A Deep Learning Model for Nowcasting Midlatitude Convective Storms

Stephanie M. Bradshaw

A Thesis submitted in partial fulfillment of

the requirements for the degree of

Master of Science

(Atmospheric and Oceanic Sciences)

at the

UNIVERSITY OF WISCONSIN-MADISON

December 2021

Thesis Declaration and Approval

I, Stephanie M. Bradshaw, declare that this Thesis titled 'A Deep Learning Model for Nowcasting Midlatitude Convective Storms' and the work presented in it are my own.

Stephanie M. Bradshaw		
Author	Signature	Date

I hereby approve and recommend for acceptance this work in partial fulfillment of the requirements for the degree of Master of Science:

Michael Pavolonis		
Committee Chair	Signature	Date

Tristan L'Ecuyer		
Faculty Member	Signature	Date

Grant Petty		
Faculty Member	Signature	Date

Abstract

A Deep Learning Model for Nowcasting Midlatitude Convective Storms

by Stephanie M. Bradshaw

Thunderstorms generate a variety of hazards, including lightning, flash flooding, damaging winds, and large hail. Short-term forecasts (0-60 minutes) of thunderstorm formation and intensity are critical for protecting life and property. In the 0-60 minute "nowcasting" timeframe, operational forecasters rely heavily on observations to identify developing convective hazards. While human expert analysis of environmental data is valuable, automated tools are needed to take better advantage of the vast quantities of environmental data now available. With the primary goal of nowcasting thunderstorm formation, a deep learning model with a U-Net convolutional neural network architecture was developed. The model utilizes geostationary satellite data from the next generation Geostationary Operational Environmental Satellite (GOES-R) Advanced Baseline Imager (ABI) to identify active convection and predict where convection will initiate in the next 60 minutes, as defined by radar. The newly developed deep learning model, which is part of the National Oceanic and Atmospheric Administration (NOAA) Probability of Severe (ProbSevere) nowcasting model suite, provides a well calibrated probabilistic solution for nowcasting convective storm formation. While the model was trained and tested for a limited spatial domain in the southern United States, it has the potential to evolve into a tool that enhances operational nowcasts.

Dedicated to my parents, Todd M. and Sheri L. Bradshaw.

Acknowledgements

This work would not have been possible without the support of my colleagues, friends, and family. I am especially grateful for the guidance of my research advisor, Mike Pavolonis, and colleague, John Cintineo, both of whom provided the framework for this research and have been involved with countless discussions regarding model development and interpretation. Additionally, thank you to the entire University of Wisconsin-Madison (UW-Madison) Atmospheric and Oceanic Science faculty for guiding me in the pursuit of knowledge in this field, especially Tristan L'Ecuyer and Grant Petty.

I would also like to acknowledge this work would not be possible without the funding provided by the National Oceanic and Atmospheric Administration (NOAA). The University of Wisconsin-Madison (UW-Madison) also supported this research and my pursuit of the degree of Master of Science. I would like to thank both organizations for their contributions.

Lastly, I could not have done this without the never-ending trove of support my partner, Alex Ortland, provides for me on a daily basis. He has been there to support me through the entirety of my pursuit of this degree, and I am grateful.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Buoyant vertical convection, thermally driven mass motion and the associated transfer of energy in the atmosphere is a fundamental driver of convective storms (thunderstorms). Once developed, these storms have many hazards associated with them including hail, high winds, lightning, and flash floods, all of which can cause infrastructure damage and, in extreme cases, loss of life. National Weather Service (NWS) forecasters issue advisories, watches, and warnings to mitigate the costs of convective storms on society by providing time for citizens, first responders, and utility workers to prepare.

Forecasters often utilize models to synthesize vast amounts of data into actionable insights. Thus, a variety of models have been developed for prediction and tracking of convective storms, hazard analysis, and prediction. Extrapolation based models with an emphasis on radar observations include the Auto-Nowcaster (Mueller et al., 2003,

1993, Roberts et al., 2012), TITAN (Thunderstorm Identification, Tracking, Analysis, and Nowcasting; (Dixon and Wiener, 1993)), CARDS (Canadian Radar Detection System; (Joe et al., 2002)), WDSS-II (Warning Decision Support System-Integrated Information; (Lakshmanan et al., 2007)), THESPA (Thunderstorm Strike Probability Nowcasting Algorithm; (Dance et al., 2010)), and CASA (Collaborative Adaptive Sensing of the Atmosphere; (Ruzanski et al., 2011)) Distributed Collaborative Adaptive Sensing Network for storm evolution and tracking. Numerical weather prediction (NWP) models have also been included in forecasting models such as GRAPES-SWIFT (Global/Regional Assimilation and Prediction System-Severe Weather Forecast tool; (Feng et al., 2007, Hu et al., 2007)). and CIWS (Corridor Integrated Weather System; (Wolfson and Clark, 2006)).

Besides radar and NWP models, geostationary satellite data have been instrumental in improving analysis and prediction capabilities for convective weather. Many studies have included the use of geostationary data in convective weather applications including Cintineo et al. (2020a, 2018), Harris et al. (2010), Lensky and Rosenfeld (2006), Mecikalski and Bedka (2006), Mecikalski et al. (2010a,b), Merk and Zinner (2013), Nisi et al. (2014), Roberts and Rutledge (2003), Rosenfeld et al. (2008), Sieglaff et al. (2011), Walker et al. (2012). All of these studies demonstrate the value of geostationary satellite data for detecting and predicting convective storms and the associated hazards.

An additional method used in model development is machine learning. Machine learning modeling methods are computerized methods for automatically finding transformations which turn data into more-useful representations for a given task (Chollet, 2018). The

work detailed in Mecikalski et al. (2015) and Cintineo et al. (2020a, 2014, 2018) are examples of machine learning applications in atmospheric sciences. Cintineo et al. (2020a, 2014, 2018) details the development of ProbSevere, which uses a naïve Bayesian classifier machine learning method to provide probabilistic forecasts of the severity of storms. ProbSevere also includes the probability of hail, the probability of high winds, and the probability of tornadoes (Cintineo et al., 2020a, 2014, 2018). Mecikalski et al. (2015) used probabilistic forecasts of thunderstorm convective initiation generated with logistic regression and random forest machine learning methods to produce fewer false alarm rates than the Geostationary Operational Environmental Satellite (GOES)-R convective initiation algorithm over the continental United States and nearshore locations by the Gulf of Mexico. However, the areas under the receiver operating characteristic curves recorded in Mecikalski et al. (2015) (with values between 0.69 and 0.83) indicate further improvements may be possible.

Recently, deep learning methods have been gaining traction in atmospheric sciences and may offer a method for improving convective initiation models. Deep learning is a subfield of machine learning because it involves searching for useful representations of input data in a specified space of possibilities using a feedback signal (Chollet, 2018). As opposed to other machine learning techniques which use only a couple of data representations, deep learning emphasizes the use of successive layers (sometimes tens to hundreds of layers) of increasingly meaningful data representations often arranged in neural networks (Chollet, 2018). Deep learning has been successful in the detection of synoptic scale fronts (Lagerquist et al., 2019), intense midlatitude convection (Cintineo et al., 2020b),

lightning (Zhou et al., 2020), and tornadoes (Lagerquist et al., 2020). Also, Lagerquist et al. (2021) demonstrates the capability of deep learning models to generate probabilistic predictions of active convection through their application of U-Net deep learning models to forecast convection with up to 120 minutes lead times using the Himawari-8 satellite and an echo-classification algorithm applied to weather radars in Taiwan. However, deep learning methodology has yet to be applied to the United States for short term (0-60 minutes) "nowcasting" of convective activity.

This study utilizes deep learning to identify active convection and predict which cumulus fields are most likely to initiate in the next 60 minutes in the United States. This deep learning model for nowcasting convective storms provides a well calibrated probabilistic solution to aid forecasters in their decision making processes and in doing so lends insight into the fundamental physical processes in convective storms.

Chapter 2

Data and Methodology

2.1 Representation of Convective Initiation

To predict convective initiation and convection in storms, it is first necessary to discuss how convective initiation has been represented from a historical standpoint as well as how it will be defined for the model. Convection itself can refer to any vertically buoyant mass motion and its associated transfer of energy. However, since forecasters value insight on where thunderstorms are most likely to form, this study primarily focuses on the early stages of the convective life cycle resulting from thunderstorm development. The transition from shallow to deep convection is often referred to as convective initiation.

Historically, convective initiation has been defined quantitatively using a radar reflectivity threshold, but the threshold values have been highly variable, ranging between about

30 dBZ and 40 dBZ . The origin of these radar reflectivity thresholds is rooted in radar based field studies characterizing storm development. For example, a field study documented in Wilson and Schreiber (1986) evaluated the nature of storm development east of the Colorado Rocky Mountains using radar. A storm was defined when a radar reflectivity threshold of 30 dBZ at about 1 kilometer above ground level (AGL) was met. Conversely, Roberts and Rutledge (2003) defined a 35 dBZ threshold to distinguish between mildly precipitating and vigorous convective storms but the authors did not cite a source in their definition, making it subjective. The 35 dBZ threshold was then used in other studies like Mueller et al. (2003) and those citing a legacy based definition of convective initiation (Mecikalski and Bedka, 2006, Mecikalski et al., 2010a,b, 2015, Mueller et al., 2003). On the high end of the convective initiation radar threshold range, 40 dBZ at the -10°C isotherm level in a storm was shown to be well correlated with electrification (Gremillion and Orville, 1999, Sieglaff et al., 2011, Zipser and Lutz, 1994). Based on the electrification mechanisms discussed in Saunders (1993), electrification most likely occurs post convective initiation.

The vertical level where the radar reflectivity threshold is applied further complicates which threshold would be appropriate for representing convective initiation for prediction with deep learning methodology. Mecikalski and Bedka (2006), Mecikalski et al. (2010b), Mecikalski et al. (2010a), Walker et al. (2012), and Mueller et al. (2003) used base level reflectivity while Mecikalski et al. (2015) used reflectivity higher in the cloud (-10°C isotherm level). Using a reflectivity at the -10°C isotherm level is supported by Browning and Atlas (1965), which determined severe storms more often develop robust

radar reflectivity cores at the -10°C level than non-severe storms. For this reason, stratiform cloud structures may be removed from the clouds of interest when using a threshold at a higher radar elevation angle. However, Zipser and Lutz (1994) indicated radar reflectivity gradually decreases with height above the freezing level, so reflectivity thresholds need to be adjusted accordingly. A threshold applied higher vertically and below freezing in a storm should decrease compared to lower vertical levels. Thus, this study defines convective initiation as the first occurrence of a radar reflectivity echo of 30 dBZ at the -10°C isotherm level. This definition is also consistent with the field study documented in Wilson and Schreiber (1986).

2.2   Deep Learning Model Structure

The goal of a deep learning model is to take a set of inputs and receive a set of outputs relevant to a particular task. In this study, the desired output is a probability map capturing convection that has initiated or will initiate in the next 60 minutes. In order to generate this desired output, deep learning models use successive layers to transform data into meaningful representations which are characterized by a set of weights (Chollet, 2018). The weights determine the contribution from each layer needed to generate output consistent with the input labels. To determine the optimal set of weights for the layers, the model undergoes a training process, which involves measuring the difference between the model output and the training targets (the desired values) with a loss (or objective) function. An optimizer is used to update the weights in a manner resulting

in the reduction of the loss value. The training process is depicted in Figure 2.1. In the

following sections, the individual aspects of the model in Figure 2.1 will be discussed.
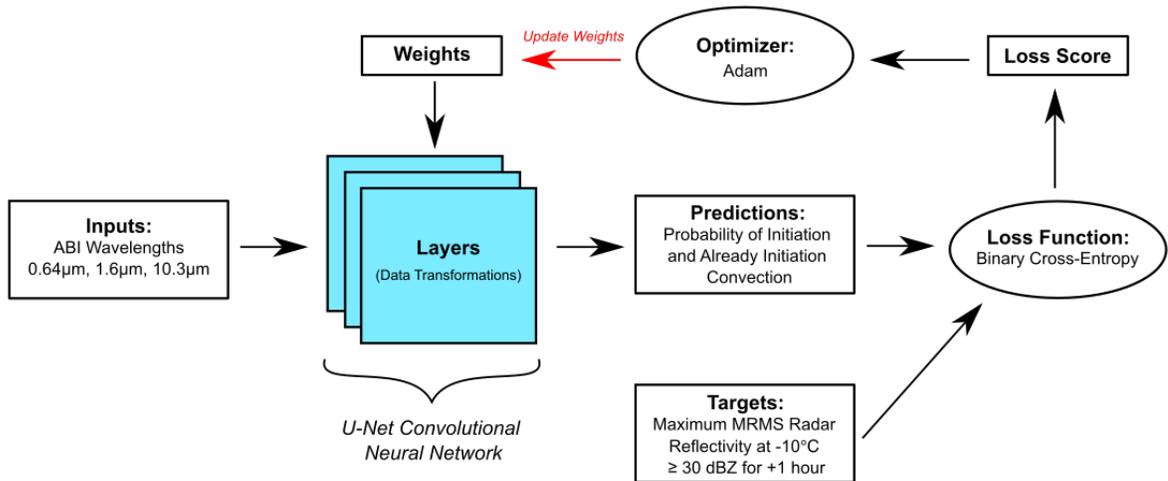

FIGURE 2.1: A diagram of the training process for developing a deep learning model for convective initiation and already initiated convection. The weight updating procedure is repeated until the loss is minimized.

2.2.1   Inputs, Outputs, and Targets

As indicated in Section 1, geostationary satellite imagery provides critical insights into the

evolution of convective clouds, including cloud top glaciation and other visual indicators

preceding or coinciding with convective initiation. Thus, geostationary satellite data are

well suited for use in a computer vision based machine learning model aimed at analyzing

and predicting convective weather. Since this model is designed to support National

Weather Service operations, data from the next generation Geostationary Operational Environmental Satellite (GOES-16), which scans the Continental United States (CONUS) at least every 5 minutes, is used as input into the model. In particular, the Advanced Baseline Imager (ABI) 0.64 μm, 1.6 μm, and 10.3 μm wavelength spectral bands (channels 2, 5, and 13) comprise the model inputs in the form of reflectance, reflectance, and brightness temperature, respectively. An example false color red-green-blue (RGB) image with these ABI bands is depicted in Figure 2.2.
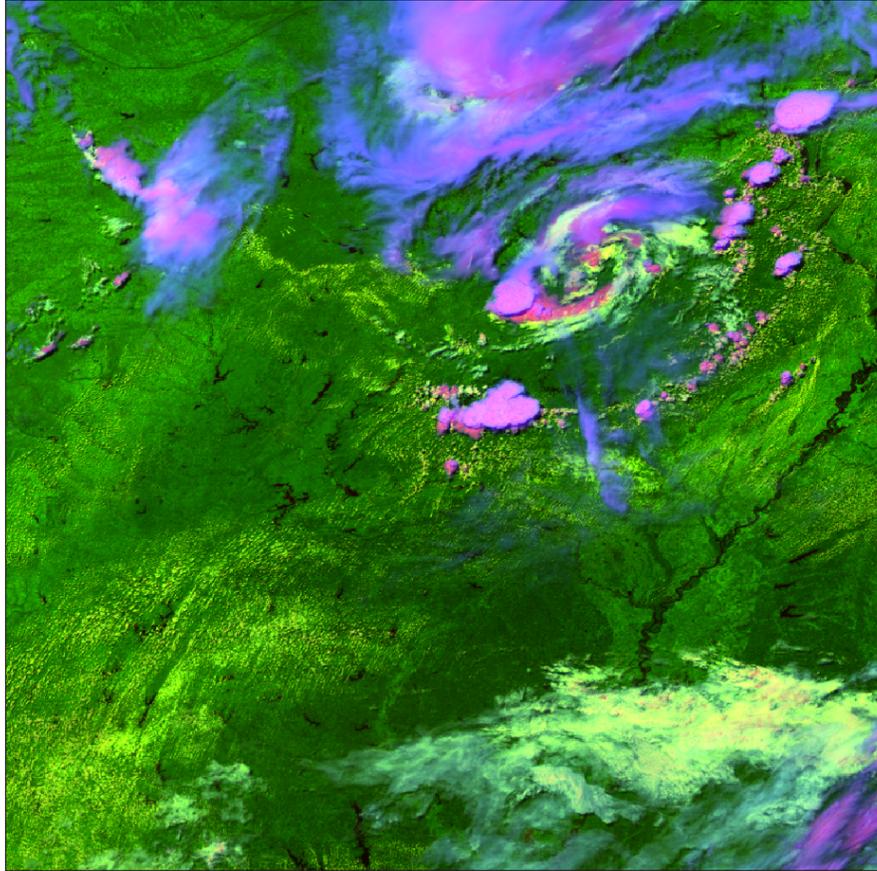
FIGURE 2.2: A false color red-green-blue (RGB) image of ABI band wavelengths 0.64 μm (red), 1.6 μm (green), and 10.3 μm (blue) for June 26, 2019 at 18:26 UTC.

The selected ABI spectral bands are collectively sensitive to pertinent features such as overshooting tops, cloud top glaciation, and cloud top height (Elsenheimer and Gravelle, 2019, Pavolonis et al., 2005). The 0.64 μm band measures reflected sunlight from the red portion of the visible spectrum at a spatial resolution of 0.5 km. At 0.5 km resolution important textual patterns, which are correlated with updraft velocity, are resolved. The 1.6 μm band measures reflected sunlight in the near infrared (IR) part of the spectrum at a spatial resolution of 1 km. This band is commonly referred to as the snow/ice band because ice absorbs more radiation at this wavelength than liquid water (Elsenheimer and Gravelle, 2019). Given the difference in absorption, the 1.6 μm band is useful for detecting cloud top glaciation (Pavolonis et al., 2005). The 10.3 μm band provides infrared window data at a spatial resolution of 2 km. The 10.3 μm band measurements are most often utilized in brightness temperature units. Commonly known as the clean IR longwave window band, the 10.3 μm band is sensitive to cloud temperature and surface properties under clear conditions.

Through the National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed, operational forecasters have expressed a clear preference for probabilistic guidance over deterministic guidance (Mecikalski et al., 2015, Siewert and Kuhlman, 2011, Terborg and Gravelle, 2012) since probabilistic guidance conveys uncertainty (Dance et al., 2010, Mecikalski et al., 2015, Steiner et al., 2010). Therefore, the outputs of the deep learning model for initiated and initiating convection are probabilistic maps, where each pixel has an associated probability value. Furthermore, for comparison to the targets and for binary statistical evaluation of the model (hits, misses, false alarms, etc.), the

convection probabilities in the probability map outputs are used as thresholds for "yes" or "no" convective initiation or initiated convection cases.

Since the definition of convective initiation is the first occurrence of a radar reflectivity echo of 30 dBZ at the -10°C isotherm level, any radar reflectivity greater than or equal to 30 dBZ at the -10°C isotherm level is taken to be initiated. Therefore, anything below this threshold is considered a null case and everything greater than or equal to this threshold is the target feature. The Multi-Radar Multi-Sensor (MRMS) product suite from NOAA and the University of Oklahoma (Zhang et al., 2016) provides quality controlled reflectivity data at the -10°C isotherm for the entire continental United States (CONUS) at 1 km resolution, making it ideal for identifying initiated convection. However, since a primary goal is to predict which cumulus fields will initiate with up to 60 minutes of lead-time, a single timestamp of MRMS radar reflectivity would only allow instantaneous identification of convection. In order to have prognostic value, the maximum MRMS radar reflectivity, in the 60 minutes following the satellite observation time, is used to generate the labeled data needed to train the machine learning model. An example of the time composite MRMS reflectivity is shown in Figure 2.3.

In Figure 2.1, the inputs, targets, and predictions (outputs) of the model are labeled to reflect the data and desired products as discussed in this section. The ABI data inputs enter the convolutional neural network, where data transformations are stored and characterized by weights. Then the outputs are compared to the labels. Based on
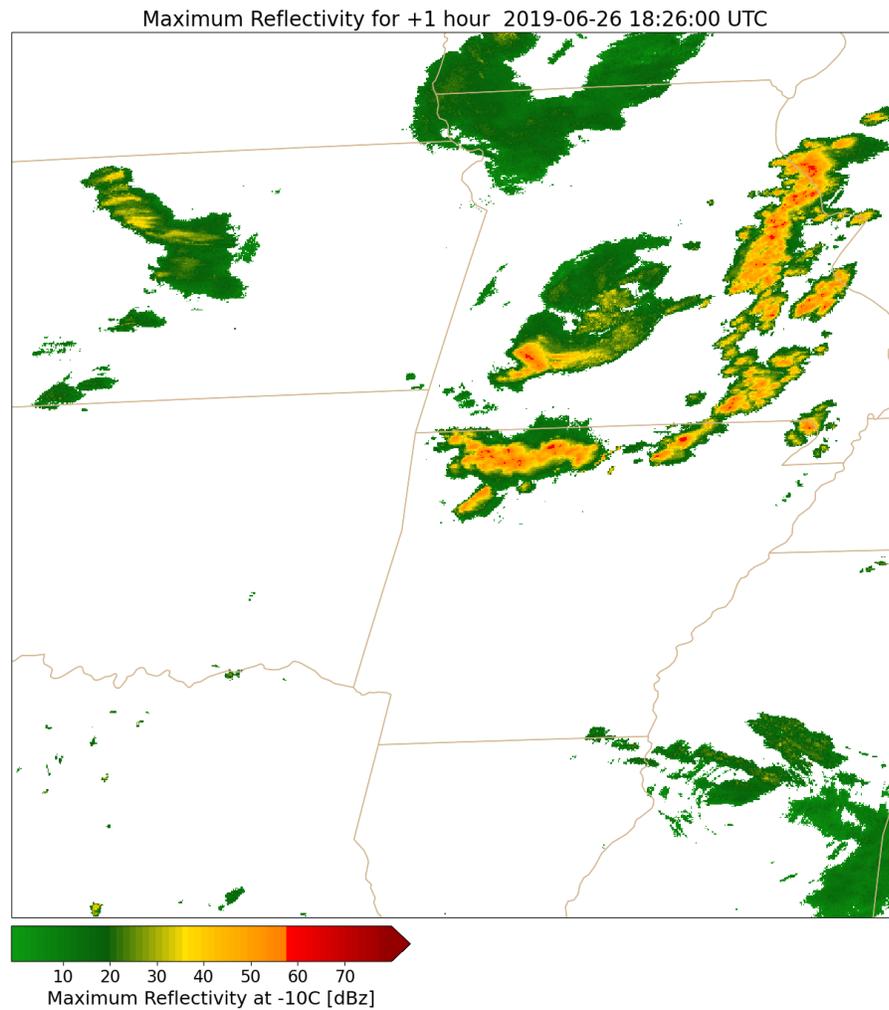
FIGURE 2.3: The maximum MRMS radar reflectivity for 1 hour into the future at the -10°C isotherm for June 26, 2019 at 18:26 UTC

the distance/loss between the probability output and the target labels, the weights are updated, and the process is repeated until the loss is minimized.

2.2.2 U-Net, a Convolutional Neural Network

There are many ways to represent the layers and associated data transformations from Figure 2.1 to best complete the desired task. For this study, the layers are arranged in

a U-Net, which is a type of convolutional neural network. U-Nets were developed by Ronneberger et al. (2015) for biomedical image segmentation and have since been used in a variety of applications. Here a U-Net is used to take ABI inputs and semantically output a segmentation map with pixel by pixel probabilistic predictions. A diagram of the U-Net set-up is included in Figure 2.4 and a description of the processes in the U-Net is included below.
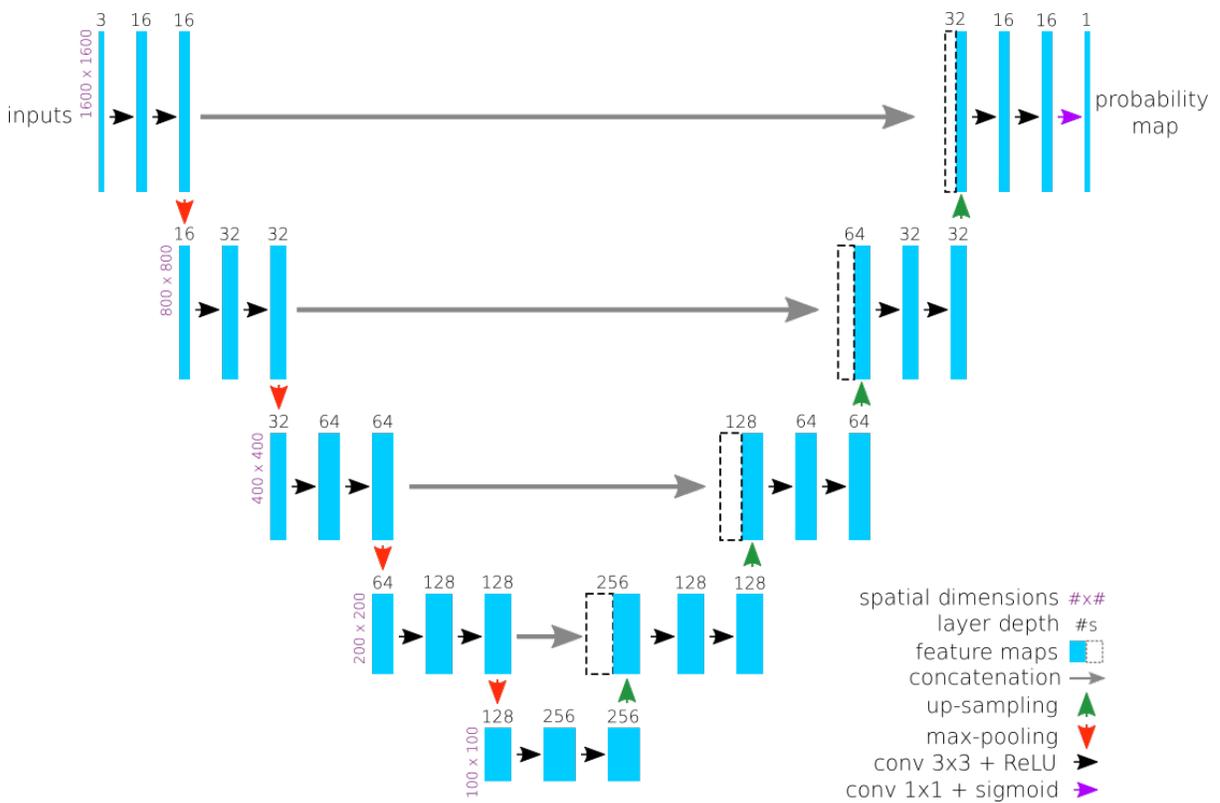
FIGURE 2.4: A depiction of the U-Net Convolutional Neural Network (data layers and transformations) for the deep learning model for convective initiation and initiated convection. Adapted for this application from the original U-Net paper by Ronneberger et al. (2015).

The inputs described previously include three satellite bands, and they enter the U-Net as three layers. These layers undergo a series of transformations represented by the arrows in

Figure 2.4. The first transformation is a two-dimensional convolution (Conv2D function in Tensorflow). During a convolution the data is split into overlapping windows where local patterns can be isolated and learned (Chollet, 2018). The size of the windows for this application are primarily three by three (except the last transformation is one by one) with a stride of one and padding with zeros to ensure the output has the same dimensions as the input. Originally the depth of the layer (also called the number of feature maps) matches the number of bands input into the model. However, after going through a transformation, these numbers no longer represent the number of inputs, instead they represent the number of filters. According to Chollet (2018) the filters encode aspects or features of the input data such as the presence of a cloud in this application. This allows for patterns within the data to be learned by the machine. The convolutions are accompanied by a specified activation function, which converts the output into a form appropriate for use as the input to the next step in the U-Net and allows for non-linearity. The majority of the convolutions in this U-Net use the rectified linear unit (ReLU) activation function to set any negative values to zero.

After going through a couple of convolutions, max-pooling (red arrows in Figure 2.4 and the MaxPooling2D function in Tensorflow) is used to aggressively downsample the data. Since the spatial dimension is large to start with, the original three by three windows view small patches of data and learn small scale patterns when going through the convolutions. To learn large scale patterns, max-pooling outputs the maximum value of each channel for a series of two by two windows (with a stride of two), which downsamples the features maps by a factor of two. Thus, the feature map sizes are cut in half during this procedure,

so, during a subsequent convolution, a window of the same size as before will view a larger area. The process of going through a couple of convolutions followed by a max-pooling transformation is repeated four times forming the encoding branch of the U-Net.

At the end of the encoding branch, the image has a low resolution. To get it back to a high resolution image so the patterns can lead to a spatially applicable, pixel by pixel result, a decoding branch with a pattern similar to the encoding branch but with up-sampling instead of max-pooling is used. Up-sampling is performed with a three-dimensional transposed convolution (Conv2DTranspose in Tensorflow) and is combined with a skip connection (concatenation by the concatenate function in Tensorflow) to retain details in the overall prediction and provide an alternate path for the gradient in backpropagation, aiding in converging on a loss value during training (Ronneberger et al., 2015). After reaching a resolution matching the original input resolution, a final convolution is performed. This convolution uses a sigmoid activation function instead of ReLU in order to constrain the outputs between zero and one, resulting in a probability map.

2.2.3   Loss Function and Optimizer

During the training process, optimal weights for the model are found by iterating through the process depicted in Figure 2.1. For each cycle through the training loop (one cycle is called an epoch), the difference between the outputs/predictions and the targets is measured with a loss function. This loss score or value is then fed into an optimizer which updates the weights of the model to see if new weight values will result in an

improved loss score. In an ideal model, the loss function will converge after numerous training epochs. The loss function for this model is binary cross-entropy since the targets are divided up into two classes (below 30 dBZ at the -10°C is considered a "no" case and everything greater than or equal to this threshold is considered a "yes" case) and the desired outputs are probabilities. This loss function is recommended for such an application by Chollet (2018) and is represented mathematically in Equation 2.1 where p_i is the predicted probability of initiated convection, y_i is the binary data label (1 if there is initiated convection and 0 otherwise) for the ith example, N is the number of examples, and ε is the binary cross-entropy in the range of $[0, \infty)$ (Cintineo et al., 2020b). Additionally, the optimizer for this model is Adam due to its computational efficiency and stability (Kingma and Ba, 2017).

$$\varepsilon = -\frac{1}{N} \sum_{i=1}^{N} [y_i log(p_i) + (1 - y_i) log(1 - p_i)] \tag{2.1}$$

2.3   Training Data

To enable more rapid experimentation with model features and hyperparameters, the initial model domain was confined to a section of the CONUS encompassing Missouri, Arkansas, and portions of surrounding states. This is the same spatial domain shown in Figure 2.3. By limiting the spatial domain, the model will be optimized for convective processes common to the domain. In addition, the initial focus is on warm season convective development occuring during the daytime (solar zenith angle is less than 85°).

Thus, the training, validation, and testing data sets were limited to May-August of 2018 and 2019. This eliminates variability due to seasonality and complications with band availability (data for visible bands are not present at night). Future work, as discussed later, will investigate elimination of the time/space constraints.

2.4 Model Validation Methodology

After training, the model is evaluated to investigate how it is performing on an independent set of data. For training the model, as mentioned in Section 2.2.1, data from May through August in 2018 and 2019 are given to the model as inputs. Within this time frame, the last five days in each month are set aside for a validation data set and are not used to train the model. Then, these dates are used within model validation, so the model is introduced to an entirely new set of data for validation. Overall, 23887 training samples and 5039 validation samples are used. The performance of the model is evaluated during validation by (1) comparing the loss values in each epoch from training to the loss during validation, (2) plotting the receiver operating characteristic (ROC) curve, a metric of performance explained further in Section 3.1, and calculating the area under this curve (AUC), (3) plotting the probability of detection, success ratio, and critical success index (CSI in Equation 2.2) on a performance diagram, and (4) plotting the conditional event frequency vs. forecast probability with a histogram counting the number of predictions in each probability bin in an attributes diagram. The model performance is discussed in detail in the next several sections.

$$CSI = \frac{\text{Hits}}{\text{Hits} + \text{False Alarms} + \text{Misses}} \qquad (2.2)$$

2.5 Model Analysis and Interpretation Methods

Deep learning models are known for being a "black box" because they map a set of inputs into a set of outputs with limited traceability. In other words, there is no direct indication of what patterns have been learned. A series of strategies, detailed below, are employed in this study to lend insight into what the model may have learned and to validate the outputs make physical sense.

2.5.1 Layer-wise Relevance Propagation

Layer-wise relevance propagation (LRP) is a technique designed to evaluate which pixels in a scene influence the prediction at a selected pixel using a decomposition of non-linear classifiers (Bach et al., 2015). The pixel-wise decomposition involves building a local redistribution rule for each neuron in a deep network and applying the rule through a backward pass through the model (Samek et al., 2016). Neurons are located within deep learning layers such as the layers shown in the U-Net for this study (Figure 2.4). They receive inputs from previous layers, are multiplied by a particular weight (learned during training), and are summed to produce an output. Thus, layer-wise relevance propagation essentially uses the weights and activations from the forward pass (trained model) and propagates backwards through the network to determine the relevance (R) of each pixel to a particular prediction (Bach et al., 2015).

Mathematically, this process is represented using Equation 2.3 detailed in Samek et al. (2016), which is known as the alpha-beta rule. Within this equation, $(x_i)_i$ are the neuron activations at layer l, $(R_j)_j$ are the relevance scores associated with the neurons at the layer $l + 1$, w_{ij} is the weight connecting neuron i to neuron j, $()^-$ denotes the negative contribution, and $()^+$ denotes the positive contribution. α and β are conservative constants upholding layer-wise conservation of relevance such that $\alpha - \beta = 1$. For example, if only positive relevance contributions are of interest, then $\alpha = 1$ and $\beta = 0$. Since both α and β are of interest in this study, $\alpha = 2$ and $\beta = 1$ in accordance with Samek et al. (2016).

$$R_i = \sum_j (\alpha \frac{(x_i w_{ij})^+}{\sum_i (x_i w_{ij})^+} - \beta \frac{(x_i w_{ij})^-}{\sum_i (x_i w_{ij})^-}) R_j \tag{2.3}$$

Within the python open source community, a python package called iNNvestigate provides functions to perform layer-wise relevance propagation methods (Alber et al., 2018). The alpha-beta rule function "LRPAlpha2Beta1" in the iNNvestigate package is used to perform layer-wise relevance propagation for this study.

2.5.2   Spectral Band Contributions

As mentioned previously, there is a clear physical rationale for each ABI band used by the model (Elsenheimer and Gravelle, 2019). To evaluate the relative importance of each band to the overall model performance, three versions of the model utilizing a subset of the three spectral bands were created and compared to the original model. Each

version eliminates a band from the input channel set. For example, one version is created with 1.6 μm and 10.3 μm (channels 5 and 13) wavelength ABI bands, which eliminates 0.64 μm (channel 2) from the inputs. The ROC curves and area under them (Section 2.4) are compared to determine which band(s) is (are) most influential (largest area under the ROC curves being most influential).

2.5.3   Comparison to LightningCast

Cintineo et al. (in prep) used a U-Net (similar to the one used in this study) to produce probabilistic lightning nowcasts from ABI data in a model known as LightningCast. By comparing the radar defined convective initiation nowcasts to LightningCast, which was trained against Geostationary Lightning Mapper (GLM) observations, the time lag between convective initiation indicators and lightning indicators in ABI imagery can be investigated.

Chapter 3

Results

3.1   Model Validation

As discussed in Section 2.2.3, the loss function is minimized through the training process (Figure 2.1) and ideally converges after many epochs. Figure 3.1 shows the loss values through the training epochs with the blue line. The loss values do decrease as the number of epochs increase and eventually converge to a value of 0.046 by epoch 6. This indicates the model has learned a set of weights leading to a minimized loss function. When the model configuration is introduced to a new set of data during validation, the model loss is relatively constant, as indicated by the orange line in Figure 3.1. Thus, overfitting does not appear to be an issue.
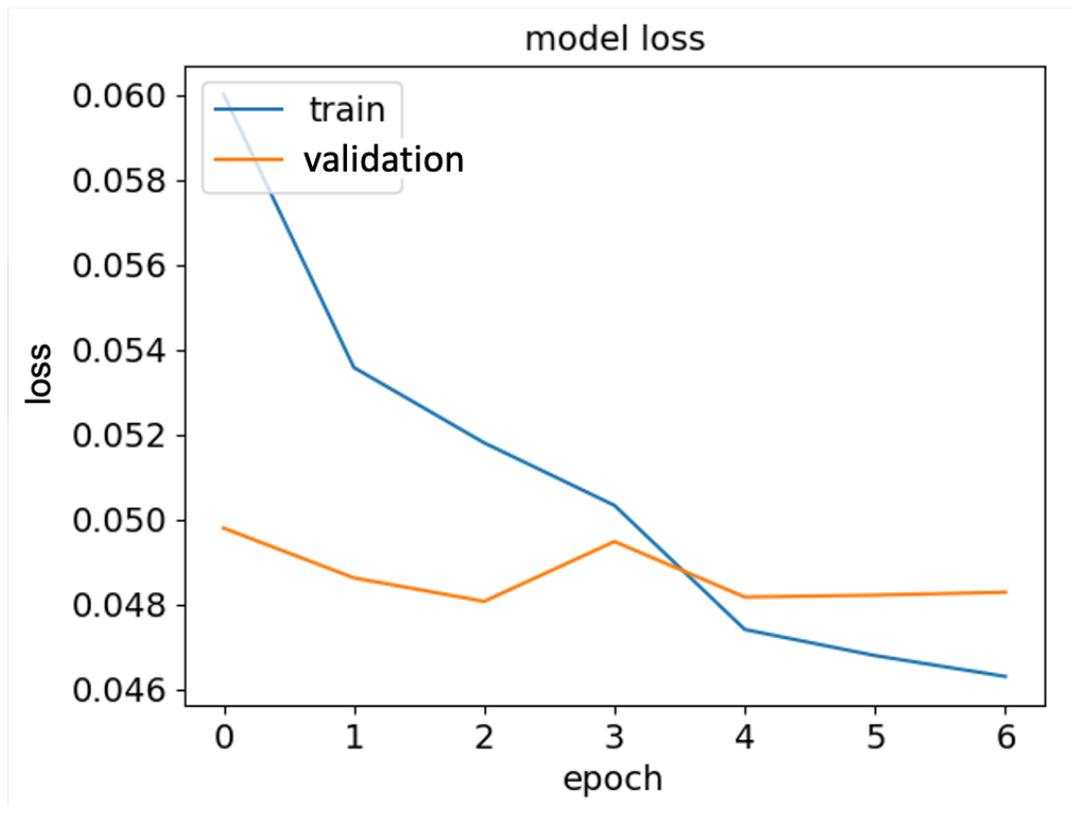
FIGURE 3.1: A graph depicting the model loss during training (blue line) and the model loss during validation (orange line).

Besides the loss, a variety of statistical methods can be used to interpret how well the model is performing on the validation dataset. As mentioned in Section 2.4, these methods include using a ROC curve, a performance diagram, and an attributes diagram. The ROC curve plots a comparison of the probability of detection and probability of false detection of each probability (model output) of convective events (e.g. 10%, 11%, etc.), defined as convection that has initiated or will initiate in the next 60 minutes (hereafter referred to as convective initiation). In a perfect model, the model would not make any incorrect predictions, so the probability of false detection would always be zero. This would lead to a straight vertical line on zero and an area under the ROC curve (AUC) of 1.0. Thus,

the closer to 1.0 the AUC is, the better. As shown in Figure 3.2, the AUC is 0.9776, indicating the model is performing well from an ROC analysis standpoint.
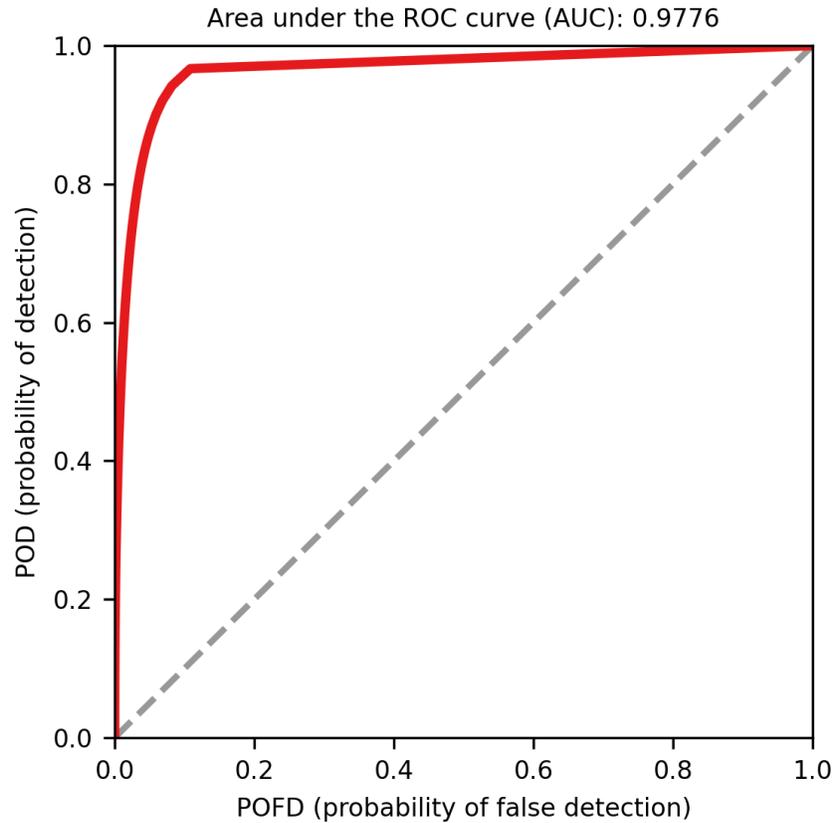
FIGURE 3.2: A graphical representation of the Receiver Operating Characteristic Curve for the convective initiation model. The probability of detection and probability of false detection values are plotted and shown in the red line. The area under this curve is 0.9776.

The performance diagram (Figure 3.3) shows a curve in red with red circles of convective initiation probabilities for the validation dataset. Low probability values are most often predicted but have the lowest success ratio, which is the number of false alarms per the total number of initiated convection predictions subtracted from one ($1 -$ false alarm ratio). The highest probabilities have the highest success ratios but are not predicted as often

(low probability of detection). Thus, when a high probability is generated by the model it is highly successful. The blue shading represents the critical success index (CSI), a verification measure of categorical forecast performance as represented by Equation 2.2 with variables computed by binarizing each probability threshold such that forecasts greater than or equal to the probability threshold, p, are equal to one ("yes" initiated convection) and the forecasts less than or equal to p are zero ("no" initiated convection). The highest CSI value of 0.38 occurs at a probability of 0.28. An ideal CSI is 1.0 because a perfect model would not have any false alarms or misses. For reference, the CSI of the convective initiation model is similar to the CSI associated with operational severe weather warnings (Cintineo et al., 2020a).

In the attributes diagram (Figure 3.4) the model would ideally follow the dotted grey line, a one-to-one relationship between forecast probability and conditional event frequency. Since the red line (convective initiation model on the validation set) curves below the one-to-one grey line, this indicates the model is slightly over-forecasting for events greater than about 40% probability.

Overall, the model performs well on the validation dataset. The loss values are consistent and the model AUC score is close to 1.0. The model most often outputs low probability values with lower success, but when it predicts a high probability of convective initiation it is highly successful. The model does over-forecast probability values greater than 40%, but this is only a slight over-forecasting. Thus, the model is consistent, reliable, and performing well.
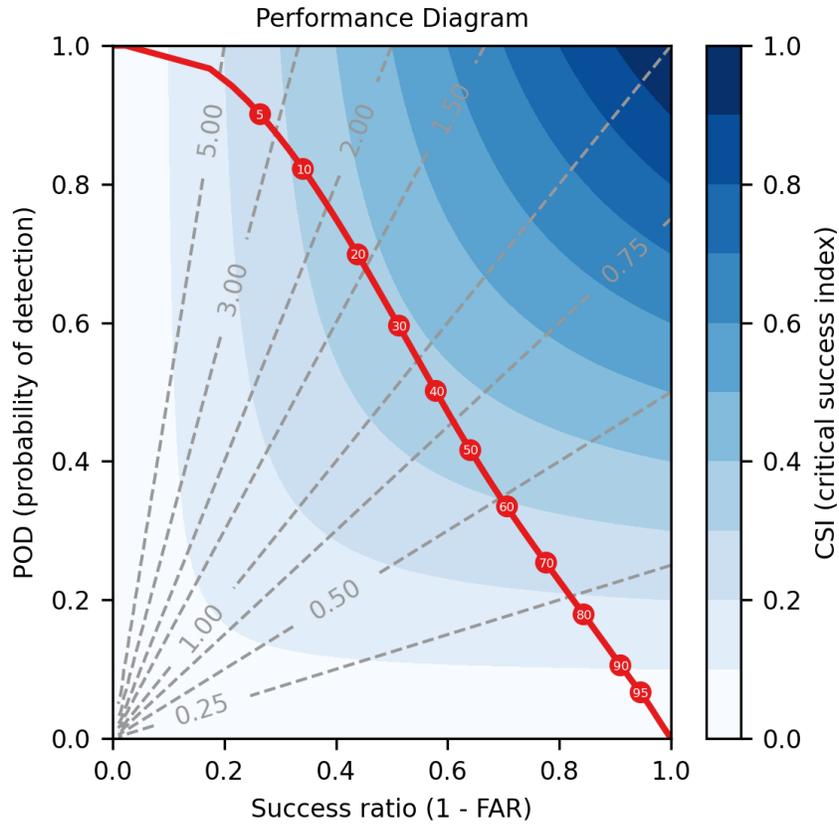
FIGURE 3.3: A representation of the model performance on a validation dataset with probability of detection on the y-axis and success ratio (1− false alarm ratio) on the x-axis. The red line represents the predicted probabilities of convective initiation with these probabilities labeled in white within the red circles on this line. The shades of blue represent the critical success index (Equation 2.2).

3.2 Model Case Studies

After validation, the model was applied to data from 2020, which is completely indepen-dent of the training and validation dataset, to visually assess how the model is performing. Presented here are two case studies: June 5, 2020 (case study one) and June 3, 2020 (case study two). Case study one was chosen because, on this particular day, a blend of mature
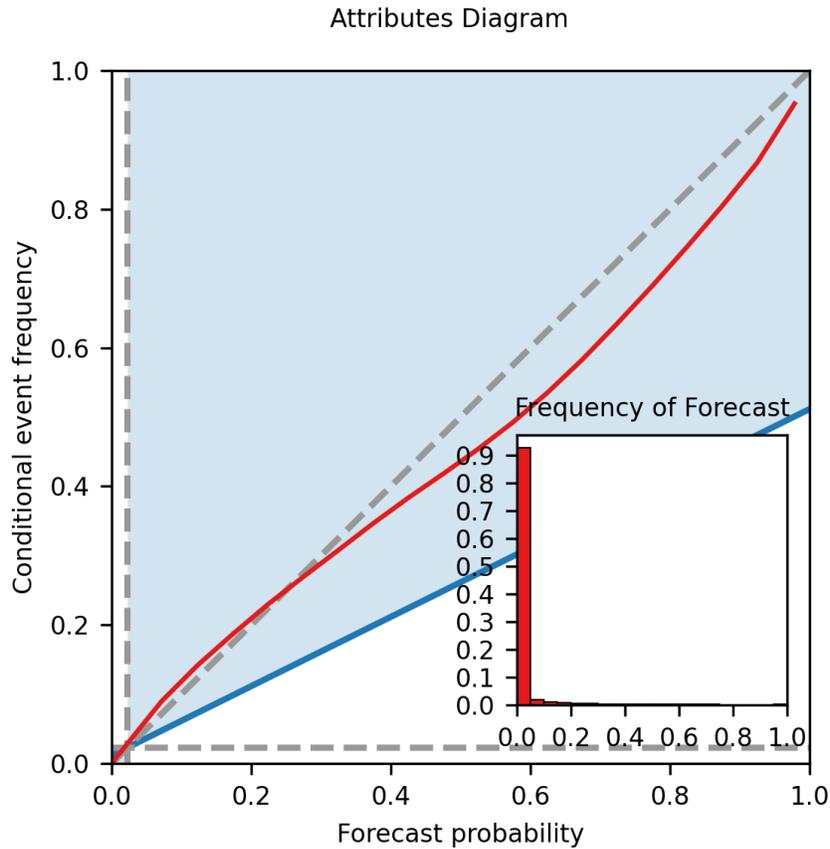
FIGURE 3.4: A depiction of how often an event is forecasted (forecast event frequency) compared to how often for a given probability bin the event occurs (conditional event frequency) for the convective initiation model on the validation dataset (in red). The inset image is a histogram of how many forecasts occur in each probability bin.

and developing convection was present in the model domain. Case study two was chosen because of the abundance of cumulus clouds, some of which developed into mature convection.

Case study one is shown in Figure 3.5. A mature convective system in middle to eastern Missouri propagates southeast towards the northeastern corner of Arkansas. The highest probability contours tend to follow the leading edge of the convective system as
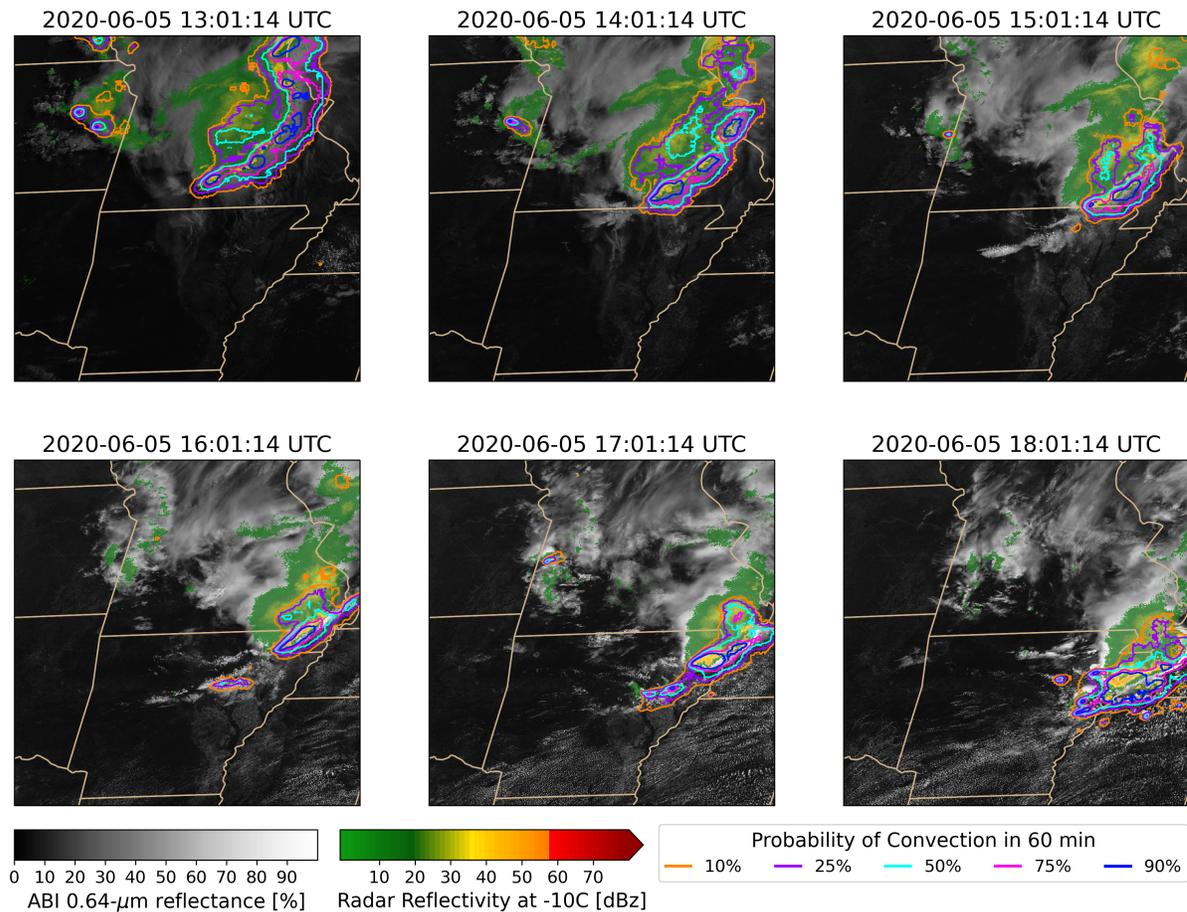
FIGURE 3.5: Case Study 1: 2020-06-05. The ABI channel 2 (0.64 μm) reflectances are displayed in black and white, the MRMS radar reflectivity at -10°C are shown in greens and yellows (second color bar from the left), and the probabilities of convective initiation are shown with colored contours. Each panel shows the same domain in hour increments.

it propagates, with some additional areas of enhanced convective initiation probability developing to the south and ahead of the mature convection (see Figure 3.5), consistent with subsequent radar data.

Case study two, shown in Figure 3.6 contains more isolated cells than case study one with the highest probabilities associated with convective precipitation cores. The time period covered by this case study is much less than the first case study because the storms
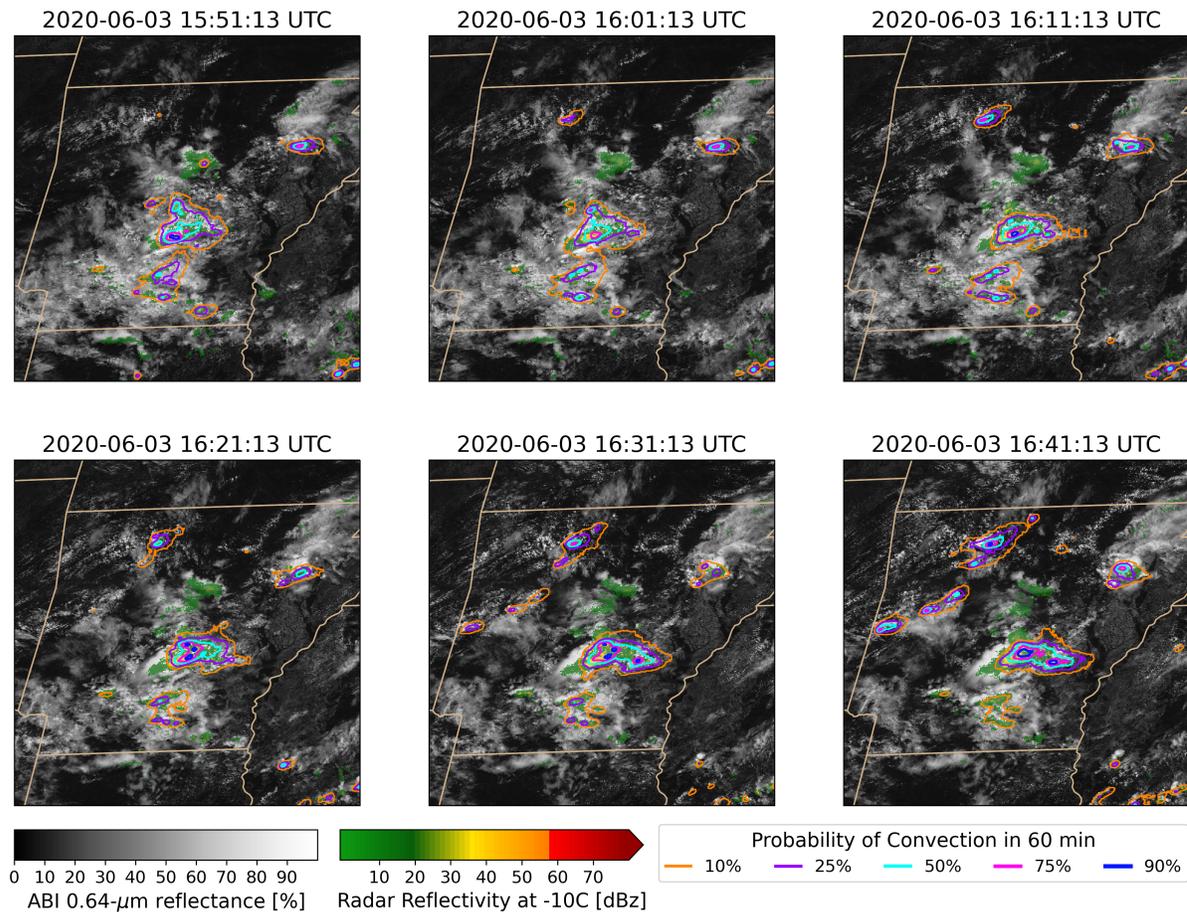
FIGURE 3.6: Case Study 2: 2020-06-03. The ABI channel 2 (0.64 μm) reflectances are displayed in black and white, the MRMS radar reflectivity at -10°C are shown in greens and yellows (second color bar from the left), and the probabilities of convective initiation are shown with colored contours. Each panel shows Arkansas in 5 minute increments (15:51:13 UTC to 16:41:13 UTC).

develop quickly and are relatively short lived. In the northwest corner of the domain, a band of storm cells develops with the first contour appearing at 15:51:13 UTC. Figure 3.7 provides a more detailed view of this developing region of convection. The first contour (10%) appears 15 minutes prior to any radar reflectivity development (with the first reflectivity values appearing at 16:06:13 UTC), and the radar signature at this time is below the convective initiation threshold. Radar reflectivity values within the convective
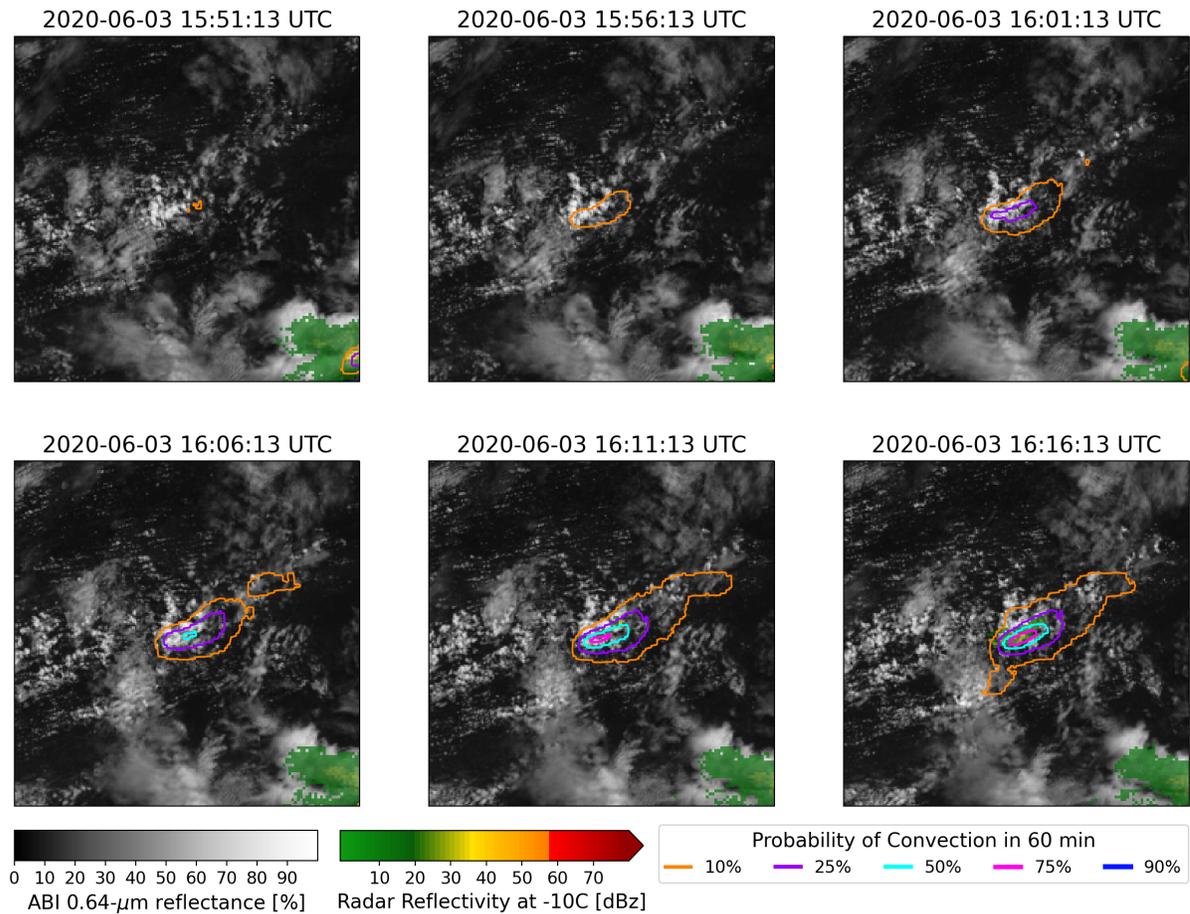
FIGURE 3.7: Case Study 2: 2020-06-03 zoomed with the center pixel at 35.99 and -93.18 degrees latitude and longitude, respectively. The ABI channel 2 (0.64 μm) reflectances are displayed in black and white, the MRMS radar reflectivity at -10°C are shown in greens and yellows (second color bar from the left), and the probabilities of convective initiation are shown with colored contours.

initiation threshold (yellows in radar color scale) start appearing in the following frame (16:11:13 UTC). Thus, the convective initiation model gives 15-20 minutes lead time on the radar reflectivity based confirmation of convective initiation.

3.3   Layer-wise Relevance Propagation

To obtain insight into what the model learned, especially in regards to identifying cumulus fields that subsequently initiate, the layer-wise relevance propagation (LRP) method, introduced in Section 2.5.1, is used. LRP determines the relevance of the pixels in an image to the prediction at a given pixel. The pixel selected for LRP from case study two, corresponds to the black star shown in Figure 3.8. Through the time frames displayed in Figure 3.7 this prediction develops from less than 20% to greater than 80% probability of convective initiation.

The results of LRP for the pixel at 35.99 degrees latitude and -93.18 degrees longitude are shown in Figure 3.9. When the storm cell starts to develop (row containing the probability of 16%), the relevant pixels are congregated around the selected pixel. In the columns for the relevance of ABI channel 2 (0.64 μm) and ABI channel 5 (1.6 μm) the pixels with the highest relevance remain on the edges of the cloud structure while the relevance for ABI channel 13 (10.3 μm) stays near the center of the storm structure, surrounding the pixel of interest. The location of the relevant pixels indicate gradients in reflectance and brightness temperature, as well as multispectral signatures, play a role in what patterns the model learns during training and uses to make predictions. This analysis demonstrates that at least some of the patterns learned by the model, such as cumulus agitation and cloud top glaciation, are consistent with patterns used by human experts when analyzing satellite imagery.
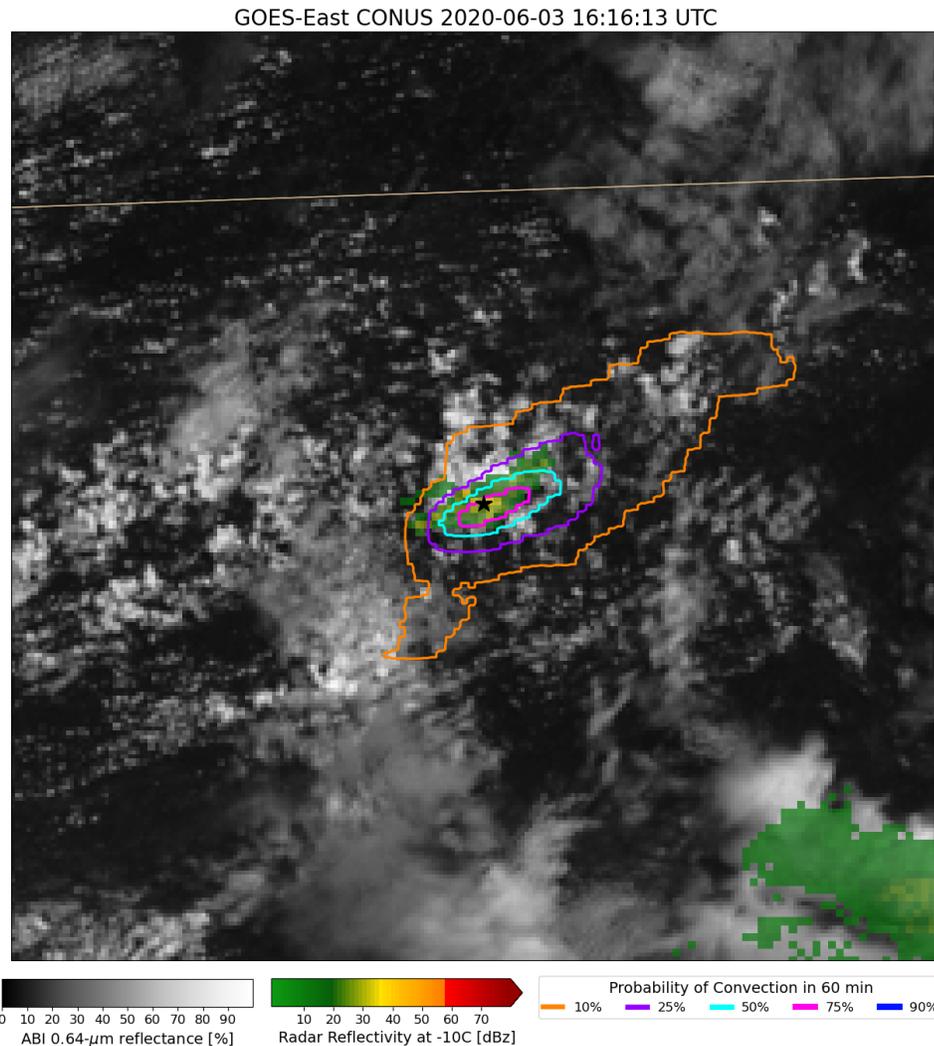
FIGURE 3.8: This image is the same as the last time frame (16:16:13 UTC) shown in Figure 3.7 with a black star to show which pixel is used for LRP. This pixel is located at 35.99 degrees latitude and -93.18 degrees longitude.

3.4 Spectral Band Contributions

To further delineate how the ABI bands contribute to how the model learns and the resulting output, multiple models are created with inputs eliminating bands. These band combinations for each model are shown in the first column of Table 3.1. The ROC scores shown in this table indicate all the model band combinations yield high ROC AUC values.

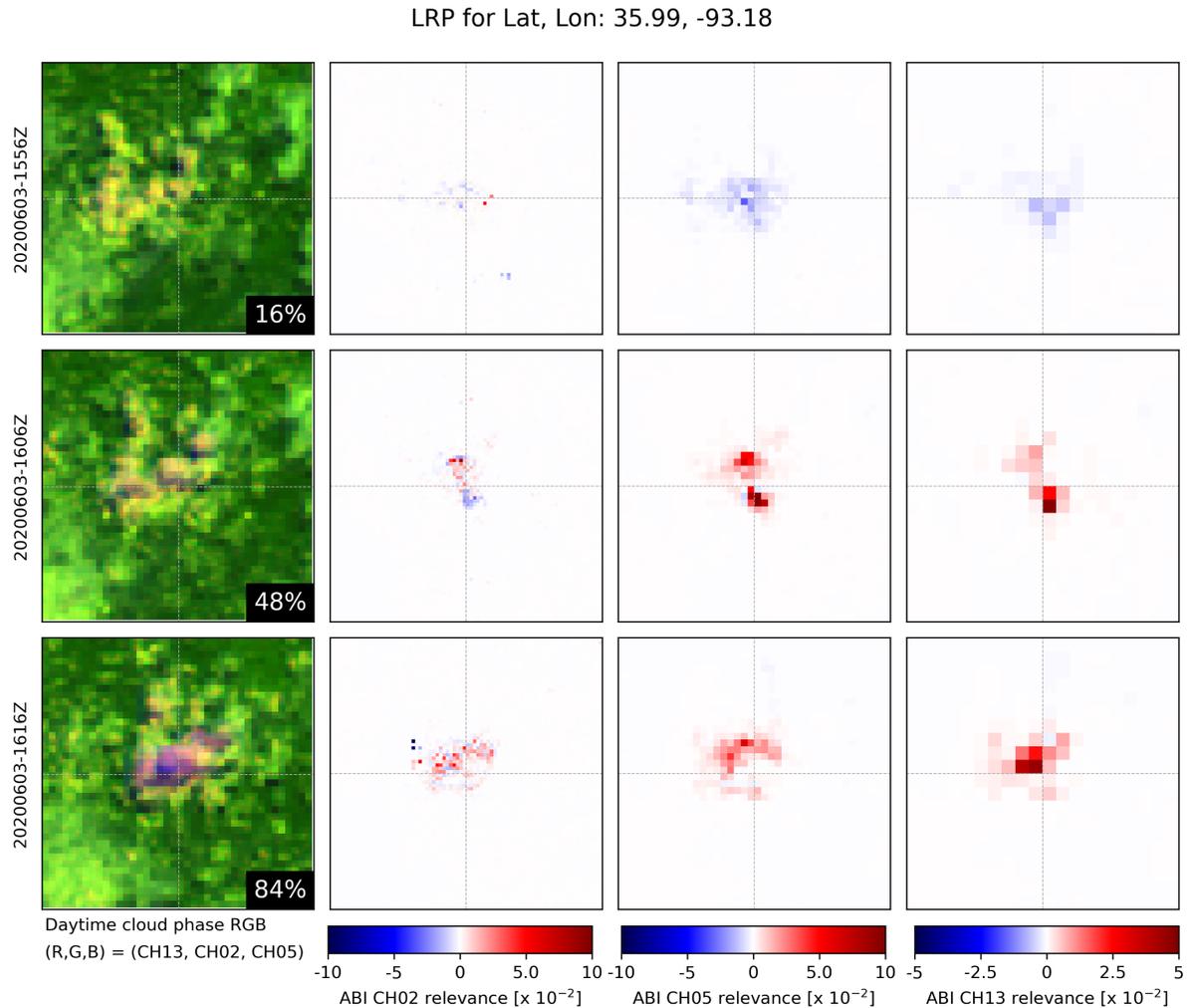FIGURE 3.9: The layer-wise relevance propagation for 35.99 and -93.18 degrees latitude and longitude respectively are shown. Each row indicates a new time step (15:56 UTC, 16:06 UTC, and 16:16 UTC on June 3, 2020). The first column on the left hand side is an RGB image similar to the one shown in Figure 2.2 with the probability of the selected pixel shown in the black box in the bottom right corner. Moving to the right from this column, the next columns depict the relevance of each input ABI band for the convective initiation model. Positive relevance is depicted in red and negative relevance is depicted in blue.

Relatively, the best ROC AUC value (closest to 1.0) is the original model with all three band inputs (0.64 μm, 1.6 μm, and 10.3 μm) then the combination eliminating channel 5 (1.6 μm). The smallest ROC AUC value occurred in the combination of channels

eliminating channel 2 (0.64 μm). This could indicate the high resolution of channel 2 may lend valuable insight into the overall prediction. Tentatively, the relative importance of the bands could be said to be as follows from most important to least: ABI bands 0.64 μm, 10.3 μm, and 1.6 μm. However, the values presented in Table 3.1 are all close to one another and more statistical testing is needed in the future to evaluate if this is significant.

ABI Band Wavelengths	ROC AUC
0.64 μm, 1.6 μm, 10.3 μm	0.97764
0.64 μm, 1.6 μm	0.97691
1.6 μm, 10.3 μm	0.97125
0.64 μm, 10.3 μm	0.97704

TABLE 3.1: This table shows the ROC AUC values for models built with various band combinations as model inputs.

Visually, the models with different band combinations can be assessed by comparing the predictions for the same date and time. This is presented in Figure 3.10. Without channel 2, the contours are broadened with less localized predictions. Since channel 2 has a fine resolution, this could indicate the fine resolution impacts the details available in the final prediction. The predictions appear similar between the other three cases. However, without channel 13, some low probability contours are omitted. For example, the farthest west probability contour in the original model result (with all three bands) is missing from the model version that lacks channel 13. Thus, cloud temperature information is likely important.

3.5   Comparison to LightningCast

In order to investigate the time lag between convective initiation indicators and lightning indicators in ABI imagery, the convective initiation predictions are compared to the LightningCast model. Both models generate predictions from ABI imagery using similar spectral channels (LightningCast utilizes one additional infrared channel). Figure 3.11 shows the first occurrence of the 10% (and 25%) LightningCast probability contour for the storm cell of interest, and discussed in depth previously, in the upper left (northwest) corner of Arkansas. This initial LightningCast signal appears at 16:01 UTC, which is 10 minutes after the convective initiation model started to generate probabilities greater than 10%, as shown in Figure 3.6. Additionally, the contours in Figure 3.11 are broader than the contours in 3.6. GLM data has a resolution of approximately 10 km which is an order of magnitude larger than the resolution of the MRMS radar reflectivity at -10°C. Similar to the relative band contributions in the previous section (Section 3.4), the contours are affected by the resolution of the data involved in training. With the current resolution of satellite lightning products, satellite based lightning models may not be able to achieve as localized of predictions as a convective initiation model can. The radar based definition of convection may also allow for longer lead times relative to a lightning based definition of convection since lightning initiation is dependent on a deep mixed phase region, which generally develops after the radar definition of convective initiation is satisfied as discussed earlier in 2.1.
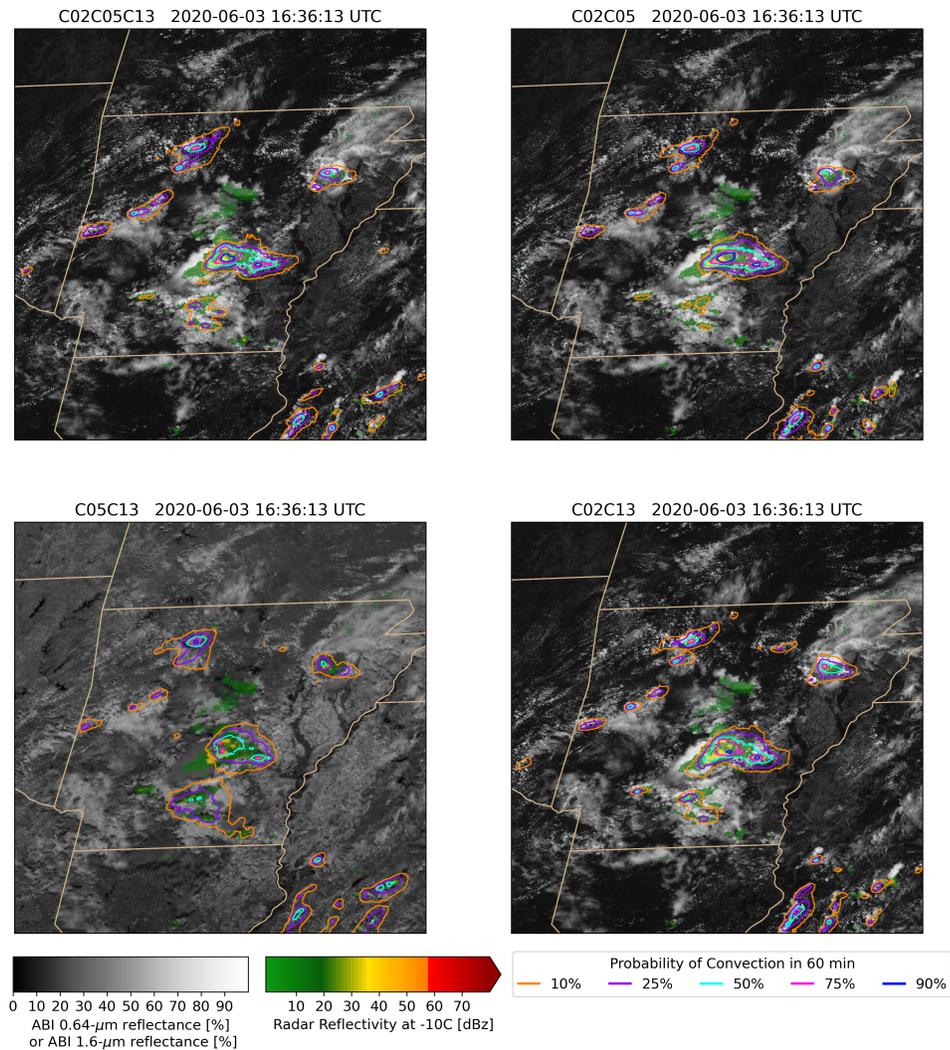
FIGURE 3.10: A comparison of model predictions for various band input combinations for June 3, 2020 at 16:36:13 UTC, with the band combinations indicated above each panel where the ABI band wavelengths are indicated by their respective channel name. C02 is ABI band of wavelength 0.64 μm, C05 is 1.6 μm, and C13 is 10.3 μm. C02 reflectances are displayed in black and white for all channel combinations with C02 included. C05 reflectances are displayed in black and white for the channel combination of C05 and C13 (lower left). The MRMS radar reflectivity at $-10°C$ are shown in greens and yellows (second color bar from the left), and the probabilities of convective initiation are shown with colored contours.

0.64 μm

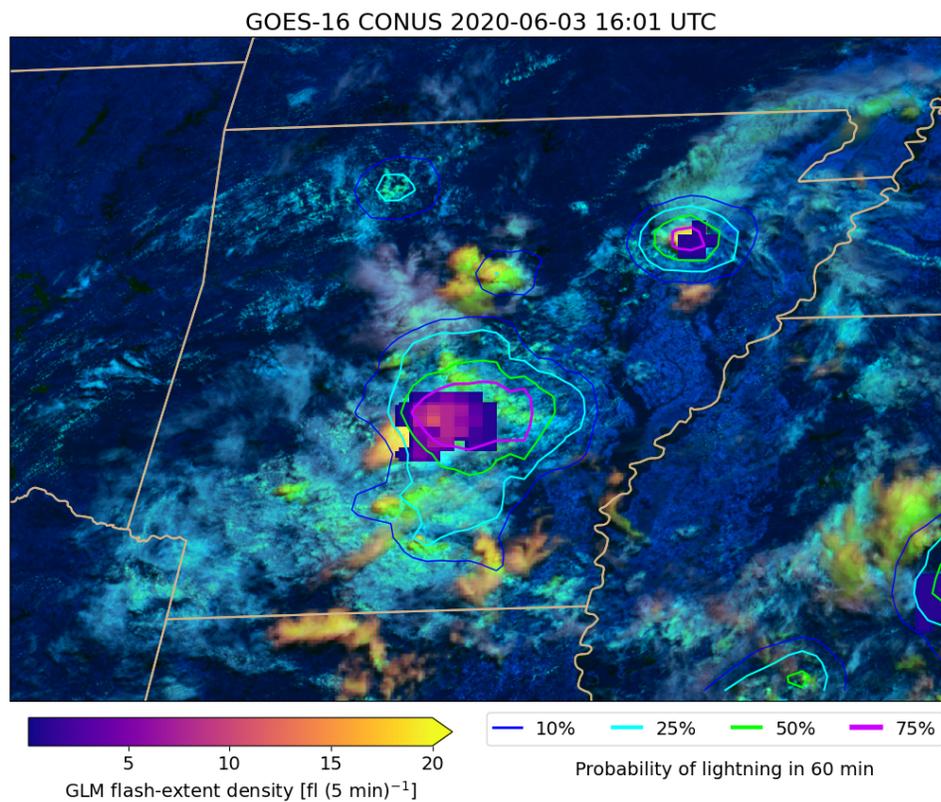FIGURE 3.11: Case Study 2: 2020-06-03. The probability of lightning in the next 60 minutes from LightningCast (Cintineo et al., in prep) are shown in contours of varied colors. Additionally, on top of an RGB image, GLM flash density is shown in colors corresponding to the color bar on the bottom left.

Chapter 4

Discussion and Conclusion

4.1 Discussion

As shown in Section 3.1, the deep learning model performed well with a converging loss value in training, a consistent loss value during validation, an area under the receiver operating characteristic curve (ROC AUC) of 0.9776, and only slight over-forecasting of probabilities greater than about 40%. For comparison, the classical machine learning models developed by Mecikalski et al. (2015) for nowcasting convective initiation yielded ROC AUC values between 0.69 and 0.83 . Though the convection model presented in this paper is not directly comparable the Mecikalski et al. (2015) models because the convection model was restricted to daytime only cases in Missouri, Arkansas, and surrounding states while including mature convection instead of only newly initiating cases,

the much larger ROC AUC score illustrates the promise of deep learning for this application. However, further work is needed to scale the model to the full continental United States and additional methodology will be necessary to ensure distributed sampling of various terrains, seasons, and climate zones.

Case study two (June 3, 2020), as presented in Section 3.2, illustrates how the model is able to identify which cumulus clouds in a larger field of cumulus will initiate about 15-20 minutes prior to the formation of a radar echo. With further development, the lead time relative to radar may enhance operational and personal decision-making (e.g. help people seek shelter before the onset of lightning and other hazards). However, lead times could depend on the labels presented to the model during training. As described in Section 2.2.1, the maximum MRMS radar reflectivity at the -10°C isotherm in the 60 minutes following the satellite observation time was used as targets or data labels during the model training process. If the maximum was taken over a different time window (perhaps 30, 90, or 120 minutes), the lead time may be further enhanced. Further experimentation is needed to ensure the maximum amount of lead time is provided without sacrificing accuracy. Anecdotally, the lead time from the deep learning based convective initiation model provided increased lead time for convective weather compared to a model developed for predicting lightning onset. The first indication of a prediction for convective weather appeared 10 minutes earlier in the deep learning model for convective initiation than the deep learning model for lightning (shown in Figure 3.11). However, this is one example. Further investigation is needed to quantify the lead time differences between lightning models and convective models as well as their spatial and temporal differences.

An explainable artificial intelligence technique, known as layer wise relevance propagation (LRP), was used to gain insight into the most influential spatial patterns in each spectral band for the predictions. A data denial experiment provided additional information on the relative importance of the three spectral bands. The location of the high relevance pixels, identified by LRP, as shown in Figure 3.9, indicated the model learned information about gradients in reflectances and brightness temperatures as well as multispectral signatures to predict convective initiation. In the spectral band omission analysis, the 0.5 km 0.64 μm band was important for maximizing the accuracy and spatial precision of the predictions. As expected, the spatial resolution of the data is important, just as it is to a human expert manually analyzing satellite imagery. Additionally, the spectral band contribution analysis suggested the importance of the spectral band inputs in descending order could be ABI wavelengths 0.64 μm, 10.3 μm, and 1.6 μm. However, more statistical testing needs to be implemented in the future to quantify this relationship fully.

Though the convective initiation model provides a well calibrated probabilistic solution to aid forecasters in their decision making process, it may be further improved by eliminating predictions associated with long lasting, mature storms. In case study one (June 5, 2020), high probability predictions appeared to lead a mature, slow moving, storm structure as it traveled from Missouri to Arkansas with some additional areas of enhanced convective initiation predictions developing ahead of the mature storm. Since the convective initiation predictions (Figure 3.5) associated with the mature storm remain close to a known storm structure for an extended period of time, removing them from the information provided to forecasters may allow for a more concise, informative model.

Thus, future work could isolate convective initiation without highlighting mature cases through manipulation of data labeling or post-processing.

4.2   Conclusion

Determining which cumulus clouds will develop into thunderstorms is a long-standing "nowcasting" challenge. While the evolution of cumulus cloud fields in geostationary satellite imagery provides forecasters with visual clues, a skillful objective analysis tool has remained elusive. To address this capability gap, a subset of machine learning, known as deep learning, was applied to Geostationary Operational Environmental Satellite (GOES-R) Advanced Baseline Imager (ABI) imagery. More specifically a custom U-Net, a type of convolutional neural network commonly applied to computer vision problems, was developed to identify active convection and to predict which cumulus fields are most likely to initiate in the next 60 minutes. The model was trained with Multi-Radar Multi-Sensor quality controlled radar reflectivity data at the -10°C isotherm using a binary threshold of 30 dBZ with anything greater than the threshold indicating the presence of convective initiation. Model validation procedures reveal it performs well with a converging loss value in training, a consistent loss value during validation, an area under the receiver operating characteristic curve of 0.9776, and only slight over-forecasting of probabilities greater than about 40%. The application of the model to two test cases indicated the predicted probability of convection agreed well with regions of radar reflectivity at the -10°C isotherm and provided between 15 to 20 minutes of lead time prior to the first radar signatures. Additional analysis revealed the model utilizes gradients in brightness

temperature, gradients in reflectance, and multispectral patterns to make predictions. Though the model was trained and tested for a limited spatial domain in the southern United States during the daytime only, future work aims to eliminate these limitations while isolating early convective initiation. With these enhancements, this deep learning model for prediction of convective initiation has the potential to evolve into a robust tool for use in operational forecasting and provide physical insight into convective initiation through further analysis of influential predictors, spatial and seasonal patterns, and comparison to lightning occurrences.

Bibliography

Alber, M., and Coauthors, 2018: innvestigate neural networks! *CoRR*, **abs/1808.04260**, URL http://arxiv.org/abs/1808.04260, 1808.04260.

Bach, S., A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, 2015: On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE*, **10**.

Browning, K. A., and D. Atlas, 1965: Initiation of precipitation in vigorous convective clouds. *Journal of Atmospheric Sciences*, **22**, 678–683, doi:10.1175/1520-0469(1965) 022⟨0678:IOPIVC⟩2.0.CO;2.

Chollet, F., 2018: *Deep Learning with Python*. Manning Publications Co.

Cintineo, J. L., M. J. Pavolonis, and J. M. Sieglaff, in prep: Probsevere lightningcast: short-term probabilistic forecasts using the goes-r advanced baseline imager and deep learning. *Weather and Forecasting*.

Cintineo, J. L., M. J. Pavolonis, J. M. Sieglaff, L. Cronce, and J. Brunner, 2020a: Noaa probsevere v2.0- probhail, probwind, and probtor. *Weather and Forecasting*, **35**, 1523–1543, doi:10.1175/WAF-D-19-0242.1.

Cintineo, J. L., M. J. Pavolonis, J. M. Sieglaff, and D. T. Lindsey, 2014: An empirical model for assessing the severe weather potential of developing convection. *Weather and Forecasting*, **29**, 639–653, doi:10.1175/WAF-D-13-00113.1.

Cintineo, J. L., M. J. Pavolonis, J. M. Sieglaff, A. Wimmers, J. Brunner, and W. Bellon, 2020b: A deep-learning model for automated detection of intense midlatitude convection using geostationary satellite images. *Weather and Forecasting*, **35**, 2567–2588, doi:10.1175/WAF-D-20-0028.1.

Cintineo, J. L., and Coauthors, 2018: The noaa/cimss probsevere model: Incorporation of total lightning and validation. *Weather and Forecasting*, **33**, 331–345, doi:10.1175/WAF-D-17-0099.1.

Dance, S., E. Ebert, and D. Scurrah, 2010: Thunderstorm strike probability nowcasting. *Weather and Forecasting*, **27**, 79–93, doi:10.1175/2009JTECHA1279.1.

Dixon, M., and G. Wiener, 1993: Titan: Thunderstorm identification, tracking, analysis, and nowcasting- a radar-based methodology. *Journal of Atmospheric and Oceanic Technology*, **10**, 785–797, doi:10.1175/1520-0426(1993)010⟨0785:TTITAA⟩2.0.CO;2.

Elsenheimer, C. B., and C. M. Gravelle, 2019: Introducing lightning threat messaging using the goes-16 day cloud phase distinction rgb composite. *Weather and Forecasting*, **34**, 1587–1600, doi:10.1175/WAF-D-19-0049.1.

Feng, Y., Y. Wang, T. Peng, and J. Yan, 2007: An algorithm on convective weather potential in the early rainy season over the pearl river delta in china. *Advances in Atmospheric Sciences*, **24**, 101–110.

Gremillion, M. S., and R. E. Orville, 1999: Thunderstorm characteristics of cloud-to-ground lightning at the kennedy space center, florida: A study of lightning initiation signatures as indicated by the wsr-88d. *Weather and Forecasting*, **14**, 640–649, doi: 10.1175/1520-0434(1999)014⟨0640:TCOCTG⟩2.0.CO;2.

Harris, R. J., J. R. Mecikalski, W. M. M. Jr., P. A. Durkee, and K. E. Nielsen, 2010: The definition of goes infrared lightning initiation interest fields. *Journal of Applied Meteorology and Climatology*, **49**, 2527–2543, doi:10.1175/2010JAMC2575.1.

Hu, S., S. Gu, X. Zhuang, and H. Luo, 2007: Automatic identification of storm cells using doppler radars. *Acta Meteorologica Sinica*, **21**, 353–365.

Joe, P., M. Falla, P. V. rijn, L. Stamadianos, T. Falla, dan Magosse, L. Ing, and J. Dobson, 2002: Radar data processing for severe weather in the national radar project of canada. *21st Conference on Severe Local Storms*, 1–4, URL http://193.7.160.230/web/hazards/amtk/CARDS/RadarProcessingForSevereWeather_joe_et_al_2002.pdf.

Kingma, D. P., and J. Ba, 2017: Adam: A method for stochastic optimization. 1412.6980.

Lagerquist, R., A. McGovern, C. R. Homeyer, D. J. G. II, and T. Smith, 2020: Deep learning on three dimensional multiscale data for next-hour tornado prediction. *Monthly Weather Review*, **148**, 2837–2861, doi:10.1175/MWR-D-19-0372.1.

Lagerquist, R., A. McGovern, and D. J. G. II, 2019: Deep learning for spatially explicit prediction of synoptic-scale fronts. *Weather and Forecasting*, **34**, 1137–1160, doi:10.1175/WAF-D-18-0183.1.

Lagerquist, R., J. Q.Stewart, I. Ebert-Uphoff, and C. Kumler, 2021: Using deep learning to nowcast the spatial coverage of convection from himawari-8 satellite data. *Monthly Weather Review*, **149 (12)**, 3897–3921, doi:10.1175/MWR-D-21-0096.1, URL https://journals.ametsoc.org/view/journals/mwre/149/12/MWR-D-21-0096.1.xml.

Lakshmanan, V., T. Smith, G. Stumpf, and K. Hondl, 2007: The warning decision support system- integrated information. *Weather and Forecasting*, **22**, 592–612, doi:10.1175/WAF1009.1.

Lensky, I. M., and D. Rosenfeld, 2006: The time-space exchangeability of satellite retrieved relations between cloud top temperature and particle effective radius. *Atmos. Chem. Phys.*, **6**, 2887–2894, doi:10.5194/acp-6-2887-2006.

Mecikalski, J. R., and K. M. Bedka, 2006: Forecasting convective initiation by monitoring the evolution of moving cumulus in daytime goes imagery. *Monthly Weather Review*, **134**, 49–78, doi:10.1175/MWR3062.1.

Mecikalski, J. R., W. M. M. Jr., M. Konig, and S. Muller, 2010a: Cloud-top properties of growing cumulus prior to convective initiation as measured by meteosat second generation. part i: Infrared fields. *Journal of Applied Meteorology and Climatology*, **49**, 521–534, doi:10.1175/2009JAMC2344.1.

Mecikalski, J. R., W. M. M. Jr., M. Konig, and S. Muller, 2010b: Cloud-top properties of growing cumulus prior to convective initiation as measured by meteosat second generation. part ii: Use of visible reflectance. *Journal of Applied Meteorology and Climatology*, **49**, 2544–2558, doi:10.1175/2010JAMC2480.1.

Mecikalski, J. R., J. K. Williams, C. P. Jewett, D. Ahijevych, A. LeRoy, and J. R. Walker, 2015: Probabilistic 0-1h convective initiation nowcasts that combine geostationary satellite observations and numerical weather prediction model data. *Journal of Applied Meteorology and Climatology*, **54**, 1039–1059, doi:10.1175/JAMC-D-14-0129.1.

Merk, D., and T. Zinner, 2013: Detection of convective initiation using meteosat seviri: implementation in and verification with the tracking and nowcasting algorithm cb-tram. *Atmos. Meas. Tech.*, **6**, 1903–1918, doi:10.5194/amt-6-1903-2013.

Mueller, C., T. Saxen, R. Roberts, J. Wilson, T. Betancourt, S. Dettling, N. Oien, and J. Yee, 2003: Ncar auto-nowcast system. *Weather and Forecasting*, **18**, 545–561, doi:10.1175/1520-0434(2003)018⟨0545:NAS⟩2.0.CO;2.

Mueller, C. K., J. W. Wilson, and N. A. Crook, 1993: The utility of sounding and mesonet data to nowcast thunderstorm initiation. *Weather and Forecasting*, **8**, 132–146, doi:10.1175/1520-0434(1993)008⟨0132:TUOSAM⟩2.0.CO;2.

Nisi, L., P. Ambrosetti, and L. Clementi, 2014: Nowcasting severe convection in the alpine region: the coalition approach. *Quarterly Journal of the Royal Meteorological Society*, **140**, 1684–1699, doi:10.1002/qj.2249.

Pavolonis, M. J., A. K. Heidinger, and T. Uttal, 2005: Daytime global cloud typing from avhrr and viirs: Algorithm description, validation, and comparisons. *Journal of Applied Meteorology and Climatology*, **44**, 804–826, doi:10.1175/JAM2236.1.

Roberts, R. D., A. R. S. Anderson, E. Nelson, B. G. Brown, J. W. Wilson, M. Pocernich, and T. Saxen, 2012: Impacts of forecaster involvement on convective storm initiation and evolution nowcasting. *Weather and Forecasting*, **27**, 1061–1089, doi: 10.1175/WAF-D-11-00087.1.

Roberts, R. D., and S. Rutledge, 2003: Nowcasting storm initiation and growth using goes-8 and wsr-88d data. *Weather and Forecasting*, **18**, 562–584, doi:10.1175/ 1520-0434(2003)018⟨0562:NSIAGU⟩2.0.CO;2.

Ronneberger, O., P. Fischer, and T. Brox, 2015: U-net: Convolutional networks for biomedical image segmentation. *CoRR*, **abs/1505.04597**, URL http://arxiv.org/abs/ 1505.04597, 1505.04597.

Rosenfeld, D., W. L. Woodley, A. Lerner, G. Kelman, and D. T. Lindsey, 2008: Satellite detection of severe convective storms by their retrieved vertical profiles of cloud particle effective radius and thermodynamic phase. *Journal of Geophysical Research: Atmospheres*, **113**, doi:10.1029/2007JD008600.

Ruzanski, E., V. Chandrasekar, and Y. Wang, 2011: The casa nowcasting system. *Journal of Atmospheric and Oceanic Technology*, **28**, 640–655, doi:10.1175/2011JTECHA1496. 1.

Samek, W., G. Montavon, A. Binder, S. Lapuschkin, and K.-R. Müller, 2016: Interpreting the predictions of complex ml models by layer-wise relevance propagation. 1611.08191.

Saunders, C. P. R., 1993: A review of thunderstorm electrification processes. *Journal of Applied Meteorology and Climatology*, **32**, 642–655, doi:10.1175/1520-0450(1993) 032⟨0642:AROTEP⟩2.0.CO;2.

Sieglaff, J. M., L. M. Cronce, W. F. Feltz, K. M. Bedka, M. J. Pavolonis, and A. K. Heidinger, 2011: Nowcasting convective storm initiation using satellite-based box-averaged cloud-top cooling and cloud-type trends. *Journal of Applied Meteorology and Climatology*, **50**, 110–126, doi:10.1175/2010JAMC2496.1.

Siewert, C., and K. Kuhlman, 2011: Hazardous weather testbed-final evaluation. Tech. rep., NOAA's Hazardous Weather Testbed. Available online at http://www.goes-r.gov/ users/docs/pg-activities/PGFR-HWT-2011-Final.pdf.

Steiner, M., R. Bateman, D. Megenhardt, Y. Liu, M. Xu, M. Pocernich, and J. Krozel, 2010: Translation of ensemble weather forecasts into probabilistic air traffic capacity impact. *Air Traffic Control Quarterly*, **18**, 229–254.

Terborg, A., and C. Gravelle, 2012: Aviation weather testbed 2012 summer experiment. Tech. rep., GOES-R desk final evaluation. Available online at https://www.goes-r.gov/users/docs/pg-activities/PGFR-AWC-2012-Final.pdf.

Walker, J. R., W. M. M. Jr., J. R. Mecikalski, and C. P. Jewett, 2012: An enhanced geostationary satellite-based convective initiation algorithm for 0-2-h nowcasting with object tracking. *Journal of Applied Meteorology and Climatology*, **51**, 1931–1949, doi: 10.1175/JAMC-D-11-0246.1.

Wilson, J. W., and W. E. Schreiber, 1986: Initiation of convective storms at radar-observed boundary-layer convergence lines. *Monthly Weather Review*, **114**, 2516–2536, doi:10.1175/1520-0493(1986)114⟨2516:IOCSAR⟩2.0.CO;2.

Wolfson, M. M., and D. A. Clark, 2006: Advanced aviation weather forecasts. *Lincoln Laboratory Journal*, **16**, 31–58.

Zhang, J., and Coauthors, 2016: Multi-radar multi-sensor (mrms) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, **97 (4)**, 621–638, doi:10.1175/BAMS-D-14-00174.1, URL https://journals.ametsoc.org/view/journals/bams/97/4/bams-d-14-00174.1.xml.

Zhou, K., Y. Zheng, W. Dong, and T. Wang, 2020: A deep learning network for cloud-to-ground lightning nowcasting with multisource data. *Journal of Atmospheric and Oceanic Technology*, **37**, 927–942, doi:10.1175/JTECH-D-19-0146.1.

Zipser, E. J., and K. R. Lutz, 1994: The vertical profile of radar reflectivity in convective cells: A strong indicator of storm intensity and lightning probability? *Monthly Weather Review*, **122**, 1751–1759, doi:10.1175/1520-0493(1994)122⟨1751:TVPORR⟩2.0.CO;2.