloud Particle Imaging (CPI) Probe
  - King Probe (Liquid Water Content [LWC])



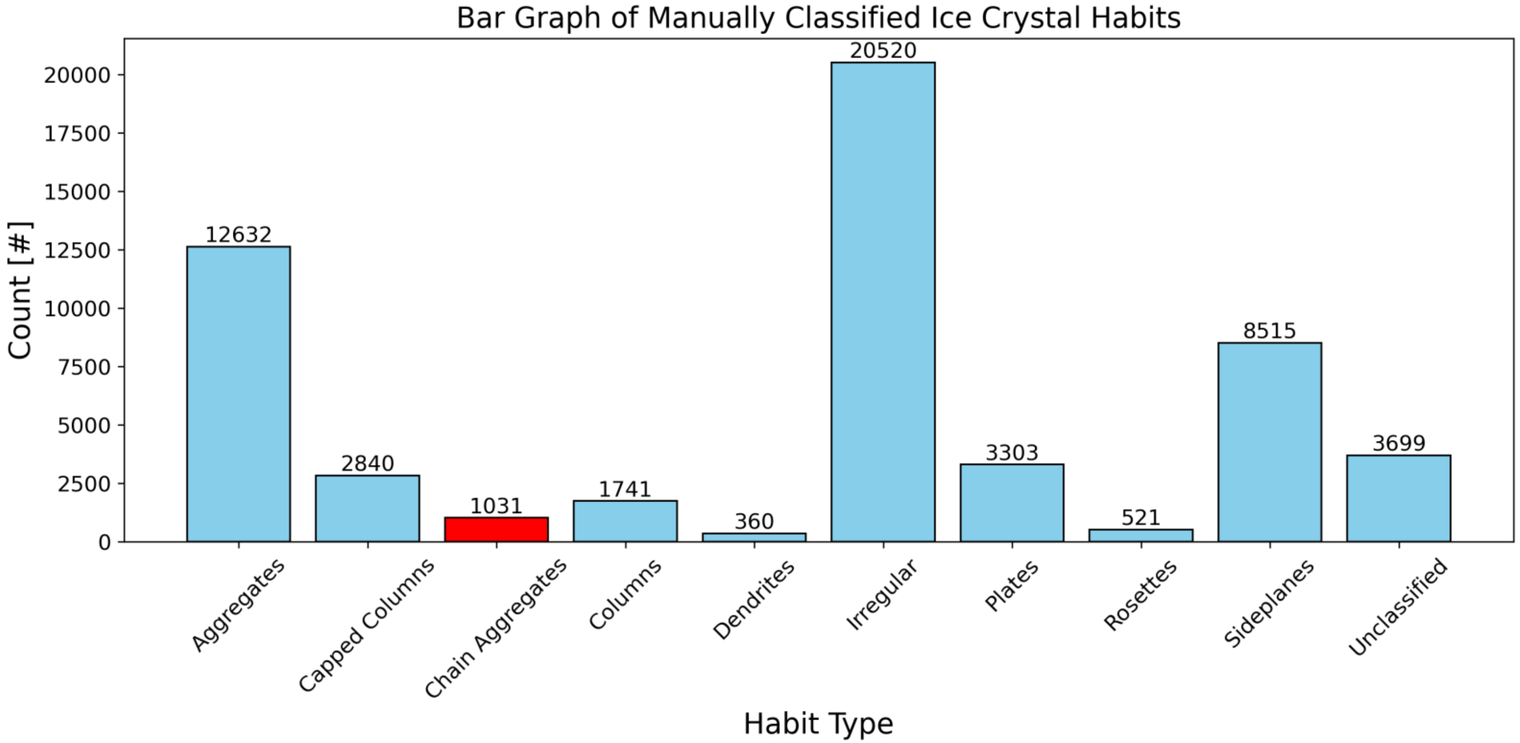
Hawkeye-CPI



Adapted from the NASA IMPACTS executive summary (https://espo.nasa.gov/impacts/).

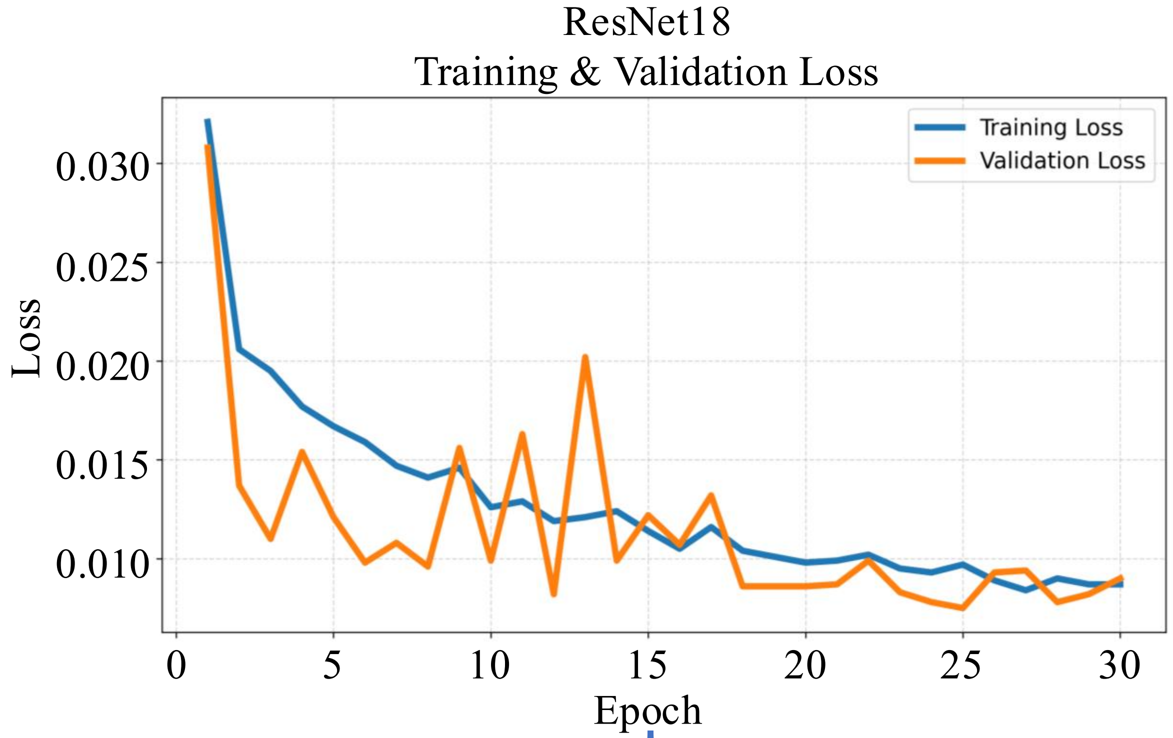
Manual Particle Classifications

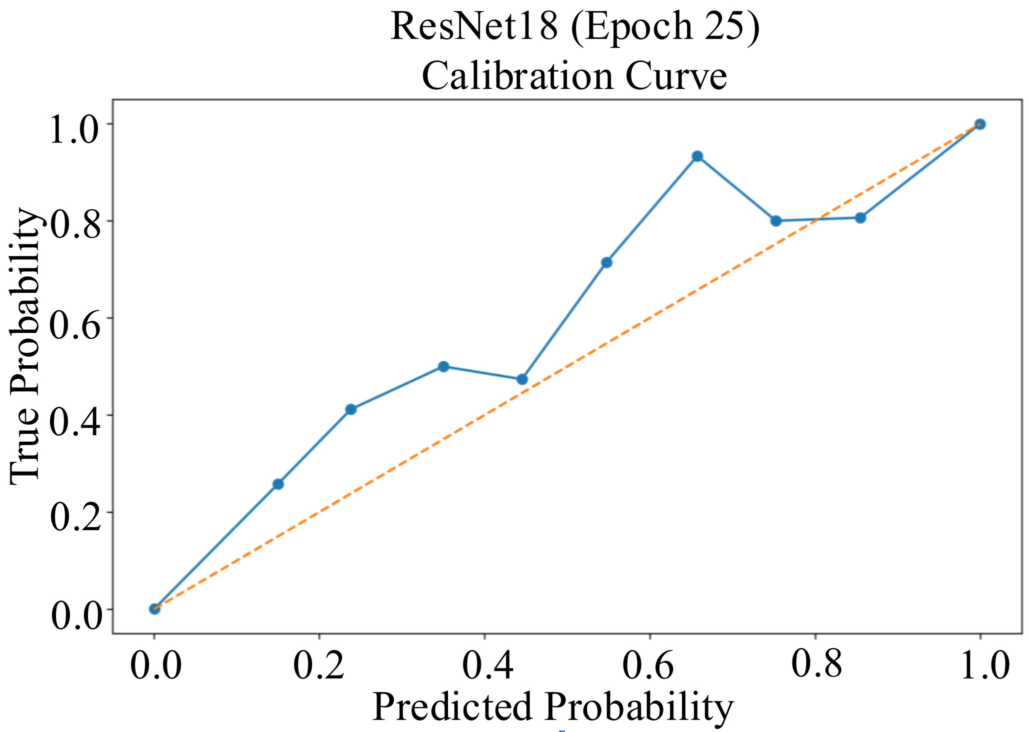
- CPI probe data from NASA IMPACTS.
- 5 Flight Segments Classified:
  Temperature Range: -36 to -5 ºC
- Total # of Particles Classified (after QA/QC): 56,193
- # of Chain Aggregates Classified (after QA/QC): 1,031



UND UNIVERSITY OF NORTH DAKOTA

CNN Training Validation

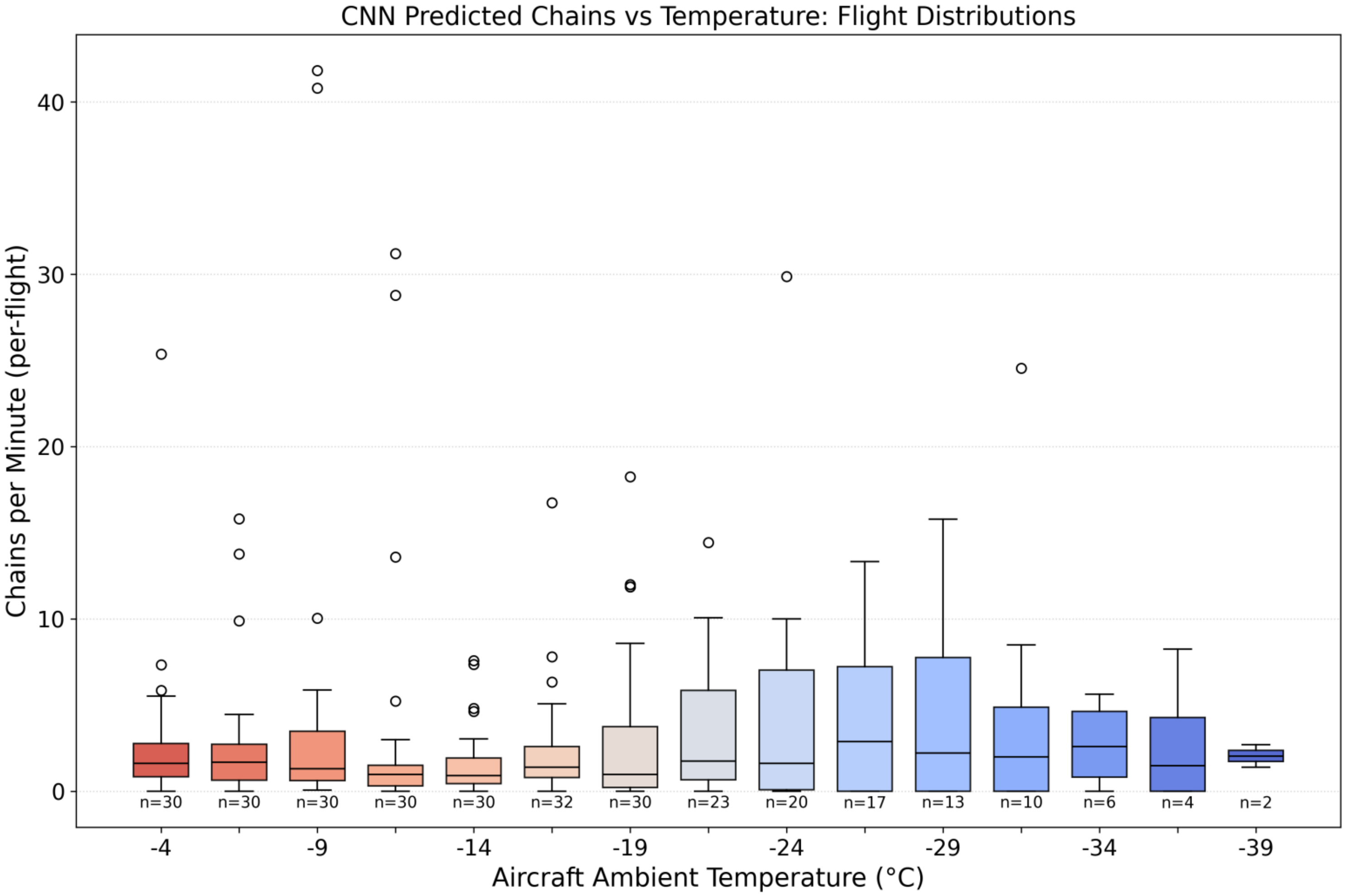




- The close tracking between training and validation loss throughout the epochs suggests **minimal overfitting**.
- The relatively **low final loss** implies the model has effectively learned discriminative features.
- The low noise and tight coupling between the curves is a **sign of a stable and well-regularized training process**.

- Curve implies that the model's **confidence scores are reliable** for both strong positives and strong negatives.
- At epoch 25, the model is well-calibrated for high-confidence predictions.
- Minor overconfidence in the mid-range, though, this model offers a good balance of accuracy and interpretability.

Distributions (Extra)



UNIVERSITY OF NORTH DAKOTA