Comparing and Evaluating Macao Flood Prediction Models

Zeyu Zhang¹, Jiayue Qiu¹, Xuefei Huang¹, Zhiming Cai², *, Linkai Zhu² and Weijun Dai²

¹School of Applied Sciences, Macao Polytechnic Institute, Macao, China
²Institute of Data Science, City University of Macau, Macao, China

*Corresponding author: P2009102@ipm.edu.mo

Abstract. Climate change causes extreme weather in Macao, especially typhoons and flooding. In this paper, some raw flood data is missing from the Macao Meteorological and Geophysical Bureau, due to some flood sensors that were damaged during Typhoon Hato in 2017 and Typhoon Mangkhut in 2018. So we use data interpolation to construct new datasets and curve fitting to simulate real inundation depth. Besides this, we explore Neural Network, Long Short-Term Memory, Random Forest, Adaptive Boosting, and Linear Regression for analyzing, comparing, and evaluating the best combinations of flood prediction models, datasets, and scenarios caused by typhoon presence in Macao. Furthermore, we apply Bayes Network to the aforementioned models to evaluate the accuracy of predicting flood situations because of typhoons. The experiment results show that the different models achieve a different performance in predicting specific scenarios.

Keywords: Typhoon, Flood Predicting, Machine Learning, Bayes Network.

1. Introduction
The floodings from typhoons not only injure and kill people but also destroy fundamental infrastructures and spread disease. Super Typhoon Hato (No.1713) brought extremely destructive winds and caused severe flooding in Macao [1]. It resulted in the tragic deaths of 10 people and injured 244 more in Macao. Also, during the passage of Super Typhoon Mangkhut (No.1822) the storm surge rose to 1.5 meters in Macao, about 21,000 homes lost power, and 7,000 homes lost internet access [2]. Therefore, it is essential to forecast floods for warning residents to increase security and make preparations.

The authors of [3] analyzed the reasons for the sharp increase of Typhoon Hato and the change process of hourly rainfall within 3 days of Typhoon Mangkhut landing in China, then assessed the risk and dynamic change of the rainstorm disaster [4]. The authors analyzed the influence of Typhoon Hato on Sanshui and Sihui of Foshan in Guangdong Province and determined the evolution of intensity [5]. The authors of [6] predicted the formation process of typhoons and improved rainfall forecast researching [7] by studying the spatial distribution of rainfall and the dynamic prediction of typhoons in Wenzhou, Zhejiang Province [8].

The samples of flood forecasting by typhoons are not enough in Macao for the following reasons. First, the water level monitoring stations were built up after 2015. Second, the sample size is low...
because Macao has only experienced two major typhoons in recent years, Typhoon Hato and Typhoon Mangkhut. Last but not least, the flooding caused some sensors to be damaged and fail to work.

In this paper, the private datasets are supplied by the Macao Meteorological and Geophysical Bureau (SMG), the Macau Maritime Administration (MMA), and the Meteorological Bureau of Shenzhen Municipality (MBSM). We use data interpolation to construct missing data. Then, we realize Random Forest, Adaboost, and Linear Regression to train dataset for categorizing the dataset by time density. All in all, we use the data of Typhoon Hato to train the Typhoon Hato model, and the trained model of Typhoon Mangkhut model separately using the data of Typhoon Hato and Typhoon Mangkhut.

![Figure 1. The flow chart of flood prediction.](image)

2. Related work
The authors built up the Neural Network model to train the data and used the LSTM model to predict the speed of the wind [9]. Meanwhile, the comprehensive learning method was introduced into the forecast, such as the AdaBoost model which added reducing sampling [10]. Through the testing result of rainstorm prediction, it was verified that the method can improve the accuracy of rainstorm prediction. Then, the authors built up the optimization model of Random Forest for training and testing the data [11]. Finally, they compared the accuracy of this model with other machine learning models. The result showed that this model has high recognition accuracy by introducing indoor localization about Linear Regression to train the history data for predicting the location of the destination [12]. The authors predicted data and decided to use the probability framework [13]. Although sometimes the value of observing is uncertain, the probability model can do this. Bayes optimization is the most effective way to solve this problem because it obtains an optimal approximate solution with less evaluation cost [14].

3. Dataset design
The private dataset comes from the Macao Meteorological and Geophysical Bureau (SMG), the Macau Maritime Administration (MMA), and the Meteorological Bureau of Shenzhen Municipality (MBSM).

Table 1 shows that the data is applied by MBSM including water level (Level), rainfall (PRES), wind direction (TG_WDIR), wind speed (WND), average wind speed (TG_WSPD), the tide of Shekou Port (SK_Tide), the tide of Shenzhen Airport (JCH_Tide), and the tide of DaPeng Peninsula (DAP Tide).

Table 2 displays the part of data of 11 water level monitoring stations supplied by SMG. First, we clean the original water level data which includes missing data reprocessing, data integration, and data
management. There are six water level monitoring stations which are the main locations of flooding and the water level in this area is low and vulnerable to reflux in Macao. Some data was missing from water level monitoring stations in Macao, so we use the interpolation method to construct new data points within known data points by analyzing the tide level data from MBSM.

| Date and Time          | Level | PRES | WND | TG WDIR | TG WSPD | SK Tide | JCH Tide | DAP Tide |
|------------------------|-------|------|-----|---------|---------|---------|----------|----------|
| 23/08/2017 9:00        | 3     | 979  | 28  | 348     | 34.6    | 2.76    | 2.97     | 2.14     |
| 23/08/2017 9:01        | 3     | 979  | 28  | 348     | 34.6    | 2.76    | 2.97     | 2.14     |
| 23/08/2017 9:02        | 3     | 979  | 28  | 348     | 34.6    | 2.76    | 2.97     | 2.14     |
| 23/08/2017 9:03        | 3     | 979  | 28  | 348     | 34.6    | 2.76    | 2.97     | 2.14     |
| 23/08/2017 9:04        | 3     | 979  | 28  | 348     | 34.6    | 2.76    | 2.97     | 2.14     |

Table 2. Typhoon Hato - part of raw data from water level monitoring stations in Macao.

| Date and Time          | Water Level(m) | LPI  | LSI  | LTA  | LPM  | LIV  | LAO  | LCH  | LPN  | LCL  |
|------------------------|----------------|------|------|------|------|------|------|------|------|------|
| 23/08/2017 9:00        | 0.0653         | -0.0021 | 0.0003 | NULL | 0.0003 | -0.0019 | -0.0030 | 0.0006 | 0.0000 |
| 23/08/2017 9:01        | 0.0662         | -0.0021 | 0.0003 | NULL | 0.0003 | -0.0018 | -0.0030 | 0.0006 | 0.0000 |
| 23/08/2017 9:02        | 0.0632         | -0.0021 | 0.0003 | NULL | 0.0003 | NULL  | -0.0030 | 0.0006 | 0.0000 |
| 23/08/2017 9:03        | 0.0628         | -0.0021 | 0.0003 | NULL | 0.0003 | 0.0000 | -0.0030 | 0.0006 | 0.0000 |
| 23/08/2017 9:04        | 0.0680         | -0.0021 | 0.0003 | NULL | 0.0003 | -0.0019 | -0.0030 | 0.0006 | 0.0000 |

4. Model selection
This section includes useful Machine Learning models to train data and the approaches to describe the relationship between data.

4.1. Neural Network
Neural Network (NN) applies signal forward propagation and error back propagation. In forward propagation, input data enter the input layer, then are processed in the hidden layer. Finally, prediction will be output from the output layer. If the real data from the output layer is distinguished from the prediction output, the error will be given feedback. The error back propagates through the output layer, hidden layer, and input layer to revise the weight value of every parameter.

4.2. Long Short-Term Memory
Long Short-Term Memory (LSTM) is designed to solve the vanishing gradient problem when trained Recurrent Neural Network (RNN). The LSTM adds three gates including input gate, output gate, and forget gate on the basis of RNN.

4.3. Random Forest
In Random Forest (RF), there is no correlation between decision trees. When building up decision trees, random sampling and complete splitting should be considered. It respectively samples the row data and column data, originally “put it back” when sample the row data. The amount of input sample is the same as output sample. After sampling, k features are chosen from all features N (k N), then the sample data is decomposed, and finally the decision trees are built up. The leaf node of the decision tree cannot split or point to the same classification. It avoids over fitting because of random sampling in the beginning.

4.4. Adaptive Boosting
Adaptive Boosting (AdaBoost) will classify the categories of training data in every training process. If the category of the sample is right, it will reduce the weighted value. Otherwise, it will add weighted
value. The accuracy of the category of the last training process will determine the weighted value of the sample in the training data set. Then, the new data set is modified by new weighted values to the next classifier to train the model. Finally, the combination of weak learners will access the higher accuracy.

4.5. Linear Regression
Linear Regression (LR) is an approach to indicate the relationship between independent variables and dependent variables. LR uses a math formula to describe a linear correlation between independent variables and dependent variables. As a result, the output value is estimated by substituting it into the math formula and the model accuracy is calculated.

4.6. Bayesian Network
Bayesian Network (BN) displays causality between events, such as the presence of a cause given the presence of an effect, and vice versa. New evidence overthrows the previous reasoning. Even though data is incomplete, reasoning can still be carried out.

5. Experiment
The water warning level definitions are in Table 3. In Table 3, the \( pl \) symbol represents predicted water level and the \( rs \) symbol represents road surface level.

| Level | Warning | Definition                        |
|-------|---------|-----------------------------------|
| 1     | Blue    | \( 0 < pl - rs < 0.5 \)           |
| 2     | Yellow  | \( 0.5 \leq pl - rs < 1 \)        |
| 3     | Orange  | \( 1 \leq pl - rs < 1.5 \)        |
| 4     | Red     | \( 1.5 \leq pl - rs < 2.5 \)      |
| 5     | Black   | \( pl - rs \geq 2.5 \)            |

5.1. NN Result
Fig. 2 shows the NN prediction of the water warning level from 11 water level monitoring stations in Macao.

5.2. LSTM Result
Fig. 3 shows the LSTM predictions of the water level of the Inner Harbor area by using the data of Typhoon Hato. Finally, the data of Typhoon Hato trains the Hato model by getting the highest
accuracy. Thus, BN is introduced into the Hato model for evaluation in LSTM Detection.

![LSTM Predicted Value vs Measured Value](image)

**Figure 3.** LSTM predicts the water level from 11 stations in Macao.

### 5.3. RF and AdaBoost Result

This experiment uses the library functions of RandomForestRegressor() and AdaBoostRegressor() to predict the water level of the Inner Harbor area, and train data using different time tenses. The comparison of results is shown in Fig. 4. The ideal model uses the data of Typhoon Mangkhut to train the Mangkhut model and the best time tension is five minutes. Finally, the Mangkhut model is evaluated using BN.

![Comparison of results](image)

**Figure 4.** Comparison of results.

### 5.4. LR Result

This experiment uses library function of LinearRegression() to predict the water level of the Inner Harbor area, and trains data using different time tenses. The results are shown in Fig. 5. The ideal model uses the data of Typhoon Mangkhut to train the Hato model, and the best time tension is 10 minutes. Finally, we evaluate the Mangkhut model using BN.

![Linear regression result](image)

**Figure 5.** Linear regression result.
5.5. BN Result

Table 4. Comparision of results.

| Model       | Accuracy | Warning Level |
|-------------|----------|---------------|
| NN          | 0.72     | 3, 4, 5       |
| LSTM        | 0.33     | 1             |
| RF          | 0.58     | 1, 2, 3       |
| AdaBoost    | 0        | NULL          |
| LR          | 0.360    | 3, 4          |

The previous experimental results show in Table 4 that Neural Network works best in predicting flooding warning levels 2, 3, 4. Random Forest Detection works best in levels 1, 2, 3. Linear Regression Detection works best in levels 3, 4. LSTM works best in levels 1, 2. Neural Network Detection works best in levels 3, 4, 5. As a result, we observe that the combination model of RF and LSTM suits low and moderate-risk areas in Fig. 6. The effect of the combination of NN and LR is better in moderate and high-risk areas. NN can be used in high-risk areas.

Figure 6. Bayesian network result.

6. Conclusion

This paper utilizes typhoon rainfall data in Macao which has post-typhoon flooding. The data is categorized by time density and realized Random Forest, Adaboost, Linear Regression. First, we use the data of Typhoon Hato to train the Typhoon Hato model, and the trained model of Typhoon Mangkhut model separately using the data of Typhoon Hato and Typhoon Mangkhut. Second, we analyze the result and build up the Macao rainfall spatial distribution model by typhoons and the dynamic change of the raining prediction model by typhoons. Third, the sample is tested by using an established prediction model to require the accuracy probability distribution. Fourth, the different parameters and permutation are utilized to boost the best method to predict the water immersion evaluation of the server flooding in Macao by using Bayes software Netica is drawn. As a result, we observe that the combination model of BP Neural Network and Linear Regression is better in middle and high-risk areas. BP Neural Network can be used in high-risk areas.

Acknowledgments

The research is supported by the joint NNSF&FDCT Project (0066/2019/AFJ) and joint MST&FDCT Project (0058/2019/AMJ).

National Natural Science Foundation of China and Macau Science and Technology Development Joint Fund under Grant 0066/2019/AFJ.
References

[1] Typhoon Hato has hit the strongest typhoon in Hong Kong, Zhuhai and Macao in 53 years. Hong Kong Economic Daily. 2017-08-23 [2017-08-25].

[2] Sum, Lok-kei (16 September 2018). “Typhoon Mangkhut: 20,000 Macau households left without power after monster storm causes severe flood- ing in casino hub”. South China Morning Post.

[3] Qin Li, Wu Qishu, Zeng Xiaotuan, et al. Analysis on the cause of the rapid intensification of convective asymmetric typhoon “Heaven” (1713) offshore [J]. Rainstorm Disaster, 2019, 38 (3): 212-220.

[4] Wang Xiaoya, Jiang Weiguo, Deng Yue, et al. Dynamic changes of hourly rainfall and dynamic assessment of hazards in areas affected by typhoon “Mangosteen” [J]. Catastrophology, 2019, 34 (3): 202-208.

[5] Huang Xianxiang, Yan Lijun, Wang Shuofu, et al. Observation and early warning analysis of the tornado process of Typhoon Mangkhut No. 1822 [J]. Journal of Tropical Meteorology, 2019, 35 (4): 458-469.

[6] MTC, DJC, SWJ, et al. Application of artificial neural networks in typhoon surge forecasting [J]. Ocean Engineering, 2007, 34 (11-12): 1757- 1768.

[7] Trafails T, Santosa B, Richman M. Learning Networks in Rainfall estimation [J]. Computational Management Science, 2005, 2 (3): 229-251.

[8] Wu Boqiang. Spatial distribution and dynamic prediction of typhoon rainfall, the main controlling factors of geological disasters in Wenzhou based on BP neural network [D]. Geological Engineering of Chang’an University, 2011.

[9] Li Bing, Zhang Yan, Liu Shi, et al. Research on short-term wind speed prediction based on LSTM [J]. Computer Simulation, 2018, 35 (11): 456- 461.

[10] Yang Yan, Yan Dongwei, Zhao Kui, Wei Ting. (2007). Application of comprehensive learning method AdaBoost in rainstorm prediction. Application of calculator system, 28 (1), 51-53. doi: 10.3969/j.issn.1003- 3254.2007.01.014.

[11] Zhao Dong, Zang Xuebai, Zhao Hongwei. (2017). Random for-est prediction method based on fruit fly optimization. Journal of Jilin University (Engineering Science Edition), 47 (2), 609- 614.doi:10.13229/j.cnki.jdxbgxb201702036.

[12] Wang Zhuxian, Guan Xingmin, Pan Xue. (2017). Research on wireless positioning technology based on local linear regression. Automation Technology and Application, 36 (12), 59-64. doi:10.3969/j.issn.1003-7241.2017.12.014

[13] Z. Ghahramani. (2015). Probabilistic machine learning and artificial intelligence. Nature, 521, 452-459.

[14] J. Snoek, H. Larochelle, R. Adams. (2012). Practical Bayesian optimiza- tion of machine learning algorithms. Advances in Neural Information Processing Systems, 21 (4), 2951-2959.