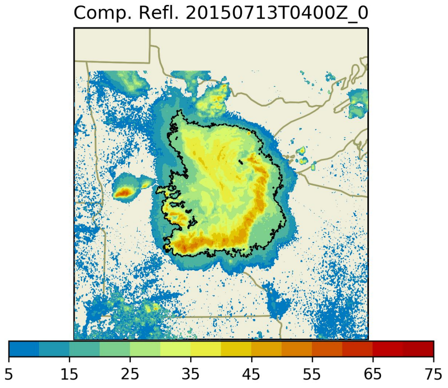
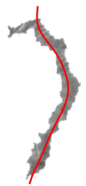
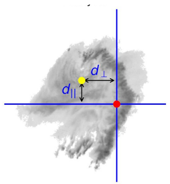
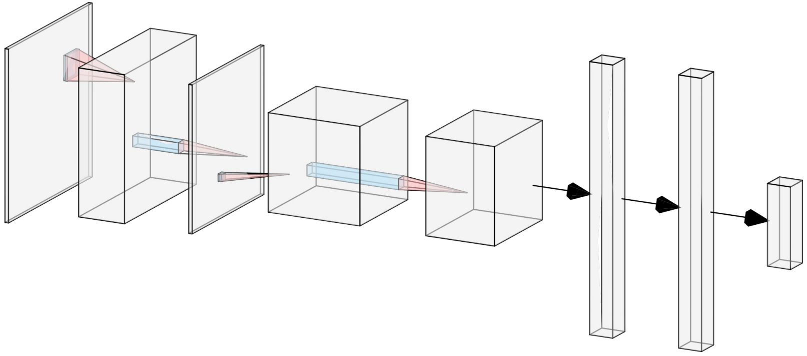
Comparison of Machine Learning Techniques for Convective Morphology Classification from Radar Imagery









Jonathan E. Thielen[1]

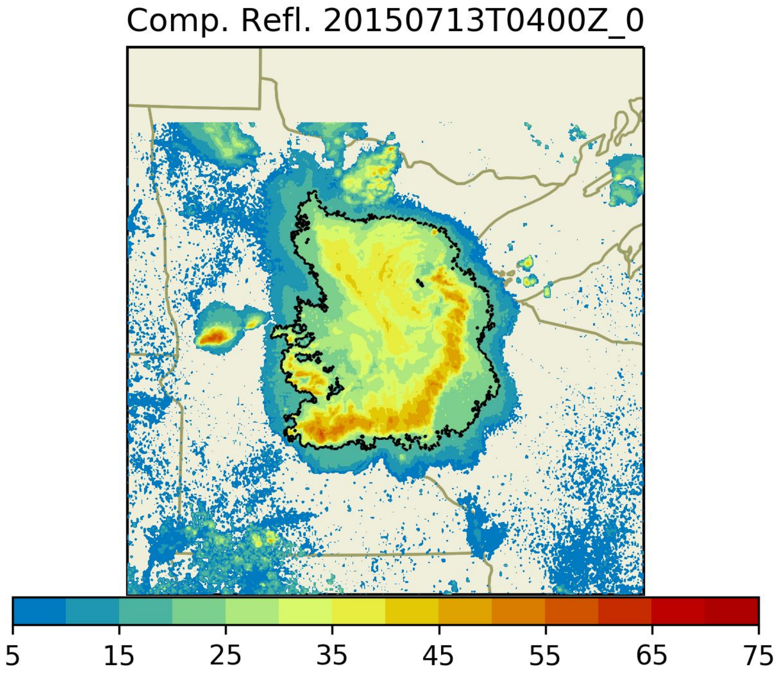
Mentors: William A. Gallus[1] and Alex M. Haberlie[2]

[1]Department of Geological and Atmospheric Sciences, Iowa State University, Ames, IA; [2]Department of Geography and Anthropology, Louisiana State University, Baton Rouge, LA

What is Convective Morphology?

What is Convective Morphology?

- Shape
- Structure
- Organization

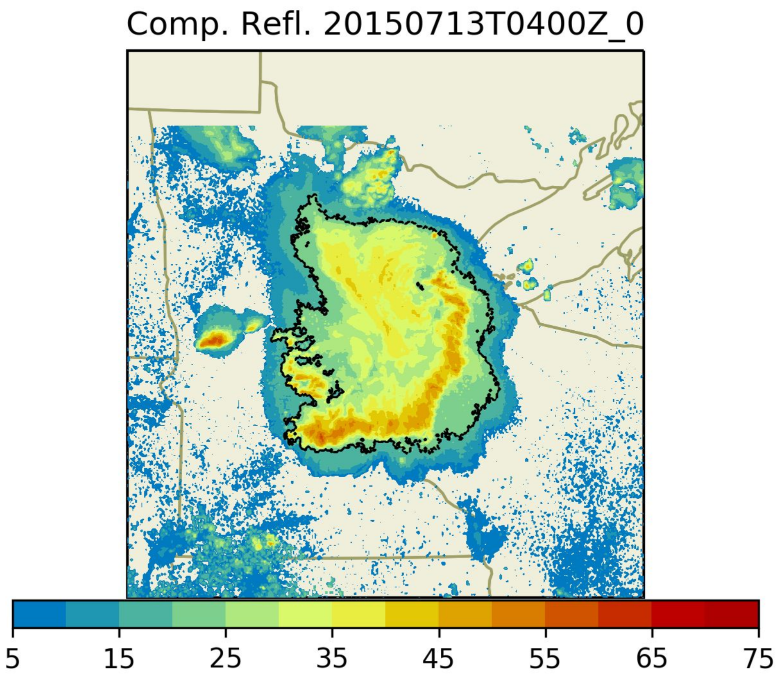


What is Convective Morphology?

- Shape
- Structure
- Organization

Importance:

- System dynamics
- Severe hazard likelihood (Gallus et al. 2008)



Classification of Convective Morphology

Classification of Convective Morphology

Subjective nine-category scheme of Gallus et al. (2008):

- Three cellular types
- Five linear types
- Non-linear

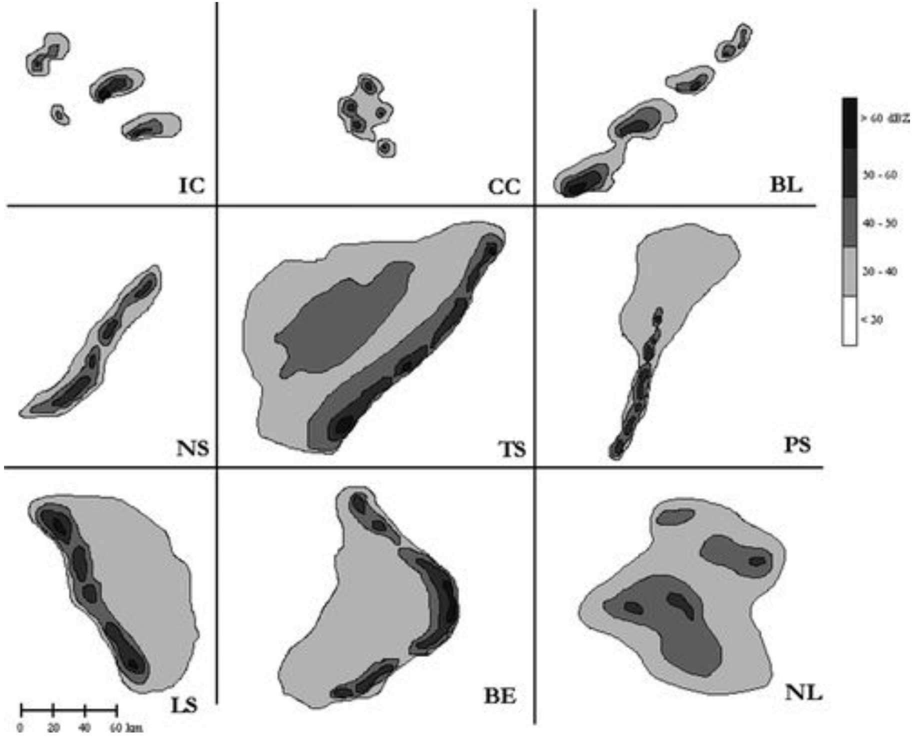


Figure from Gallus et al. (2008)

Classification of Convective Morphology

Subjective nine-category scheme of Gallus et al. (2008):

- Three cellular types
- Five linear types
- Non-linear

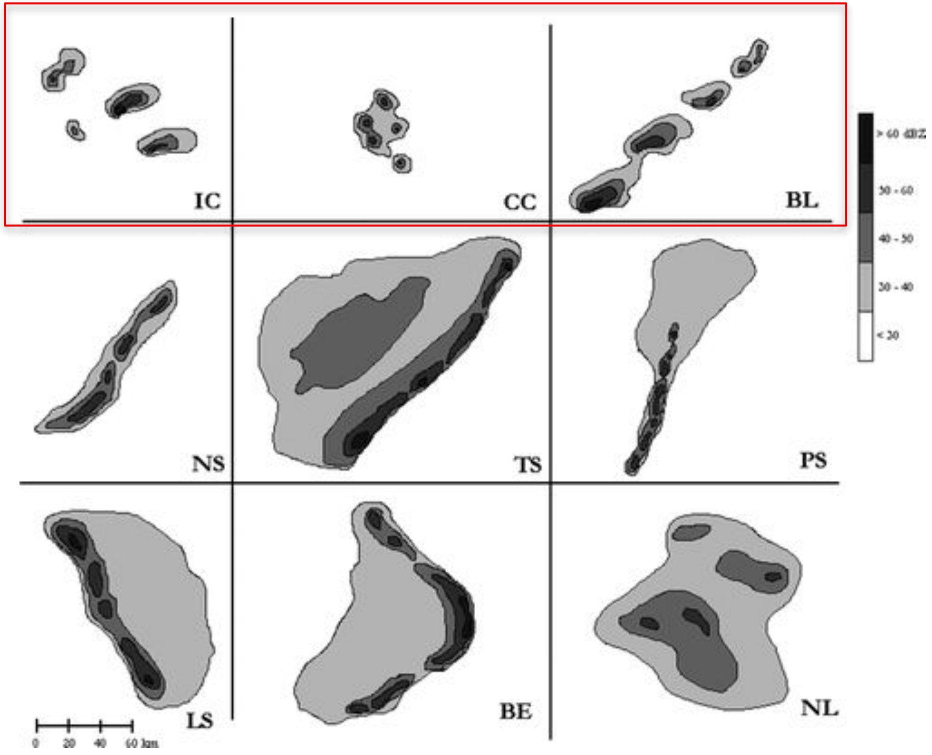


Figure from Gallus et al. (2008)

Classification of Convective Morphology

Subjective nine-category scheme of Gallus et al. (2008):

- Three cellular types
- Five linear types
- Non-linear

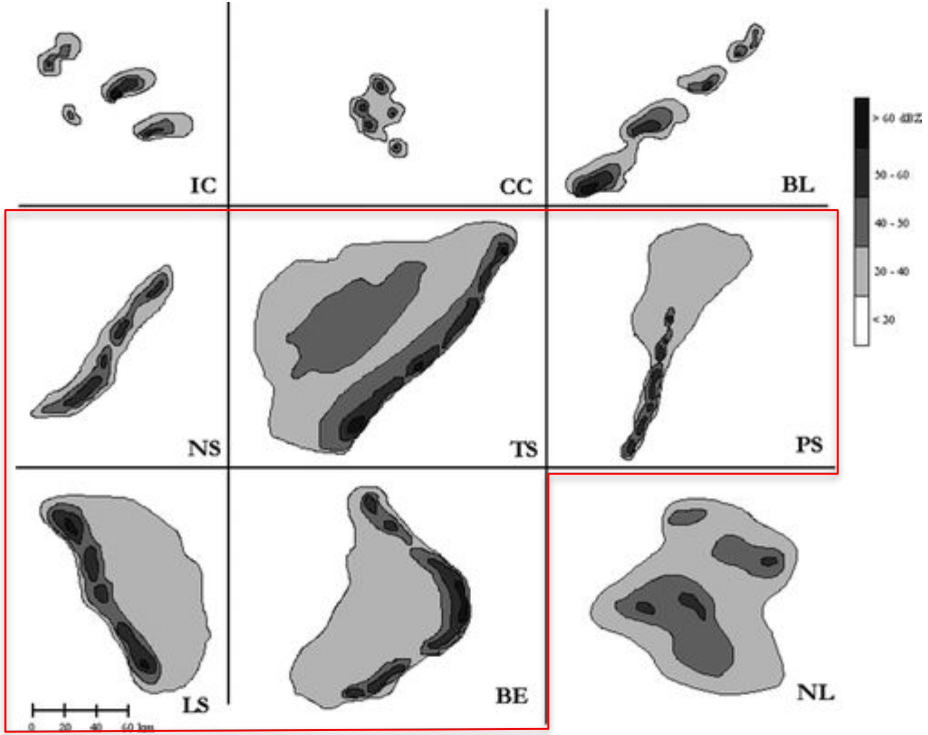


Figure from Gallus et al. (2008)

Classification of Convective Morphology

Subjective nine-category scheme of Gallus et al. (2008):

- Three cellular types
- Five linear types
- Non-linear

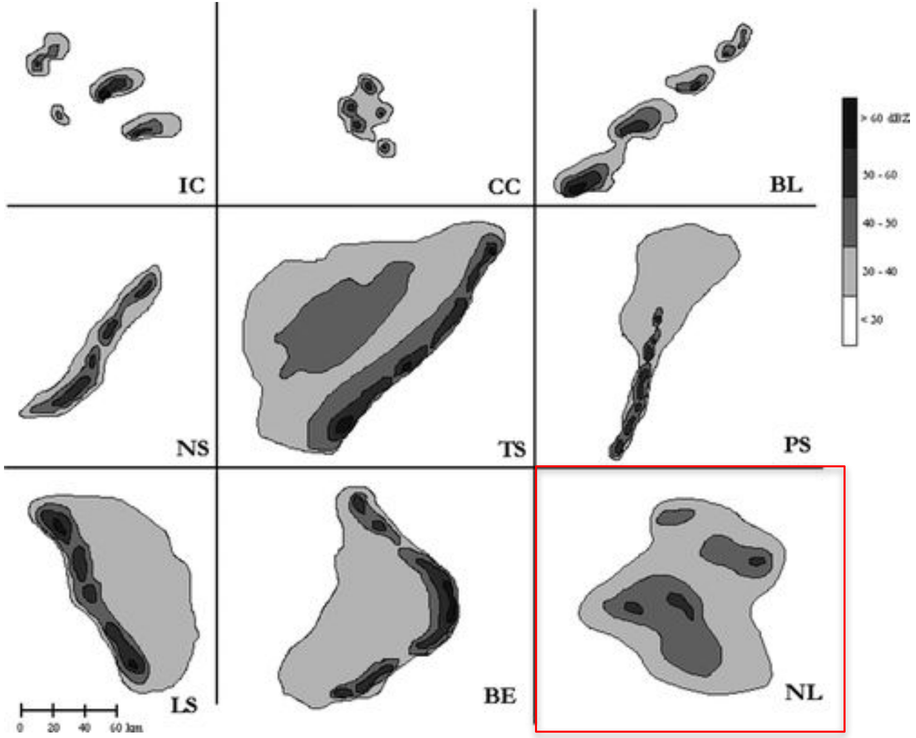


Figure from Gallus et al. (2008)

Classification of Convective Morphology

Subjective nine-category scheme of Gallus et al. (2008):

- Three cellular types
- Five linear types
- Non-linear

Manual vs. Automated:

- Sample sizes (Thielen et al. 2018)
- Consistency (Corfidi et al. 2016)

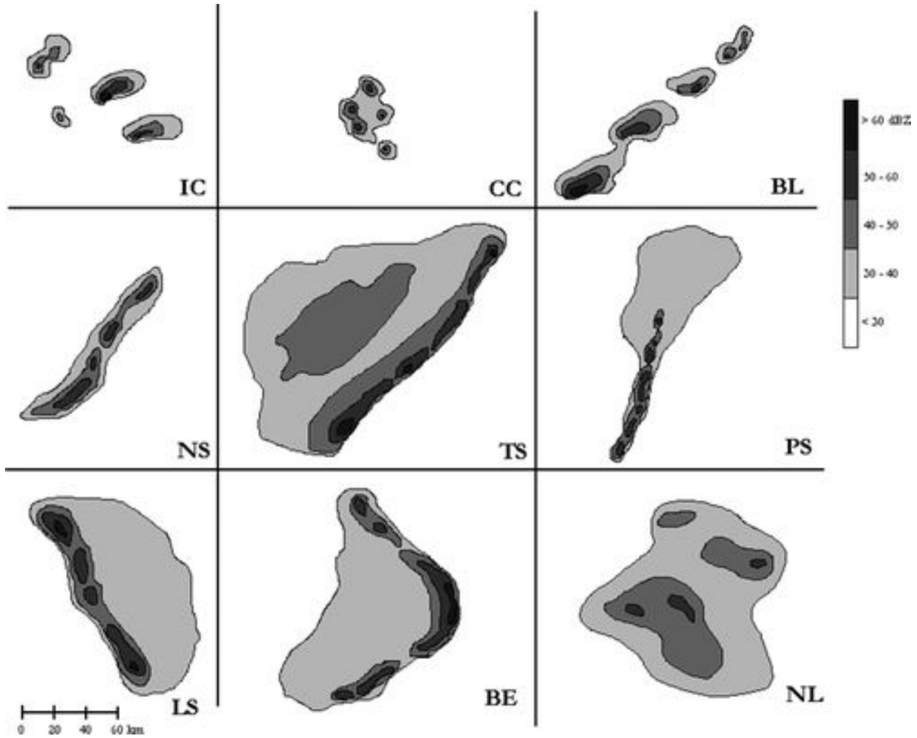


Figure from Gallus et al. (2008)

Machine Learning

Creating and using mathematical models built on training data to perform a task, rather than explicitly programming the task

Machine Learning

Creating and using mathematical models built on training data to perform a task, rather than explicitly programming the task

- Ensembles of decision trees
  - Tree gives classification from “branching” decisions on feature parameters
  - Combinations of many trees
  - Used in past studies with broader classifications (Gagne et al. 2009; Lack and Fox 2012; Haberlie and Ashley 2018a)

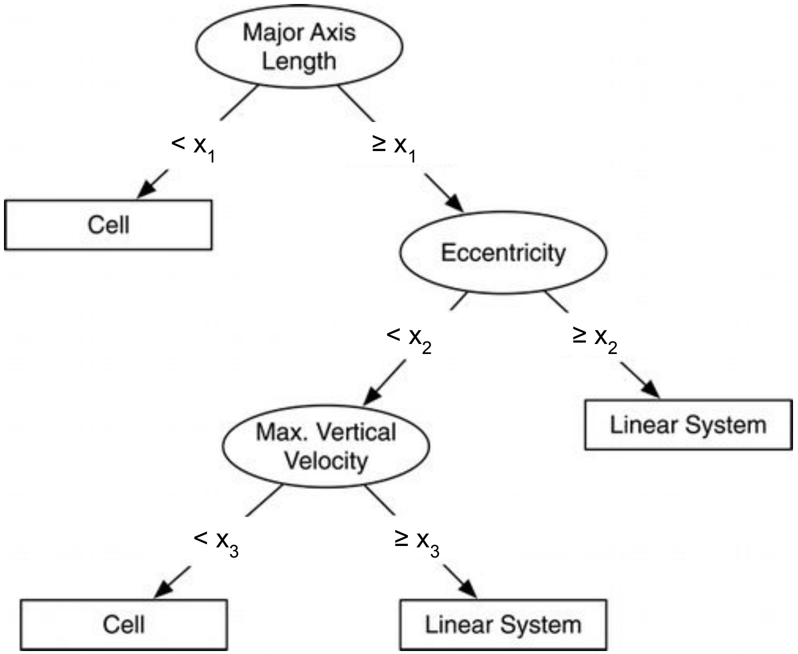
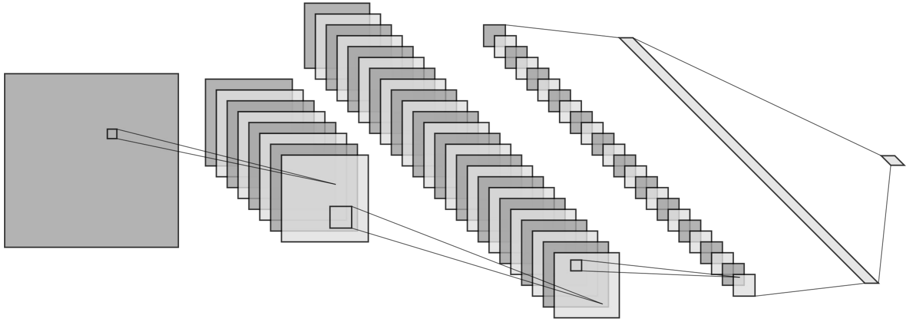


Figure from Gange et al. (2009), with modification

Machine Learning

Creating and using mathematical models built on training data to perform a task, rather than explicitly programming the task

- Convolutional Neural Networks (CNNs)
  - Patterned after the visual cortex
  - Train on images directly
  - Many possible architectures
  - Promising early results with storm classification (McGovern et al. 2017; Haberlie and Ashley 2018b)



Goals of This Study

- Develop methodologies that extend on past attempts
- Assess performance in classifying morphology
  - Ensembles of decision trees (parameter-based)
  - CNNs (image-based)
- Is either accurate enough to be relied on in future research?

Methodology: An Overview of the Process

Initial Source Data

Random Sample; Filter for MCS Occurrence

Extract Composite Reflectivity

Extract Slices (HA18a) and Corresponding Morphological Parameters

Label by Hand

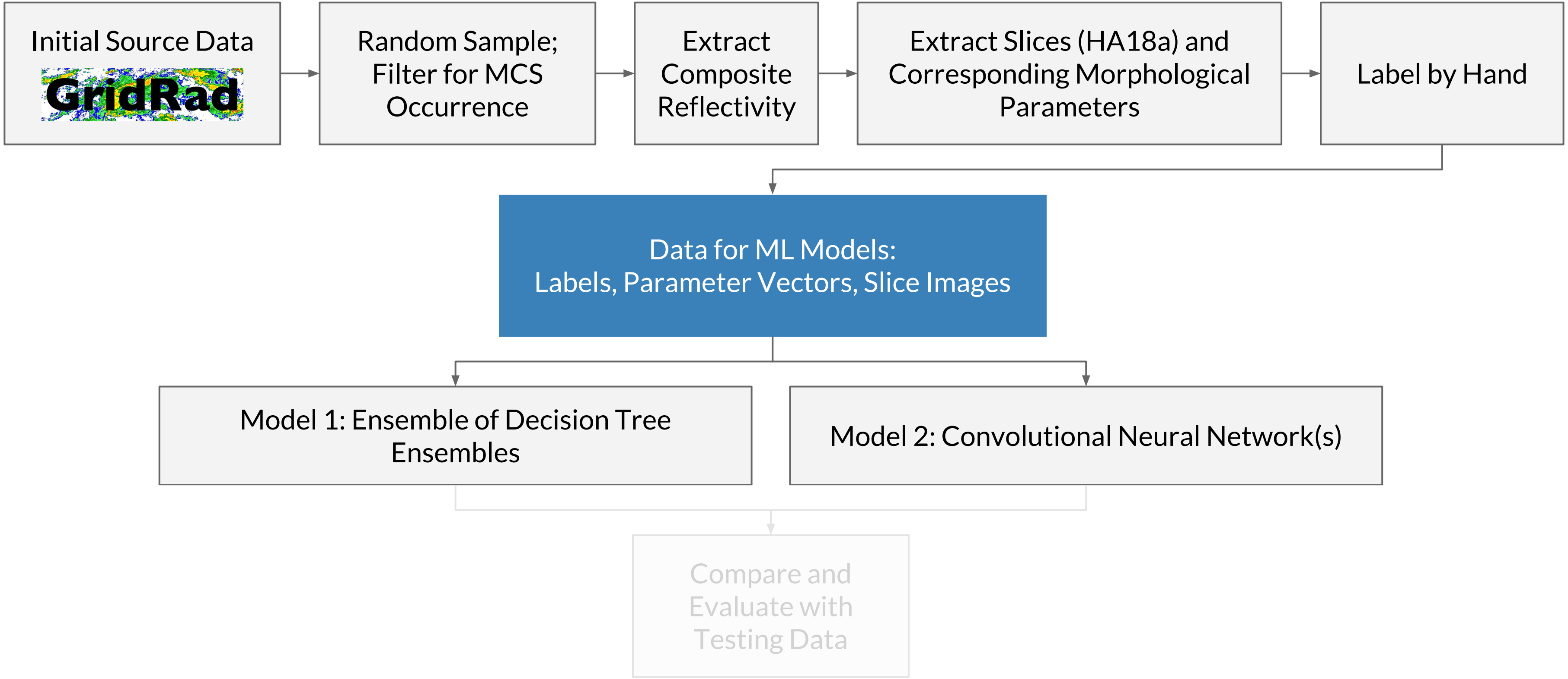
Data for ML Models:
Labels, Parameter Vectors, Slice Images

Model 1: Ensemble of Decision Tree Ensembles

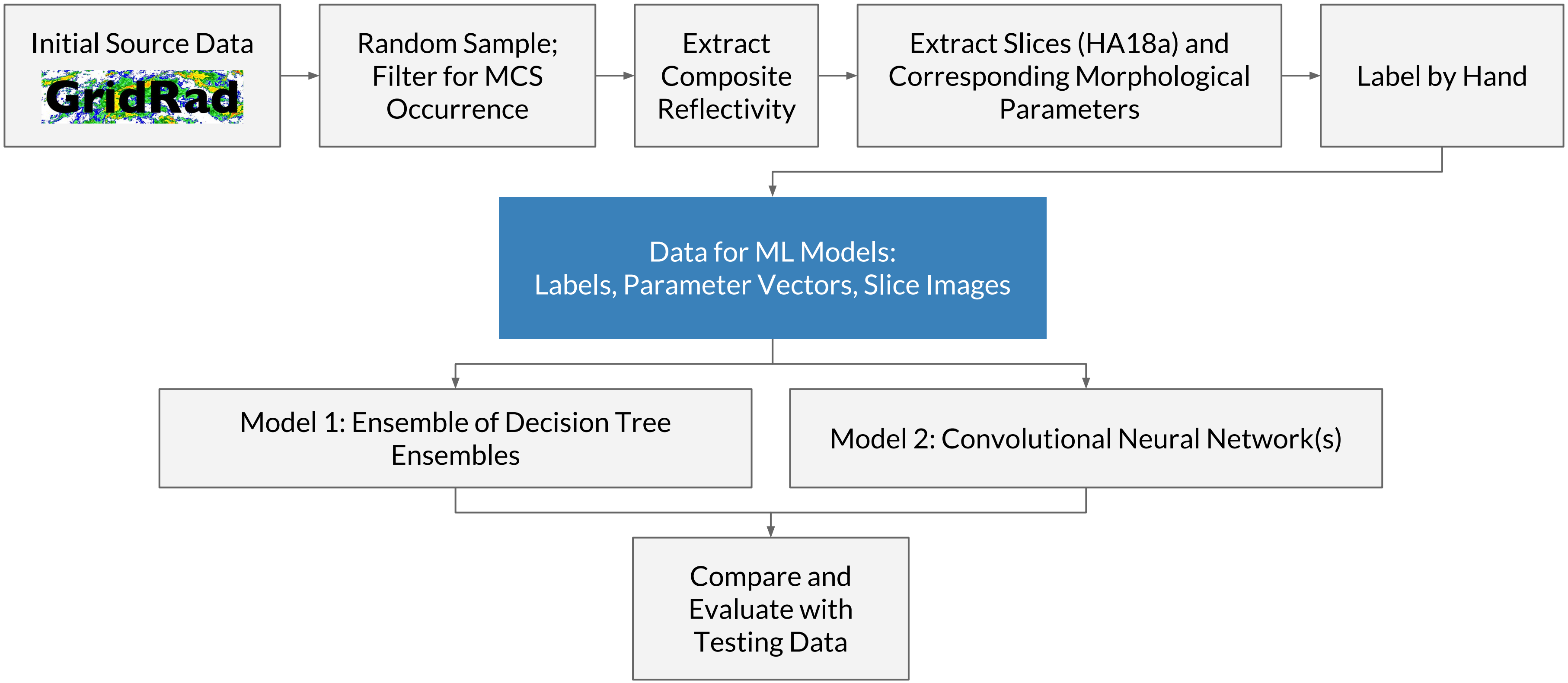
Model 2: Convolutional Neural Network(s)

Compare and Evaluate with Testing Data

Methodology: An Overview of the Process



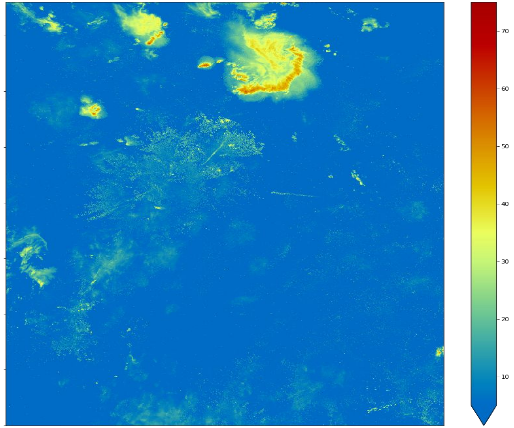
Methodology: An Overview of the Process



GridRad Data: Where it all starts

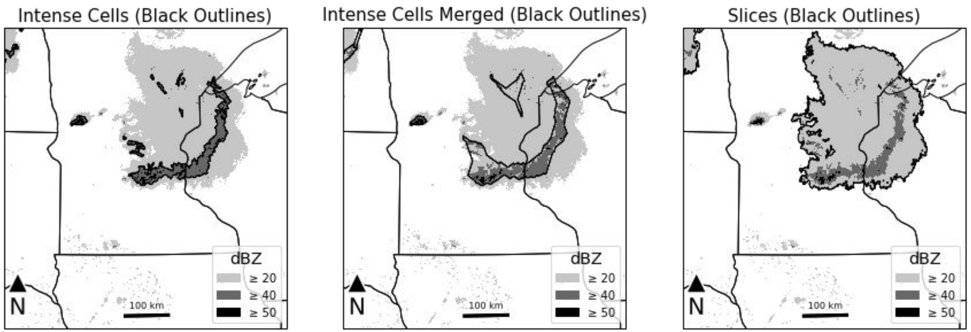
- 3D Gridded NEXRAD Data (Bowman and Homeyer 2017)
- Resolution: 0.02° x 0.02° horizontal, 1 km vertical, 1 hr temporal
- Available 2004-2016
- Subsetting on
    - Warm season (May-Sep)
    - Central U.S.
    - Convective system present
- Using composite (column-maximum) reflectivity regridded to 2 km Lambert Conformal grid

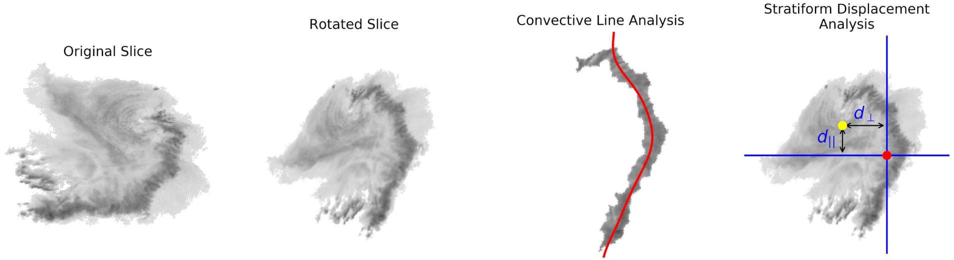
Study Domain



Slice Extraction

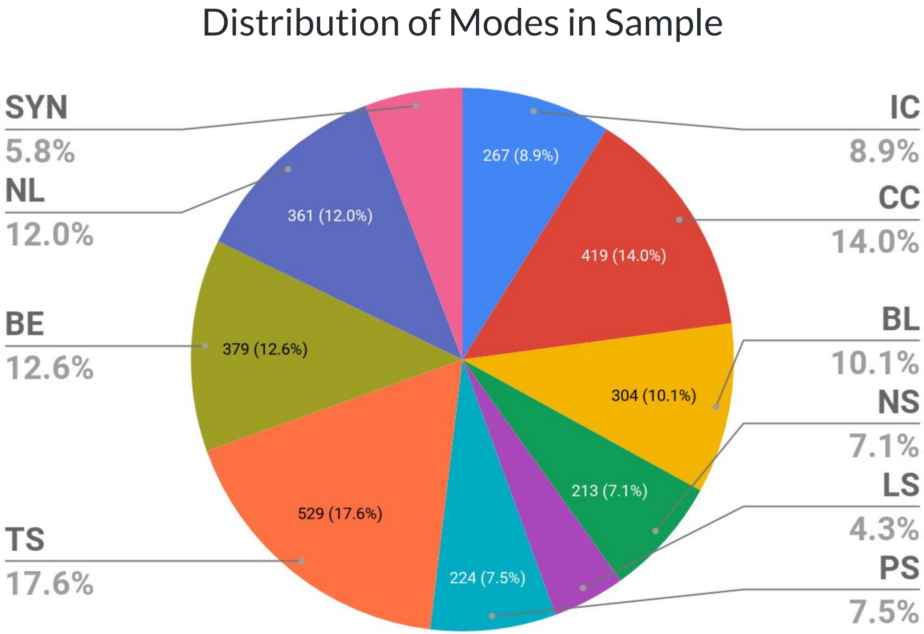
- Extract candidate system slices from mosaic (Haberlie and Ashley 2018a)
- Derive geometric and intensity parameters of system regions
    - Areas/lengths/widths
    - Intensity statistics
    - Convective line curvature
    - Stratiform displacement
    - Connectivity on triangular mesh of cell maxima
- Image analysis via scikit-image



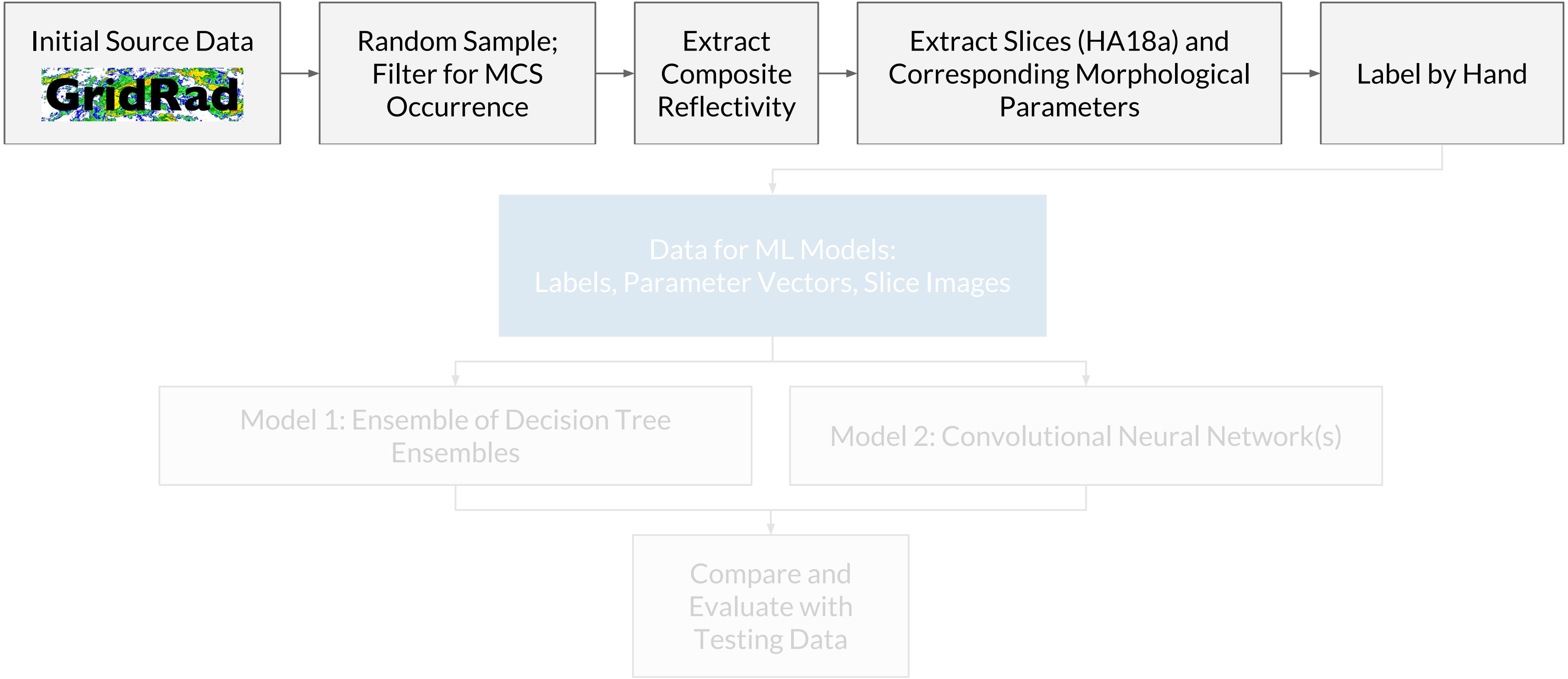


Summary of Manual Classification Results

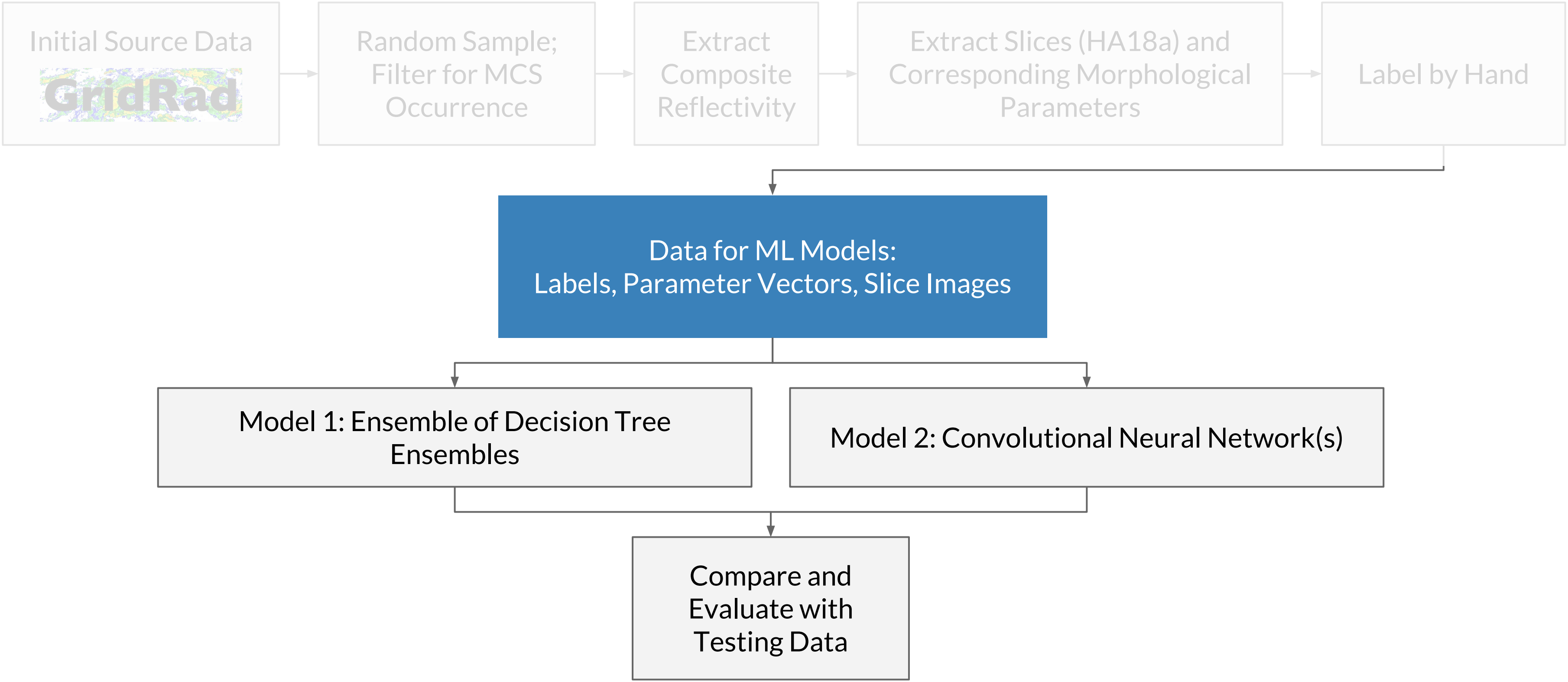
- 3000 extracted and hand-labeled systems
- Use nine modes of Gallus et al. (2008) and SYN type of Haberlie and Ashley (2018a)
  - Subjective classification
  - No temporal continuity
- Split into two subsets:
  - Training data (2007-2016): 2397
  - Testing data (2004-2006): 603



Process Overview



Process Overview



Classification Models

- Eight ensembles of decision trees (scikit-learn)
    - Four types
        - Random forest (RFC)
        - Gradient boosted trees (GBC)
        - XGBoost (XGBC)
        - Ensemble of these three (ENS)
    - Two feature sets
        - Original (Haberlie and Ashley 2018a)
        - All (with 24 added)

Classification Models

- Eight ensembles of decision trees (scikit-learn)
  - Four types
    - Random forest (RFC)
    - Gradient boosted trees (GBC)
    - XGBoost (XGBC)
    - Ensemble of these three (ENS)
  - Two feature sets
    - Original (Haberlie and Ashley 2018a)
    - All (with 24 added)
- Three CNNs (Keras with TensorFlow backend)
  - Fixed input image size
  - Centered on primary convective line
  - Differ on image resizing
    - Scale to 128 by 128 px (Scaled)
    - Crop to 128 by 128 px (Chopped)
    - Upscale to 4 km, then crop to 128 by 128 px (4km Chopped)

Accuracy Scores

Accuracy: proportion of classifications that are correct

Classifier	*Accuracy (%)*
RFC-Original	47.76
GBC-Original	43.62
XGBC-Original	46.93
ENS-Original	45.44
RFC-All	57.05
GBC-All	56.55
XGBC-All	56.05
ENS-All	**59.37**
CNN-Scaled	44.78
CNN-Chopped	30.85
CNN-4km-Chopped	38.14

Accuracy Scores

Accuracy: proportion of classifications that are correct

- Ensembles of decision trees using all feature parameters performed best

Classifier	*Accuracy (%)*
RFC-Original	47.76
GBC-Original	43.62
XGBC-Original	46.93
ENS-Original	45.44
RFC-All	57.05
GBC-All	56.55
XGBC-All	56.05
ENS-All	**59.37**
CNN-Scaled	44.78
CNN-Chopped	30.85
CNN-4km-Chopped	38.14

Accuracy Scores

Accuracy: proportion of classifications that are correct

- Ensembles of decision trees using all feature parameters performed best
- Additional parameters enhanced results

Classifier	*Accuracy (%)*
RFC-Original	47.76
GBC-Original	43.62
XGBC-Original	46.93
ENS-Original	45.44
RFC-All	57.05
GBC-All	56.55
XGBC-All	56.05
ENS-All	**59.37**
CNN-Scaled	44.78
CNN-Chopped	30.85
CNN-4km-Chopped	38.14

Accuracy Scores

Accuracy: proportion of classifications that are correct

- Ensembles of decision trees using all feature parameters performed best
- Additional parameters enhanced results
- CNNs generally performed worst

Classifier	*Accuracy (%)*
RFC-Original	47.76
GBC-Original	43.62
XGBC-Original	46.93
ENS-Original	45.44
RFC-All	57.05
GBC-All	56.55
XGBC-All	56.05
ENS-All	**59.37**
CNN-Scaled	44.78
CNN-Chopped	30.85
CNN-4km-Chopped	38.14

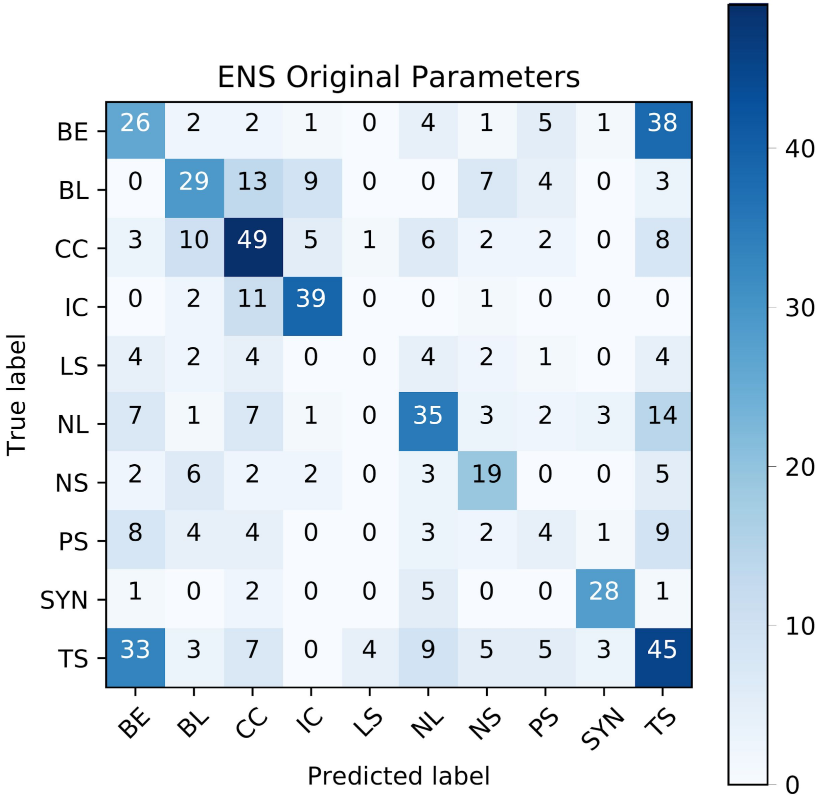
Accuracy Scores

Accuracy: proportion of classifications that are correct

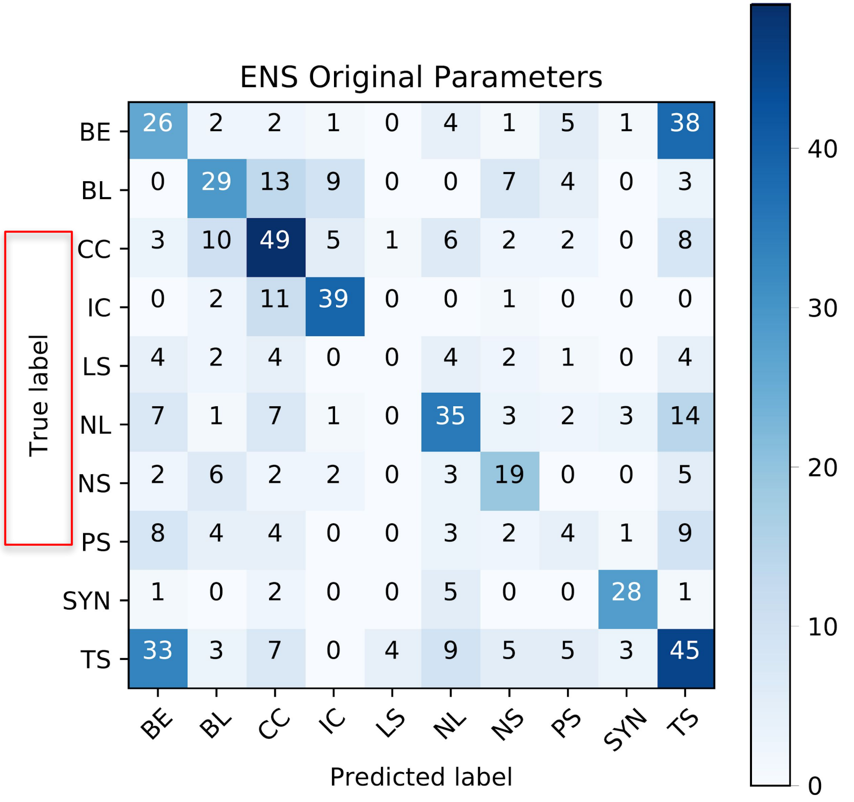
- Ensembles of decision trees using all feature parameters performed best
- Additional parameters enhanced results
- CNNs generally performed worst
- All results below threshold

Classifier	*Accuracy (%)*
RFC-Original	47.76
GBC-Original	43.62
XGBC-Original	46.93
ENS-Original	45.44
RFC-All	57.05
GBC-All	56.55
XGBC-All	56.05
ENS-All	**59.37**
CNN-Scaled	44.78
CNN-Chopped	30.85
CNN-4km-Chopped	38.14

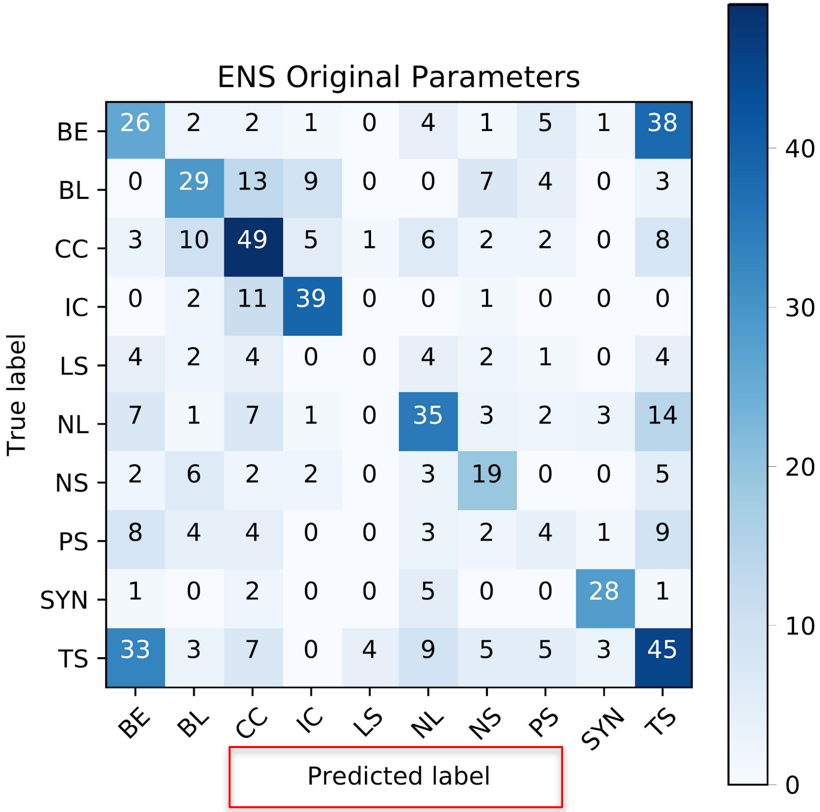
Confusion Matrices: ENS-Original



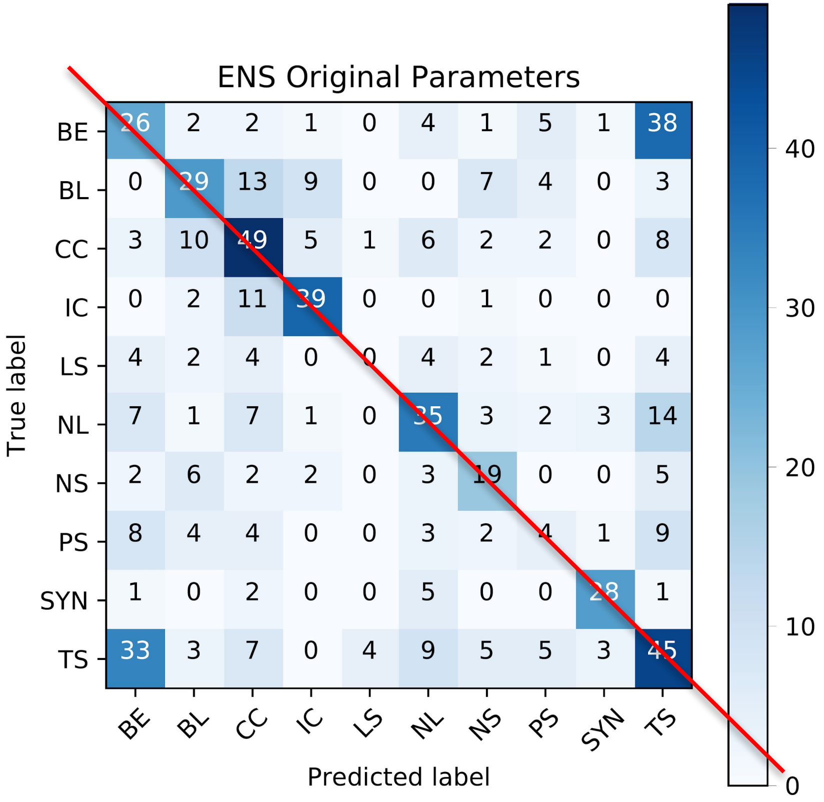
Confusion Matrices: ENS-Original



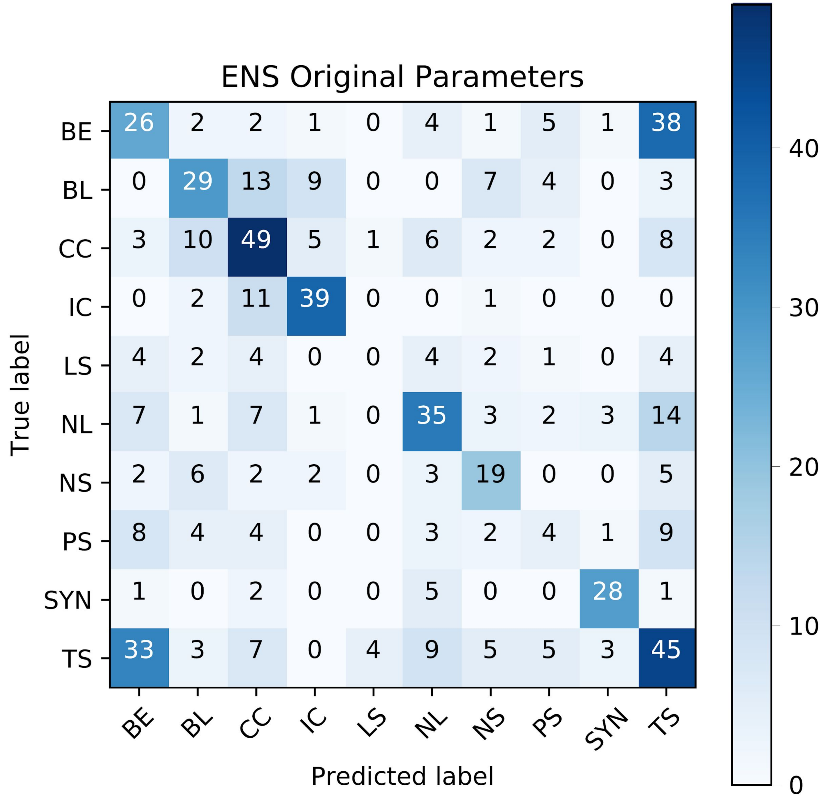
Confusion Matrices: ENS-Original



Confusion Matrices: ENS-Original

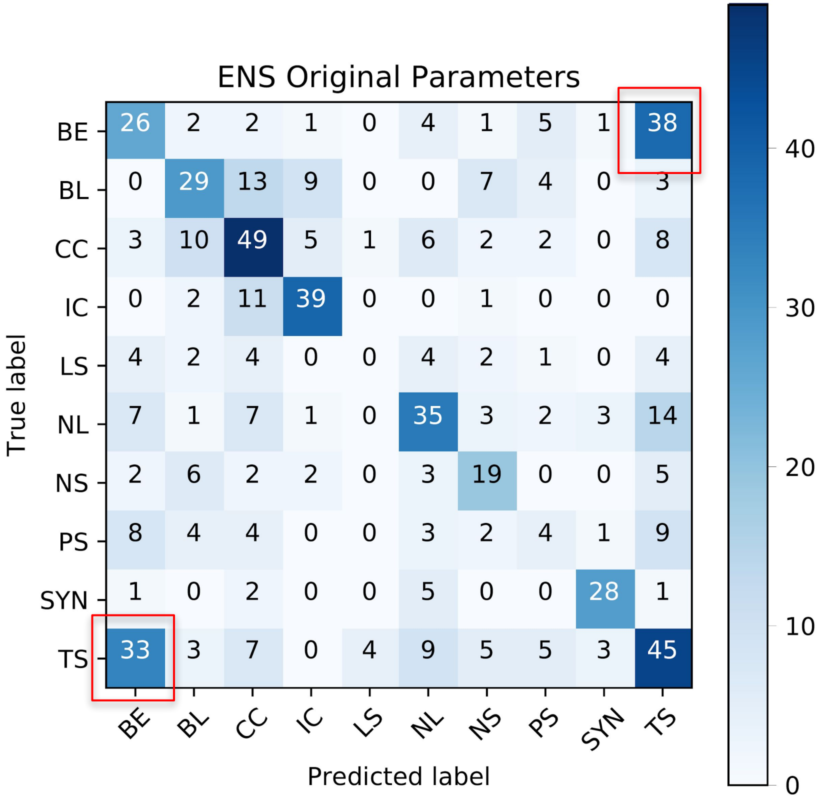


Confusion Matrices: ENS-Original



Confusion Matrices: ENS-Original

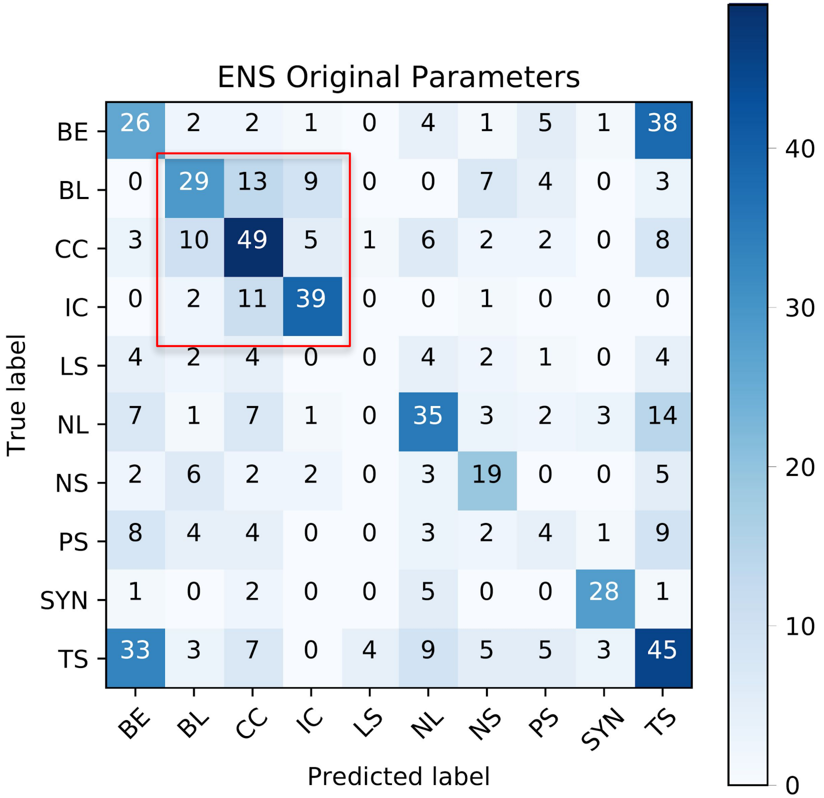
- BE/TS mismatches
- Cellular mode mismatches
- No LS-LS and few PS-PS
- Best recall for CC and SYN



	BE	BL	CC	IC	LS	NL	NS	PS	SYN	TS
BE	26	2	2	1	0	4	1	5	1	38
BL	0	29	13	9	0	0	7	4	0	3
CC	3	10	49	5	1	6	2	2	0	8
IC	0	2	11	39	0	0	1	0	0	0
LS	4	2	4	0	0	4	2	1	0	4
NL	7	1	7	1	0	35	3	2	3	14
NS	2	6	2	2	0	3	19	0	0	5
PS	8	4	4	0	0	3	2	4	1	9
SYN	1	0	2	0	0	5	0	0	28	1
TS	33	3	7	0	4	9	5	5	3	45

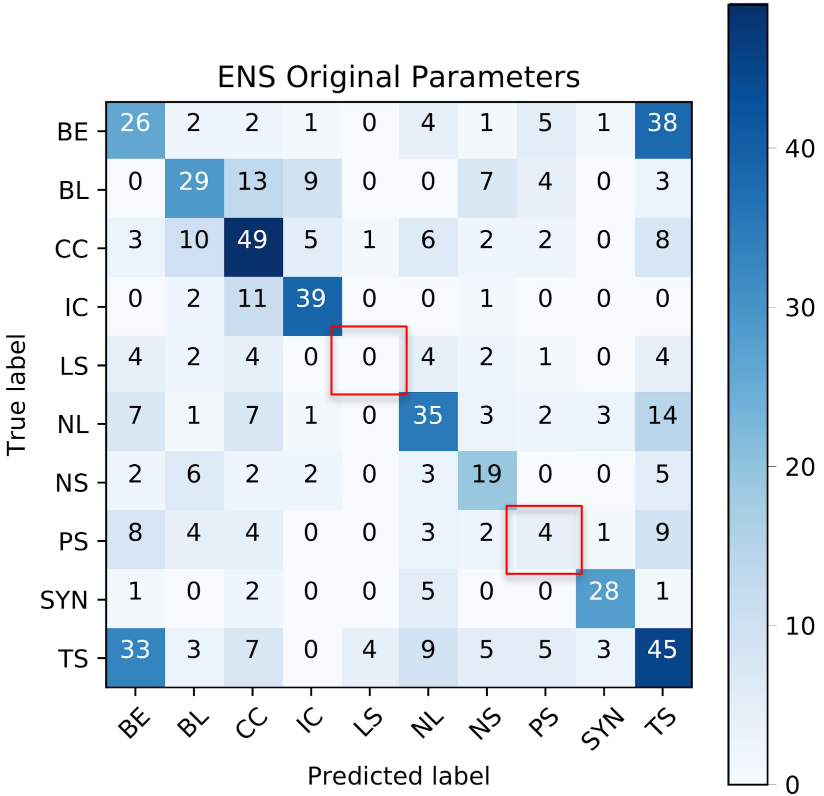
Confusion Matrices: ENS-Original

- BE/TS mismatches
- Cellular mode mismatches
- No LS-LS and few PS-PS
- Best recall for CC and SYN



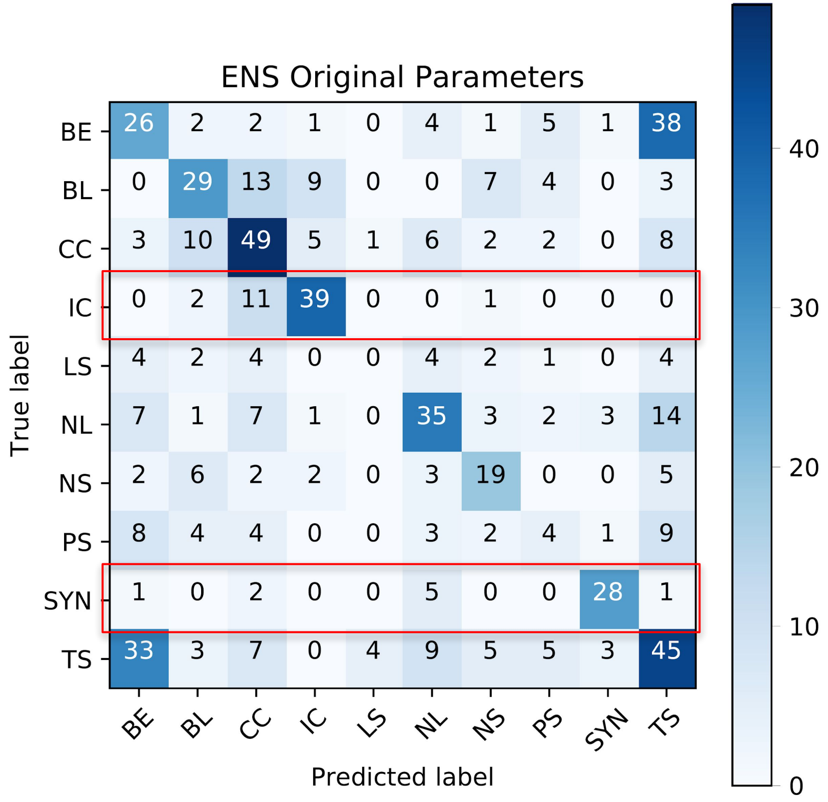
Confusion Matrices: ENS-Original

- BE/TS mismatches
- Cellular mode mismatches
- No LS-LS and few PS-PS
- Best recall for CC and SYN



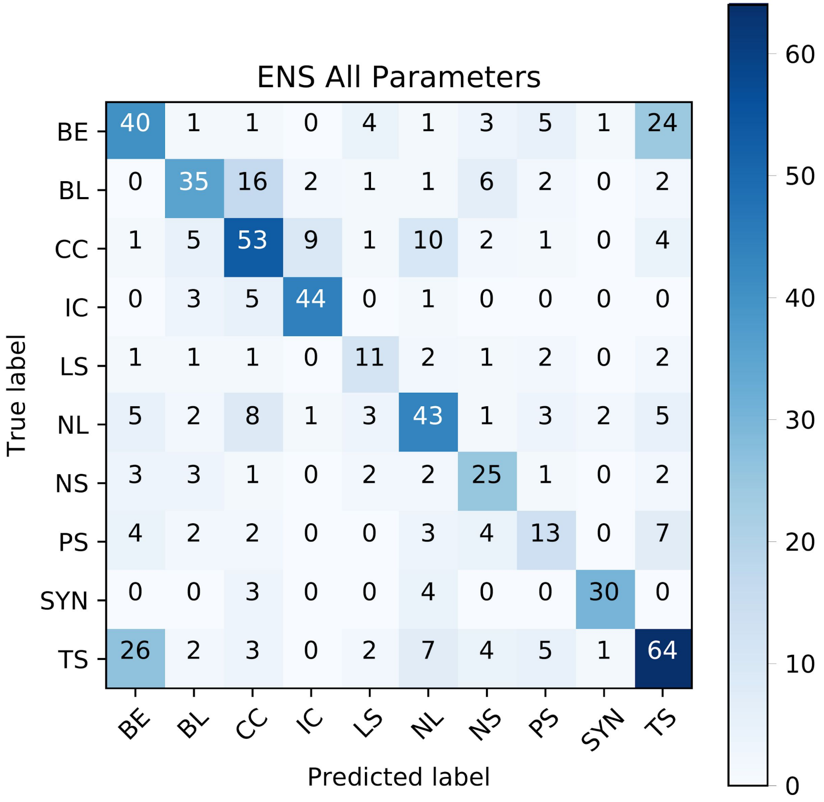
Confusion Matrices: ENS-Original

- BE/TS mismatches
- Cellular mode mismatches
- No LS-LS and few PS-PS
- Best recall for CC and SYN



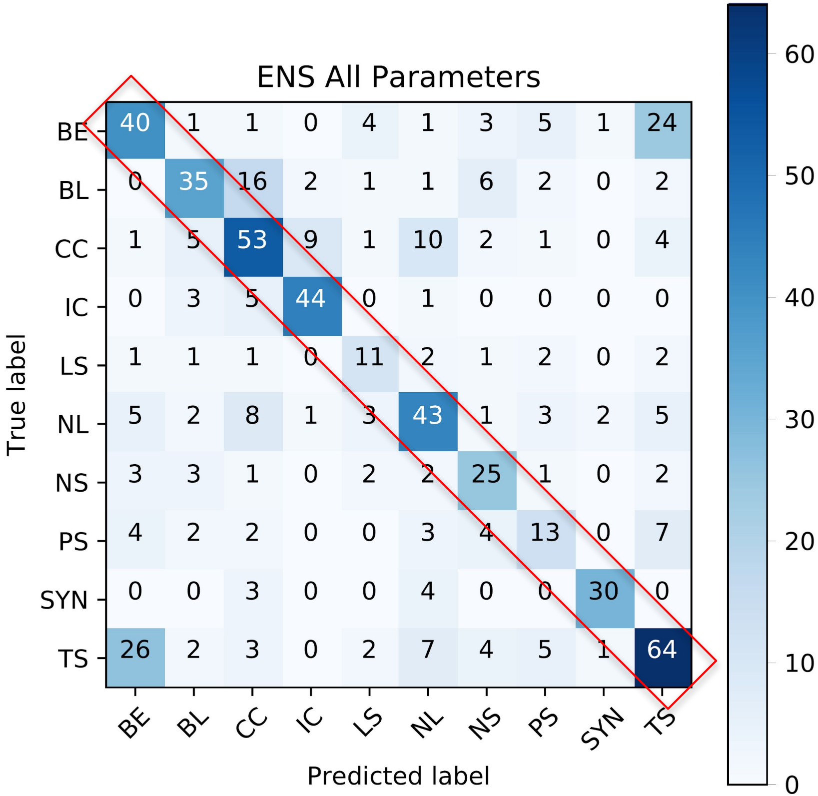
Confusion Matrices: ENS-All

- Improved results for all modes
- BE/TS issue remains (curvature parameters)
- Some Cellular and NL/CC issues remain (linearity or connectivity parameters)
- How many of these remaining mismatches are borderline cases?



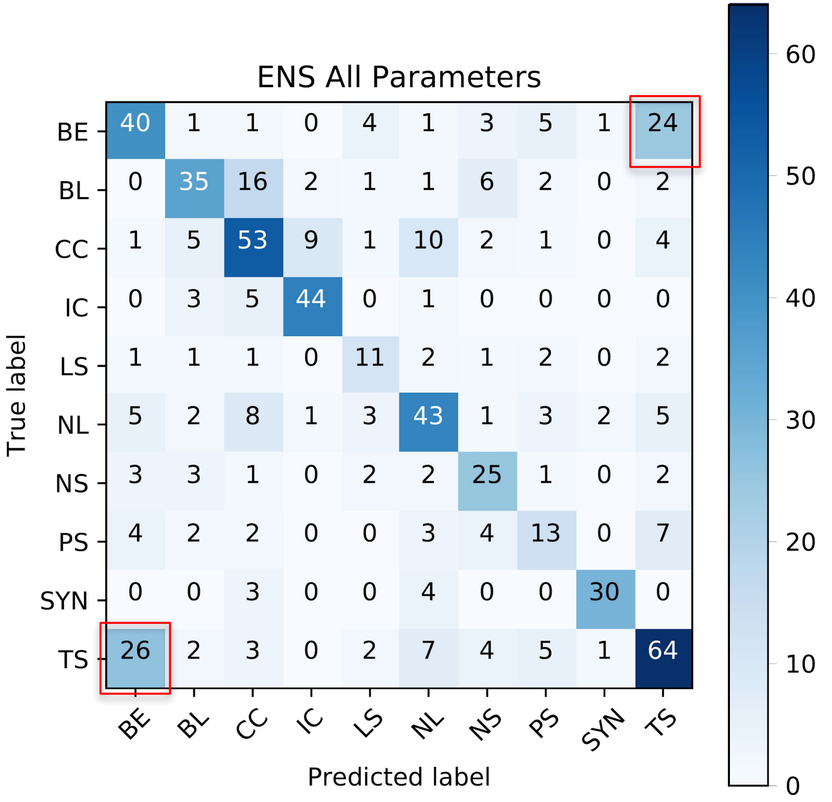
Confusion Matrices: ENS-All

- Improved results for all modes
- BE/TS issue remains (curvature parameters)
- Some Cellular and NL/CC issues remain (linearity or connectivity parameters)
- How many of these remaining mismatches are borderline cases?



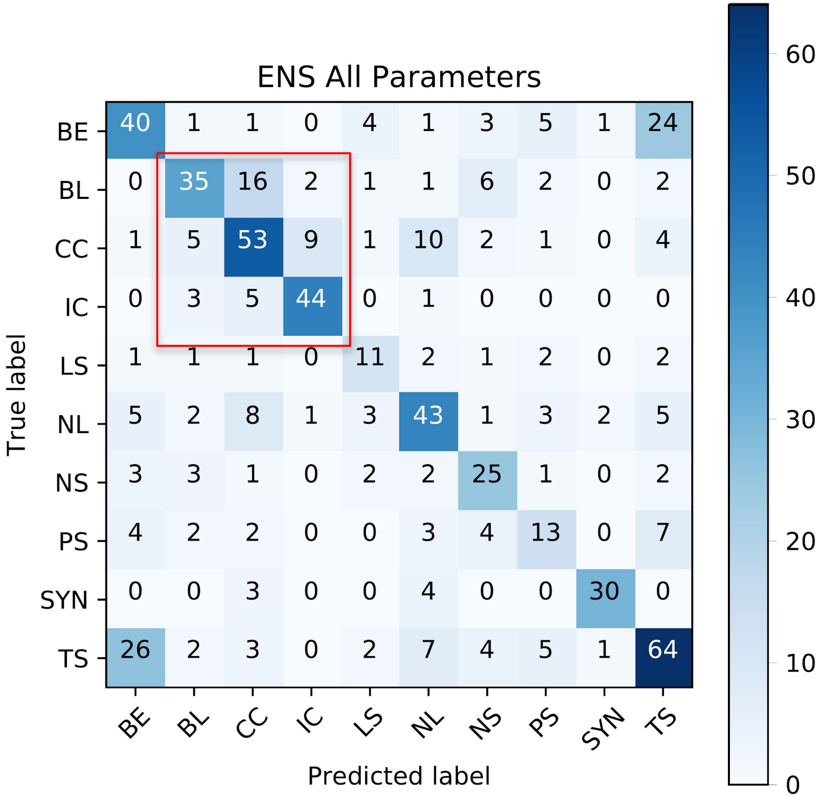
Confusion Matrices: ENS-All

- Improved results for all modes
- BE/TS issue remains (curvature parameters)
- Some Cellular and NL/CC issues remain (linearity or connectivity parameters)
- How many of these remaining mismatches are borderline cases?



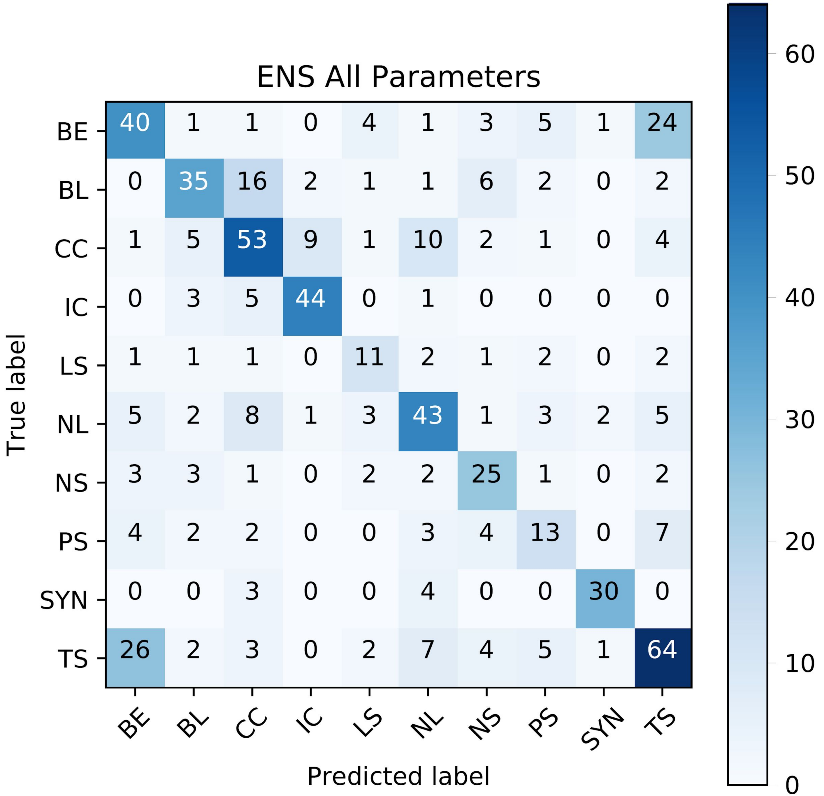
Confusion Matrices: ENS-All

- Improved results for all modes
- BE/TS issue remains (curvature parameters)
- Some Cellular and NL/CC issues remain (linearity or connectivity parameters)
- How many of these remaining mismatches are borderline cases?



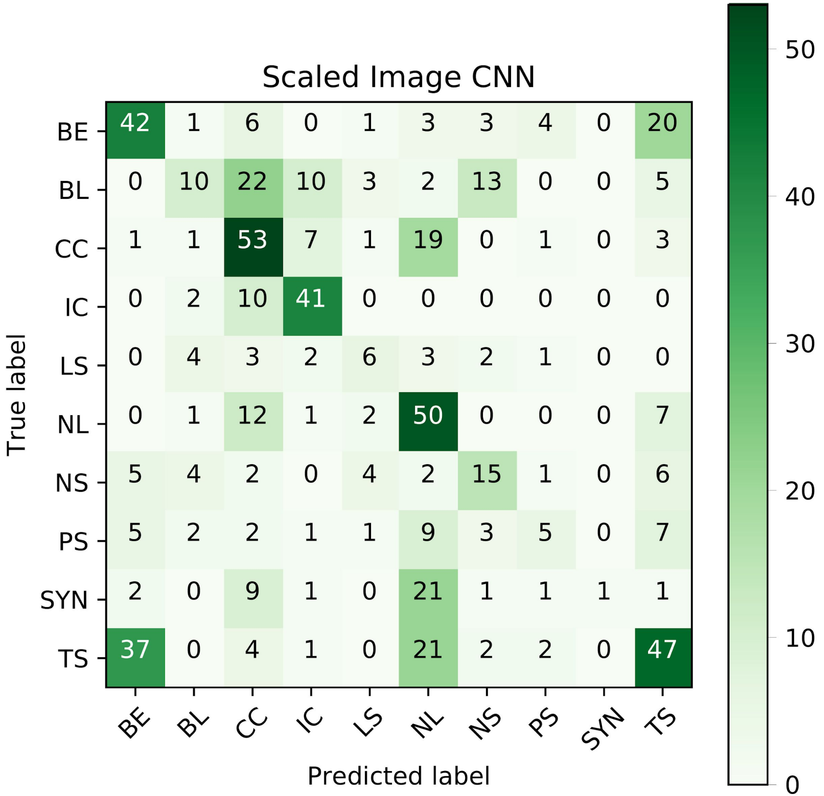
Confusion Matrices: ENS-All

- Improved results for all modes
- BE/TS issue remains (curvature parameters)
- Some Cellular and NL/CC issues remain (linearity or connectivity parameters)
- How many of these remaining mismatches are borderline cases?



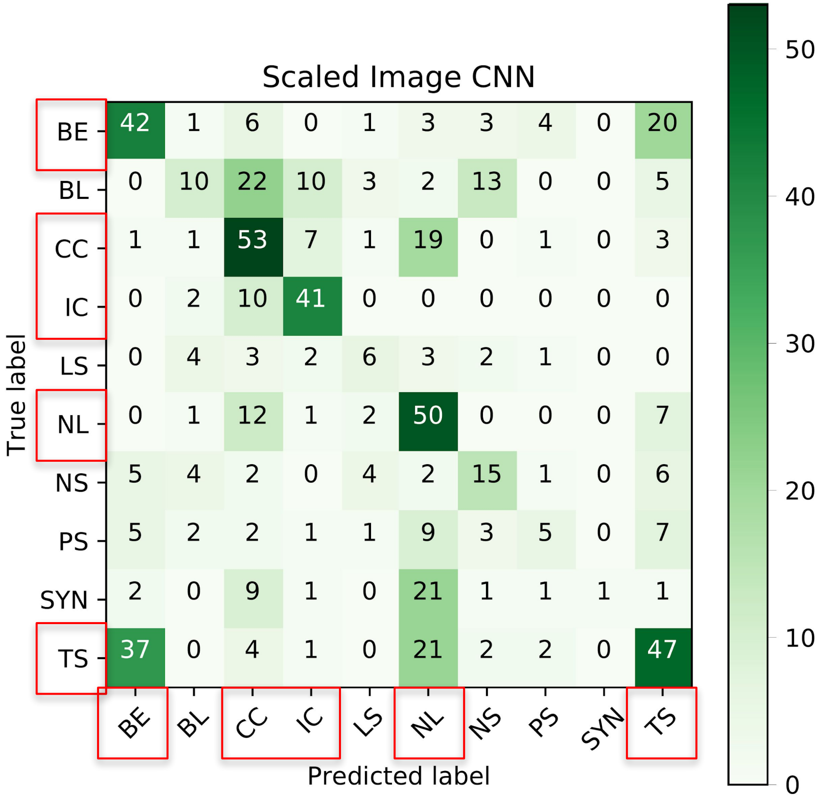
Confusion Matrices: CNN-Scaled

- With common modes (BE, CC, IC, NL, and TS):
  - High correct counts
  - Over-prediction
- With less common modes:
  - Very inconsistent results
- Potential causes:
  - Loss of scale-dependent features
  - Insufficient training data
  - Issues in network structure (only one used)



Confusion Matrices: CNN-Scaled

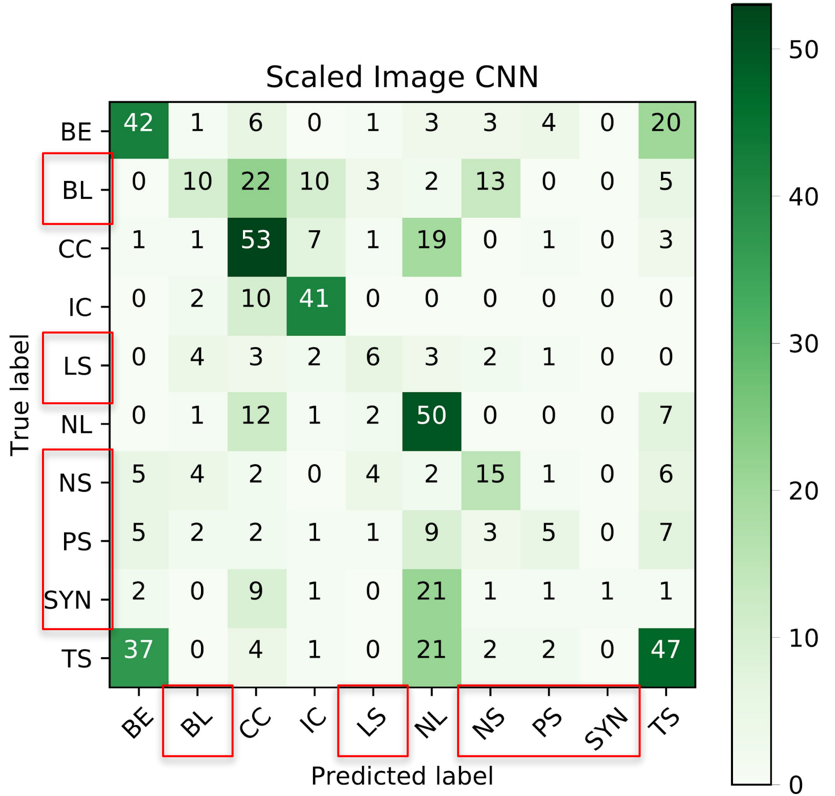
- With common modes (BE, CC, IC, NL, and TS):
  - High correct counts
  - Over-prediction
- With less common modes:
  - Very inconsistent results
- Potential causes:
  - Loss of scale-dependent features
  - Insufficient training data
  - Issues in network structure (only one used)



True label \ Predicted label	BE	BL	CC	IC	LS	NL	NS	PS	SYN	TS
BE	42	1	6	0	1	3	3	4	0	20
BL	0	10	22	10	3	2	13	0	0	5
CC	1	1	53	7	1	19	0	1	0	3
IC	0	2	10	41	0	0	0	0	0	0
LS	0	4	3	2	6	3	2	1	0	0
NL	0	1	12	1	2	50	0	0	0	7
NS	5	4	2	0	4	2	15	1	0	6
PS	5	2	2	1	1	9	3	5	0	7
SYN	2	0	9	1	0	21	1	1	1	1
TS	37	0	4	1	0	21	2	2	0	47

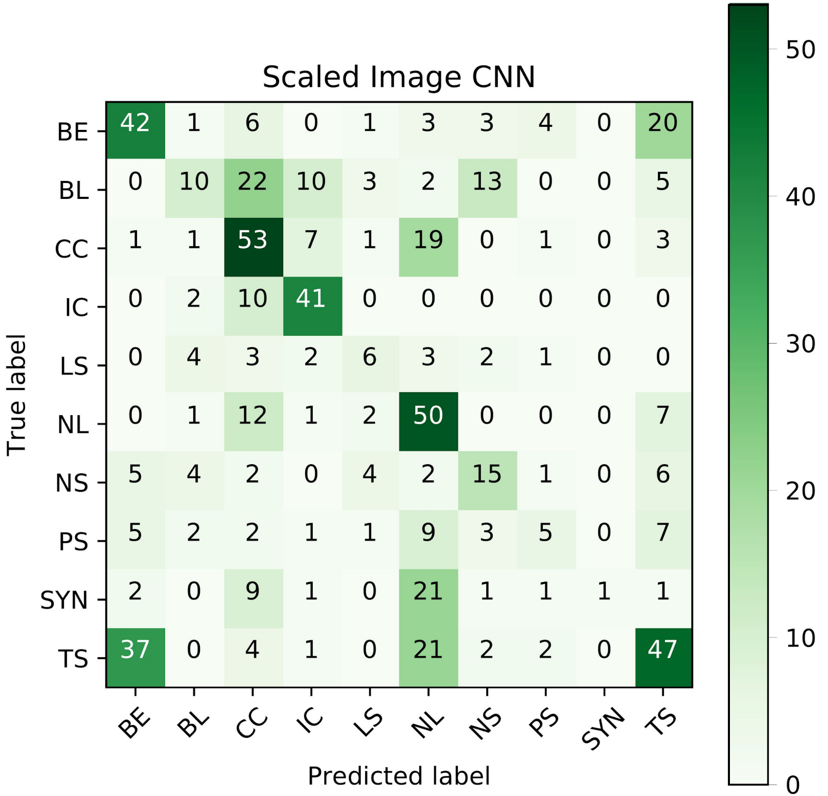
Confusion Matrices: CNN-Scaled

- With common modes (BE, CC, IC, NL, and TS):
  - High correct counts
  - Over-prediction
- With less common modes:
  - Very inconsistent results
- Potential causes:
  - Loss of scale-dependent features
  - Insufficient training data
  - Issues in network structure (only one used)



Confusion Matrices: CNN-Scaled

- With common modes (BE, CC, IC, NL, and TS):
  - High correct counts
  - Over-prediction
- With less common modes:
  - Very inconsistent results
- Potential causes:
  - Loss of scale-dependent features
  - Insufficient training data
  - Issues in network structure (only one used)



Conclusions

- Ensemble of ensembles of decision trees (with all parameters) performed best
  - Extra parameters improved results
  - Accuracy still insufficient for applications
- CNNs generally struggled
  - Lack of data (both dataset and input image sizes)
  - Lack of robustness/complexity (computational limits)
- Future work will
  - Analyze classifier mismatches in detail
  - Utilize larger samples
  - Investigate improvements in morphology feature extraction
  - Explore more robust/complex CNNs
  - Examine probabilistic classifications

Acknowledgements

- Mentors Dr. William Gallus and Dr. Alex Haberlie for guidance and input on this project
- Melissa Piper for assistance in filtering cases for occurrence of convective systems
- Machine learning sessions at the 2018 Unidata Users Workshop for inspiration and training
- Google Colaboratory for computational resources
- Supporting Python packages: NumPy, matplotlib, scipy, xarray, pandas, Cartopy, pyproj, xESMF, and MetPy

References

Baldwin, M. E., J. S. Kain, and S. Lakshmivarahan, 2005: Development of an Automated Classification Procedure for Rainfall Systems. Mon. Wea. Rev., 133, 844–862, https://doi.org/10.1175/MWR2892.1.

Biggerstaff, M. I., and S. A. Listemaa, 2000: An Improved Scheme for Convective/Stratiform Echo Classification Using Radar Reflectivity. J. Appl. Meteor., 39, 2129–2150, https://doi.org/10.1175/1520-0450(2001)040%3C2129:AISFCS%3E2.0.CO;2.

Bowman, K. P., and C. R. Homeyer, 2017: GridRad - Three-Dimensional Gridded NEXRAD WSR-88D Radar Data. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO. [Available online at https://doi.org/10.5065/D6NK3CR7.] Accessed 2 September 2018.

Chollet, F., 2015: Keras. [Available online at https://keras.io.] Accessed 2 September 2018.

Corfidi, S. F., M. C. Coniglio, A. E. Cohen, and C. M. Mead, 2016: A proposed revision to the definition of ‘‘derecho.’’ Bull. Amer. Meteor. Soc., 97, 935–949, https://doi.org/10.1175/BAMS-D-14-00254.1.

Chen, T., and C. Guestrin, 2016: XGBoost: A scalable tree boosting system. Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, San Francisco, CA, Association for Computing Machinery, 785–794, https://dl.acm.org/citation.cfm?id=2939785.

Fowle, M. A. and P. J. Roebber, 2003: Short-Range (0–48 h) Numerical Prediction of Convective Occurrence, Mode, and Location. Wea. Forecasting, 18, 782–794, https://doi.org/10.1175/1520-0434(2003)018<0782:SHNPOC>2.0.CO;2.

Fritsch, J. M., R. J. Kane, and C. R. Chelius, 1986: The Contribution of Mesoscale Convective Weather Systems to the Warm-Season Precipitation in the United States. J. Climate Appl. Meteor., 25, 1333–1345, https://doi.org/10.1175/1520-0450(1986)025<1333:TCOMCW>2.0.CO;2.

Gagne, D. J., A. McGovern, and J. Brotzge, 2009: Classification of Convective Areas Using Decision Trees. J. Atmos. Oceanic Technol., 26, 1341–1353, https://doi.org/10.1175/2008JTECHA1205.1.

Gallus, W. A., N. A. Snook, and E. V. Johnson, 2008: Spring and Summer Severe Weather Reports over the Midwest as a Function of Convective Mode: A Preliminary Study. Wea. Forecasting, 23, 101–113, https://doi.org/10.1175/2007WAF2006120.1.

Geerts, B., and Coauthors, 2017: The 2015 Plains Elevated Convection at Night Field Project. Bull. Amer. Meteor. Soc., 98, 767–786, https://doi.org/10.1175/BAMS-D-15-00257.1.

Haberlie, A. M. and W. S. Ashley, 2018: A Method for Identifying Midlatitude Mesoscale Convective Systems in Radar Mosaics. Part I: Segmentation and Classification. J. Appl. Meteor. Climatol., 57, 1575–1598, https://doi.org/10.1175/JAMC-D-17-0293.1.

Herman, G. R., and R. S. Schumacher, 2018: Money Doesn't Grow on Trees, but Forecasts Do: Forecasting Extreme Precipitation with Random Forests. Mon. Wea. Rev., 146, 1571-1600, https://doi.org/10.1175/MWR-D-17-0250.1.

Herman, G. R., and R. S. Schumacher, 2018: “Dendrology” in Numerical Weather Prediction: What Random Forests and Logistic Regression Tell Us about Forecasting Extreme Precipitation. Mon. Wea. Rev., 146, 1785-1812, https://doi.org/10.1175/MWR-D-17-0307.1.

Houze, R. A., Jr., 2004: Mesoscale convective systems. Rev. Geophys., 42, RG4003, https://doi.org/10.1029/2004RG000150.

Jirak, I.L. and W.R. Cotton, 2007: Observational Analysis of the Predictability of Mesoscale Convective Systems. Wea. Forecasting, 22, 813–838, https://doi.org/10.1175/WAF1012.1.

Kamani, M. M., F. Farhat, S. Wistar, and J. Z. Wang, 2017: Skeleton matching with applications in severe weather detection. Appl. Soft Comput., 70, 1154-1166, https://doi.org/10.1016/j.asoc.2017.05.037.

Lack, S. A., and N. I. Fox, 2012: Development of an automated approach for identifying convective storm type using reflectivity-derived and near-storm environment data. Atmos. Research, 116, 67–81, https://doi.org/10.1016/J.ATMOSRES.2012.02.009.

Lakshmanan, V. and T. Smith, 2009: Data Mining Storm Attributes from Spatial Grids. J. Atmos. Oceanic Technol., 26, 2353–2365, https://doi.org/10.1175/2009JTECHA1257.1.

McGovern, A., K. L. Elmore, D. J. Gagne, S.E. Haupt, C.D. Karstens, R. Lagerquist, T. Smith, and J.K. Williams, 2017: Using Artificial Intelligence to Improve Real-Time Decision-Making for High-Impact Weather. Bull. Amer. Meteor. Soc., 98, 2073–2090, https://doi.org/10.1175/BAMS-D-16-0123.1

Parker, M.D. and R.H. Johnson, 2000: Organizational Modes of Midlatitude Mesoscale Convective Systems. Mon. Wea. Rev., 128, 3413–3436, https://doi.org/10.1175/1520-0493(2001)129<3413:OMOMMC>2.0.CO;2

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine Learning in Python. J. Machine Learning Research, 12, 2825–2830.

Thielen, J. E., W. A. Gallus, and B. J. Squitieri, 2018: Microphysical and Horizontal Grid Spacing Influences on WRF Forecasts of Stratiform Rain Regions and General Convective Morphology Evolution in Nocturnal MCSs. 25th Conf. on Numer. Wea. Predict., Denver, CO, Amer. Meteor. Soc., 10B.6, https://ams.confex.com/ams/29WAF25NWP/webprogram/Paper344877.html.

van der Walt, S., J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, 2014: scikit-image: image processing in Python. PeerJ, 2:e453, https://doi.org/10.7717/peerj.453.

Questions?

Appendix

Lists of Feature Parameters

- Total Area
- Intense Area
- Convective Area
- Intense-Total Area Ratio
- Convective-Total Area Ratio
- Intense-Convective Area Ratio
- Convex Area
- Solidity
- Major Axis Length
- Minor Axis Length
- Eccentricity
- Mean Intensity
- Max Intensity
- Intensity Variance

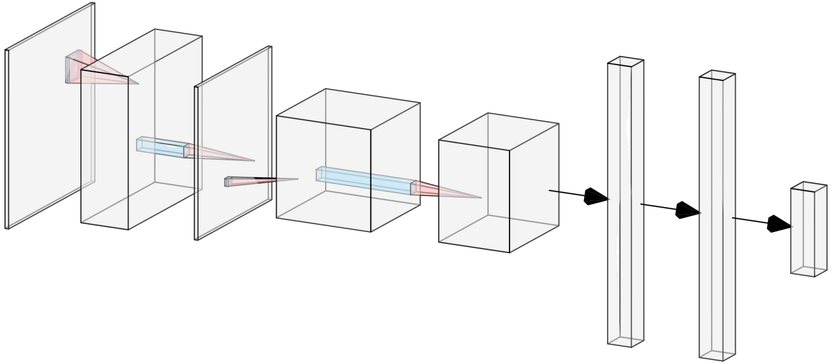
- Convective Solidity
- Normal Stratiform Displacement
- Parallel Stratiform Displacement
- Normalized Cell Count
- Mean Characteristic Curvature
- Max Characteristic Curvature
- Max Mean Characteristic Curvature
- Convective Length
- Convective Width
- Convective Length-Width Ratio
- Stratiform Width
- System-Convective Length Ratio
- Stratiform-Convective Width Ratio
- Delaunay Edges
- Edge Proportion with Minimum at
    - None
    - Stratiform
    - Convective
- Edge Proportion with Average at
    - None
    - Stratiform
    - Convective
    - Intense
- Edge Mean Length
- Cell Centroid Spread
- Cell Centroid R^2

Distribution of Cases

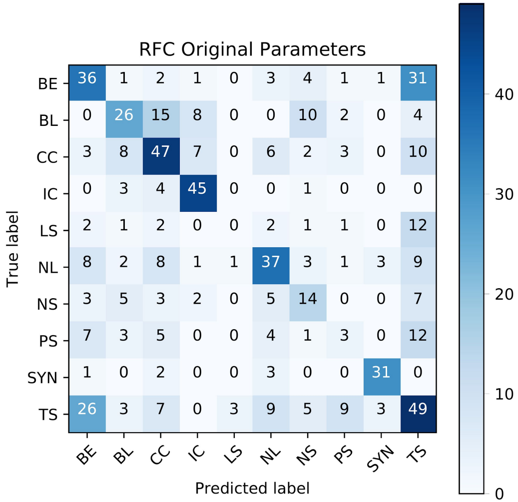
Mode	*Testing (2004-2006)*	*Training (2007-2016)*	*Total*	*Percentage of Total*
IC	53	214	267	8.90
CC	86	333	419	13.97
BL	65	239	304	10.13
NS	39	174	213	7.10
LS	21	109	130	4.33
PS	35	189	224	7.47
TS	114	415	529	17.63
BE	80	299	379	12.63
NL	73	288	361	12.03
SYN	37	137	174	5.80
Total	603	2397	3000	

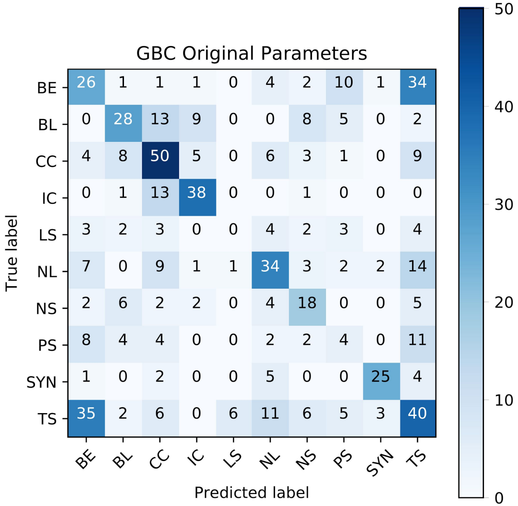
CNN Structure

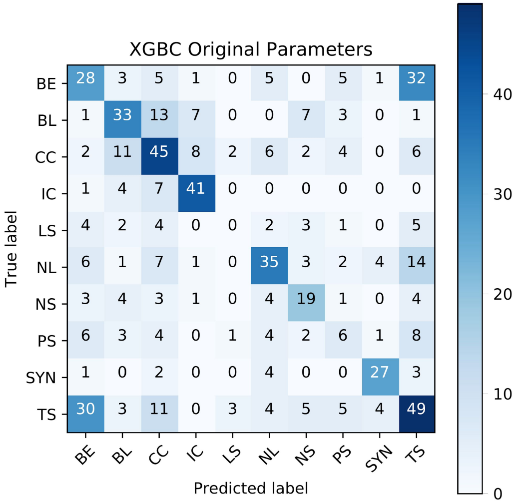
Layer Type	*# Features*	*Filter Size*	*Stride*	*Activation*	*Dropout*
Convolutional	64	7 x 7	--	ReLu	--
Convolutional	64	3 x 3	--	ReLu	--
Max Pooling	--	2 x 2	2 x 2	--	--
Convolutional	128	3 x 3	--	ReLu	--
Convolutional	128	3 x 3	--	ReLu	--
Max Pooling	--	2 x 2	2 x 2	--	--
Convolutional	256	3 x 3	--	ReLu	--
Convolutional	256	3 x 3	--	ReLu	--
Max Pooling	--	2 x 2	2 x 2	--	--
Convolutional	512	3 x 3	--	ReLu	--
Convolutional	512	3 x 3	--	ReLu	--
Max Pooling	--	2 x 2	2 x 2	--	--
Flatten	--	--	--	--	--
Dense	4096	--	--	ReLu	Dropout (0.3)
Dense	4096	--	--	ReLu	Dropout (0.3)
Dense	10	--	--	Softmax	--



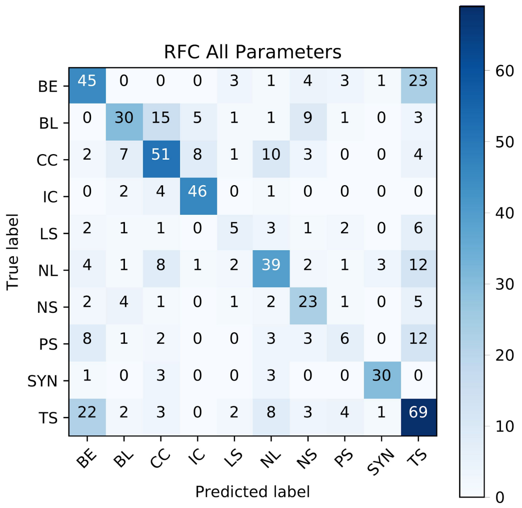
Supplemental Results: Confusion Matrices

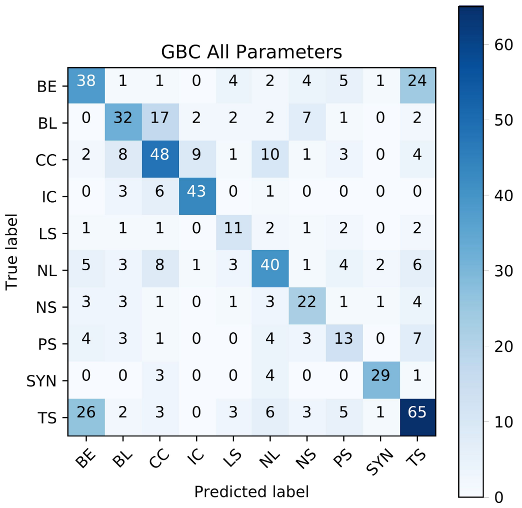


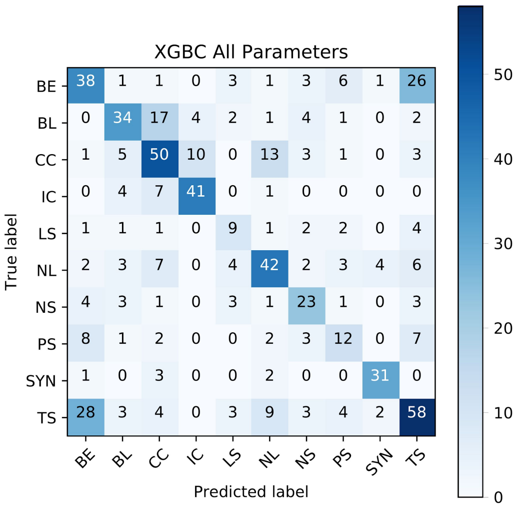




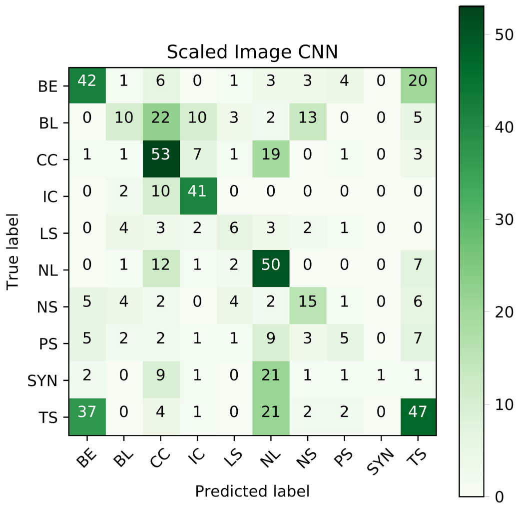
Supplemental Results: Confusion Matrices

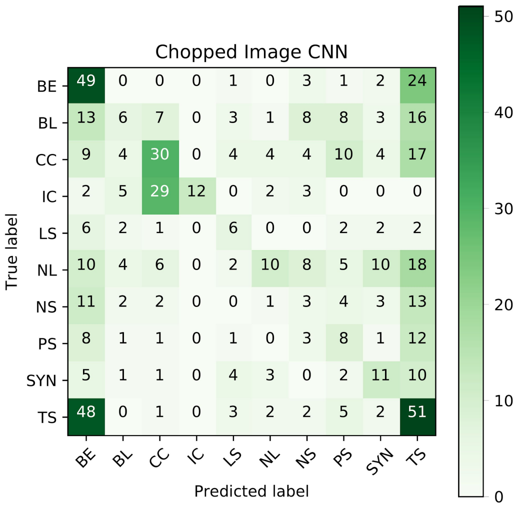


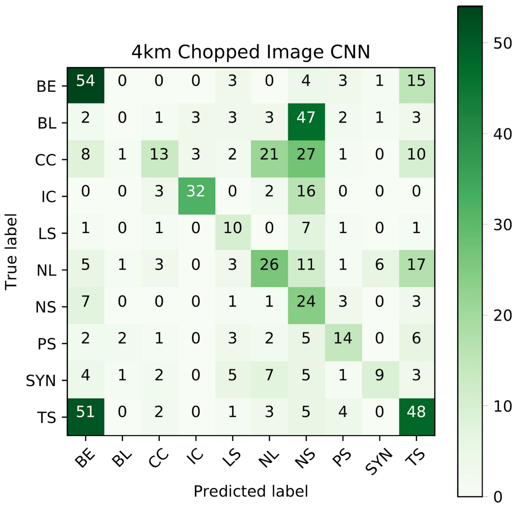




Supplemental Results: Confusion Matrices







Examples of Highest Confidence Results