Tropical Cyclone Intensity Forecasting Using Deep Learning

Rose Atuah, Martin Pineda, Daniel McKirgan, Qianlong Wang, Michael McGuire
Department of Computer and Information Sciences
Towson University, Towson, MD 21252, USA
ratuah1@students.towson.edu, mpined4@students.towson.edu, dmckir1@students.towson.edu, qwang@towson.edu, and mmcguire@towson.edu

2023 International Conference on Machine Learning and Applications (ICMLA) | 979-8-3503-4534-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICMLA58977.2023.00187

***Abstract*—Tropical cyclones can produce devastating effects on humans, animals, and the environment. Globally, it has been a recurring problem across different continents. This has necessitated conducting research on different aspects relating to Tropical cyclones. To contribute to this research domain, in this study, three Deep Learning (DL) models were developed to predict Tropical Cyclone (TC) intensity. The study used the Hursat and Bestrack datasets from the United States National Oceanic and Atmospheric Administration and employed Convolutional Neural Networks (CNN), Longest Short-term Memory (LSTM), and a combination of CNN and LSTM (CNN-LSTM) to predict TC intensity. Results obtained from the study show that the LSTM model achieved the best results although the difference between the three models was not wide. Contributions from this study can aid in reducing damages to life and properties associated with ropical Cyclones by improving the prediction of TC intensity.**

***Index Terms*—Deep Learning, CNN, LSTM, Hybrid, Cyclone Intensity, Prediction**

I. Introduction and Related Work

The intensity of a Tropical Cyclone TC according to [1], is defined as the per-minute maximum sustained winds speed at 10m over the surface of not less than 74mph. The prediction of this natural phenomenon is deemed a herculean but critical task for both stakeholders and researchers as this knowledge can help to devise a means to curtail the devastation and loss of life and property. Recent improvements in data collection technologies have enabled capturing more data at a higher level of complexity rendering the traditional mathematical-based numerical weather prediction system used in simulating short-term climatic conditions days in advance and the statistical regression methods used for long-term predictions inadequate for the prediction of TC intensities [2], [3], [4], [5].

Machine learning methods have recorded reputable achievements in prediction tasks in recent times. ML techniques have greatly been used for weather-related research. It has been used for both short and long-term TC predictions using data divided into a suitable ratio of training, testing, and validation. Trained models are tested and validated to ascertain model conformity with real data. When applied to TC intensity prediction, ML methods have shown to be able to predict TC intensity with moderate success [6]–[8] however, ML models suffer from the difficulty in addressing the complex relationships between the ever-changing mechanism of hurricane formation termed as genesis, the path it takes known as track and the level of severity known as intensity [9]–[11].

Deep Learning (DL) is a subset of Artificial Intelligence that centers on using neural networks to create decisions based on previously available data. DL works explicitly where datasets are multidimensional, complex, and voluminous and has seen increased use in a number of domains including, cancer detection, document analysis and recognition, healthcare, object detection, speech recognition, image classification and recognition, pedestrian detection, natural language processing, voice activity detection, etc. [12], [13]

Several studies have applied a number of methods to predict variables related to tropical cyclones (TC). In a study by [14] on short-time ionospheric total electron content, the authors compared DL and statistical techniques where LSTM and Seq2seq were modeled as DL techniques, and ARIMA was modeled as a statistical technique. Results from the study show that LSTM outperformed both Seq2seq and ARIMA models. In a related study, [15] proposed an LSTM model for forecasting storms making landfall within a 24-hour period in China. Finally, and most related to this study, is [16] where a Tropical Cyclone TC intensity prediction framework based on neural network called TCPred was proposed as an early warning system for emergency decision-making. In the study, they compared a number of deep learning models in their ability to predict TC intensity based on a number of atmospheric and oceanic datasets, where a convolutional GRU model achieved the best performance.

Addressing the problem of sparse data in Tropical cyclone analysis during the formation stage, [17] proposed a multi-modal multitask learning system to predict Hurricane genesis. The model outperformed standard models achieving high accuracy.

The hybridization of algorithms is a common trend adopted by researchers in order to produce a more robust model that can potentially draw on the strength of the different algorithms. In a study by [18] a CNN-LSTM based on 2D and 3D CNN was hybridized to produce a model that would be used to establish the relationship between the features of typhoon formation for better forecasting. Rigorous experiments featuring a wide range of data including Western Pacific, Eastern Pacific and Northern Atlantic Oceans sourced from IBTracs and ERA were conducted. Training, testing, and validation were done using past data on meteorology in the Pacific and Atlantic Oceans with results yielding improved accuracy.

Most similar to this paper, [19] developed a DL-based multilayer perceptron (MLP) TC intensity prediction model by conducting experiments using the SHIPS dataset. The leave-one-year-out LOYO scheme was adopted to evaluate the performance of the MLP which was found to outperform other models when compared.

Although the application of deep learning algorithms on TC prediction tasks has achieved encouraging performance, there are still shortcomings in terms of approaches employed by existing studies because the prediction based on time-series approach using LSTM and CNN is yet to be proposed. Based on existing studies, both LSTM and CNN as individual or hybridized algorithms have proven to be efficient by recording remarkable performance in several TC prediction related tasks as evidently reported in [14], [15], [16], [17], and [18]. Therefore, in this research, the performance of CNN,LSTM,CNN-LSTM modles are tested in the task of TC intensity prediction based on time-series approach. The models are trained and tested to predict windspeed as the target value using both Hursat and windspeed corresponding values from Best Track available at the National Centres for Environmental Information(NCEI) [20] and SHIPS datasets.

The rest of the paper is organized as follows. Section 2 presents the research methodology and the proposed architectures and evaluation techniques adopted. Section 3 shows the results of the experiments conducted and a discussion of the results from each experiment followed by section 4 which summarizes the performance of the models in comparison to similar works and future research directions.

II. Methodology

This section outlines the different steps of the proposed approach. In stage one, a set of pre-processing activities performed on the dataset to ensure that it conforms to our proposed standard i.e. having a set of predictors and a target(wind speed) attribute. In the modeling and evaluation stage, the deep learning approaches are trained and tested on their performance in TC intensity prediction. In the evaluation stage the performance of each model is evaluated based on the standard prediction evaluation metrics including the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Each of these stages is presented individually in the subsequent sections.

A. Data

The data used in this study is obtained from the hurricane satellite (Hursat) archive and the BestTrack hurricane dataset both available at the National Centres for Environmental Information (NCEI) [20] The Hursat dataset is a popular dataset that has been used by several researchers' hurricane and tropical cycle research activities. It contains data from 1978 to 2016 in Network Common Data Form (NC) format. Five years of data were used in this study i.e. 2012, 2013, 2014, 2015, and 2016.

Also, a combined dataset (SHIPS, GTCD and NOAA) used by [19] was applied to our proposed model in comparison with the model by [19]. This is done so that a fair comparison can be conducted on the same dataset by our proposed model and the model by [19].

1) Data Description: Observational datasets from different sources were used in this study. This section provides a description of each of these datasets.

- International Best Track Archive for Climate Stewardship(IBTrACS): This dataset is composed of more than 3,000 tropical and extra-tropical storm tracks since 1979, extracted from the NOAA database. The dataset provides global tropical cyclone best track data in a centralized location to help in understanding the distribution, frequency, and intensity of tropical cyclones worldwide. Produced by multiple governmental agencies, it includes attributes such as position, wind, and pressure [21]
- Hurricane Satellite: The HURSAT dataset provides geostationary satellite imagery on TCs in the IBTrACS dataset [22]. Primarily developed to produce a homogeneous analysis of TC intensity through time [23]. In this study, HURSAT version 6 is used. It covers the years from 1979 to 2015, with hundreds of images from geostationary satellites analyzed and stored in NetCDF file format.
- Statistical Hurricane Intensity Prediction (SHIPS): This is a statistical-dynamical model that is employed operationally at the National Hurricane Center (NHC) of the U.S. National Weather Services. It is based on a multiple linear regression approach that features persistence, climatological, and synoptic predictors [24]. These predictors such as zonal and meridional wind, shear, vorticity, and divergence. etc. are derived from the global forecasting system(GFS) of the National Centers for Environment Prediction (NCEP).

2) Data Preprocessing: The pre-processing activities performed on the dataset are: feature extraction, data merging, and data transformation. These techniques are used to transform data into forms that are easy to proccess into DL tasks which enables among other things, improvements in accuracy, prevent overfitting, and interpretability of the resulting model.

- Feature Extraction:In the HURSAT dataset, TC images are contained in the middle of each image where the vital information about the intensity is located. Consequently, these images were cropped to a 50 by 50 pixel dimension to remove the outer part of the TC from the image. The outcome is converted into a CSV file that contains 2500 pixel value for each TC in the dataset, and would finally be merged with the windspeed from the best-track dataset.
- Data Merging: In this stage, the features extracted from the HURSAT satellite imagery are merged with their corresponding windspeed from the bestrack dataset in IBTrACC. Each satellite image record only provides extracted pixel values and does not provide the corresponding windspeed of the TC, therefore, the windspeed

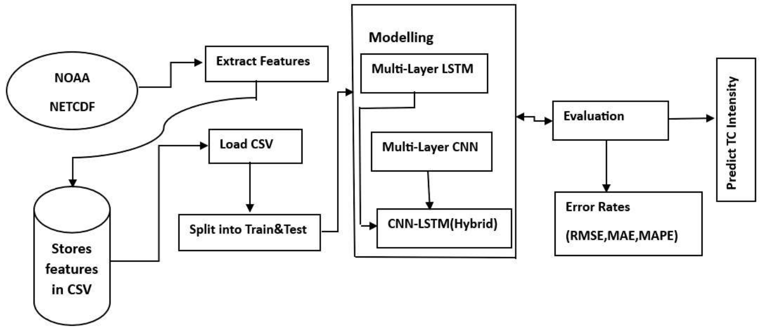

Fig. 1. Research Architecture

of each TC is retrieved from the bestrack dataset and is correspondingly appended (and set as the target value) with the TC record extracted from the HURSAT dataset. The outcome of this adds one column to the existing 2500 extracted features as elaborated in the feature extraction section.

- Data Transformation: Reframing multivariate time series into multi-step supervised learning problems. This would help in exploring different framings of a time series to know which of the framings can give us better results. The outcome of this framing would be a data frame with columns each suitably named by both the variable number and the time step.

B. Modelling

After pre-processing the dataset, three deep learning algorithms were trained to predict TC intensity. This includes, CNN, LSTM, and a hybrid CNN-LSTM model. The modeling is based on the prediction of wind speed in accordance with the Saffir-Simpson Hurricane wind scale [25].

1) Network Design and Training:

- Convolutional Neural Network (CNN): The CNN algorithm is made of several layers that continuously extract abstract features of input data, then match the features to a target based on specified tasks such as prediction or classification [26]. Generically, every layer is made of several neurons that calculate the weighted combination of inputs linearly and then train the model on the dataset. A nonlinear activation function is used to optimize the model parameters. The nonlinear activation function used in this study to map relationships between the input and output variables is the Rectified Linear Unit (ReLU). ReLU, defined as f(x) = max(0, x) where x represent the input to the neuron. The function returns positive values for any positive input but returns zero for any negative input thus prevents the exponential growth of computations required to manage the neural network. This was done to efficiently minimize inherent neural network gradient disappearance problems and error backpropagation.
The CNN modelling for this study as shown in Figure 2 is based on the following; input of the model of 2501x1

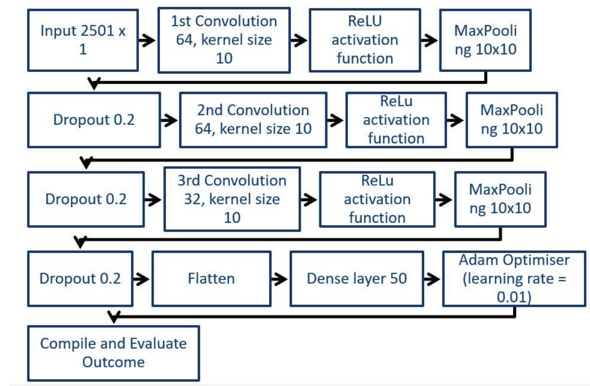

Fig. 2. Proposed CNN Architecture

tensor, three convolutional layers where the first, second and third layers has 64,64 and 32 layers respectively, kernel size of 10 and ReLU as activation funtion. The first,second and third pooling layers apply a 10 max pooling, the fully connected layer has 50 units, a Dropout rate of 0.2, and finally a single output unit.

2) Long Short-Term Memory (LSTM): : Long-Short-Term Memory is a type of Recurrent Neural Network-RNN. It is made up of a memory block, which contains one or more memory cells and three adaptive generative layers shared by all cells in the block that regulate the flow of information into and out of the cells [27]. The cells recollect values over arbitrary time intervals and it is possible to trim the input, forget, and output units to ascertain the type of data that should be stored and manipulated. This allows the LSTM to output selective information about present and future time intervals [28], [29].

The LSTM model used for this study as shown in Figure 3 is based on the following: the model input is a 2501x1, 3 LSTM layers of which the first two layers with 64 hidden layers, and the last one with 32 hidden layers. The model uses ReLU as the activation function. All LSTM layers are followed by dropout layers in which some neurons' contributions to subsequent layers are excluded from the

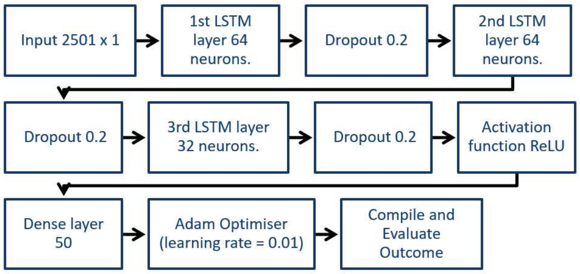

Fig. 3. Proposed LSTM Architecture

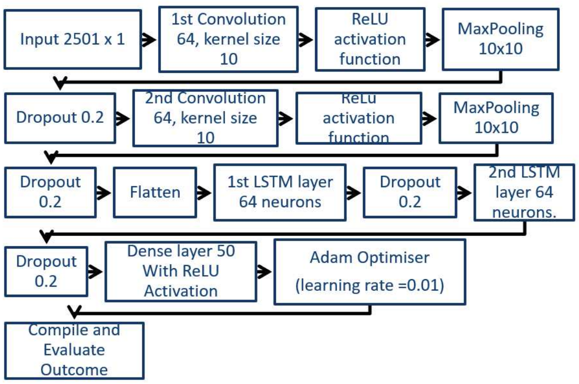

Fig. 4. Proposed CNN-LSTM Architecture

training data to avoid overfitting. A dropout rate of 0.2, and a single output unit are adopted.

3) The Hybrid CNN-LSTM: : The CNN-LSTM modelling for this study as shown in Figure 4 is based on the following: an input of the model of 2501x1 tensor, 2 CNN and 2 LSTM layers of each two layers with 64 hidden layers respectively, ReLU as the activation function, all layers are followed by Dropout rate of 0.2. The network is structured based on a single output unit which is the target i.e. windspeed to predict the TC Intensity. CNN is for image analysis and LSTM is for time series analysis. The reason we combine both (Hybrid) is that we believe the Hurricane intensity can be detected from the imagery data and also has a dependency on time domain.

- Dropout: Dropout regularization rate of 0.2 was applied to randomly ignore a specified percentage of the input data so as to avoid the model from overfitting, was applied for all the models.

The input layer for the CNN, LSTM, and hybrid CNN-LSTM models includes 2500 features, based on the dimension of the satellite imagery of 50 x 50 pixels, where each input feature represents an extracted pixel value. Also, each of the three models has a single output which is the wind speed. As elaborated earlier, wind speed can be used as the value to be predicted to determine the intensity of a TC.

C. Evaluation

To evaluate the performance of each model in TC intensity prediction, the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) are used to compare predicted values with observed values. All the metrics(RMSE, MAE, and MAPE) represnt error rates. For each model prediction, this means that the lower the error rate the better the prediction. These evaluation metrics collectively play an important role in evaluating model performance. RMSE as shown in Equation 1 according to [30] is used in evaluating the closeness of the predicted values against the observed value. MAE as shown in Equation 2 according to [31] is used to measure the average magnitude of the absolute errors between the predicted value and actual values. MAPE as a measure of predicting accuracy according to [32] calculates the average of the percentage error as shown in Equation 3.

$$RMSE = \sqrt{\frac{1}{n}\sum_{1}^{n}(D_{pre} - D_{act})^2} \tag{1}$$

$$MAE = \frac{1}{n}\sum_{1}^{n}|D_{pre} - D_{act}| \tag{2}$$

$$MAPE = \frac{100}{n}\sum_{1}^{n}|\frac{D_{pre} - D_{act}}{D_{act}}| \tag{3}$$

Where (D_{act}) is the actual variable, (D_{pre}) is the predicted variable, and n is the amount of collected data.

III. Results

The trained DL models were evaluated on their ability to predict TC intensity by using independent data that was not included in the training process. The RMSE, MAE, and MAPE outcome were used to evaluate the performance of each model taking into consideration that all data based on NetCDF format remain the same and does not vary. Several experiments were conducted to ascertain the optimal parameters that produce the best results from each model. The results of the experiments are shown in Table 1.

A. Experiment 1

Results obtained from experiment 1 based on 100 epochs, pool size of 10, kernel size of 10 and, 64 neurons for the 3 models are presented in Table 1 The results based on RMSE, MAE, and MAPE for the 3 models on both training validation and testing for each model.

It can be observed from Table 1 that for the RMSE measured in kts) train validation error rates, CNN-LSTM is the lowest with 14.082, followed by the LSTM with 15.42 and CNN with 18.184, while on testing LSTM recorded the lowest error rate of 5.488, followed by CNN-LSTM with 7.298 and lastly CNN with 12.528. Similarly on MAE and MAPE also measured in kts, the CNN-LSTM recorded the lowest error rates on train validation with 10.784 and 0.048 respectively, followed by the LSTM with 11.605 and 0.051 respectively, and lastly CNN with the highest train validation of 13.328 and 0.057 on MAE and MAPE respectively. On testing based on MAE and MAPE the LSTM recorded the least error rate of 4.928 and 0.022 respectively, followed by the CNN-LSTM with 6.766 followed by the CNN with 12.167 both on MAE, and CNN-LSTM with 0.031 MAPE which is less than CNN MAPE of 0.055 by 0.024. On train and validation loss it can be observed that the training and validation loss decreases to a reasonable point of stability with a

TABLE I
MODEL EVALUATION RESULTS

Parameters	DATASET:HURSAT, POOLSIZE:10, KERNELSIZE:10, NEURONS:64,64,32, OPTIMIZER:ADAM					
Epochs:100	RMSE		MAE		MAPE	
MODEL	TRAINING	TESTING	TRAINING	TESTING	TRAINING	TESTING
CNN-LSTM	14.082	7.298	10784	6.766	0.048	0.031
CNN	18.184	12.528	13.328	12.167	0.057	0.055
LSTM	**15.428**	**5.488**	**11.605**	**4.928**	**0.051**	**0.022**
Epochs:75	RMSE		MAE		MAPE	
CNN-LSTM	14.832	5.655	11.519	5.217	0.051	0.24
CNN	17.908	13.631	12.956	12.838	0.055	0.058
LSTM	**15.434**	**4.450**	**11.792**	**4.049**	**0.052**	**0.019**
Parameters	DATASET:SHIPS, POOLSIZE 2, KERNELSIZE 2, NEURONS 64,64,32, OPTIMIZER: ADAM					
Epochs:100	RMSE		MAE		MAPE	
CNN-LSTM	3.668	3.529	3.500	3.387	3.783	3.368
CNN	7.349	7.292	7.319	7.273	5.6077	5.013
LSTM	**3.668**	**3.529**	**3.500**	**3.387**	**3.781**	**3.366**
Epochs:75	RMSE		MAE		MAPE	
CNN-LSTM	3.668	3.529	3.500	3.388	3.785	3.369
CNN	8.828	8.773	8.803	8.757	7.427	6.075
LSTM	**3.669**	**3.532**	**3.500**	**3.393**	**3.795**	**3.377**

minimal generalization gap of 0.02kt on average, thus demonstrating a reasonable fit.

B. Experiment 2

Results obtained from experiment 2 based on 75 epochs It can be observed from Table 1 that for the RMSE train validation error rates, CNN-LSTM scored the lowest with 14.832, followed by the LSTM with 15.434 and CNN with 17.908, On testing, LSTM recorded the lowest error rate of 4.450 followed by CNN-LSTM with 5.655 and lastly CNN with 13.631. Comparably on MAE and MAPE, the CNN-LSTM recorded the lowest error rates on train validation with 11.519 and 0.051 respectively, followed by the LSTM train validation of 11.792 and 0.052 respectively, and lastly CNN with the highest train validation of 12.956 and 0.055 on MAE and MAPE respectively. On testing based on MAE and MAPE the LSTM recorded the least error rate of 4.049 and 0.019 respectively, followed by the CNN-LSTM with 5.217 and 0.024, followed by the CNN with 12.838 and 0.058 both on MAE, and MAPE.

C. Comparison with State-of-the-Art

The study by [19] is based on Deep Learning Multilayer perceptron using leave one year out LOYO and neural network optimization and to our knowledge, is the only one that approached TC intensity as a prediction problem rather than a classification problem. We implemented our models CNN, LSTM, and CNN-LSTM using this dataset and compared the performance outcome because their work focused on the current TC intensity as the most importance predictor scoring value of 0.101. The experimental setup for their study is based on an adaptive learning rate, ReLu activation function, L2 regularization with alpha 0.0005 and Adam optimization. The study recorded RMSE and MAE values of 11.07, and 8.22 respectively of truly independent values as if in a real-time mode.

The result comparison as shown in Table 2 reveals that the LSTM model outperformed the previous study on RMSE and MAE. Similarly, the CNN-LSTM from this study outperformed the previous studies on RMSE and MAE It is observed that the LSTM model from this study outperformed the MLP approach in [19] thus highlighting the predictive accuracy of the LSTM over CNN on Tropical Cyclone Intensity. This achievement is attributed to the use of Adam Optimiser, kernel size applied as shown in the proposed multilayered architecture approach in figures 2,3, and 4. Furthermore, the proposed model approached the prediction as a time series problem, thus, not only focusing on value of predictors but also the possible occurences before each prediction. This is why the model achieved accurate results on both datasets.

TABLE II
COMPARISON WITH PREVIOUS WORK

APPROACH	RMSE	MAE	MAPE
MLP [19]	11.07	8.22	-
CNN-LSTM	5.655	5.217	0.024
CNN	13.631	12.838	0.0058
LSTM	**4.450**	**4.049**	**0.0019**

IV. DISCUSSION AND CONCLUSION

Results obtained from the evaluation on the training set show that the LSTM model performed best (by recording lower errors) in terms of all metrics i.e. RMSE, MAE, MAPE in the majority of conducted experiments. This was followed by the hybrid CNN-LSTM model, and lastly, the CNN performed worse than the other 2 models. However, despite the LSTM recording the best accuracy, the CNN-LSTM performed significantly close to the LSTM, but in terms of speed is the slowest compared to the other 2 models. LSTM had the fastest performance followed by the CNN model. The conclusions drawn from analyzing the results obtained from the testing are:

- Performance difference was observed by the respective CNN, LSTM, and CNN-LSTM models on the two datasets, where on the SHIPS dataset, the models performed better on RMSE and MAE only, but for MAPE performance, the models performed better on the HURSAT dataset. This performance difference by the models can be attributed to the variation of the datasets i.e. length, and number of attributes. However, despite the performance difference on RMSE, MAE and MAPE, it was observed that the performance pattern for each model remained the same i.e LSTM, CNN, and CNN-LSTM achieved better testing results than training results on both datasets.
- Among the three models experimented on in this study, the LSTM model produced the best results, followed closely by the hybrid (combination of CNN and LSTM) and lastly the CNN.

- Despite having a minimal error rate compared to CNN, and the LSTM models, the hybrid CNN-LSTM model is the slowest, thus being more computationally expensive. This is followed by the CNN and lastly the LSTM model. The LSTM model having the lowest prediction error is the fastest of the models then followed by the CNN.
- The difference between all the three models in terms of error mostly ranges between 1 to 6 kts, this implies that despite the difference in terms of the three model's performance, they all perform well as the gap between them in terms of error rate is not wide.

Considering that the modeling process plays a vital role in the produced models' performance, the approach employed by this study in modeling as a Data Mining problem(by extracting features from NOAA NetCDF to CSV) has yielded positive outcomes by producing better results compared to studies such as [19]. However, a major limitation of the employed approach is the time delay between obtaining the NETCDF files to the prediction outcome. The several steps involved such as the extraction of features from the dataset into CSV and approaching the prediction as a time-series problem consumes more time than approaching the prediction as a computer vision problem where data is extracted directly from the satellite image. In the future, we intend to further improve this study's approach by applying it on multiple data streams to predict TC intensity along with Minimum Central Pressure(MCP) and TC track. Also, we intend to test other types of deep learning models to evaluate how they perform on the above prediction problems. In particular, we would like to determine how closely related the TC intensity is with the TC track. This would improve the ability to efficiently predict both TC intensity and TC track would not only help with warning on Intensity but also on the track of the TC.

REFERENCES

[1] T. Schott, C. Landsea, G. Hafele, J. Lorens, A. Taylor, H. Thurm, B. Ward, M. Willis, and W. Zaleski, "The saffir-simpson hurricane wind scale," *NOAA/National Weather Service [Internet]*, pp. 1–4, 2012.

[2] S. Alessandrini, L. D. Monache, C. M. Rozoff, and W. E. Lewis, "Probabilistic prediction of tropical cyclone intensity with an analog ensemble," *Monthly Weather Review*, vol. 146, pp. 1723–1744, 6 2018.

[3] J. P. Kossin, K. R. Knapp, T. L. Olander, and C. S. Velden, "Global increase in major tropical cyclone exceedance probability over the past four decades," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, 6 2020.

[4] S. Kotal, S. Roy Bhowmik, P. Kundu, and A. Das Kumar, "A statistical cyclone intensity prediction (scip) model for the bay of bengal," *Journal of earth system science*, vol. 117, pp. 157–168, 2008.

[5] J. Xu and Y. Wang, "A statistical analysis on the dependence of tropical cyclone intensification rate on the storm intensity and size in the north atlantic," *Weather and Forecasting*, vol. 30, no. 3, pp. 692–701, 2015.

[6] M. B. Richman, L. M. Leslie, H. A. Ramsay, and P. J. Klotzbach, "Reducing tropical cyclone prediction errors using machine learning approaches," vol. 114. Elsevier B.V., 2017, pp. 314–323.

[7] S. Kumar, A. Dube, R. Ashrit, and A. K. Mitra, "A machine learning (ml)-based approach to improve tropical cyclone intensity prediction of ncmrwf ensemble prediction system," *Pure and Applied Geophysics*, vol. 180, pp. 261–275, 1 2023.

[8] M. Kim, M. S. Park, J. Im, S. Park, and M. I. Lee, "Machine learning approaches for detecting tropical cyclone formation using satellite data," *Remote Sensing*, vol. 11, 5 2019.

[9] R. Bose, A. L. Pintar, and E. Simiu, "Simulation of atlantic hurricane tracks and features: A deep learning approach," 8 2022. [Online]. Available: http://arxiv.org/abs/2209.06901

[10] W. Xu, K. Balaguru, A. August, N. Lalo, N. Hodas, M. DeMaria, and D. Judi, "Deep learning experiments for tropical cyclone intensity forecasts," *Weather and Forecasting*, vol. 36, no. 4, pp. 1453–1470, 2021.

[11] T. Haiden, M. Janousek, F. Vitart, Z. B. Bouallègue, L. Ferranti, F. Prates, and D. Richardson, *Evaluation of ECMWF forecasts, including the 2018 upgrade*. European Centre for Medium Range Weather Forecasts Reading, UK, 2018.

[12] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A survey of deep learning and its applications: A new paradigm to machine learning," *Archives of Computational Methods in Engineering*, vol. 27, pp. 1071–1092, 9 2020.

[13] I. H. Sarker, "Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions," *SN Computer Science*, vol. 2, 11 2021.

[14] R. Tang, F. Zeng, Z. Chen, J. S. Wang, C. M. Huang, and Z. Wu, "The comparison of predicting storm-time ionospheric tec by three methods: Arima, lstm, and seq2seq," *Atmosphere*, vol. 11, 4 2020.

[15] P. Ban, L. Guo, Z. Zhao, S. Sun, T. Xu, Z. Xu, and F. Sun, "A regional ionospheric storm forecasting method using a deep learning algorithm: Lstm," *Space Weather*, vol. 21, 3 2023.

[16] Z. Zhang, X. Yang, L. Shi, B. Wang, Z. Du, F. Zhang, and R. Liu, "A neural network framework for fine-grained tropical cyclone intensity prediction," *Knowledge-Based Systems*, vol. 241, 4 2022.

[17] M. Pineda, Q. Wang, W. Liao, M. McGuire, and W. Yu, "A multi-model multi-task learning system for hurricane genesis prediction," in *International Conference on Software Engineering Research and Applications*. Springer, 2022, pp. 113–129.

[18] R. Chen, X. Wang, W. Zhang, X. Zhu, A. Li, and C. Yang, "A hybrid cnn-lstm model for typhoon formation forecasting," *GeoInformatica*, vol. 23, pp. 375–396, 7 2019.

[19] W. Xu, K. Balaguru, A. August, N. Lalo, N. Hodas, M. DeMaria, and D. Judi, "Deep learning experiments for tropical cyclone intensity forecasts," *Weather and Forecasting*, vol. 36, no. 4, pp. 1453 – 1470, 2021. [Online]. Available: https://journals.ametsoc.org/view/journals/wefo/36/4/WAF-D-20-0104.1.xml

[20] C. W. Landsea and J. L. Franklin, "Atlantic hurricane database uncertainty and presentation of a new database format," *Monthly Weather Review*, vol. 141, no. 10, pp. 3576–3592, 2013.

[21] K. R. Knapp, M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, "The international best track archive for climate stewardship (ibtracs) unifying tropical cyclone data," *Bulletin of the American Meteorological Society*, vol. 91, no. 3, pp. 363–376, 2010.

[22] K. R. Knapp, S. Ansari, C. L. Bain, M. A. Bourassa, M. J. Dickinson, C. Funk, C. N. Helms, C. C. Hennon, C. D. Holmes, G. J. Huffman, J. P. Kossin, H.-T. Lee, A. Loew, and G. Magnusdottir, "Globally gridded satellite observations for climate studies," *Bulletin of the American Meteorological Society*, vol. 92, no. 7, pp. 893 – 907, 2011.

[23] M. Xiao, "Change in the occurrence frequency of landfalling and non-landfalling tropical cyclones over the northwest pacific," *Journal of Climate*, vol. 34, no. 8, pp. 3145 – 3155, 2021. [Online]. Available: https://journals.ametsoc.org/view/journals/clim/34/8/JCLI-D-20-0647.1.xml

[24] M. DeMaria, M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan, "Further improvements to the statistical hurricane intensity prediction scheme (ships)," *Weather and Forecasting*, vol. 20, no. 4, pp. 531 – 543, 2005.

[25] H. T. Taylor, B. Ward, M. Willis, and W. Zaleski, "The saffir-simpson hurricane wind scale," *Atmospheric Administration: Washington, DC, USA*, 2010.

[26] B.-F. Chen, B. Chen, H.-T. Lin, and R. L. Elsberry, "Estimating tropical cyclone intensity by satellite imagery utilizing convolutional neural networks," *Weather and Forecasting*, vol. 34, no. 2, pp. 447–465, 2019.

[27] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with lstm recurrent networks," *Journal of machine learning research*, vol. 3, no. Aug, pp. 115–143, 2002.

[28] S. Alemany, J. Beltran, A. Perez, and S. Ganzfried, "Predicting hurricane trajectories using a recurrent neural network," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 468–475.

[29] C. Bernard, "Vqa: Visual question answering."

[30] X.-H. Le, H. V. Ho, G. Lee, and S. Jung, "Application of long short-term memory (lstm) neural network for flood forecasting," *Water*, vol. 11, no. 7, p. 1387, 2019.

[31] A. Jierula, S. Wang, T.-M. Oh, and P. Wang, "Study on accuracy metrics for evaluating the predictions of damage locations in deep piles using artificial neural networks with acoustic emission data," *applied sciences*, vol. 11, no. 5, p. 2314, 2021.

[32] A. Alassadi and T. Ivanauskas, "Classification performance between machine learning and traditional programming in java," 2019.