Title: Cyclone Forecasting Using Machine Learning

Prepared By: Md Mahamud Hasan

Data Science with Python

Batch 2

Prepared For: **Dr. Tania Islam**

Assistant Professor

Department of Computer Science and Engineering

University of Barishal

Date: 16^{th} November, 2024.

Cyclone Forecasting Using Machine Learning

Abstract: Meteorologists have long been concerned about cyclones, and over the past century, several studies have been conducted on their axisymmetric structures, dynamic mechanisms, and forecasting methods. This study illustrates the continuous advancement and the numerous lingering issues. As a type of artificial intelligence, machine learning has been approved by many researchers believe they can offer a fresh approach to addressing tropical cyclone bottlenecks, predictions, whether utilizing a model that is solely based on data or enhancing numerical models by adding machine learning. This review presents advancements based on machine learning in genesis forecasts, track forecasts, intensity forecasts, extreme weather forecasts associated with tropical cyclones (such as strong winds and rainstorms, and their disastrous impacts), storm surge forecasts, and numerical forecast models by analyzing and summarizing the difficulties of tropical cyclone forecasting in recent years and successful cases of machine learning methods in these aspects. These can all be seen as challenges as well as opportunities. The chance is that, as of right now, both the full capability of machine learning and a significant amount of multi-source data have not been properly employed to increase the accuracy of tropical cyclone forecasting. Because tropical cyclones differ from regular weather events and oceanographic processes, have intricate dynamic mechanisms, and are readily influenced by a wide range of factors, it can be challenging to guarantee the predictable period and stability of tropical cyclone prediction.

1. Introduction: Cyclone are regarded as extreme weather events, along with gales, rainstorms, and storm surges, which can cause huge losses in coastal areas worldwide. In the past century, numerous meteorologists and warning centers devoted themselves to this study and made progress in observational technology, intensification physics; interactions of the atmospheric environment, the atmospheric boundary layer and air-sea interface, the ocean responses, and forecasting techniques. However, many problems with predictive skills remain, particularly with the cyclone genesis, intensity, and risk forecasts. Generally, the most popular tropical cyclone dynamical forecast models have a relatively low accuracy, which is mainly due to the inaccurate vortex initialization of cyclone, incomplete representation of complex physical processes, and coarse resolution. There are studies that show that insufficient representations of the air–sea energy exchange under very high wind speed conditions would hinder simulating the intensity of cyclone more effectively. In addition, there is also a clear view that upper ocean feedback has important effects on cyclone, but few operational numerical forecast models take it into consideration, which also reduces the performance of the models. Additionally, other methods, such as statistical models, also are unable to deal with the complex and nonlinear relationship between cyclone-related predictors; thus, their forecast results need to be further improved. In order to solve these problems of traditional methods, scientists began to consider using machine learning (ML) to explore satellite, radar, in-situ data, cyclone. to improve the forecast skills of cyclone in recent years. Machine learning algorithms, as a means of artificial intelligence (AI), can be divided into three categories according to their applications: feature selection, clustering, and regression or classification. Feature selection algorithms can eliminate irrelevant attributes through attribute selection to increase the effectiveness of the tasks, and then improve the accuracy of the models. For example, a typical Tucker decomposition method can solve the spatio-temporal problems that the traditional tensor decomposition algorithm cannot deal with. A clustering algorithm is one of the earliest methods used in pattern recognition and data mining and can automatically divide a sample dataset into multiple categories. This has a wide range of applications in big data analysis. Typical clustering algorithms include the finite-mixed model (FMM), hierarchical clustering , and K-means algorithm. As for classification or regression, one representative algorithm is support vector machine (SVM) for classification [16] and support vector regression (SVR) for regression, which can effectively deal with nonlinear problems by defining kernel functions. In addition, decision tree (DT) is another typical algorithm that can mine and display the rules of classification, with high accuracy. A majority of works done with those mapping tasks are well performed with artificial neural networks (ANNs), which are considered as universal approximators for complex nonlinear mappings. Since Hinton, a leading scholar of machine learning, put forward the deep neural network model in 2006, a new era of deep learning was opened. A deep neural network that contains many hidden neural layers and is excellent in feature learning, can overcome the difficulty of training through layer-by-layer initialization, and can achieve overall optimization of the network. The classic networks include convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Compared with the traditional machine learning algorithm, deep learning (DL) has advantages in high-dimensional data, which is more suitable for complex applications. Therefore, choosing the appropriate machine learning algorithm for different data and different needs can solve practical problems more effectively. There are still many difficulties in tropical cyclone forecasts, such as an insufficient understanding of the physical mechanisms and the complex interactions with the ocean and surrounding atmosphere environment. All of these will hinder the prediction of cyclone genesis, tracks, intensity, and associated disastrous weather. Machine learning proved able to provide a new way to improve the accuracy and efficiency of cyclone prediction. Although there may be difficulties in the current stage of machine learning in long-lead-time forecasts, in the development of a generic and interpretable ML-based cyclone forecast model, and in improving the numerical cyclone model itself, there is still a bright future for machine learning in cyclone forecasts because the explosive growth of multi-source data and efficient machine learning algorithms have not been well utilized. The contribution of this paper is to analyze and summarize the successful cases of machine learning in tropical cyclone forecast modeling in recent years, and then introduce them separately according to different predictive purposes. Further opportunities and challenges of machine learning in cyclone forecasts are described at the end, and we aim to improve the status of cyclone forecasts.

2. Machine Learning

2.1. A Brief Introduction to Machine Learning: Machine learning is a series of computer programs, and its core task is to build mathematical models using statistics to make inferences from samples. Given a model that defines certain parameters, learning is the execution of a computer program that uses training data or experiences to optimize the parameters of the model. The model can predict the future state, or describe the knowledge from the data, or both. The practical process of machine learning methods (see Figure 1) can be summarized as follows: (1) Define a problem to an unknown mapping f and set a hypothetical set H of the solving model. (2) Collect and organize a training set D with a finite set. (3) Specify the loss function for the model. (4) Select the learning algorithm A. (5) Obtain the parameters that make the loss function feCycloneh the pole hour and choose them as the optimal parameters of the model. (6) Save this model g with the optimal parameters, and use it to make predictions and analysis of new data. Machine learning algorithms can also be divided into several categories according to the learning tasks, such as prediction, feature selection, and dimensionality reduction. As this review focuses on CYCLONE forecast modeling, only predictive algorithms will be described here. Generally, if the goal of the model is to predict discrete values, this kind of learning task is called "classification"; if it is to predict continuous values instead, this learning task is called "regression". In addition, learning tasks can also be broadly classified into "supervised learning" and "unsupervised learning" depending on whether the training data are labeled or not, with classification and regression representing the former and clustering representing the latter. The prediction task intends to establish a mapping f from the input space X to the output space Y, $f : X \rightarrow Y$, and f depends on a vector of nonlinear (in general) parameters, $w : y = f(x, w)$. The parameters w are obtained in the process of training, which, for the classification or regression/mapping problem, is an optimization of the performance criterion (e.g., a minimization of the mean square error). Of course, a machine learning algorithm itself may have additional parameters (hyperparameters), such as the number of hidden neurons and learning rates for neural networks. The selection of hyperparameters also plays a crucial role in training an appropriate machine learning-based forecast model.

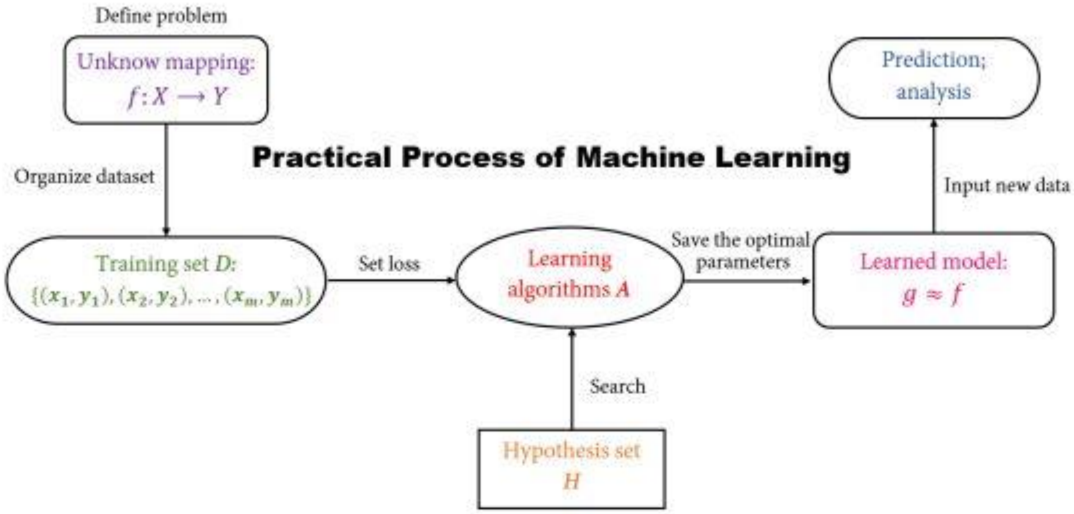


Figure 1. This is the practical process of machine learning.

2.2. An Overview of Machine Learning in Cyclone Forecasts Cyclone forecasts focus on the prediction of the central location and intensity of Cyclone, as well as on the effects of catastrophic weather when they make landfall or come close to shore, and the forecast techniques are mainly empirical forecasting, statistical forecasting, and numerical forecasting models. In recent years, path prediction has made huge progress, apart from some abnormal paths, normal path prediction can achieve relatively accurate results. Existing techniques for predicting the genesis and intensity of Cyclone are still very limited, and the reasons include (1) Cyclone genesis and intensity are not as well defined as Cyclone location, and (2) the physical processes involved are more complex and difficult to describe precisely by statistical models or dynamic equations. In addition, the forecasts of high winds, rainstorms, and storm surges in Cyclone - affected areas highly rely on accurate trajectories and intensity forecasts, which makes their current forecasts even more worrisome. For machine learning, as mentioned in Section 2.1, the major application is to perform predictions. This type of method discovering rules from data and unconstrained by physics is particularly suitable for solving problems where the physical mechanisms are unclear, such as Cyclone changes. Therefore, if there are sufficient historical Cyclone samples and a large amount of relevant meteorological and oceanic data, machine learning is expected to accurately predict Cyclone events. Of course, the explosive growth of satellite data, observational data, and re-analysis data offers tremendous opportunities for machine learning in Cyclone forecasts. As shown in Figure 2, the applications of machine learning in Cyclone forecasts can be divided into five aspects. Regarding Cyclone genesis forecasts, the final goals are to generate probabilistic forecasts of a fixed region in real time and quantitative forecasts in the time and place of cyclone genesis, so as to better monitor the tropical ocean. However, at this stage, machine learning is only capable of predicting whether the precursors can evolve into Cyclone, and the seasonal frequency of Cyclone genesis in each area, which corresponds to a classification task and regression task in machine learning, respectively. Thus, researchers primarily use several typical algorithms, including DT, logistic regression (LR), SVM, and ensemble algorithms, like AdaBoost and random forest (RF), for Cyclone genesis prediction. These ensemble algorithms are theoretically better than a single algorithm; however, they still need to be judged on a case-by-case basis. Additionally, deep learning algorithms, such as multi-layer perceptrons (MLPs) and CNNs, which can better fit complex functions and process image data, also play an important role in improving Cyclone genesis forecasting techniques. For Cyclone track forecasts, machine learning-based models are commonly derived from statistical learning methods, i.e., using the characteristics of the Cyclone itself and the associated meteorological and oceanic variables to predict the position of Cyclone, and this is considered nonlinear mapping. MLP and RNN, as the researchers' first choice, were also proven to be more effective than traditional methods. In addition, CYCLONEs change in both time and space; therefore, the spatio-temporal models, like ConvLSTM, have also achieved good results in this prediction task. In addition to those classic ideas, there is also the attempt of using generative adversarial networks (GANs) to generate predicted CYCLONE cloud images and then locate the CYCLONE center on them, or the use of deep belief networks (DBNs) and clustering algorithms like FMM to search for historically similar paths for forecasting. Each of these methods achieved good experimental results in their papers, but it remains to be seen whether better results can be achieved in operational forecasting than with existing techniques. DT, as an algorithm capable of mining rules, can be well used to mine the predictive factors and rules for CYCLONE landfall and recurvature, thus laying the foundation for track forecast modeling in the future.

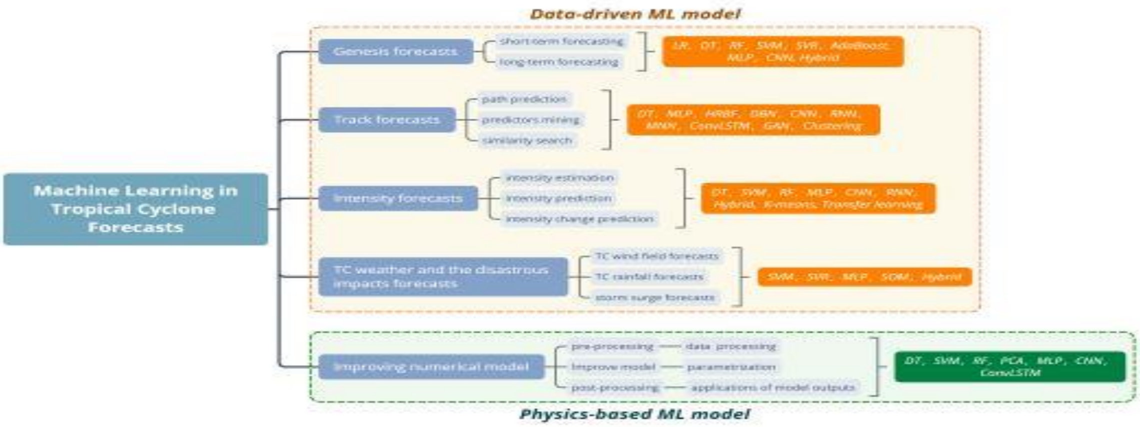


Figure 2. An organization chart of cases involving machine learning in tropical cyclone forecasts. The abbreviations used in this figure are as follows: logistic regression (LR), decision tree (DT), random forest (RF), support vector machine (SVM), support vector regression (SVR), multi-layer perceptron (MLP), convolutional neural network (CNN), generative adversarial network (GAN), recurrent neural network (RNN), hybrid radial basis function network (HRBF), self-organizing map (SOM), principal component analysis (PCA), convolutional long short-term memory network (ConvLSTM).

Apart from the pure data-driven machine learning methods mentioned above for CYCLONE genesis, tracks, intensity, and disastrous weather impact forecasts, there is another way to improve forecast results by developing a physics-based machine learning model. Although the existing techniques are not sufficient to comprehensively improve numerical forecast models through machine learning, there have been some successful cases regarding this topic. Here, we will briefly divide them into three categories. The first is pre-processing, which includes the quality control of data used to construct the initial field of the model. For example, SVM can be used to determine if there is a CYCLONE region in the data, either by eliminating it, or by performing special processing to improve the quality of the data. The second is the improvement of the model itself, including model error correction and an improved parameterization scheme. Existing studies used only MLP and CNN to quantify sea surface temperature cooling (SSCYCLONE) induced by CYCLONEs in Weather Research and Forecasting (WRF) to improve the numerical forecasts of CYCLONE intensity, or they used improvements in parameterizing the CYCLONE wind field based on RF and principal component analysis (PCA). The third category is post-processing, which includes model output corrections and applying for numerical products. However, only applications of numerical model products for predicting CYCLONE genesis, paths, intensity, rapid intensification, eCyclone., have been determined currently, and, to the best of our knowledge, no studies on revising the CYCLONE numerical forecast results have been found. Therefore, there are only some preliminary studies for machine learning in this aspect, and there is still room for improvement.

3. Methodology

3.1 Data Collection

Gathered historical cyclone data from sources like NOAA and IMD, including cyclone tracks, wind speeds and satellite imagery and other important features. In this project data were collected from Kaggle. Data looked like:

	name	year	month	day	hour	lat	long	status	category	wind	pressure	tropicalst	hurricane_force_diameter
1	Amy	1975	6	27	0	27.5	-79	tropical de	NA	25	1013	NA	NA
2	Amy	1975	6	27	6	28.5	-79	tropical de	NA	25	1013	NA	NA
3	Amy	1975	6	27	12	29.5	-79	tropical de	NA	25	1013	NA	NA
4	Amy	1975	6	27	18	30.5	-79	tropical de	NA	25	1013	NA	NA
5	Amy	1975	6	28	0	31.5	-78.8	tropical de	NA	25	1012	NA	NA
6	Amy	1975	6	28	6	32.4	-78.7	tropical de	NA	25	1012	NA	NA
7	Amy	1975	6	28	12	33.3	-78	tropical de	NA	25	1011	NA	NA
8	Amy	1975	6	28	18	34	-77	tropical de	NA	30	1006	NA	NA
9	Amy	1975	6	29	0	34.4	-75.8	tropical st	NA	35	1004	NA	NA
10	Amy	1975	6	29	6	34	-74.8	tropical st	NA	40	1002	NA	NA
11	Amy	1975	6	29	12	33.8	-73.8	tropical st	NA	45	1000	NA	NA
12	Amy	1975	6	29	18	33.8	-72.8	tropical st	NA	50	998	NA	NA
13	Amy	1975	6	30	0	34.3	-71.6	tropical st	NA	50	998	NA	NA
14	Amy	1975	6	30	6	35.6	-70.8	tropical st	NA	55	998	NA	NA
15	Amy	1975	6	30	12	35.9	-70.5	tropical st	NA	60	987	NA	NA
16	Amy	1975	6	30	18	36.2	-70.2	tropical st	NA	60	987	NA	NA
17	Amy	1975	7	1	0	36.2	-69.8	tropical st	NA	60	984	NA	NA
18	Amy	1975	7	1	6	36.2	-69.4	tropical st	NA	60	984	NA	NA
19	Amy	1975	7	1	12	36.2	-68.3	tropical st	NA	60	984	NA	NA
20	Amy	1975	7	1	18	36.7	-67.2	tropical st	NA	60	984	NA	NA
21	Amy	1975	7	2	0	37.4	-66.7	tropical st	NA	60	984	NA	NA

storms

Ready

3.2 Data Calling

For loading data in our our model this code was written and ouCycloneome looked like this

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score, confusion_matrix,
classification_report, f1_score
from sklearn.preprocessing import LabelEncoder
from sklearn.impute import SimpleImputer
import seaborn as sns

data = pd.read_csv('storms.csv')
print(data)
```

```
       Unnamed: 0   name  year  month  day  hour   lat  long  \
0               1    Amy  1975      6   27     0  27.5 -79.0
1               2    Amy  1975      6   27     6  28.5 -79.0
2               3    Amy  1975      6   27    12  29.5 -79.0
3               4    Amy  1975      6   27    18  30.5 -79.0
4               5    Amy  1975      6   28     0  31.5 -78.8
...           ...    ...   ...    ...  ...   ...   ...   ...
19061       19062  Wanda  2021     11    7     0  37.4 -37.4
19062       19063  Wanda  2021     11    7     6  38.1 -36.4
19063       19064  Wanda  2021     11    7    12  39.2 -34.9
19064       19065  Wanda  2021     11    7    18  40.9 -32.8
19065       19066  Wanda  2021     11    8     0  43.2 -29.7

                    status  category  wind  pressure  \
0      tropical depression       NaN    25      1013
1      tropical depression       NaN    25      1013
2      tropical depression       NaN    25      1013
3      tropical depression       NaN    25      1013
4      tropical depression       NaN    25      1012
```

```
       tropicalstorm_force_diameter  hurricane_force_diameter
0                               NaN                       NaN
1                               NaN                       NaN
2                               NaN                       NaN
3                               NaN                       NaN
4                               NaN                       NaN
...                             ...                       ...
19061                          60.0                       0.0
19062                          60.0                       0.0
19063                          90.0                       0.0
19064                          90.0                       0.0
19065                          70.0                       0.0

[19066 rows x 14 columns]
```

3.3 Data Preprocessing

For this project, we have to see the info about data, so if there were any missing values for required data.

```
1   data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19066 entries, 0 to 19065
Data columns (total 14 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Unnamed: 0                    19066 non-null  int64
 1   name                          19066 non-null  object
 2   year                          19066 non-null  int64
 3   month                         19066 non-null  int64
 4   day                           19066 non-null  int64
 5   hour                          19066 non-null  int64
 6   lat                           19066 non-null  float64
 7   long                          19066 non-null  float64
 8   status                        19066 non-null  object
 9   category                      4684 non-null   float64
 10  wind                          19066 non-null  int64
 11  pressure                      19066 non-null  int64
 12  tropicalstorm_force_diameter  9554 non-null   float64
 13  hurricane_force_diameter      9554 non-null   float64
dtypes: float64(5), int64(7), object(2)
memory usage: 2.0+ MB
```

for knowing various statistical value this description is done:

```
1   data.describe()
```

	Unnamed: 0	year	month	day	hour	lat	long	category	wind	pressure	tropicalstorm_force_
count	19066.000000	19066.000000	19066.000000	19066.000000	19066.000000	19066.000000	19066.000000	4684.000000	19066.000000	19066.000000	955
mean	9533.500000	2002.278926	8.698626	15.781968	9.094042	26.994252	-61.524300	1.898377	50.017413	993.554390	14
std	5504.024452	12.556517	1.352956	8.878563	6.733683	10.414302	21.062519	1.150590	25.501030	18.737342	15
min	1.000000	1975.000000	1.000000	1.000000	0.000000	7.000000	-109.300000	1.000000	10.000000	882.000000	
25%	4767.250000	1993.000000	8.000000	8.000000	5.000000	18.400000	-78.700000	1.000000	30.000000	987.000000	
50%	9533.500000	2004.000000	9.000000	16.000000	12.000000	26.600000	-62.250000	1.000000	45.000000	1000.000000	11
75%	14299.750000	2012.000000	9.000000	24.000000	18.000000	33.700000	-45.600000	3.000000	65.000000	1007.000000	22
max	19066.000000	2021.000000	12.000000	31.000000	23.000000	70.700000	13.500000	5.000000	165.000000	1024.000000	144

3.4 Model Selection

Evaluate various machine learning algorithms (e.g., Random Forest, SVM, Neural Networks) for prediction. For this model I selected Gaussian Naïve Bayes model.

```python
data = data[['long', 'lat', 'hour', 'wind', 'pressure', 'status']]

numeric_columns = ['long', 'lat', 'hour', 'wind', 'pressure']
categorical_columns = ['status']

label_encoder = LabelEncoder()
data['status'] = label_encoder.fit_transform(data['status'])
imputer = SimpleImputer(strategy='mean')
data[numeric_columns] = imputer.fit_transform(data[numeric_columns])

X = data[['long', 'lat', 'hour', 'wind', 'pressure']]
y = data['status']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=0)

model = GaussianNB()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy}")

exp_data = [[0, 1, 1, 0, 1]]
predicted_label = model.predict(exp_data)
predicted_status = label_encoder.inverse_transform(predicted_label)

print(f"Predicted status: {predicted_status[0]}")
```

3.5 Model Training and Testing

Split the dataset into training (80%) and testing (20%) sets, train the model, and evaluate its performance using standard metrics.

3.6 Project Outcome

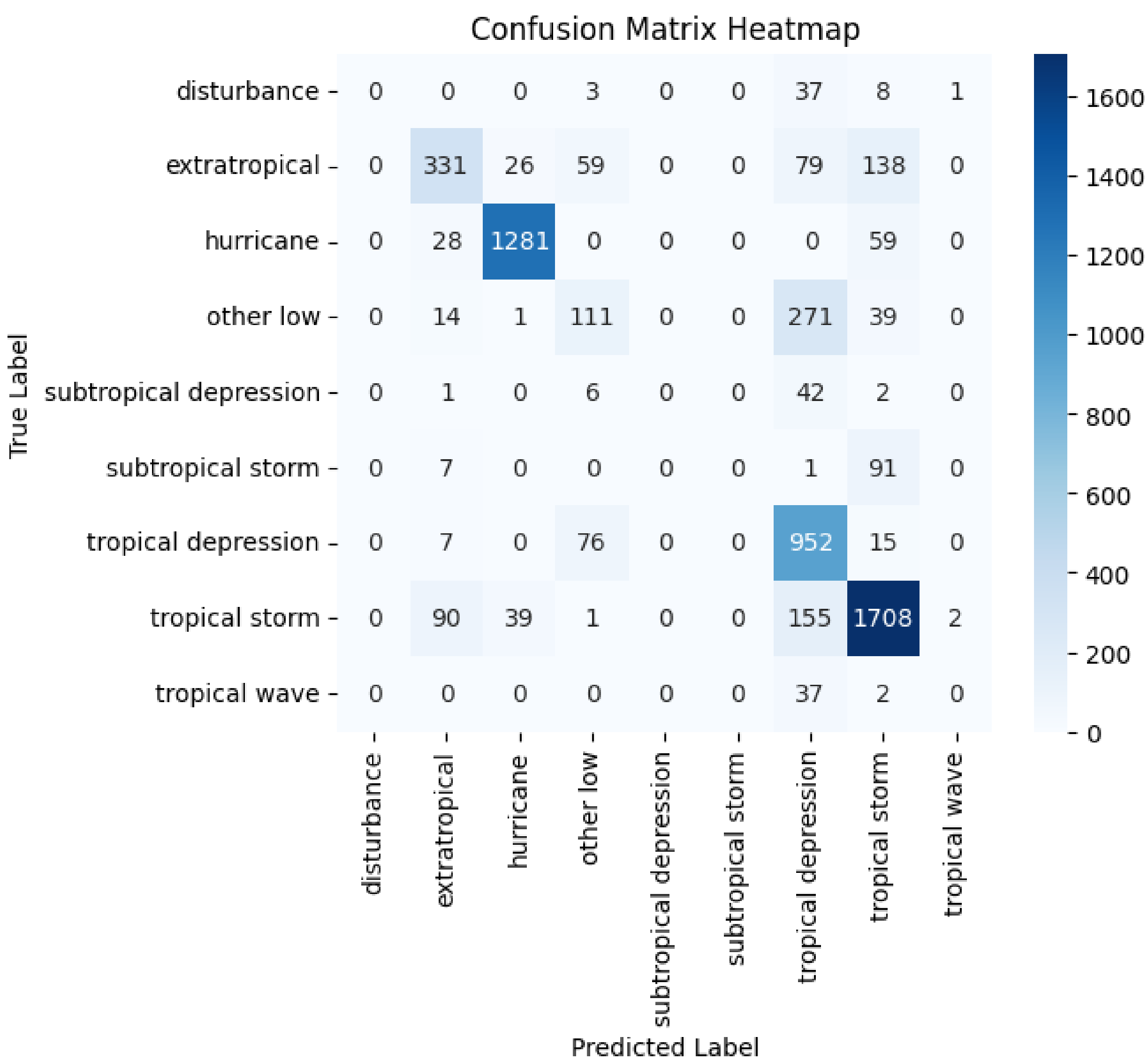
```
Accuracy: 0.7662587412587413
Predicted status: hurricane
```

3.6 Confusion matrix heatmap

A confusion matrix is a performance measurement tool used in classification problems to evaluate the accuracy of a machine learning model.

```
cm = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:")
print(cm)
```

```
Confusion Matrix:
[[   0    0    0    3    0    0   37    8    1]
 [   0  331   26   59    0    0   79  138    0]
 [   0   28 1281    0    0    0    0   59    0]
 [   0   14    1  111    0    0  271   39    0]
 [   0    1    0    6    0    0   42    2    0]
 [   0    7    0    0    0    0    1   91    0]
 [   0    7    0   76    0    0  952   15    0]
 [   0   90   39    1    0    0  155 1708    2]
 [   0    0    0    0    0    0   37    2    0]]
```



4. Discussion Machine learning has achieved progress in different aspects, including cyclone genesis, tracking, intensity, cyclone weather, and impact forecasts; however, many difficult and unsolvable problems remain in this field. This new way of thinking and new method of coping with key issues are both opportunities and challenges for cyclone forecasters and researchers.

4.1. Opportunities (1) At present, the large-scale satellite data, reanalysis data, and observations data have not been fully developed and utilized, and machine learning has proven able to effectively detect and study various problems in remote sensing. Therefore, the prediction of cyclone based on multi-source data, especially real-time data from satellites, is a promising topic. (2) There are many bottlenecks in cyclone forecasts, such as the quantitative forecasts of cyclogenesis, the prediction of cyclone with anomalous paths or rapid intensification, the intense precipitation caused by cyclone, and cyclone wind field forecasts. Although there have been some preliminary attempts, they could not yet meet the requirements of operational forecasts, and these require further exploration in the future. (3) It cannot be denied that numerical forecasts are still the dominant means of cyclone forecasting at this stage, and their importance should not be ignored. To improve the performance of numerical models in cyclone forecasts, machine learning could be used for an integration with numerical models, including improving the parameterization scheme, replacing the sub-models represented by empirical formulas, or revising the deviation of outputs of the numerical models. (4) Although numerical models are currently irreplaceable, they are expensive to run and have many physical processes that cannot be expressed by equations. Error propagation during model solving is also a big problem that contributes to poor forecasting. Therefore, the development of a pure data-driven cyclone prediction system that ensures high efficiency and low cost, while providing more accurate and stable cyclone predictions, is also the focus of future research by researchers.

4.2. Challenges (1) The inexplicability of machine learning, especially deep learning, has been discredited by many experts in the meteorological field for its ability to predict cyclone in a realistic and stable manner, because they consider that machine learning only discovers rules hidden within the data, and is detached from the real physical rules. How to make machine learning-based predictive models more stable and the prediction process more convincing is a new challenge. (2) Machine learning in existing studies typically only provides short-lead-time predictions, or the accuracy does not meet expectations when making long-lead-time predictions. How to make longer and more accurate predictions is a bottleneck that needs to be overcome in the near future. (3) A cyclone is a rare but extremely fast-changing and complex system with a large part of its lifetime over oceans, so for cyclone, in-situ observations are scarce. Although there are a great deal of reanalysis data and satellite observations, as well as some airbone reconnaissance data available to study cyclone, this is far from being enough to help us understand how real cyclone happen and change. Therefore, to prove whether machine learning can capture the internal rules inside cyclone based on limited data to make reliable predictions will still require extensive experiments and observations. (4) The majority of machine learning-based cyclone predictive models at this stage are supervised learning methods, and cyclone, an extreme weather phenomenon that cannot be quantitatively described in the real world, cannot be used directly as labels. How to reasonably construct labeled data and training datasets to train machine learning models to achieve predictive goals is also a question worth pondering. How unsupervised and semi-supervised learning methods should be effectively used in cyclone forecasting also requires further research.

5. Conclusions Cyclones have been a concern of meteorologists for more than 100 years. Numerous scholars have conducted in-depth studies on key issues, such as the structure, dynamics, and forecasting techniques. Machine learning is derived from statistical methods that can automatically discover relevant rules from large-scale data for detection, analysis, prediction, cyclone. The application of machine learning for the key problems of cyclone provides a new way of thinking to address many bottlenecks in this field. Techniques based on a pure data-driven approach and using machine learning to improve numerical models have both been shown by a large number of studies to provide huge contributions to improving cyclone predictions. Although existing research has made some progress in genesis forecasts, path prediction, intensity prediction, cyclone weather prediction, and improving numerical forecast models by integrating machine learning, there are still many aspects that remain to be studied, which we regard as both an opportunity and a challenge. The opportunity is that the potential of machine learning has not yet been exploited, and large-scale data are still underutilized. The challenge is that cyclone are different from normal weather phenomena and oceanic physical processes in that they are complex, subject to many factors, and it is difficult to obtain comprehensive in situ observations inside cyclone. We can conclude that machine learning in cyclone forecasts is both promising and challenging, which means that it requires researchers to have a good understanding of cyclone dynamics as well as a knowledge of machine learning in order to discover the key problems faced and to solve them by building suitable machine learning models. By analyzing and summarizing the studies of machine learning in cyclone forecasts over the years, we hope this review can provide readers with insight into this research and lay a foundation for future works regarding machine learning in cyclone forecast modelling.

6. References

Emanuel, K. 100 Years of Progress in Tropical Cyclone Research. Meteorol. Monogr. 2018, 59, 15.1–15.68. [CrossRef]

Ma, L.-M. Research Progress on China typhoon numerical prediction models and associated major techniques. Prog. Geophys. 2014, 29, 1013–1022. [CrossRef]

Tropical Cyclone Forecast Model. Available online: https://en.wikipedia.org/wiki/Tropical_cyclone_ forecast_model (accessed on 1 May 2020).

Moon, I.J.; Ginis, I.; Hara, T.; Thomas, B. A Physics-Based Parameterization of Air-Sea Momentum Flux at High Wind Speeds and Its Impact on Hurricane Intensity Predictions. Mon. Weather Rev. 2007, 135, 2869–2878. [CrossRef]

Bender, M.A.; Ginis, I.; Kurihara, Y.J. Numerical simulations of tropical cyclone-ocean interaction with a high-resolution coupled model. J. Geophys. Res. Atmos. 1993, 98, 23245–23263. [CrossRef]

Lee, C.Y.; Tippett, M.K.; Camargo, S.J.; Sobel, A.H. Probabilistic Multiple Linear Regression Modeling for Tropical Cyclone Intensity. Mon. Wea. Rev. 2015, 143, 933–954. [CrossRef]

Wang, Y.; Rao, Y.; Tan, Z.M.; Schönemann, D. A Statistical Analysis of the Effects of Vertical Wind Shear on Tropical Cyclone Intensity Change over the Western North Pacific. Mon. Weather Rev. 2015, 143, 3434–3453. [CrossRef]

DeMaria, M.; Kaplan, J. An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern North Pacific basins. Weather Forecast. 1999, 14, 326–337. [CrossRef]

DeMaria, M.; Kaplan, J.J.W. A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin. Weather Forecast. 1994, 9, 209–220. [CrossRef]

Demaria, M.; Mainelli, M.; Shay, L.K.; Knaff, J.A.; Kaplan, J. Further Improvements to the Statistical Hurricane Intensity Prediction Scheme (SHIPS). Weather Forecast. 2005, 20, 531–543. [CrossRef]

Machine Learning. Available online: https://en.wikipedia.org/wiki/Machine_learning (accessed on 9 April 2020).

Kim, Y.D.; Choi, S. Nonnegative tucker decomposition. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 1–8.

Melnykov, V.; Maitra, R. Finite mixture models and model-based clustering. Stat. Surv. 2010, 4, 80–116. [CrossRef]

Johnson, S.C. Hierarchical clustering schemes. Psychometrika 1967, 32, 241–254. [CrossRef] [PubMed]

Krishna, K.; Murty, M.N. Genetic K-means algorithm. IEEE Trans. Syst. Man Cybern. Part B (Cybernetics) 1999, 29, 433–439. [CrossRef] [PubMed]