Real-Time Forecasting of Marine Visibility with Self-Adaptive Deep Learning*

William Girard
Computer and Information Science Department
University of Massachusetts Dartmouth
Dartmouth, MA 02747, USA
wgirard@umassd.edu

Haiping Xu
Computer and Information Science Department
University of Massachusetts Dartmouth
Dartmouth, MA 02747, USA
hxu@umassd.edu

Donghui Yan
Mathematics Department
University of Massachusetts Dartmouth
Dartmouth, MA 02747, USA
dyan@umassd.edu

***Abstract*—Marine visibility is a crucial factor for safe navigation and operational efficiency at sea, as poor visibility can lead to costly accidents and pose significant safety risks. While deep learning models, such as artificial neural networks (ANNs), have shown effectiveness in forecasting land visibility, predicting marine visibility presents unique challenges due to its complexity and variability. In this paper, we introduce self-adaptive deep learning (SADL) models to enhance visibility prediction across varying time periods and distances from ships. Our approach employs real-time training models that continuously learn and adapt based on incoming data, enabling dynamic and comprehensive visibility forecasts. We present case studies to validate the effectiveness of the SADL models and demonstrate their accuracy across diverse locations, time frames, and scenarios. The results show the models' capability to predict visibility accurately and in a timely manner, thereby effectively improving marine safety.**

Keywords—marine visibility, self-adaptive deep learning, visibility forecasting, real-time training, model fine-tuning

I. INTRODUCTION

Weather forecasting has become a cornerstone of modern society, with its scientific origins tracing back to the invention of measuring instruments, such as the mercury barometer, in the mid-17th century. Innovations like the establishing of weather stations in the early 19th century, the advent of radar in the 1940s, and the deployment of satellites in the 1960s have significantly advanced the field [1]. Despite these developments, weather forecasting remains an essential area of research, particularly in the context of climate change. The increasing frequency of extreme weather events, such as heatwaves, heavy precipitation, and tropical cyclones, has further complicated weather forecasting efforts [2]. Visibility, a key parameter in weather forecasting, refers to the maximum horizontal distance at which objects can be seen. Low visibility can have significant impacts, including reduced airport capacity, flight delays and disruptions, marine navigational hazards, impaired road traffic safety, and decreased search-and-rescue effectiveness. In marine environments, poor visibility can lead to costly accidents and pose significant safety risks, making it a crucial factor for safe navigation and operational efficiency. To address this challenge and the rapidly changing marine conditions, we introduce a forecasting approach for marine visibility.

Deep learning models, such as artificial neural networks (ANN), have proven effective in predicting land-based visibility [3]. However, predicting marine visibility presents unique challenges due to its inherent complexity and variability. Traditional techniques like ANNs often lack the adaptability required for fast-paced applications, such as marine visibility forecasting, which needs to take into account rapidly changing environmental conditions and the absence of weather stations on the open ocean. To address these challenges, we propose a framework for real-time marine visibility forecasting. The framework employs a cluster of self-adaptive deep learning (SADL) models to make predictions at both current and remote locations over varying time periods. By providing accurate real-time visibility predictions, captains or navigators can proactively reroute ships or prepare for low visibility conditions, thereby reducing the risk of accidents. The effectiveness of the SADL models is demonstrated through diverse case studies, showcasing their accuracy and adaptability. These models are based on the multilayer perceptron (MLP) deep neural network architecture, comprising an input layer, hidden layers, and an output layer, with each layer containing a set of perception elements known as neurons. The SADL models utilize a stochastic gradient descent optimizer, employ backpropagation to iteratively refine weights, and incorporate real-time training to deliver timely and reliable predictions.

The proposed real-time framework using SADL models consists of two model clusters: one for the current location and another for remote locations. This separation stems from the difference in available input features. The remote location models rely solely on satellite data, drone data, and simulated data derived from the current location, while the current location models utilize a wider range of on-board sensors, capturing local data such as temperature, wind direction and wind speed. The selected input features are traditional weather measurements and do not include variables such as latitude and longitude coordinates and seasonal information. This approach aims to demonstrate the effectiveness of a general visibility model with high adaptability, relying only on weather parameters. The current location model cluster makes predictions for four future time intervals: 15, 30, 45, and 60 minutes. Similarly, the remote location cluster predicts for the same time intervals, but also forecasts for four cardinal directions at distances of 5, 10, 15, and 20 miles from the ship. While this approach requires continuous training and predictions, the use of an adaptive training window with small-batch training ensures that predictions are generated with both reasonable speed and accuracy. This study highlights the advantages of using SADL model clusters to predict visibility across varying time intervals and locations in the ocean, demonstrating their adaptability and effectiveness in dynamic marine environments.

II. RELATED WORK

There have been several research efforts focused on the effects of low-visibility conditions and visibility prediction. For example, Abdel-Aty et al. highlighted the dangers of low-visibility conditions such as fog or smoke (FS) while driving, particularly during the night [4]. They performed a temporal

* This material is based on work supported by Office of Naval Research (ONR) Grant, MUST IV, University of Massachusetts Dartmouth.

distribution analysis on the available data, concluding that the early morning hours from 5 a.m. to 8 a.m., from December to February, are the most likely times for FS-related crashes. Jonnalagadda and Hashemi introduced an Auto-Regressive Recurrent Neural Network (ARRNN) model for land-based visibility prediction [3]. They demonstrated the validity of their approach compared to other types of RNNs, such as LSTM and basic RNNs. Niu et al. attempted to predict land-based visibility with three decision tree-based models: XGBoost, LightGBM, and Random Forest [5]. Using these models, they performed visibility predictions for future hours at five selected weather stations. The approaches mentioned above demonstrate how land-based visibility predictions can be achieved using weather data from weather stations. In contrast, our approach addresses the challenges posed by limited data on the open ocean and applies real-time training techniques to predict marine visibility, where traditional approaches fall short in terms of adaptability.

In recent years, significant research has been conducted in the field of marine weather and environments. Muttil and Chau presented two machine learning (ML) techniques, ANN and genetic programming (GP), for predicting algal blooms [6]. By focusing on feature ranking, they concluded that an auto-regressive nature or persistence in the algal bloom dynamics may be related to the long flushing time in semi-enclosed coastal waters. Kim et al. presented a convolutional LSTM model for predicting weather parameters [7]. This convolutional LSTM uses time-series images as inputs to predict eight different ocean weather parameters: surface temperature, wave height, wave period, wave direction, wind speed, and current ship speed. Krestenitis et al. used deep convolutional neural networks (DCNN) to identify oil spills from satellite images [8]. Their results suggest that DCNN segmentation models, trained and evaluated on the provided dataset, can be used to implement efficient oil spill detectors. The approaches mentioned above focus on predicting marine weather and environmental conditions for a fixed location. In contrast, our approach aims to predict weather conditions, particularly marine visibility, not only for the current location but also for remote locations on a moving vessel.

There have also been research efforts that employ real-time techniques in deep learning. Singhal and Ahmad introduced a deep learning facial recognition system for university attendance [9]. They utilized CNNs with real-time video processing to improve model accuracy. Ford et al. introduced a real-time self-adaptive classifier (RT-SAC) to classify suspicious online bidders [10]. Given the real-time nature of the online auction environment, they employed a moving window approach, achieving reasonable detection accuracy. Girard et al. developed a deep learning model trained on a moving window of data to predict hydropower generation [11]. By utilizing real-time ANNs, the models accurately predict hydropower generation on a daily, weekly, and monthly basis. In this paper, we extend the real-time training architecture discussed above and develop clusters of deep learning models to predict marine visibility in real time across multiple time periods and remote locations.

III. Real-Time Forecasting of Marine Visibility

A. *A Framework for Real-Time Visibility Prediction*

Traditional deep learning methods typically focus on predicting static outcomes, such as gross domestic product estimates, facial recognition results, or disease diagnoses based on patient data. In contrast, sea travel occurs in highly dynamic environments where weather conditions can change rapidly within minutes. Furthermore, the scarcity of weather stations in oceanic regions limits the availability of input data for weather prediction. To address these challenges, we propose a framework for real-time marine visibility forecasting that incorporates two types of models: current location models and remote location models. By predicting visibility at both the current location and across various remote locations, the approach provides ship captains and navigators with a highly accurate visibility "picture" of the surrounding area. This supports real-time decision-making and helps reduce the risk of accidents. As mentioned earlier, the current location model cluster consists of four deep learning models, each forecasting visibility 15, 30, 45, and 60 minutes into the future. Similarly, the remote location model cluster includes four models that predict visibility over the same time intervals. However, each model also generates predictions for four cardinal directions (north, east, south, and west) at distances of 5, 10, 15, and 20 miles from the vessel. This results in 16 predictions per time step for each remote model. Collectively, these predictions create a comprehensive visibility "radar" map around the vessel. Since the sea-based vessel may continuously move through the water, it is critical to use deep learning models that can adapt to changes in both location and time. Figure 1 presents a high-level framework for real-time visibility prediction.

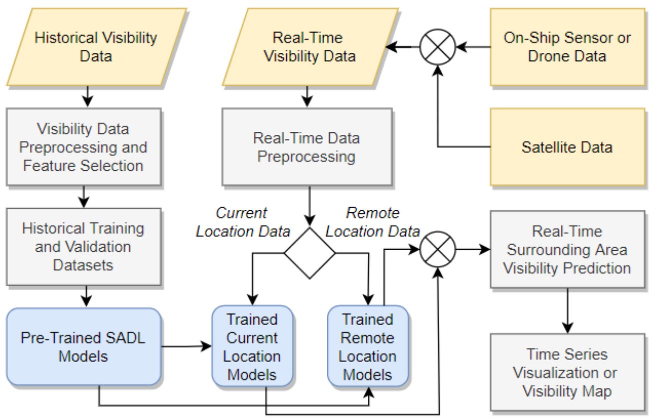


Fig. 1. A framework for real-time visibility prediction

As shown in Fig. 1, the SADL models must first be pre-trained using historical visibility data to ensure high testing accuracy [10], [11]. In our real-time approach, the visibility data consists of satellite and on-ship sensor or drone data. This real-time data is then preprocessed and sent to either the current or remote location models. It is important to note that the current location models can leverage more input features than the remote location models, as on-ship sensor data is unavailable for remote locations. All SADL models are then trained in real time to make the necessary marine visibility predictions. Finally, these real-time predictions are collected, visualized as a time series, evaluated for accuracy, and in future work, displayed on a live "radar" map.

B. *Feature Selection*

In our approach, the output feature, or label, is visibility, defined as the farthest horizontal distance at which objects remain visible. For instance, on a dense foggy day, visibility can drop to as low as 0.03 miles, while on a clear day, it can extend up to 12 miles [12]. The features used in the model were selected for their strong correlation with visibility, including temperature, wind speed, and humidity levels, all of which are known to impact visibility. Relevant input features are critical for making accurate predictions using deep learning, as irrelevant or harmful features can lead to bias and poor predictions [13], [14]. Table I lists several examples of input features that can be captured by satellites, on-ship sensors, or drones.

TABLE I. EXAMPLES OF INPUT FEATURES

Parameter	Unit	Source	Short Description
Precipitation	Millimeter	Satellite	Gridded rainfall measurement
Cloud Cover	Percentage	Satellite	Percentage of unit grid covered by clouds
Dry-Bulb Temperature	Fahrenheit	Satellite/ On-ship sensor/drone	Gridded temperature measurement
Relative Humidity	Percentage	On-ship sensor/drone	Ratio of absolute humidity to maximum possible humidity
Surface Pressure	Inch of Mercury	On-ship sensor/drone	Proportional to the mass of air over the location
Wind Speed	Mile/hour	On-ship sensor/drone	Wind speed for current or remote locations
Wind Direction	Degree	On-ship sensor/drone	Wind direction for current or remote locations

Satellite features are typically simpler in complexity and more widely available compared to on-ship sensor features; therefore, they are key features in remote location datasets. On-ship sensor and drone feature data are obtained from instruments on the sea-faring vessel. These features are more complex and require specialized equipment. As a result, on-ship sensor data is only available in current location datasets, while drone data is obtained by deploying drones to nearby remote locations, such as within a mile.

IV. SELF-ADAPTIVE DEEP LEARNING MODELS

SADL models are designed for dynamic, fast-paced applications, especially where the learning environment may evolve with each prediction. Training a model solely on a historical dataset for real-time visibility predictions would significantly reduce accuracy. Instead, our approach employs small batch training and evaluation to ensure that the SADL models adapt to the learning environment in real time.

A. *Adaptive SADL Models*

The SADL model update process is shown in Figure 2. As illustrated, recent visibility data is processed and added to the adaptive training window, enabling small-batch real-time training with a manageable number of training data points. Simultaneously, older, less relevant data is removed from the adaptive training window and moved to the historical data pool for future model pre-training. Note that small-batch real-time training ensures fast computation times. Once the real-time training is complete, the SADL model is updated and used to make visibility predictions for visual display.

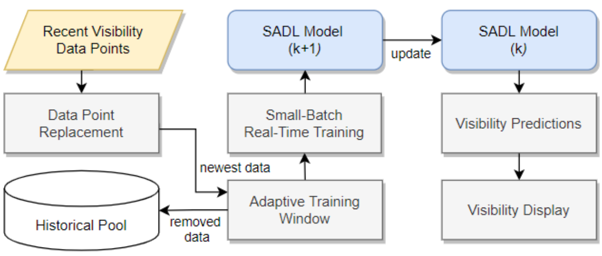


Fig. 2. The SADL model update process

B. *SADL Model Pre-Training*

The SADL models are ANNs created using the Keras API from TensorFlow, which provides high-level functions for building sequential models, commonly known as feedforward neural networks (FNNs). Previous research has shown that pre-training on a large subset of historical data is essential for accurate predictions [10], [11]. In this implementation, the SADL models are compiled with mean squared error (MSE) as the loss function and mean absolute error (MAE) as an additional metric. These metrics are used solely during the pre-training phase. The SADL models are pre-trained using Keras's fit function with a standardized 20 percent validation split. The selected epochs and batch size are 250 and 128, respectively. Figure 3 shows the pre-training results for one of the tested models, the 15-minute current location model.

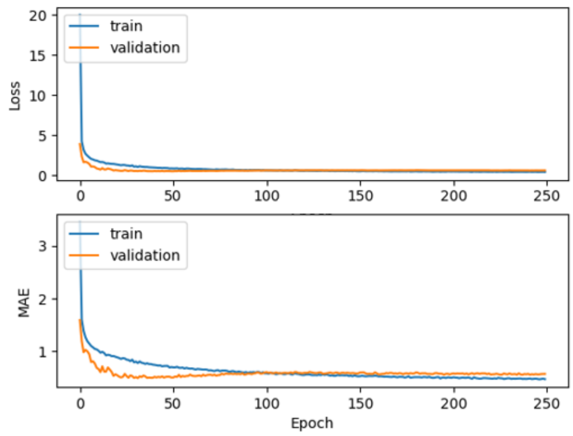


Fig. 3. Pre-training MAE and loss graphs for an SADL model

In Fig. 3, the top graph plots the loss (MSE) against the training epochs, while the bottom graph plots the MAE against the epochs. As shown, the training and validation lines are well-fitted in both graphs, with the validation line slightly higher than the training line once the number of epochs reaches 250. These graphs demonstrate effective pre-training, indicating that the model is ready for real-time training.

C. *Real-Time Training of SADL Models*

In fast-paced environments such as weather forecasting and online auctions, traditional deep learning approaches are often not viable. By training a deep learning model on only the most recent data points, it can be fine-tuned over time, leading to improved model performance. In this section, we describe the real-time (or incremental) training process. The adaptive training window is a small collection of data points consisting of both recent data that the model has already trained on and new data that the model has not yet seen. The SADL model trains on the entire window, which is defined by a suitable window size. The real-time training approach using a moving training window has proven successful in previous work [10][11]. In this paper, we expand the approach to include predictions across multiple time scales and locations. Figure 4 shows an example of the adaptive training window with a window size of 6.

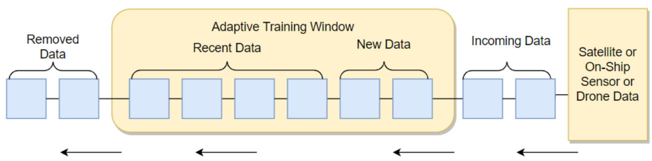


Fig. 4. An example of adaptive training window with a window size of 6

Both the current and remote SADL models predict visibility, with the former focusing on four future time intervals (15, 30, 45, and 60 minutes) and the latter extending to 16 remote locations across four directions. To evaluate the accuracy of these predictions, we define the prediction accuracy for a single visibility prediction as $(1 - |v_a - v_p| / max(v_a, v_p))*100$, where v_a is the actual visibility value, and v_p is the predicted visibility value. The prediction accuracy of a remote SADL model is calculated as the average accuracy across the different remote locations. Algorithm 1 outlines the real-time training logic and visibility prediction for a list of SADL models. Since each SADL model has its own adaptive training window and incoming data, all models can be trained and make predictions concurrently.

Algorithm 1 Concurrent Real-Time Training and Prediction

Input: A list of n SADL models Φ_i for x_i-minutes predictions, each with an adaptive training window W_i of size τ_i, where $1 \le i \le n$; a prediction interval y, a sampling interval z, and a number of remote locations nr
Output: A list of new SADL models Φ_i and predicted visibility v_i

Initialize data collection time $t = 0$
while $t < y$
  obtain a new labeled data point for the current location and a new labeled data point for each remote location
  $t = t + z$
Number of new data points (current location) $m_c = y / z$
Number of new data points (remote locations) $m_r = nr * (y / z)$
for each model Φ_i in the list of SADL models, where $1 \le i \le n$
  Create a new thread and do the following:
    Let m be m_c or m_r based on model type (current/remote)
    if $m \ge \tau_i$, replace all data in W_i by τ_i most recent data
    else add m new data to W_i and remove m old data from W_i
    Fine-tune SADL model Φ_i using all data points in W_i and obtain fine-tuned model Φ_i'
    Let current SADL model Φ_i be Φ_i'
    if model Φ_i is for the current location:
      Use the most recently collected new data point to predict visibility v_i of the current location in x_i -minute
    else if model Φ_i is for remote locations:
      Use the most recently collected nr new data points to predict visibility v_i of all remote locations in x_i -minute
wait until all threads have completed
return a list of new SADL models Φ_i and predicted visibility v_i

The algorithm begins by collecting newly labeled data points based on the prediction interval y and the sampling interval z. The variables m_c and m_r represent the number of new data points obtained for the current location and the remote locations, respectively. These new data points are added to the adaptive training window, while an equal number of older data points are removed. All SADL models, each with their own adaptive training window, are fine-tuned in parallel, significantly improving computation time. The most recent features are used to make the necessary visibility predictions. A current location SADL model performs a single prediction, while a remote location model performs predictions for each remote location. Finally, a list of SADL models and their predicted visibility is returned.

V. Fine-Tuning Model Hyperparameters

A. Dataset Preprocessing and Creation

As shown in Table I, we identified relevant weather-related features for the task of sea-based visibility prediction. Some of the data, such as visibility, were obtained from the ICOADS dataset, provided through NOAA's data portal [15]. Cloud cover and precipitation data, derived from satellite observations, were obtained from the Copernicus data portal, providing global gridded monthly and daily datasets from 1979 to the present [16]. All datasets were properly cleaned using the Pandas library from Python. For demonstration purposes, the data must be simulated into small time intervals, such as five-minute intervals. Every five-minutes, a new training data point is collected and added to the adaptive training window.

To simulate marine weather conditions and demonstrate the effectiveness of our real-time approach, we developed four current location datasets, one for each of the following prediction intervals: 15, 30, 45, and 60 minutes. When creating the datasets, the label (i.e., visibility) must be time-shifted. For example, in a training dataset for the 15-minute prediction model, the label for a data point observed at time t refers to the visibility observed at time t+15. The current location datasets were created from existing data and contain a total of 8,188 rows, spanning from January 1, 2019, to January 15, 2019. Similarly, we developed four remote location datasets corresponding to 15, 30, 45, and 60-minute predictions. However, since each remote location model predicts visibility across 16 remote locations, the remote location datasets are much larger. These datasets are based on existing data and contain a total of 131,008 rows, spanning from January 1, 2019, to January 15, 2019, and cover four directions (north, east, south, and west), each with four different distances.

B. Manual Tuning of Hyperparameters

Hyperparameters are crucial for fine-tuning models to achieve accurate predictions. For the SADL models, the hyperparameters include the number of neurons per hidden layer, dropout rate, batch size, learning rate, optimizer, epochs for small batch training, and the activation function. These hyperparameters were tested using the GridSearchCV function from scikit-learn. This function evaluates many possible combinations of hyperparameters to find the optimal set. Only selected individual tests were conducted, as testing an extensive range of hyperparameters with GridSearchCV is highly time intensive. The number of epochs was also manually tested. Table II shows examples of the tested values and the chosen values for hyperparameter tuning.

TABLE II. Examples of Hyperparameter Tuning

Hyperparameter	Tested Values	Chosen Value
HL #1 Neurons	32, 64, 128, 256	64
HL #2 Neurons	16, 32, 64, 128	32
HL #3 Neurons	8, 16, 32, 64	16
Dropout Rate	0%, 10%, 20%	10%
Learning Rate	0.01, 0.005, 0.001, 0.0005	0.005
Optimizer	Adam, SGD, RMSProp	SGD
Activation Function	Leaky ReLu, Relu, Sigmoid	Leaky ReLu
Epochs (small batch)	20, 25, 30, 35, 40, 45, 50	20

C. Window Size Selection

A small window size can fit the model well to recent data points but may struggle to capture long-term trends. In contrast, a large window size allows for capturing long-term patterns but can be detrimental to the model, especially in volatile environments like marine visibility, which is highly influenced by dynamic and unpredictable factors. Based on previous work that performed detailed analysis to select the optimal window size [11], we conducted experiments on various datasets to determine the most suitable window sizes for different models. Table III shows the selected window sizes and the corresponding prediction accuracy for both current location and remote location models. Note that, since more data is available for remote locations, the window sizes for remote location models are larger than those for current location models.

TABLE III. WINDOW SIZES FOR CURRENT/REMOTE LOCATIONS

Time Interval	Current Location		Remote Locations	
	Window Size	Accuracy	Window Size	Accuracy
15 minutes	12	96.25%	128	96.49%
30 minutes	12	93.07%	128	95.12%
45 minutes	12	96.19%	64	96.39%
60 minutes	12	91.01%	72	95.94%

VI. CASE STUDY

To demonstrate the feasibility and effectiveness of our approach, we present several case studies. We first examine general cases of marine weather during the daytime for both current location and remote location visibility predictions. We then test on the day-night cycle and adapt the model to account for lower visibility conditions at night. Finally, we simulate a mock storm to assess the models' adaptability.

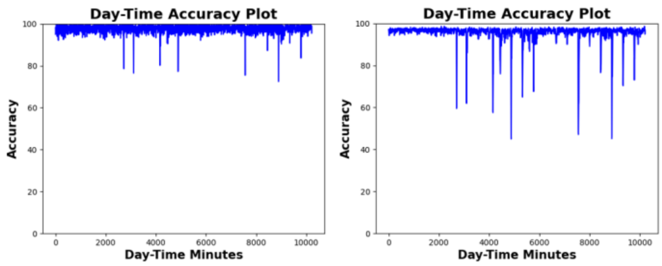
A. Daytime Accuracy Analysis

The SADL current location and remote location models were tested upon daytime data points from the current location and remote location datasets, respectively. Daytime data points were specifically chosen to assess the models' ability to accurately track trends, such as foggy or rainy weather. Table IV shows the observed testing accuracy for all four current location models and all four remote location models. Since daytime visibility tends to be relatively consistent, we have set the training interval to 30 minutes and the sampling interval to 5 minutes.

TABLE IV. DAY-TIME ACCURACY FOR CURRENT/REMOTE LOCATIONS

Time Interval	*Daytime Accuracy*	
	Current Location	*Remote Location*
15 minutes	96.25%	96.49%
30 minutes	93.07%	95.12%
45 minutes	96.19%	96.39%
60 minutes	91.01%	95.94%

As observed from Table IV, all models achieve a prediction accuracy of over 90 percent. The 15-minute models are the most accurate, though the other models also show similar performance. Figure 5 presents a breakdown of the prediction accuracy for the 15-minute current location and 15-minute remote location models on a per-point basis.



(a) Current location (b) Remote locations

Fig. 5. Daytime prediction accuracy of the 15-minute models

As shown in Fig. 5, the accuracy of visibility predictions is plotted over a period of 10,000 minutes of daytime, with each prediction occurring every five minutes. The 15-minute models are chosen for illustration due to their relatively higher accuracy. As observed, most predictions from the current location model fall within the 93-100 percent range, with an average accuracy of 96.25 percent. While there are a few notable misses, no individual predictions fall below 75 percent accuracy. This consistency over the 10,000 minutes of daytime data indicates that the current location model is well-trained and can provide reliable real-time visibility predictions. For the 15-minute remote location model, prediction accuracy is averaged across the 16 remote locations. Despite some inaccuracies, only around 10 predictions fall outside the expected range over 10,000-minute daytime, and none drop below 40 percent. Overall, the remote location model demonstrates reliable performance in predicting real-time visibility across various distances over an extended period.

B. Simulations of Day-Night Cycle

In this case study, we demonstrate how a remote location 15-minute model accurately predicts the day-night cycle over a two-day period. Lower visibility is typically observed at dawn, night, and dusk, and the model should be capable of predicting these conditions. Since the SADL models are feed-forward ANNs, they may not effectively learn long-term patterns. Instead, they learn the complex relationship between the input features and the output feature over shorter periods of time. To address this issue, we introduce an additional feature, called time indicator. During dawn, night, or dusk, a lower value is generated for this feature, while during the day, a higher value is assigned. This feature interacts with the other input features, ensuring that even if the time parameter indicates daytime, the model can still predict low visibility in the presence of factors such as rain or clouds. Figure 6 shows the prediction results from the 15-minute remote location model for a distance of 5 miles to the north.

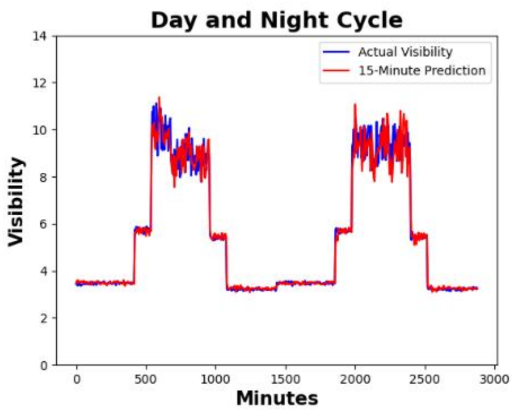


Fig. 6. Real-time visibilty prediction of day-night cycle at a remote location

As shown in Fig. 6, the 15-minute prediction at a 5-mile distance in the north closely follows the actual visibility. The first segment shows low visibility during nighttime, followed by a slight increase at dawn, and then a rise in daytime visibility. The visibility decreases again during dusk, and the entire cycle repeats. This graph demonstrates the ability of our real-time prediction models to capture visibility trends, such as the day-night cycle.

C. Storm Simulation Analysis

While overall accuracy is a valid metric for model evaluation, it is also important to examine special cases. In this study, we focus on predicting the visibility of a developing storm using the remote location models. The simulated storm is located in the north direction, 5 miles away from the vessel. Storm weather is characterized by high wind speeds, rainfall, cloud cover, and low pressure. We simulated a 360-minute storm scenario to challenge the SADL models, as storms often present unusual conditions with rapidly changing visibility. To address this challenge, we adjusted the prediction interval to 5 minutes and the sampling interval to 30 seconds, allowing the model to train on more storm data in real time. Additionally, we incorporated simulated storm data into the training datasets to help the model better learn the storm scenario. Figure 7 shows the real-time prediction results for the simulated storm, located 5 miles to the north, using the 15-minute remote location model.

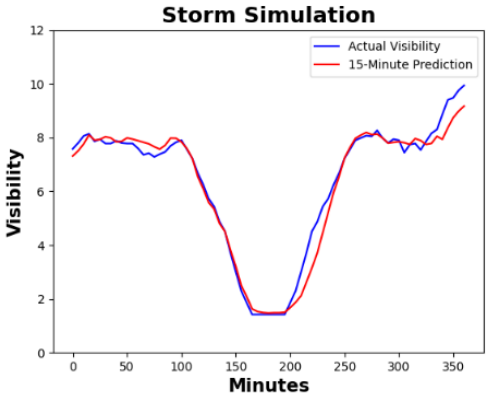


Fig. 7. Real-time prediction of a simulated storm at a remote location

Similar to the 15-minute remote location model, the 30, 45, and 60-minute models also closely track the actual visibility, though with slight delays. This suggests that the models have sufficiently learned to predict future visibility, especially with sufficient storm data, along with much shorter prediction and sampling intervals. As shown in Fig. 7, the storm begins with a drop in visibility, reaching its lowest point at 160 minutes. The storm continues until it fades, and visibility gradually increases. There are more missed predictions for the storm when visibility starts to increase. Since the storm remains at very low visibility for less than an hour, we believe the models tend to overfit to this data and are slightly slow to adapt. Nevertheless, the vast majority of predictions align closely with the actual visibility, even in this special scenario.

VII. Conclusions And Future Research

This research aims to address the challenging field of dynamic weather prediction at sea. The lack of traditional equipment in ocean and the rapidly changing environments necessitates real-time training and prediction. Traditional approaches, where a deep learning model is trained on a large dataset and then used for prediction, fail to perform accurately in such dynamic environments. In contrast, in our SADL approach, both training and prediction are conducted in real time, enabling the model to remain up-to-date and responsive. After presenting the optimized hyperparameters, including window sizes, we evaluated our approach in various scenarios. The current and remote location models performed well in typical cases including daytime and day-night cycles, with all models achieving reasonable accuracy. Predicting storms, however, remains a difficult challenge, with noticeable delays in accuracy. Despite this, we believe our approach serves as a strong starting point for future research in this area.

Future work will involve the implementation of a dynamic "radar" map and a demonstration of a mock ship journey. We also aim to incorporate artificial intelligence (AI) technology to select the optimal ship path, based on the most recently predicted visibility, assisting the captain with navigation. Additionally, we plan to experiment with datasets from different seasons, such as the monsoon period. This may require additional training on storm data, potentially even during the real-time training loop. If storms are less frequent, the model may need to focus more on storm datasets to retain the storm pattern. Conversely, if storms are more frequent, the model can update this knowledge less often. Finally, we plan to automate the process of dynamically determining model parameters, including prediction and sampling intervals, to enhance model performance in pattern learning.

References

[1] J. J. Cahir, "Weather forecasting," *Britannica*, Dec. 2024 [Online]. Available: https://www.britannica.com/science/weather-forecasting

[2] IPCC, "Climate change 2021: the physical science basis," *IPCC Sixth Assessment Report*, Working Group 1, 2021 [Online]. Available: https://www.ipcc.ch/report/ar6/wg1/

[3] J. Jonnalagadda and M. Hashemi, "Forecasting atmospheric visibility using auto regressive recurrent neural network," In *Proceedings of the 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, Las Vegas, NV, USA, 2020, pp. 209-215.

[4] M. Abdel-Aty, A.-A. Ekram, H. Huang, and K. Choi, "A study on crashes related to visibility obstruction due to fog and smoke," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1730-1737, Sep. 2011.

[5] W. Niu, B. Gong, X. Mao, H. Zhang, H. Wang, et al., "A study on short-term visibility prediction model in Jiangsu province based on random forest," In *Proceeding of the 2024 9th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, pp. 53-58, Apr. 2024.

[6] N. Muttil and K.-W. Chau, "Machine-learning paradigms for selecting ecologically significant input variables," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 6, pp. 735–744, Sep. 2007.

[7] K.-S. Kim, J.-B. Lee, M.-I. Roh, K.-M. Han, and G.-H. Lee, "Prediction of ocean weather based on denoising autoencoder and convolutional LSTM," *Journal of Marine Science and Engineering*, vol. 8, no. 10, 805, pp. 1-24, Oct. 2020.

[8] M. Krestenitis, G. Orfanidis, K. Ioannidis, K. Avgerinakis, S. Vrochidis, and Y. Kompatsiaris., "Oil spill identification from satellite images using deep neural networks," *Remote Sensing*, vol. 11, no. 15, 1762, pp. 1-22, Jul. 2019.

[9] M. Singhal and G. Ahmad, "Deep learning based real time face recognition for university attendance system," In *Proceedings of the 2023 International Symposium on Devices, Circuits and Systems (ISDCS)*, pp. 1–4, May 2023.

[10] B. J. Ford, H. Xu, and I. Valova, "A real-time self-adaptive classifier for identifying suspicious bidders in online auctions," *The Computer Journal*, vol. 56, no. 5, pp. 646–663, Mar. 2012.

[11] W. Girard, H. Xu, and D. Yan, "Real-time evolving deep learning models for predicting hydropower generation," In *Proceedings of the 3rd IEEE International Conference on Computing and Machine Intelligence (ICMI 2024)*, Mt Pleasant, Michigan, USA, April 13-14, 2024, pp. 1-6.

[12] T. H. Fang, Y.-G. Kim, I.-Y. Gong, S. Park, and A.-Y. Kim, "Development of performance measures based on visibility for effective placement of aids to navigation," *International Journal of Naval Architecture and Ocean Engineering*, vol. 7, no. 3, pp. 640–653, May 2015.

[13] I. Garrido-Muñoz , A. Montejo-Ráez , F. Martínez-Santiago , and L. A. Ureña-López , "A survey on bias in deep NLP," *Applied Sciences*, vol. 11, no. 7, p. 3184, Apr. 2021.

[14] G. Vardi, "On the implicit bias in deep-learning algorithms," *Communications of the ACM*, vol. 66, no. 6, pp. 86-93, June 2023.

[15] NOAA, "International comprehensive ocean-atmosphere data set," *ICOADS*, U.S. National Oceanic and Atmospheric Administration, Apr. 2024 [Online]. Available: https://icoads.noaa.gov/

[16] Copernicus, "Europe's eyes on earth," *Copernicus Data Access*, Oct. 2024 [Online]. Available: https://www.copernicus.eu/en/access-data