

# Forecasting Hurricane Formation and Meteorological Behavior using Artificial Intelligence Models

Patrick Milks

*Faculty of Arts and Science, Capilano University, North Vancouver, BC V7J3H5, Canada*

Completed December 12<sup>th</sup>, 2024

---

## Abstract

This research explores the application of transformer models in predicting hurricane genesis, trajectory, and intensity changes. A three-model system was developed to address different stages of a hurricane's lifecycle: (1) a hurricane genesis model to estimate the likelihood of formation, (2) an initial conditions model to predict post-formation characteristics, and (3) a trajectory model to forecast the storm's lifespan and intensity changes until dissipation or landfall. These models were designed to work in a seamlessly pipeline. The study leveraged the NOAA's HURDAT hurricane and the NOAA's AVHRR Pathfinder oceanic conditions databases as training datasets for the TabTransformer models, implemented in PyTorch. While the combined system yielded mixed results, individual models demonstrated mixed to high accuracy. Despite these promising findings, the research identified areas for improvement, including refined data preprocessing and addressing possible sources of error. Future work aims to enhance model performance, generating more accurate and reliable hurricane predictions. This research's results highlight the potential for further advancements in the use of transformers for weather forecasting applications and contribute to the growing body of literature on artificial intelligence in meteorology.

---

## 1. Introduction

Hurricanes are among the most destructive natural disasters. Hurricanes cause widespread devastation through powerful winds, torrential rainfall, and storm surges, leading to the destruction of homes, infrastructure, and critical services (Chavas et al. 2013). Their impacts include massive economic losses, displacement of communities, and loss of life, often disproportionately affecting vulnerable populations. Moreover, their behavior is being profoundly altered by climate change. Increasing global temperatures, rising oceans, and shifting atmospheric patterns have contributed to hurricanes becoming more intense, more frequent, and less predictable (Holland et al., 2014). These changes have expanded the temporal and geographical range of hurricane formation, increasing the risk to human lives, infrastructure, and economies. The growing variability in hurricane activity underscores the urgent need for innovative forecasting tools capable of addressing these climate-driven challenges.

Hurricane forecasting has experienced significant improvements thanks to advances in meteorological instrumentation, satellite technology, and computational methods. Early forecasting relied on rough physical models and broad mathematical formulations, while the advent of machine learning introduced more sophisticated methods for analyzing historical data. Current research has demonstrated the effectiveness of artificial intelligence with Convolutional Neural Networks (CNNs) handling tabular meteorological data and Recurrent Neural Networks (RNNs) processing satellite imagery effectively against hurricane behaviour. However, these methods largely rely on the presence of already-formed storms, relegating their utility from predicting hurricane genesis and early formation dynamics.

Transformers, first introduced in the *"Attention Is All You Need"* paper (Vaswani et al., 2023), have revolutionized deep learning by leveraging attention mechanisms to capture long-range dependencies in data. Initially designed for natural language processing, transformers have proven remarkably versatile, expanding into other domains such as time-series analysis and tabular data. Unlike CNNs and RNNs, transformers process entire sequences simultaneously. This parallelization enables much faster training times and the ability to handle much longer sequences than RNNs. The self-attention mechanism in transformers also enables the model to consider the entire data sequence simultaneously, eliminating the need for recurrence or hidden vectors. Instead, positional encoding maintains information about the position of each element in the sequence using new attention heads.

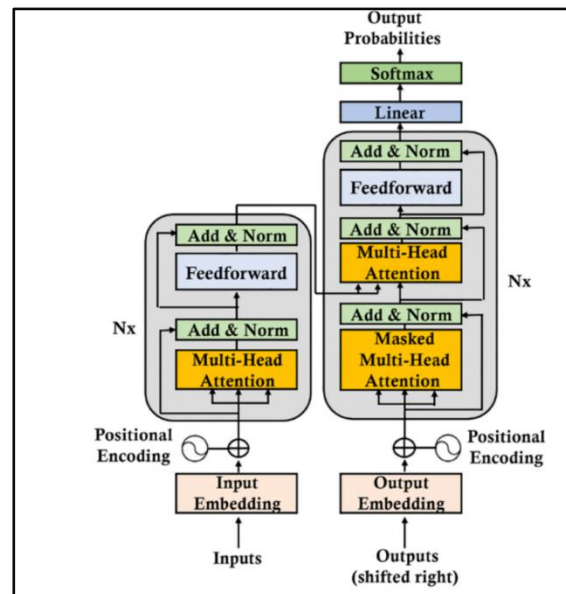


Figure 1: General transformer architecture

Accurate hurricane forecasting is of paramount importance, offering critical insights for disaster preparedness and policy-making. Early and precise predictions can inform evacuation plans, resource allocation, and infrastructure fortification, potentially saving lives and mitigating economic losses. Precise forecasting can enable policymakers to make informed seasonal decisions on disaster preparedness, resource allocation and infrastructure investment to mitigate potential impact. By leveraging transformers' capabilities in order to predict hurricane genesis, formation, and intensity directly from historical meteorological data, this project aims to bridge existing gaps in forecasting and provide a more robust tool for proactive disaster management.

This research outlines the development of a transformer-based model to analyze historical meteorological data and forecast hurricanes with increased accuracy as well as including genesis and early formation dynamics, offering a transformative approach to preparing for these increasingly unpredictable storms.

## 2. Research Question

This research hopes to answer the following question:

*Can modern machine learning architecture, transformers, accurately predict real-time and future hurricane formation, trajectory and intensity using historical meteorological and hurricane data?*

To address the research question, error rates will be evaluated against the current NHC NOAA hurricane error measurements for intensity and trajectory, as these are the metrics they monitor (“National Hurricane Center Forecast Verification,” n.d.). For this reason, the formation and birth models will not be included in determining the outcome of the research question.

### 2.1. Hypothesis

Null Hypothesis ( $H_0$ ): Transformers do not perform better than existing forecasting methods for predicting hurricanes, determined relative to the NOAA’s error trends.

Alternative Hypothesis ( $H_A$ ): Transformers perform better than existing forecasting methods for predicting hurricanes, determined relative to the NOAA’s error trends.

I hypothesize that the model will exhibit higher error margins compared to the current NHC NOAA error measurements, as their approach integrates multiple methodologies and accounts for a broader range of variables than the experimental models.

Outside the scope of the research question, I predict that the trajectory model will be more accurate than Alemany et al.'s, as transformers represent an advancement over the CNN framework utilized in their study and the input data used in both is largely the same (Alemany et al., 2019).

## 3. Literature Review

### I. Introduction

Hurricanes are intense storms characterized by powerful wind sheers, heavy rainfall and widespread flooding often resulting in major damage in coastal communities, devastating homes and businesses. Residents are forced to evacuate and those unable to do so put their lives in jeopardy. The impact of hurricanes can be lasting, leading to economic downturns, environmental disasters, long-term health consequences and increased marginalization of minority groups. Early detection systems are crucial to identifying the affected localities, determining the appropriate emergency response and issuing warning and alerts.

The modelling of hurricane formation and projection of their trajectories has been thoroughly researched for decades but it is a complex and challenging task due to the number of interrelated factors, the erratic and chaotic nature of weather forecasting and limited

observational data at the site of their births. Nonetheless, efforts dating back to the 1950s have been made to utilize mathematical and computational techniques to predict hurricanes (Dorst, 2007).

Today's modern approaches employ artificial intelligence systems, ranging from hybrid AI-assisted mathematical models to multilayered machine learning neural networks. With hurricanes trending to become more frequent and more devastating as a result of climate change, the need for a high accuracy prediction model is paramount for policymakers, emergency and relief services as well as residents.

This research will cover the current state of hurricane forecasting performed with artificial intelligence and the predictive accuracy of existing models.

#### *a. Current Research - Damage and Economic Estimation*

Much of the research performed with artificial intelligence related to hurricanes is done on assessing the storm's aftermath and economic damage rather than predicting its pre-landfall behaviour. This task has received much attention due to the extensive human labour required, the lengthiness of the assessment process and misidentification of disaster damage. Artificial intelligence for hurricane damage estimation has been used in two prevailing methods: 1) convolutional neural networks examining satellite imagery of impacted regions (Calton et al., 2021) and 2) data mining of damage statistics to reveal variable relationships through AI models such as Decision Trees, Naïve Bayes and Neural Network Clustering (Nawari, 2012).

While this is only adjacent to research performed on predicting hurricane behaviour, it provides important context as to where efforts are currently allocated and what presence AI has in the study of hurricanes.

#### *b. Neural Networks*

Moving more specifically into hurricane forecasting using AI, neural networks have been the prevailing machine learning model due to the multidimensionality of the learning data and their increased pattern and relation perception at deeper layers. More precisely researchers have landed on deep neural networks, which contain multilayered artificial neural networks between the input and output layers, allowing it to capture higher levels of patterns.

The following is a brief explanation of neural networks necessary to understand the specific techniques used in hurricane forecasting.

A simple, single-layered artificial neural network is represented in the above. Weights are the connections between nodes and carry a value. Biases are values assigned to all non-input layer nodes that attempt to capture unforeseen factors. Each neuron, a node within the hidden layers, has an activation function dependent on the input value, weights and bias. The inputs will either fail to meet activation requirements or achieve them and continue down onto a further layer with a new assigned value calculated based on the chosen activation function. The process is repeated until assigning values to the output nodes.

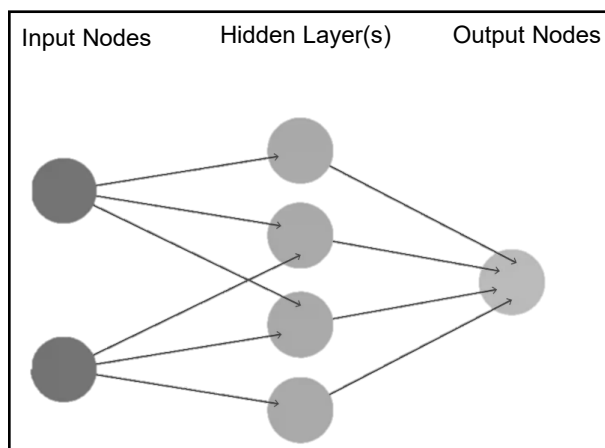


Figure 2: Simple artificial neural network with single hidden layer

Once the training process is completed and weights and biases have been assigned, the model is prepared to become predictive.

Three common classifications of deep neural networks that see the most use today:

- 1) Convolutional Neural Networks (CNN)
- 2) Recurrent Neural Networks (RNN)
- 3) Transformers

The first two have already seen use in hurricane forecasting.

### c. Convolutional Neural Networks

Convolutional Neural Networks are a class of deep neural networks, typically utilized in computer vision. The AI system automatically extracts features from images for specific tasks such as image classification and face authentication. Different layers do so by executing convolution operations, each performing nonlinear activation functions on narrower image subsets determined by the outputs of previous layers.

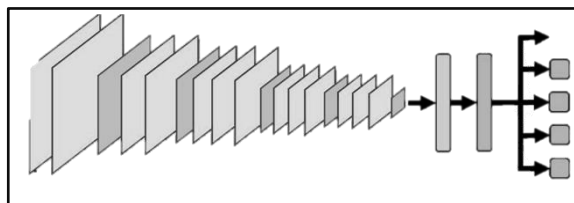


Figure 3: Convolution Neural Network architecture with convolution and pooling layers

Between convolution layers, there are pooling layers that map features into smaller regions. CNNs are feedforward neural networks, meaning

information only moves forward through hidden nodes. Their performance has steadily improved with certain models such as ResNet surpassing the 5% error rate of human vision (Yin et al., 2017).

CNNs coupled with satellite imagery have been used in identifying hurricane trajectories, particularly of their post-landfall movement through the remaining path of damage and to determine the storm's intensity (Guo, 2021).

The issue with using CNNs to explain hurricane characteristics is their reliance on satellite imagery. Imaging storms only provides data on storms that have formed to a significant enough extent to be identifiable from orbit. This could omit crucial information provided from smaller tropical storms while they either transition into hurricanes or die out. While they have a powerful capacity to map out trajectories from past storms, which can be used predictively in models. CNN in this capacity also require a standardization of satellite imagery along a specific pitch and geolocation for pattern to be accurately recognized. Most essentially, however, in relevance to the research question posed, CNNs are impractical in predicting hurricane births as the meteorological precursors are not visible.

#### d. Recurrent Neural Networks

Recurrent Neural Networks are nonlinear dynamic deep neural networks used to represent complex sequential relationships between variables, like in spatiotemporal processes. Due to their flexibility with such data and other data with temporal dependencies, RNNs have been tasked with natural language processing including language modelling, sentiment analysis and music generation. RNNs are fully connected and feed results back into the network, dissimilar to CNNs. The input of RNNs consist of the current input as well as that of previous samples, meaning the connections between nodes form a directed graph along a temporal sequence. Furthermore, neurons store internal memories of the computation history from previous samples. Different RNN models have been constructed to specify neuron memory length and directionality of dependencies according to their task (Yin et al., 2017).

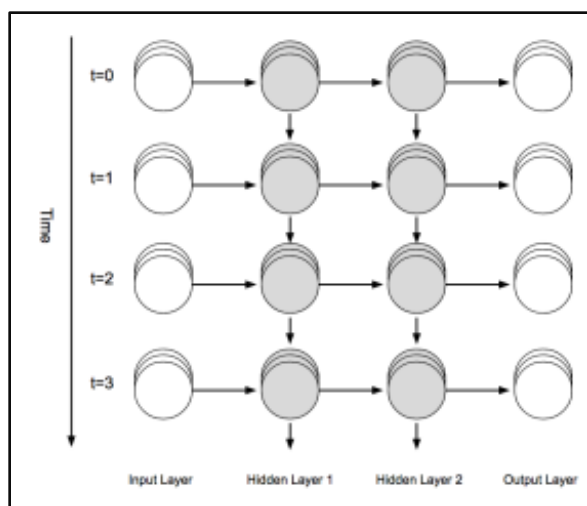


Figure 4: Simple Recurrent Neural Network architecture with hidden layers

RNN models have been used with the specific database intended to be employed by this research and the 6-hour incremental hurricane data provided sequential temporal data. This made parameters relative rather than absolute. To analyze hurricane behaviour, RNNs are trained on a grid model, distinguished by geographic coordinates, to track hurricane trajectories. Researchers have also selected specific hyperparameters for hurricane predicting RNNs, choosing the number of hidden layers and employing particular nodes (Alemany et al., 2019).

Research done on RNN models in conjunction with hurricane prediction is much more extensive than with CNNs and is the consensus model for predicting hurricane movement. Work previously done confirms the feasibility of this research with the NOAA HURDAT database and provides a solid basis for predicting hurricane births. It also demonstrates to what extent such models are capable of being accurate and that errors arise in how the data is parsed and passed into the model. Crucially however, RNN's suffer from scalability issues and the vanishing gradient problem, which describes a model's inability to properly capture long-range dependencies due to exponentially decreasing changes in weight during backpropagation (Hochreiter, 1998). These shortcomings can make RNN's unable to fully grasp relationships between variables.

#### *e. Transformers*

Transformers, introduced in the seminal paper "*Attention is All You Need*" by Vaswani et al., revolutionized machine learning by addressing limitations of earlier models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Unlike CNNs, which are primarily designed for spatial data like images, and RNNs, which can handle sequential data processing but suffers from the vanishing gradient problem and poor scalability, transformers rely entirely on the attention mechanism to model relationships between input data elements.

At the core of the transformer architecture is the self-attention mechanism, which enables the model to weigh the importance of each element in the input sequence when training. This mechanism is implemented in layers that contain multi-head attention neural networks. Multi-head attention allows the model to focus on different aspects of the input data simultaneously, capturing relationships and dependencies that other frameworks might struggle to represent, especially for long-range dependencies in sequential data, like time series (Zimerman and Wolf, 2023). The parallelization of transformer processing, handling entire input sequences at once, also makes them significantly faster to train, making the amount of data they can handle indirectly greater.

Transformers are now widely used across domains, including natural language processing, computer vision, and even weather forecasting. Hittawe et al. developed a transformer model to predict Red Sea conditions from historical data that outperformed existing models with an  $R^2$  surpassing 99 (Hittawe et al., 2024), Nguyen showed their efficacy and reliability in medium-range weather forecasting, outperforming current methods beyond 7 days (Nguyen et al., 2024), and researchers like Arifin et al. (Arifin et al., 2024) and Hasan (Hasan, 2024) have used transformer models in meteorological forecasting relevant to agriculture and climate science.

### *f. TabTransformer*

An existing transformer architecture that could train off historical NOAA data is the TabTransformer. The TabTransformer in PyTorch, implemented by “lucidrain” (Huang et al., 2020), is a transformer-based model tailored for tabular data. This architecture consists of a column embedding layer, multiple transformer layers, and a final multi-layer perceptron (MLP). It handles both categorical and continuous features efficiently. TabTransformer models can have their parameters specified like the number of continuous or categorical features, binary or continuous outputs, model dimensions and other hyperparameters such as the number of transformer layers, attention heads and the learning and dropout rate. Huang et al.’s paper recommends a variety of parameter settings for general use.

## **II. Conclusion**

Upon conducting research on the current state of AI-enabled hurricane prediction models, many aspects of my research question were answered while others were unresolved. It was found that AI has been used to predict weather patterns and hurricane trajectories, whereas other characteristics such as anticipating intensity changes and forecasting hurricane births was touched upon less in published work.

Research also showed that the preferred AI model used by experimenters was a deep neural network model for its efficiency in handling complex multidimensional problems with chaotic underlying relationships, specifically recurrent neural networks that specialize in sequential temporal data. Many modern projects utilize the NOAA’s HURDAT database for their hurricane data and plot hurricane movement along an oceanic geolocation coordinate grid. However, none were found to integrate with historical sea surface temperature data, like that from the NOAA’s AVHRR Pathfinder. This may provide a new avenue for hurricane analysis particularly on what constitutes a hurricane to be born. Published papers also showed that the computational power to train such models is reasonable for any researchers and that shallow depths of learning can be sufficient.

As transformers have already shown their effectiveness in weather forecasting tasks, and previous neural networks have proven successful in hurricane prediction, a logical next step is to test transformers, such as the TabTransformer’s, performance in hurricane forecasting. It also provides an opportunity to conduct experimentation on predicting hurricane births as well as their trajectory and intensity changes.

---

## **4. Research Design & Methodology**

This project aims to implement a three-model transformer-based system to improve hurricane forecasting capabilities, focusing on predicting hurricane genesis, estimating initial conditions, and forecasting trajectories. Before explaining the methodologies for these models, it is essential to outline the approach to parsing and cleaning the training data.



The first data source will be the NOAA's HURDAT database, which contains records of all observed hurricanes from the late 1880s to 2023. This dataset was used in other forecasting research like Alemany et al.'s. This data will be parsed into Hurricane class objects, each containing detailed attributes of the storm's lifecycle. Among these attributes will be held all recorded entries of the hurricane, stored as Entry class objects. Each entry will provide key details at six-hour intervals unless the storm exhibits noteworthy behavior, tracking of its behaviour. Complementary sea surface and meteorological data will be obtained from NOAA's AVHRR Pathfinder database, which offers daily daytime and nighttime measurements from 1981 through 2023. To align this data with hurricane activity, a custom web-scraping script will extract relevant files corresponding to active hurricane days, defined as the earliest and latest yearly entries recorded in the HURDAT database, found to be between the 2<sup>nd</sup> of June and the 25<sup>th</sup> of December. This approach ensures that only pertinent ocean and weather conditions are analyzed.

A series of preprocessing functions will be performed on the meteorological data, including linearly interpolating missing temperature values, specifying longitude-latitude grid intervals, and filtering out non-oceanic data. These steps aim to reduce computational load while preserving the quality and relevance of the data. Any hurricane or meteorological records with incomplete information will be excluded to maintain dataset integrity. The data will also be filtered to include only hurricane-active determined regions of the ocean, focusing resources on areas where hurricanes are most likely to form and habituate, found to be within the latitude and longitudes of 7N-70N and 137W-14E.

The cleaned and preprocessed data will be partitioned to train each model effectively. For the hurricane genesis model, a grid will be created for each day, with latitude and longitude points marked as 1s for locations with hurricane birth and 0s elsewhere. These binary values will serve as the model's output. For the initial conditions model, the first recorded entry of each hurricane will be extracted to define the "birth conditions," which will be supplemented with complementary sea surface and weather conditions at the time and location of their genesis. Lastly, for the trajectory forecasting model, all consecutive hurricane entries will be used to predict the storm's next step also tied with sea conditions at both entry locations. Models with binary outputs versus continuous outputs and multivariate versus single-variable inputs will require tailored transformer parameters, such as loss functions, embedding sizes, and attention head configurations, to optimize performance for their specific prediction tasks.

All three models will be built using lucidrains' PyTorch Tabular Transformer (Huang et al., 2020), which processes multivariate tabular data with temporal and spatial dependencies. A 65%-35% training-to-testing data split will be employed, adhering to standard practices (Pawluszek-Filipiak and Borkowski, 2020). Each model will include error validation during every training epoch to ensure robust performance. The three models will work in tandem to create a continuous timeline of potential and existing storms using available data. For parameter such as embedding dimension, depth, number of heads, attention dropout and feed forward dropout rate, Huang et al.'s recommended settings are used.

The hurricane genesis model will be evaluated according to the binary cross-entropy loss function, measuring the difference between predicted probabilities and actual labels. Specifically, it calculates the negative log likelihood of the true labels given the predicted probabilities, incrementally penalizing predictions the farther they are from actual values. Its formula is the following:

$$L = -\frac{1}{N} \sum_{i=1}^N y_i \log(p_i) + (1 - y_i) \log(1 - p_i)$$

*Equation 1: Binary cross-entropy loss function*

where:

- $N$  is the number of data points,
- $y_i$  is the actual binary label,
- $p_i$  is the predicted probability of a 1 (i.e. hurricane birth)
- 

The two regression models, the initial conditions and trajectory forecasting models, alternatively use the mean squared loss function. It measures the average squared difference between the predicted values and actual values. The squared differences ensure that larger errors are penalized more heavily, encouraging the model to focus on minimizing larger deviations. The formula for MSE is:

$$L = \frac{1}{N} \sum_{i=1}^N (y_i - y_p)^2$$

*Equation 2: Mean squared loss function*

where:

- $N$  is the number of data points,
- $y_i$  is the actual value,
- $y_p$  is the predicted value

A standard scaler from sci-kit was used to standardize the inputs of all models and for the outputs of models that were multivariate and continuous like those of the trajectory and initial conditions models. When a scaler was involved, it had to also be used to “unscale” the outputs of individual predictions.

Link to entire model repositories: <https://github.com/pmilks/SCI400-Project-Milks>

### 5. Results

The performance of the models was mixed. For the trajectory model, tracking a hurricane’s movements and change in characteristics over a single 6-hour interval, the transformer’s loss model showed a logarithmic improvement over training epochs with a scaled mean squared error loss tending towards ~0.04.

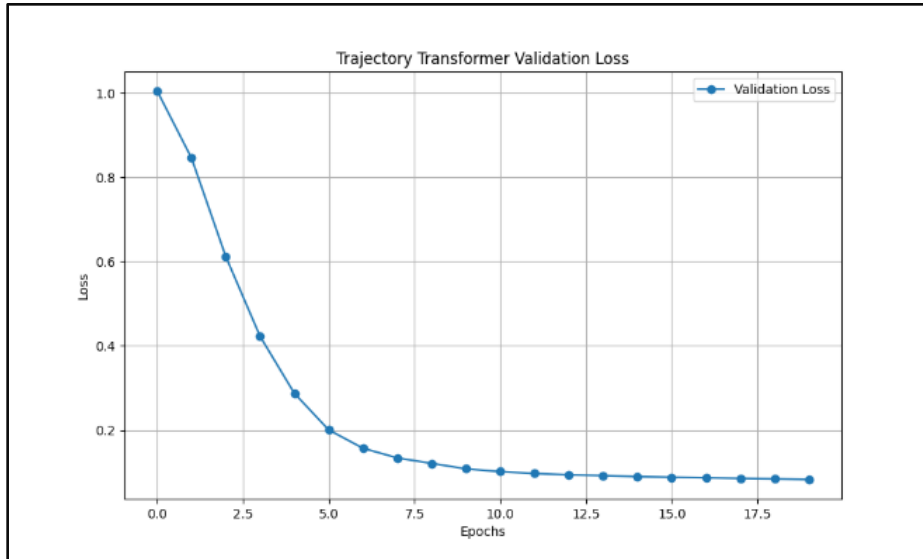


Figure 5: Trajectory Transformer's loss validation

The initial conditions model also exhibited a positive logarithmic learning curve through its mean squared error loss data but tended towards a much higher ~0.92 over 20 epochs. As this data is also standardized, a loss of 0.92 is highly significant.

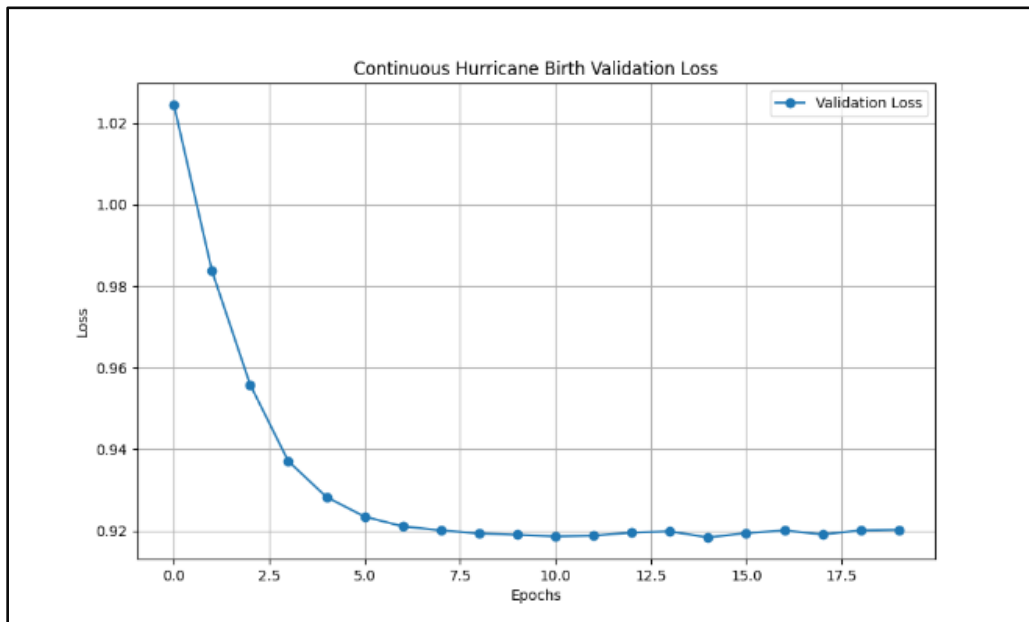


Figure 6: Initial Conditions Transformer's loss validation

As for the third and final, birth model, that used a binary cross-entropy function, its loss data performed differently than the previous two. The loss tended towards between a lower 0.04 and 0.05 but did not exhibit the same logarithmic learning curve. Additional epochs also did not necessarily tend to a more accurate model with occasional increases in loss over generations.

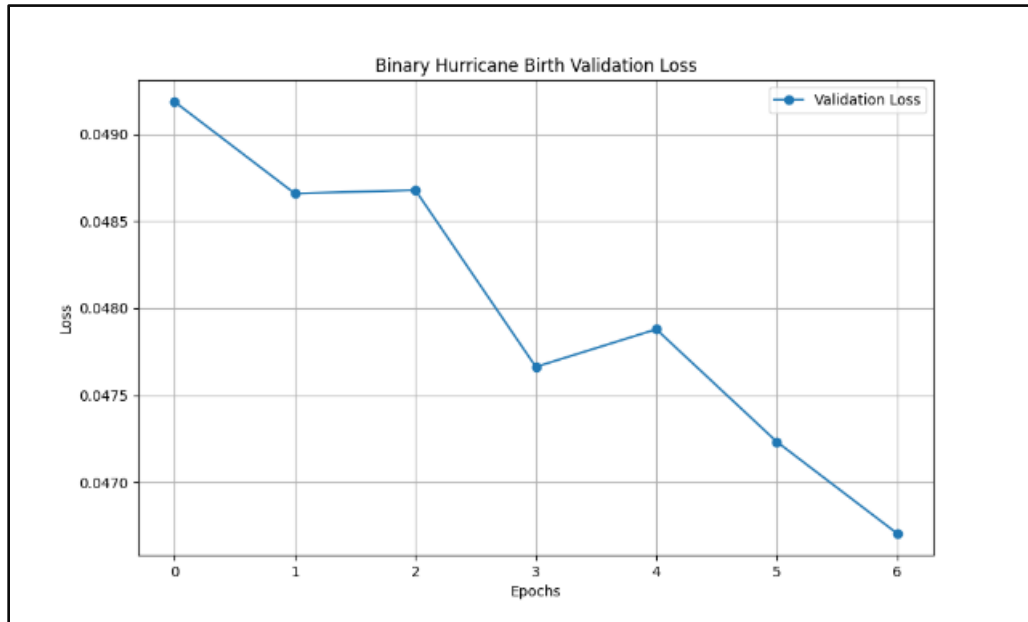


Figure 7: Hurricane Genesis loss validation

When the following three models were tested wholly to determine whether a predicted hurricane could mimic the birth, initial conditions and lifecycle of its corresponding actual hurricane, meaningful variance was observed. Hurricane Nicole in 2022 was chosen as an arbitrary comparison due to its recency, meaning it had complete data, and its singular landfall.

The first step was to determine the likelihood of Nicole’s formation on November 6<sup>th</sup>, 2022, given that date oceanic data. A probability heatmap on the day claimed Nicole’s formation was unlikely relative to other oceanic locations on the day and the averaged probability of Nicole’s genesis was ~0.4%. The November 6<sup>th</sup>, 2022, hurricane genesis probability heatmap demonstrates zones of higher probability in the warmer waters near shores, within the Gulf of Mexico and sparsely through out the Atlantic Ocean. There are aberrant data points with values

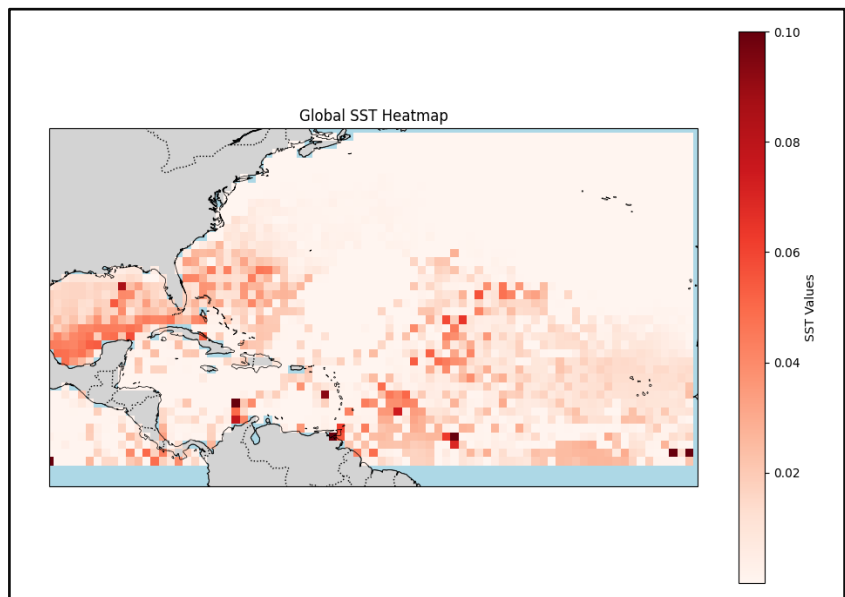


Figure 8: Atlantic heatmap of hurricane genesis probability estimated by the initial conditions model

far exceeding the day’s norm and others with values differing dramatically from its neighbouring locations. Another concern is that model seems to uniformly overestimate the probability of hurricane formation with this day’s heatmap alone predicting multiple hurricanes.

Secondly, using Hurricane Nicole’s real genesis location’s oceanic data, the initial conditions model predicted its birth characteristics.

	Max. Sustained Winds (mph)	Min. Pressure (mbar)	Hurricane Force Winds Radius (nm)
Actual	30	1005	100
Predicted	29.37	1008.61	67.8
Difference	-0.63	3.61	-32.2

Table 1: Differences in predicted and actual conditions of Hurricane Nicole using the initial conditions model

The average initial conditions of a predicted Hurricane Nicole, according to its genesis location’s oceanic data, was roughly accurate when determining the maximum wind speed and minimum pressure but drastically underestimated the hurricane force winds radius (HFWR). Notably, the predicted value for the HFWR was below the minimum HURDAT measurement, meaning the predicted Nicole would avoid hurricane classification altogether.

With Nicole’s first initial predicted conditions, a continuous looping of the trajectory model was used to predict Nicole’s lifespan until first landfall, a location with no oceanic data, or on her dissipation. A hurricane was considered dead when two of the maximum sustained winds, minimum pressure and HFWR were below HURDAT hurricanes corresponding average final reading. For maximum sustained wind speed this was 31.32mph, while minimum pressure and hurricane force winds radius were 1002.82mbar and 67.89nm respectively. Starting with the initial predicted conditions, each new predicted change in trajectory and intensity was used as the input for the following prediction.

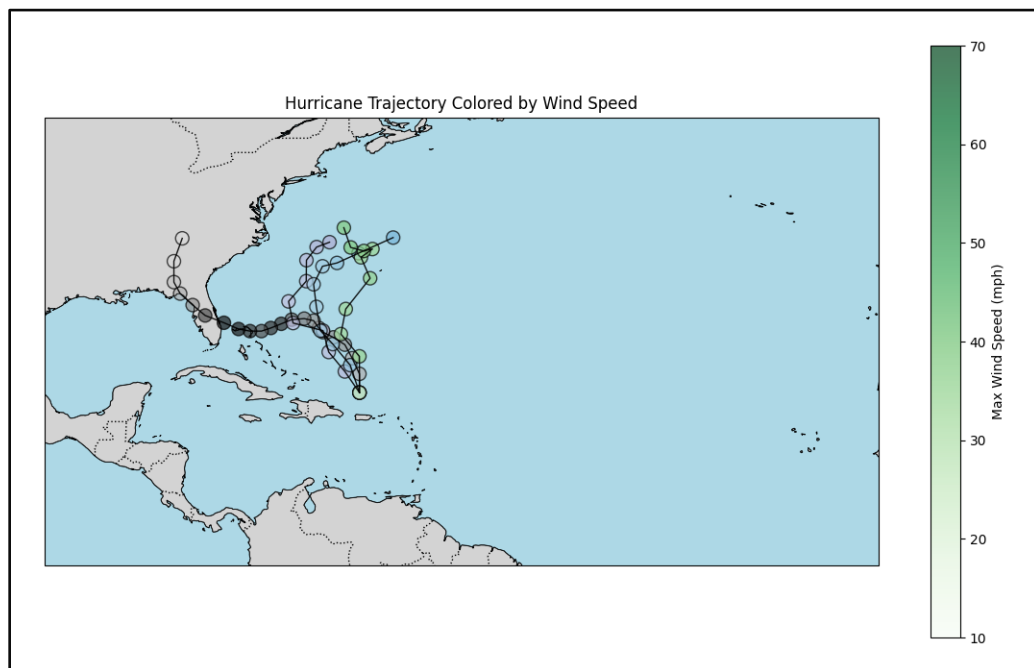


Figure 9: Sample of three predicted paths of Hurricane Nicole estimated with consecutive use of the trajectory model with their intensity indicated with color boldness

The trajectory model had difficulty predicting paths similar to that of the actual Hurricane Nicole. Over numerous tests, predicted Nicoles roughly matched actual Nicole in her initial movements, but would eventually break off. While individual tests varied drastically with many dissipating within the first few steps and others refusing to do so after consecutive entries where they lay motionless at sea, no predict Hurricane Nicole made landfall. Predicted Hurricane Nicoles also failed to experience significant increases or decreases in their intensity measurements, stalling and staying near their initial conditions.

	Date	Lat	Lon	SST	SSTd	hurMaxWind	hurMinPressure	hurRadius	Status
0	2022-11-06 12:00:00	20.600000	-66.800000	301.71	0.5	29.370000	1008.610000	98.420000	0
1	2022-11-06 20:00:00	22.099034	-68.545113	297.83	-3.3	32.174787	1007.277256	100.013557	A
2	2022-11-07 04:00:00	23.665636	-69.553129	299.94	-1.2	31.456520	1007.355943	107.698910	A
3	2022-11-07 12:00:00	25.535724	-70.317293	287.50	12.6	31.038106	1006.874256	114.262138	A
4	2022-11-07 20:00:00	25.591341	-73.368493	295.46	-4.6	33.382137	1004.332194	102.830290	A
5	2022-11-08 04:00:00	28.124120	-75.253462	271.72	-2.6	31.707246	1004.857941	113.478772	A
6	2022-11-08 12:00:00	29.156220	-70.940630	295.15	-3.7	35.182640	1002.932632	103.632372	A
7	2022-11-08 20:00:00	31.459520	-74.568351	274.03	1.7	31.170185	1002.730724	117.634101	D

Table 2: Entries of a sampled predicted Nicole's lifespan of 7 steps using the trajectory model

From a sample of three path of predicted Hurricane Nicoles, the absolute error in latitude-longitude degrees steadily increases as more entries are generated, indicating a possible compounding effect of initial errors or a broader error in the data processing or training method.

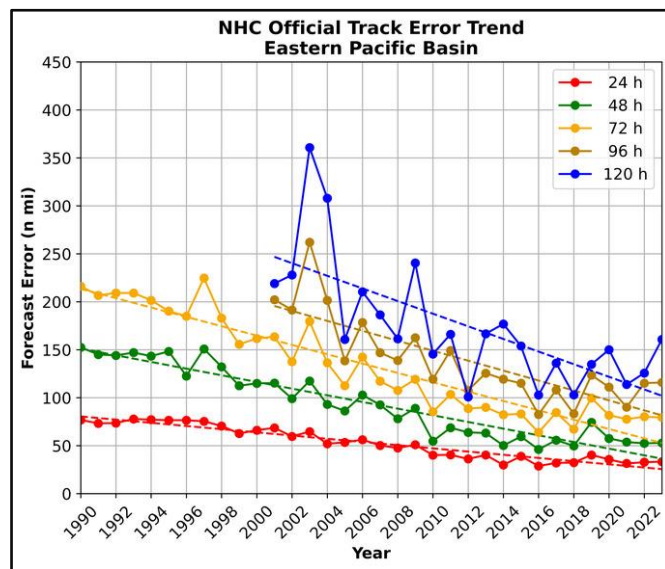
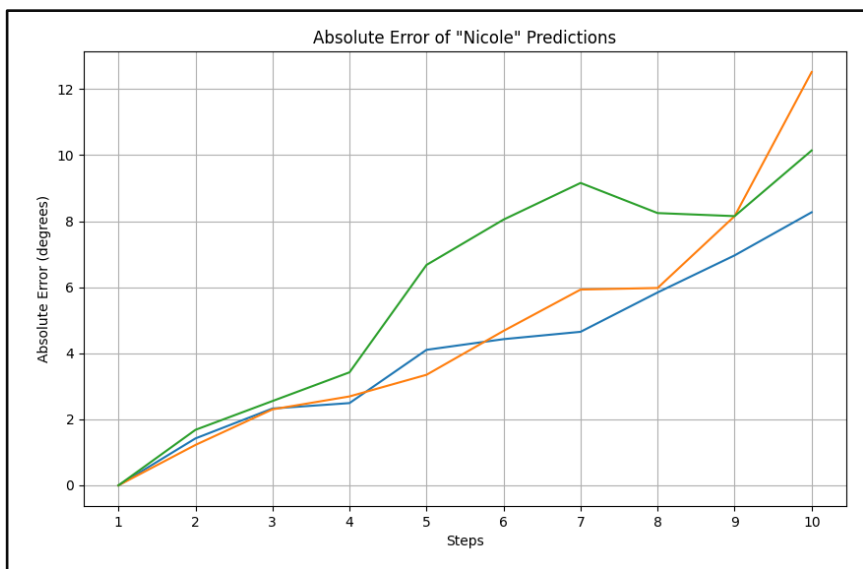


Figure 11: Absolute error of predicted Hurricane Nicole paths measured in absolute degrees

Figure 10: NHC Official Historical Tracking Error Trends over different forecast ranges, measured in miles ("National Hurricane Center Forecast Verification.", n.d.)

When comparing the absolute tracking errors of the predicted Nicoles and the NHC's Official tracking error trends, the average tracking errors from the trajectory model at 24 hours and 48 hours surpass the NHC's current tracking errors even if the longitudinal-latitude absolute degrees of separation are converted to their minimum mile value, calculated as the longitudinal difference at the highest expected latitude, 70N. If the deviation is measured at this minimum scenario, 1 degree is roughly equal to 23.65 miles.

Formula for longitudinal degrees-to-miles conversion, relative to latitude:

$$Distance = \cos(Lat_{rads}) \cdot \frac{Circumference_{Earth}}{360}$$

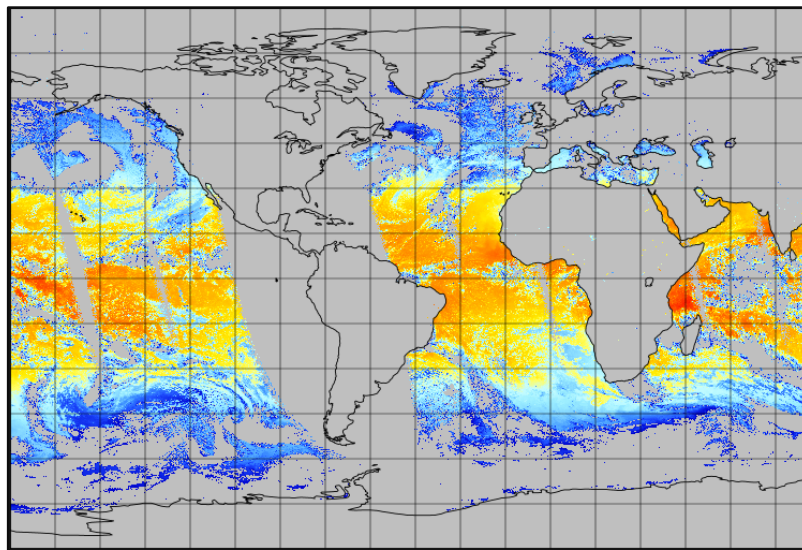
*Equation 3: Longitudinal degree to mile conversion according to latitude*

At 24 hours, or step 4, the predicted storms tended towards ~3 degrees of error, equating to a minimum deviation of ~70.95 miles, above the NHC's 24-hour ~40 mile tracking error. The same is the case at 48 hours, or step 8, with the trajectory model's and NHC's tracking errors being 165.55 miles and ~50 miles respectively. As the tracking errors of predicted Hurricane Nicoles grow continuously at a higher rate than the NHC's tracking errors, the trajectory model trends to continue being less accurate than the NHC's tracking errors as forecasting period increases.

## 6. Conclusion

The implementation of the three hurricane prediction models, hurricane genesis, initial conditions, and trajectory, revealed both strengths and significant limitations. While the models were designed to work consecutively, the outcomes indicate that integrating their predictions into a single pipeline was not successful in achieving accurate, reliable outputs.

Individually, the hurricane genesis model showed notable promise in identifying regions with higher or lower likelihoods of hurricane formation. This capability demonstrates that the model successfully learned broad patterns from the training data, suggesting a degree of robustness in identifying key meteorological indicators. However, the presence of aberrant data points, deviating heavily from neighboring predictions or reporting doubtfully high probabilities, highlighted inconsistencies. A probable explanation for these anomalies could lie in the interpolated temperature data used to fill "dead zones" of the AVHRR Pathfinder satellite. Such dead zones,



*Figure 12: AVHRR Pathfinder SST data for December 30th, 2022, with broad regions of missing data*

caused by intermittent satellite visibility, can create discontinuities confusing the model and producing erroneous predictions. Moreover, the model exhibited a consistent overestimation of hurricane genesis probabilities. This bias may stem from the inherent structure and distribution of the training data. The training process involved iterating over the entire oceanic region daily and associating hurricane births with grid cells limited by 1-degree intervals. By assigning hurricane genesis conditions to an entire grid cell spanning hundreds of square miles, the model may have spread the range of favorable conditions. This oversimplification could result in the model assigning higher probabilities to regions that do not meet the necessary conditions for hurricane formation.

To address this limitation, further experimentation with finer spatial resolutions is necessary. Reducing the grid intervals to fractions of a degree could improve the data, allowing the model to learn more local relationships and reduce overestimation errors. Such adjustments would require more computational resources but could yield improvements in predictive accuracy.

Another notable finding was the model's tendency to favor regions adjacent to shorelines for hurricane genesis. While hurricanes occasionally form close to coastal areas, they typically develop in deeper ocean waters where favorable conditions are more likely to persist. Incorporating oceanic depth data could help the model better understand the oceanographic constraints of hurricane formation. Despite these limitations, the hurricane genesis model still holds potential as a tool for identifying regions of interest where hurricane formation is more likely. By flagging such regions, the model could serve as an early-warning system, prompting further analysis using established models or observational data to confirm or refine predictions.

The performance of the initial conditions model reveals several areas that require adjustments and refinement to improve its predictive accuracy. At present, it is challenging to determine the model's success. A critical issue lies in the use of the mean squared error (MSE) loss function to evaluate the model. For the initial conditions model, the inputs and outputs were scaled to account for magnitude differences among variables. However, the scaling method used may not have been optimal for this application. A min-max feature scaler would have been more appropriate, as it rescales the data and preserves the relationships between minimum and maximum values. In the context of predicting hurricane's first conditions, this is essential for capturing the lower threshold values of determinants, critical for a system to be classified as a storm. By ensuring that minimum values are represented proportionally, a min-max scaler could improve the model's ability to identify early-stage hurricanes.

Another significant challenge stems from the training dataset, which was the same as that used for the hurricane genesis model. As previously discussed, this dataset likely introduced biases due to the imbalance between hurricane birth and non-birth conditions. This imbalance may have skewed the model's understanding of the conditions necessary for hurricane formation, leading to inaccuracies. Additionally, the inclusion of the hurricane-force wind radius (HFWR) parameter likely posed an issue. It has only recently been incorporated into the HURDAT database and, consequently, HFWR values are absent in older hurricane data. However, even for modern hurricanes, during a hurricane's transitional stage between tropical storm and hurricane, either early or late in its life, there are no hurricane force winds, meaning the entries were omitted in an effort to have complete datasets. The inclusion of early transition stage entries would be crucial in best training a dataset focused on determining what a hurricane's first conditions might be. The omission these transitional storms may have reduced



the model's ability to accurately capture relationships between early-stage hurricanes and their determinants.

The predicted initial conditions during individual tests highlighted this issue. While the model predicted maximum wind speed and minimum pressure moderately well, its performance on HFWR was significantly poorer. This discrepancy suggests that the HFWR parameter introduced noise into the training process. Removing this variable altogether may allow the model to focus on better-defined features, improving prediction accuracy. Ideally this would be reflected by shift the model's loss validation's curve downwards, maintaining the logarithmic improvement indicating growth while lowering the floor towards scaled zero. In light of this analysis, a retraining of the model without the HFWR variable using a more reflective dataset and min-max scaler is recommended to assess its performance and applicability.

The performance of the trajectory model initially seemed the most promising. The mean squared error (MSE) loss graph indicated effective learning, with an exponential decrease in error tending toward a stable and minute minimum. The model, according to this metric, even demonstrated superior accuracy in single-step predictions when compared to the grid-based RNN trajectory tracking model developed by Alemany et al.

	Mean Squared Loss
Grid-Based RNN ( <i>Alemany et al.</i> )	0.07447
Transformer Trajectory Model ( <i>Milks</i> )	0.04589

*Table 3: Compared between mean squared loss between Alemany et al.'s Grid-Based RNN and this research's Trajectory Transformer model both predicting hurricane trajectory*

However, significant limitations became evident when the trajectory model was employed consecutively to predict a hurricane's entire lifespan. While initial predictions showed high accuracy, errors accumulated steadily with each consecutive step, resulting in an absolute tracking error consistently exceeding current National Hurricane Center (NHC) error margins. Another critical limitation was the model's inability to simulate hurricane intensity changes effectively. Weak storms frequently dissipated early, rarely achieving growth into stronger hurricanes or stagnated for steps far beyond the average hurricane's lifespan, exhibiting no significant movements or change in intensity. Furthermore, storms predicted by the model never made landfall. A plausible explanation for this issue lies in the training data. The model was trained exclusively on oceanic data, as landfall entries lack the sea surface data necessary for feature association. Consequently, a hurricane's first landfall was designated as their final entries, neglecting scenarios where storms re-emerge over open waters or made multiple landfalls across different landmasses. This omission likely created a bias where the model indirectly prioritized hurricane persistence over water, failing to recognize or predict landfall events. To address these limitations, a reworking of the methodology would be necessary to incorporate landfall data.

The inclusion of the hurricane-force wind radius (HFWR) parameter, for the same reasons as discussed in the initial conditions model, may also have contributed to inaccuracies in the trajectory model. Removing this variable from the training process could eliminate nonrepresentative dependencies and correct relationships between other variables and changes in hurricanes. A further likely source of error was in the designation of storm death,

which occurred when two of maximum wind speed, minimum pressure, and HFWR fell below their threshold of predefined minimum hurricane values. This approach contradicts the NHC standard of relying predominantly on maximum sustained wind speed for hurricane classification. Adjustments to these thresholds and a more nuanced definition of storm termination would require both more testing and input from meteorological expertise but could address the issue of pre-emptive storm dissipation and stagnation. Furthermore, separate evaluations of the model's performance near land, over open oceans, and across varying storm strengths would be essential to assess its accuracy comprehensively. Incorporating atmospheric pressure data into the models input data would likely lead to improvements in their predictive accuracy, as pressure is a key indicator of hurricane formation, intensification, and dissipation. Its inclusion could address relational gaps that are present in the models and provide context to improve others. At the time of the making of this research no such historical global atmospheric pressure database exists.

In conclusion, due to errors revealed during repeated use of the trajectory model exceeding current NOAA margins, this research fails to reject the null hypothesis.

While the results of these models are mixed, their development, alongside other research in meteorology leveraging transformers, demonstrates that transformer-based architectures hold considerable promise in predicting complex meteorological events, including natural disasters like hurricanes. Transformers excel at capturing temporal and spatial relationships within vast datasets, a critical advantage for forecasting highly dynamic and chaotic systems such as hurricanes. As climate change accelerates, reshaping the meteorological landscape faster than humans can adapt, the demand for cutting-edge neural network approaches grows ever more urgent. These models, particularly transformers, offer the flexibility and computational power to analyze evolving patterns that may otherwise escape conventional statistical techniques. Their use in natural disaster prediction, especially in the case of hurricanes, could help keep more people safe.

## **7. Acknowledgments**

I would like to express my gratitude to Capilano University for providing the staff, resources, and support that made this project possible. Special thanks go to Professor Jason Madar for his invaluable guidance on artificial intelligence models and transformers. I would also like to extend my appreciation to the NOAA, whose open-source databases and ongoing maintenance enable research in this field.

## 8. Sources

- Aleman, Sheila, Jonathan Beltran, Adrian Perez, and Sam Ganzfried. 2019. "Predicting Hurricane Trajectories Using a Recurrent Neural Network." *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (01): 468–75. <https://doi.org/10.1609/aaai.v33i01.3301468>.
- Arifin, Yulyani, Ilvico Sonata, Maryani, and Elizabeth Paskahlia Gunawan. 2024. "Weather Prediction in Agriculture Yields with Transformer Model." *Procedia Computer Science*, 9th International Conference on Computer Science and Computational Intelligence 2024 (ICCS CI 2024), 245 (January):750–58. <https://doi.org/10.1016/j.procs.2024.10.301>.
- Calton, Landon, and Zhangping Wei. 2022. "Using Artificial Neural Network Models to Assess Hurricane Damage through Transfer Learning." *Applied Sciences* 12 (3): 1466. <https://doi.org/10.3390/app12031466>.
- Chavas, Daniel, Emmi Yonekura, Christina Karamperidou, Nicholas Cavanaugh, and Katherine Serafin. 2013. "U.S. Hurricanes and Economic Damage: Extreme Value Perspective." *Natural Hazards Review* 14 (4): 237–46. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000102](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000102).
- Dorst, Neal M. 2007. "The National Hurricane Research Project: 50 Years of Research, Rough Rides, and Name Changes," October. <https://doi.org/10.1175/BAMS-88-10-1566>.
- Gorishniy, Yury, Ivan Rubachev, and Artem Babenko. 2023. "On Embeddings for Numerical Features in Tabular Deep Learning." arXiv. <https://doi.org/10.48550/arXiv.2203.05556>.
- Guo, Tujie. 2021. "Hurricane Damage Prediction Based on Convolutional Neural Network Models." In *2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, 298–302. <https://doi.org/10.1109/ICAICE54393.2021.00065>.
- Hittawe, Mohamad Mazen, Fouzi Harrou, Mohammed Amine Togou, Ying Sun, and Omar Knio. 2024. "Time-Series Weather Prediction in the Red Sea Using Ensemble Transformers." *Applied Soft Computing* 164 (October):111926. <https://doi.org/10.1016/j.asoc.2024.111926>.
- Hochreiter, Sepp. 1998. "The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 06 (02): 107–16. <https://doi.org/10.1142/S0218488598000094>.
- Holland, G., & Bruyère, C. L. (2014). Recent intense hurricane response to global climate change. *Climate Dynamics*, 42(3), 617–627. <https://doi.org/10.1007/s00382-013-1713-0>
- Huang, Xin, Ashish Khetan, Milan Cvitkovic, and Zohar Karnin. 2020. "TabTransformer: Tabular Data Modeling Using Contextual Embeddings." arXiv. <https://doi.org/10.48550/arXiv.2012.06678>.
- "National Hurricane Center Forecast Verification." n.d. Accessed December 12, 2024. <https://www.nhc.noaa.gov/verification/verify5.shtml>.
- Nawari, N. O. 2012. "The Role of Data Mining Techniques in the Prediction of Hurricane Damages," April, 1–10. [https://doi.org/10.1061/41016\(314\)315](https://doi.org/10.1061/41016(314)315).
- "NHC Data Archive." 2024. September 17, 2024. <https://www.nhc.noaa.gov/data/>.

- Nguyen, Tung, Rohan Shah, Hritik Bansal, Troy Arcomano, Sandeep Madireddy, Romit Maulik, Veerabhadra Kotamarthi, Ian Foster, and Aditya Grover. 2024. "Scaling Transformers for Skillful and Reliable Medium-Range Weather Forecasting." In . <https://openreview.net/forum?id=qtcYYLSkQ9>.
- Pawluszek-Filipiak, Kamila, and Andrzej Borkowski. 2020. "On the Importance of Train–Test Split Ratio of Datasets in Automatic Landslide Detection by Supervised Classification." *Remote Sensing* 12 (18): 3054. <https://doi.org/10.3390/rs12183054>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (No. arXiv:1706.03762). arXiv. <https://doi.org/10.48550/arXiv.1706.03762>
- Yin, Wenpeng, Katharina Kann, Mo Yu, and Hinrich Schütze. 2017. "Comparative Study of CNN and RNN for Natural Language Processing." arXiv. <http://arxiv.org/abs/1702.01923>.
- Zimmerman, Itamar, and Lior Wolf. 2023. "On the Long Range Abilities of Transformers." arXiv. <https://doi.org/10.48550/arXiv.2311.16620>.