Machine Learning - Hail Awareness Spatial Analysis Toolkit (HASAT)

Graduate Student Haoruo Fu (fu361@purdue.edu) Student Advisor: Chien-Tsung Lu (ctlu@purdue.edu) Joseph Hupy (jhupy@purdue.edu) Purdue University

Abstract

The National Airspace System (NAS) is a complex network of air traffic control, navigation, and communication systems that support the safe and efficient flow of air traffic in the United States. Severe weather conditions like hails have been a significant threat to flight safety. Hail consists of frozen precipitation that can cause damage to aircraft, disrupt flight operations, and pose a threat to the safety of passengers and crews. Aviation organizations have implemented various safety measures to minimize hail-related risks, such as improved weather forecasting, enhanced aircraft design, and better aeronautical decision-making (ADM) training.

This research used Esri's ArcGIS as a mapping software for geospatial analysis of the severe weather's impact on the NAS. A set of Machine Learning Neural Network algorithm has been used for hail events perdition. The Machine Learning process is conducted by MathWorks's MATLAB. The output showed that the central United States has been impact by hail events the most. The Machine Learning hail prediction showed that most of the predicted hail events are in the center of selected states. Better Neural Network algorithm with more variables are recommended for future study in order to increase the accuracy of the hail prediction.

Introduction

Twenty-three percent of all accidents were also related to weather (Department of Transportation [DOT], 2010). According to the Federal Aviation Administration (FAA) statistics, the weather is the cause of approximately 70 percent of the delays in the National Airspace System (NAS) before September 11, 2001.

Severe weather can happen at any time and may affect travel conditions for pilots and passengers. High winds, flooding, heavy downpour, and power outage can happen during severe weather events. Airports are often closed to the public during natural disasters, and flight paths are rerouted. The flight can be affected around the entire country.

For commercial and General Aviation (GA) pilots, it is necessary to be aware of Temporary Flight Restrictions (TFRs) and Notices to Air Missions (NOTAMs) for updated information related to severe weather. The FAA has provided various tips for severe operations, such as operating with two pilots, operating with traffic avoidance systems, and preparing for mechanical problems during operations (FAA, 2022).

Literature Review

Hail is a form of precipitation consisting of solid ice that forms inside a thunderstorm updraft. The detection of hail can be classified as the measurement of convective cells in cumulonimbus clouds.

Modern methods, algorithms, and automated weather radar programs can compute the direction and speed of the thunderstorm movement. Advanced weather radar can also recognize the target object's categories and estimate the hails' damage (Abshaev et al., 2010). Weather satellites can capture images of cloud formations and temperature changes in the atmosphere.

Meteorologists can use this data to identify areas of potential hail formation and track the movement of storms.

A geographic information system (GIS) is a system that creates, manages, analyzes, and maps all types of data (Esri, n.d.). The use of the GIS framework can connect the outage cost to weather-related data, such as the location and the duration of severe weather (Chen et al., 2015).

Machine learning is increasingly being used in weather forecasting to improve the accuracy and efficiency of predictions.

Neural networks are a type of machine learning algorithm that are increasingly being used in weather forecasting to improve the accuracy of predictions.

Research Methodology

Figure 1 shows the process of this study.

For the GIS analysis, Class B, C, and D airspaces were selected because they contain the heaviest air traffic in the United States.

For ArcGIS demographical indication, all hail events with size of diameter over 3 inches will be selected for the analysis. For the Machine Learning Neural Network hail prediction analysis. Feed-Forward Neural Network will be used in this prediction model. The Neural Network will be created with one hidden layer of 10 neurons. The Levenberg-Marquardt algorithm will be used for training the Neural Network (**Figure 2**).

After the Machine Learning Neural Network process, all the output data will be transferred into ArcGIS for better visualization.

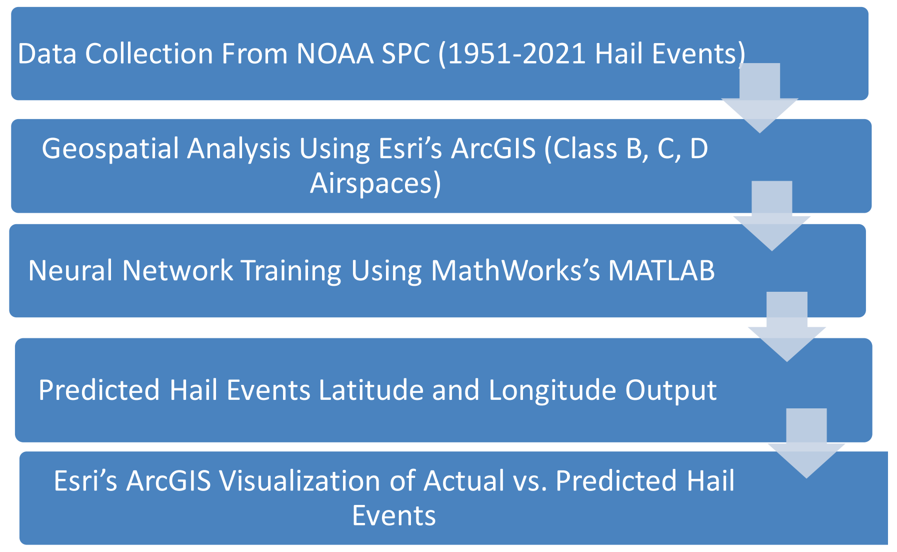

Figure 1: Research Approach of the Study

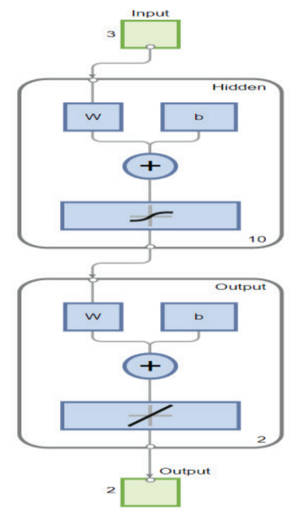

Figure 2: Feed-Forward Neural Network

PURDUE UNIVERSITY

School of Aviation and Transportation Technology
POLYTECHNIC INSTITUTE

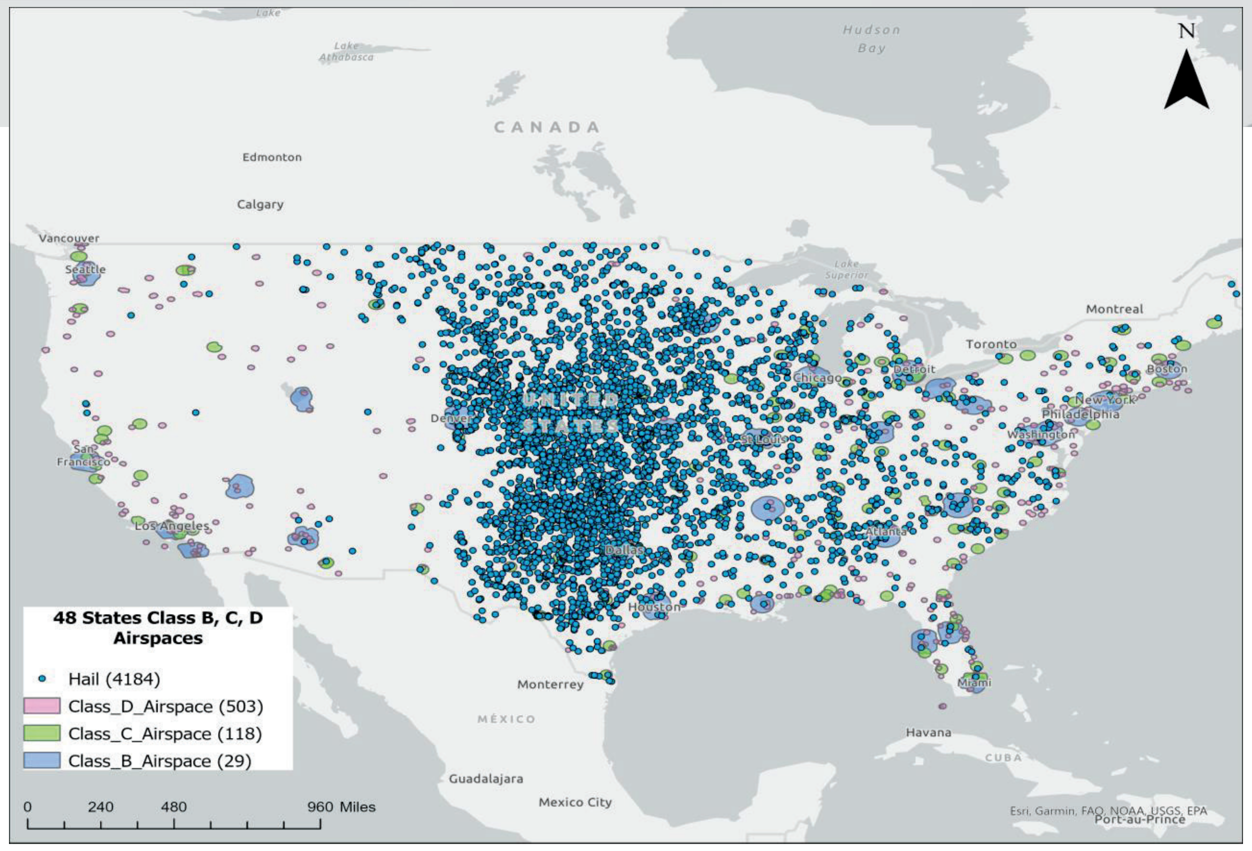

Figure 3: 48 States Airspaces and Hail Events

Findings

Based on NOAA's data, from 1951 to 2021, there were a total of 4,184 hail events with size of diameter over 3.00 inches. **Figure 3** represents the impact of hail events (>3 inches) on Class B, C, and D airspaces in the contiguous United States. Central United States is the epicenter of the hail events. Some airspaces, such as Dallas Class B airspace, have been significantly hit by hail events. During this analysis, the path of the hail was excluded as most of the hails occurred within static distance. In other words, most hails do not act like tornadoes traveling long distances.

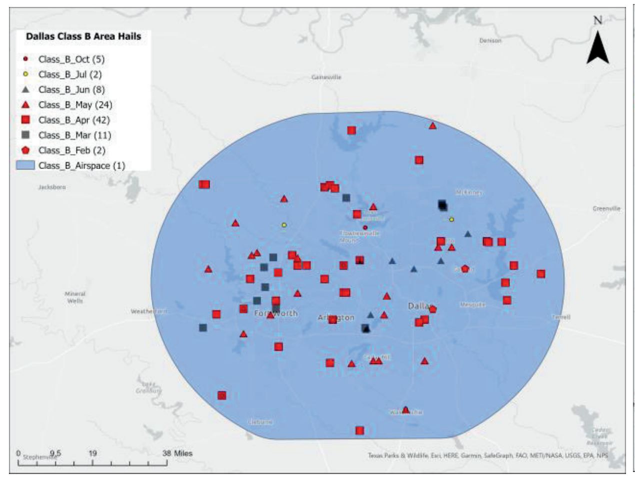

Figure 4: Dallas Class B Hail Events

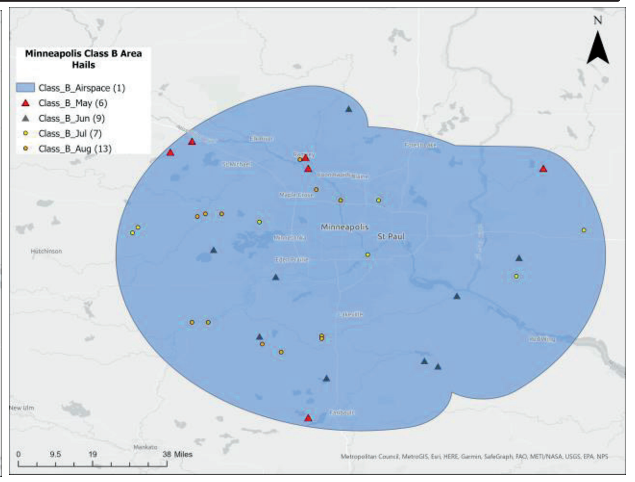

Figure 5: Minnesota Class B Hail Events

Figure 4 represents the geographical hail impact on Dallas Class B airspace in months. There were no hail events (>3 inches) observed during wintertime (January, November, and December). Most hail events recorded during April (42), May (24), and March (11). In contrast, **Figure 5** represents the geographical hail impact on Minnesota Class B airspace, indicated that there are only four months have recorded hail events with size greater than 3 inches: May (6), June (9), July (7), and August (13). Southern states have recorded severe hail events earlier in a year as compared to northern states.

Figure 6: Wichita Class C Hail Events

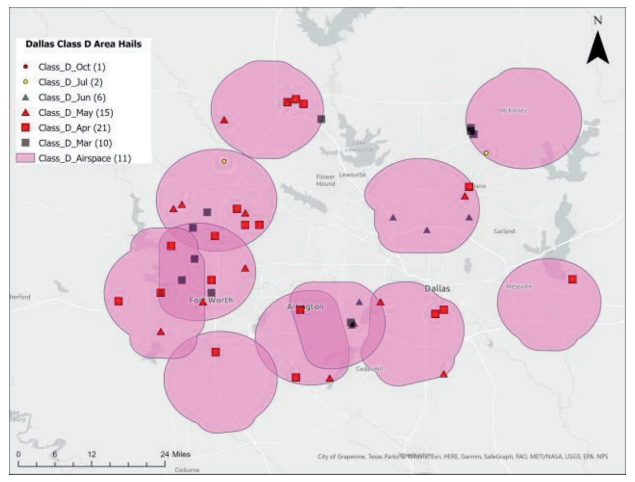

Figure 7: Dallas Class D Hail Events

Figure 6 represents the geographical hail impact on Wichita Class C airspace in months. Class C Airspace may be smaller than class B airspace. However, many essential international airports lay under class C airspace. Within Wichita Class C airspace, June have recorded the most hail events (9). **Figure 7** represents the geographical hail impact on Class D airspaces around Dallas area, The configuration of each Class D airspace area is individually tailored and when instrument procedures are published, the airspace will normally be designed to contain the procedures. Within 11 Class D airspaces around Dallas area, April have recorded the most severe hail events (21), May have followed the next with 15 recorded severe hail events.

Neural Network Hail Prediction Analysis

A total of three states: Dallas, Kansas, and Minnesota, have been tested for the hail longitude and latitude prediction. For each Machine Learning process, MATLAB provides the following graphs for visualization: Training State, Best Validation, and Error Histogram. All the predicted hail events are visualized using Esri's ArcGIS Pro based on latitude and longitude. In this part of the study, only hail event with hail size of larger than 1.5 inch were selected for the analysis.

The Training State (**Figure 8**) diagrams contain three main graphs:

- The G represents the direction and magnitude of the change that needs to be made to the model's parameters to reduce the loss function.
- The Best Validation (Mu) is used in conjunction with the gradient to adjust the weight update size dynamically during training. If the gradient is large, mu is decreased to take smaller steps, and if the gradient is small, mu is increased to take larger steps.
- The Validation Checks shows the number of validation checks performed during the training process. Validation Checks are used to monitor the performance of the network on a validation set and to prevent overfitting.

Besides, Training State, two additional figures will also be presented by the program: The best validation performance (**Figure 9**)graph shows how the training of the neural network is progressing over time, measured by the mean squared error (MSE) on the training, validation, and test sets. Error histogram (**Figure 10**) is used to visualize the distribution of errors in a prediction model. It shows the number of predictions that fall within certain ranges of errors. The 20 bins represent 20 different ranges of errors.

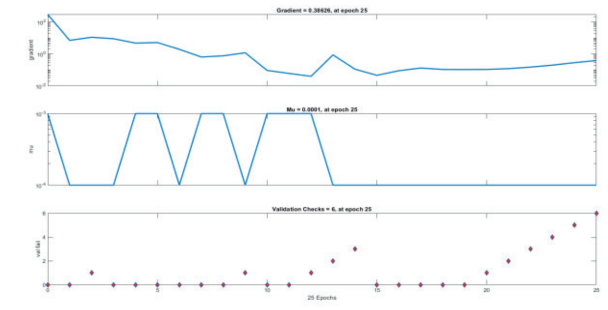

Figure 8: Texas Training State

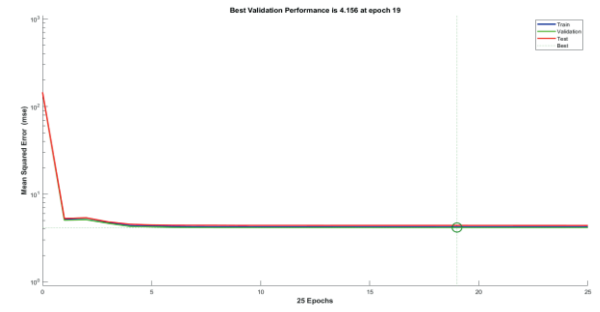

Figure 9: Texas Best Validation Performance

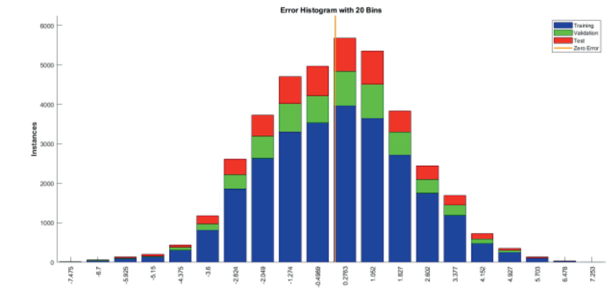

Figure 10: Texas Hail Error Histogram

State	Latitude RMSE	Longitude RMSE	Approx. Distance Variance
Texas	1.881513	2.222319	130-135 miles
Kansas	0.826516	1.951182	50-102 miles
Minnesota	1.374824	1.315647	57-94 miles

Table 1: Latitude/Longitude Difference

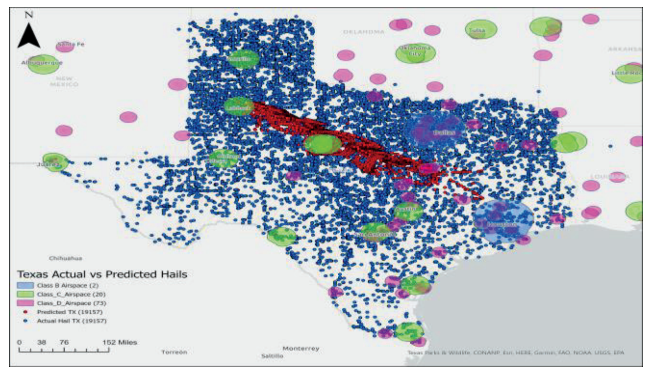

Figure 11: Texas Hail Prediction

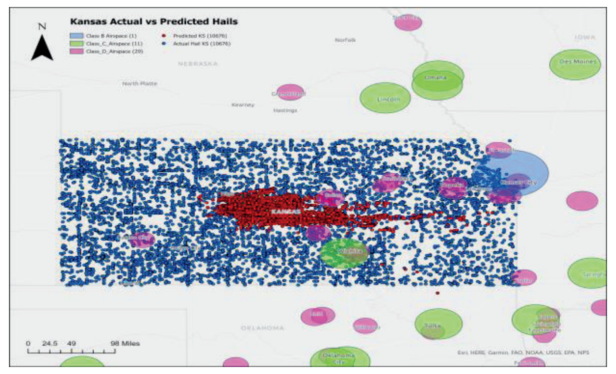

Figure 12: Kansas Hail Prediction

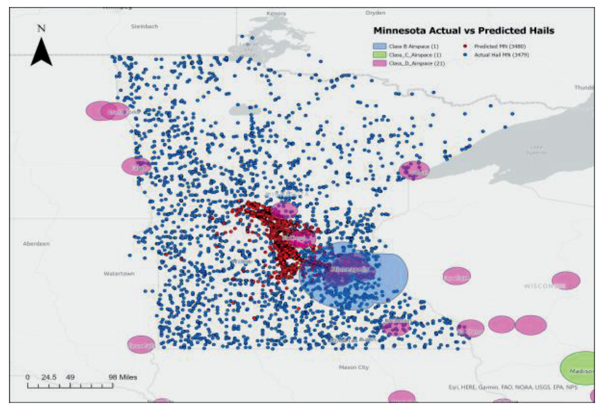

Figure 13: Minnesota Hail Prediction

Table 1 shows the Root Mean Square Error (RMSE) of the predicted hail events' latitude and longitude. After the conversion, the result showed that Kansas has the best distance variance via latitude (~50 miles) and Minnesota has the best distance variance via longitude (~ 94miles). **Figure 11, 12, and 13** shows the comparison between actual and predicted hail events in Texas, Kansas, and Minnesota. Most of the predicted hail events are located at the center of each state. One possible reason would be the Neural Network is trained based on historical events, including the locations and times where hail occurred. Based on the location pattern of historical events, the current Neural Network used in this study is likely to predict the hail event at the center. A better Neural Network algorithm is recommended in the future study to increase the accuracy of the prediction model.

Conclusion

The study analyzed the impact of hail events to the NAS in a geographical view. The geospatial analysis showed that central United States is the epicenter of severe hail events. Most hail events occurred from March to August every year. In addition, this study used Neural Network as a Machine Learning hail event forecast tool to predict the hail events. The result showed that most of the predicted hail events are located in the center of selected states.

For future analysis, more accurate Neural Network model is recommended. Additional patterns such as temperature, humidity, and wind shear can also be added to the analysis. In addition, using Cross-validation can help evaluate the performance of Neural Network analysis to get more accurate estimate of the outputs.