



OPEN

Harnessing LSTM and XGBoost algorithms for storm prediction

Ayyoub Frifra^{1,2}, Mohamed Maanan^{1✉}, Mehdi Maanan² & Hassan Rhinane²

Storms can cause significant damage, severe social disturbance and loss of human life, but predicting them is challenging due to their infrequent occurrence. To overcome this problem, a novel deep learning and machine learning approach based on long short-term memory (LSTM) and Extreme Gradient Boosting (XGBoost) was applied to predict storm characteristics and occurrence in Western France. A combination of data from buoys and a storm database between 1996 and 2020 was processed for model training and testing. The models were trained and validated with the dataset from January 1996 to December 2015 and the trained models were then used to predict storm characteristics and occurrence from January 2016 to December 2020. The LSTM model used to predict storm characteristics showed great accuracy in forecasting temperature and pressure, with challenges observed in capturing extreme values for wave height and wind speed. The trained XGBoost model, on the other hand, performed extremely well in predicting storm occurrence. The methodology adopted can help reduce the impact of storms on humans and objects.

Storms and extreme extratropical cyclones represent a significant risk to human life and the environment^{1–5}. They are the primary natural hazard affecting Western and Central Europe and frequently have a devastating impact on human activity and infrastructure^{6–8}. Most countries have experienced a rise in economic losses due to storms^{9,10}. However, extratropical cyclones are also necessary as they are responsible for the majority of precipitation in large parts of the world; over 70% of the total precipitation in Western and Central Europe and much of North America, for example, is associated with the passage of extratropical cyclones¹¹.

In Western France, several storms have resulted in significant human and material loss. On December 26 and 28, 1999, Storms Lothar and Martin caused considerable damage and many fatalities; 92 people were killed, and 3.5 million households were left without electricity for several weeks¹². In addition, much damage was done to forests and buildings, with roofs either totally or partially blown off¹³. Another tragic event, Storm Xynthia, occurred on February 28, 2010, leading to heavy loss of life and devastating damage to infrastructure. The combination of the storm and a high tide caused several sea walls to rupture, leading to extensive flooding in low-lying coastal zones, more than two billion euros' worth of damage, and 47 fatalities¹⁴.

The increased likelihood of extreme weather events and their risk¹⁵ have received considerable attention in recent decades^{16–18}. Predicting these events is highly challenging; Ensemble Prediction Systems (EPS) can generate short-term forecasts, for example the EPS at the European Centre for Medium-Range Weather Forecast (ECMWF) which predicted storms Anatol, Lothar, and Martin in December 1999¹⁹. Predicting storms with lead times of more than a couple of weeks can be achieved with seasonal forecast systems, such as the DEMETER (Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction) and ENSEMBLES (ENSEMBLE-based predictions of climate changes and their impacts) projects, which produce reliable forecasts of storms over the North Atlantic and Europe²⁰, and modern seasonal forecast systems from the ECMWF and the Met Office Hadley Centre, used to forecast the frequency of storms over the Northern Hemisphere²¹. Storms can also be predicted by forecasting a specific variable associated with storm events, as is the case with the local sea wave model of the German Weather Service, which can predict severe winter storms connected with extraordinarily high waves²².

The use of machine learning (ML) and deep learning (DL) models is increasingly common in the field of extreme weather events. Their main function is to predict storm characteristics; artificial neural networks combined with wavelet analysis, for example, are used to predict extreme wave height during storms²³. Evaluation of the results has indicated that ML and DL solutions provide better results than traditional methods in terms of computational expense and accuracy of predictions²⁴, and can serve as an alternative tool to conventional models for forecasting storms²⁵.

¹UMR 6554 CNRS LETG-Nantes Laboratory, Institute of Geography and Planning, Nantes University, 44312 Nantes, France. ²Geosciences Laboratory, Faculty of Sciences Ain Chock, Hassan II University of Casablanca, 20100 Casablanca, Morocco. ✉email: mohamed.maanan@univ-nantes.fr

In recent years, recurrent neural networks (RNNs) have attracted considerable attention as a time-series prediction tool^{26,27}. Despite the ability of RNNs to understand and capture short-term dependencies, they experience difficulty with long-term dependencies due to the vanishing gradient problem²⁸; the backpropagation error disappears or vanishes in the earlier inputs after propagating through several steps. Consequently, the standard RNN cannot learn effectively, and the information from earlier stages is ignored in the prediction task²⁹. To overcome this drawback and improve the memorization of RNNs, a new model, the LSTM algorithm, was developed as an extension to RNNs, and can learn long-term dependencies from time series data and resolve the vanishing gradient problem³⁰. The LSTM algorithm has performed very well with a large variety of issues. It has successfully resolved many sequence learning problems and time-series predictions, such as speech recognition³¹, machine translation³², and wind speed prediction^{33,34}. The LSTM algorithm has also been used for predicting extreme weather events, for example the storm waves induced by two winter storms in the US North Atlantic region²⁵. The model was trained using buoy measurements and the findings demonstrated that the LSTM model can accurately predict the characteristics associated with storm events²⁵. In a second study, the LSTM model proved its efficiency in predicting storm characteristics by forecasting significant wave height and wave period at storm peaks, based on data from two offshore buoys. The results demonstrate its ability to capture all major wave events, including storm peaks, and serve as a tool for early storm warning²⁴.

XGBoost is another robust ML algorithm widely used in many applications and given rave reviews by ML practitioners³⁵. This algorithm is a scalable ML-based technique for the tree boosting method introduced by Friedman³⁶. XGBoost, developed by Chen and Guestrin³⁷, has proven its efficiency in various classification and regression time series problems, such as rare-event classification; Ranjan et al. used highly imbalanced data from a pulp and paper mill as input for XGBoost to forecast paper break events³⁸. XGBoost is a flexible ML method capable of dealing with the non-linearity of time series with its efficient self-learning ability³⁹. It performs extremely well in time series prediction, with adequate computing time and memory resource usage⁴⁰. The XGBoost model has also been used to predict wind-wave conditions at storm peaks based on hourly wind, significant wave height, and peak wave period observations at buoy stations. The findings indicated the ability of XGBoost to predict wave dynamics under extreme conditions²⁴.

Prior research has investigated the use of LSTM and XGBoost models in storm prediction, as shown by studies conducted by Hu et al.²⁴ and Ian et al.⁴¹. Nevertheless, this study presents an innovative technique that deviates from current approaches. Prior investigations have mostly concentrated on using LSTM and XGBoost independently and comparing their performance in storm surge prediction, usually considering it as a regression problem. In contrast, the proposed approach in this paper cooperatively combines both models to forecast storm characteristics and occurrence. Specifically, LSTM was used for regression-based prediction of various storm characteristics, while XGBoost was used for classifying the days of storm occurrence. This integrated approach allows for a more thorough investigation of storm prediction. In addition, previous research has often focused on forecasting a limited number of storm variables, such as wave height or water level. This paper expands the forecast scope to include a wider variety of storm variables, such as wave height, wave period, wind speed, temperature, pressure, and humidity. In summary, this study builds upon existing literature by including a broader spectrum of storm characteristics and using an innovative and comprehensive method for forecasting storms, which will help enhance preparation and mitigation strategies.

With the main goal of providing a new data-driven approach to forecasting storm occurrence and characteristics, this article also aims to investigate the effectiveness of two of the best models currently available, XGBoost and LSTM, in terms of providing valuable predictions of rare events, such as storms. It then presents first the methods, followed by the results, discussion, and then conclusions.

Methods

The research roadmap involved two methods: LSTM and XGBoost. The first is designed to predict the different characteristics of storms, and the second to predict the occurrence of storms based on their characteristics.

Study site and data

The site for this study was the western coast of France, comprising four regions: Normandy, Brittany, Pays de la Loire, and Nouvelle-Aquitaine (Fig. 1). These regions have experienced unexpected storms, resulting in many human and material losses^{12,14}, highlighting the need for historical reconstruction of past storms. The study area has therefore been the subject of numerous studies to identify past storm dynamics, their relation to climatological mechanisms, their trajectories, and the spatialization of damage^{42,43}. These studies have identified two main storm paths: The first estimated from west to east and impacting a restricted area of the French Atlantic coast, the second more extensive, with a southwest-to-northeast trajectory affecting a larger section of the coast⁴². These results concur with the classification method known as Drevetton, which proposes a classification of storms based on the origin of the associated depression, either Atlantic or Mediterranean; those associated with low-pressure systems over the Atlantic are more frequent than those from the Mediterranean⁴⁴. Two classes of storm directly affect the Atlantic coast. The first relates to storms generated from depressions over the British Isles. The storm tracks, therefore, pass through Brittany then move up towards northern France, affecting either the north of France or the entire country. The other storm class relates to those generated from depressions over the Bay of Biscay. The trajectory of this second category of storms moves inland via the Pays de la Loire—Poitou—Charentes and continues towards the east of France and then Germany, affecting the northern part of France or the whole country. The studies conducted by Castelle et al.^{45,46} have greatly contributed to enhancing the comprehension of storm impacts and creating predictive tools for reducing coastal hazards in the study area. For example, their study of the effects of intense winter storms in 2013–2014 provides insights on the erosion patterns and the development of megacusps embayments⁴⁵. This research focuses on the correlation between storm wave attributes,

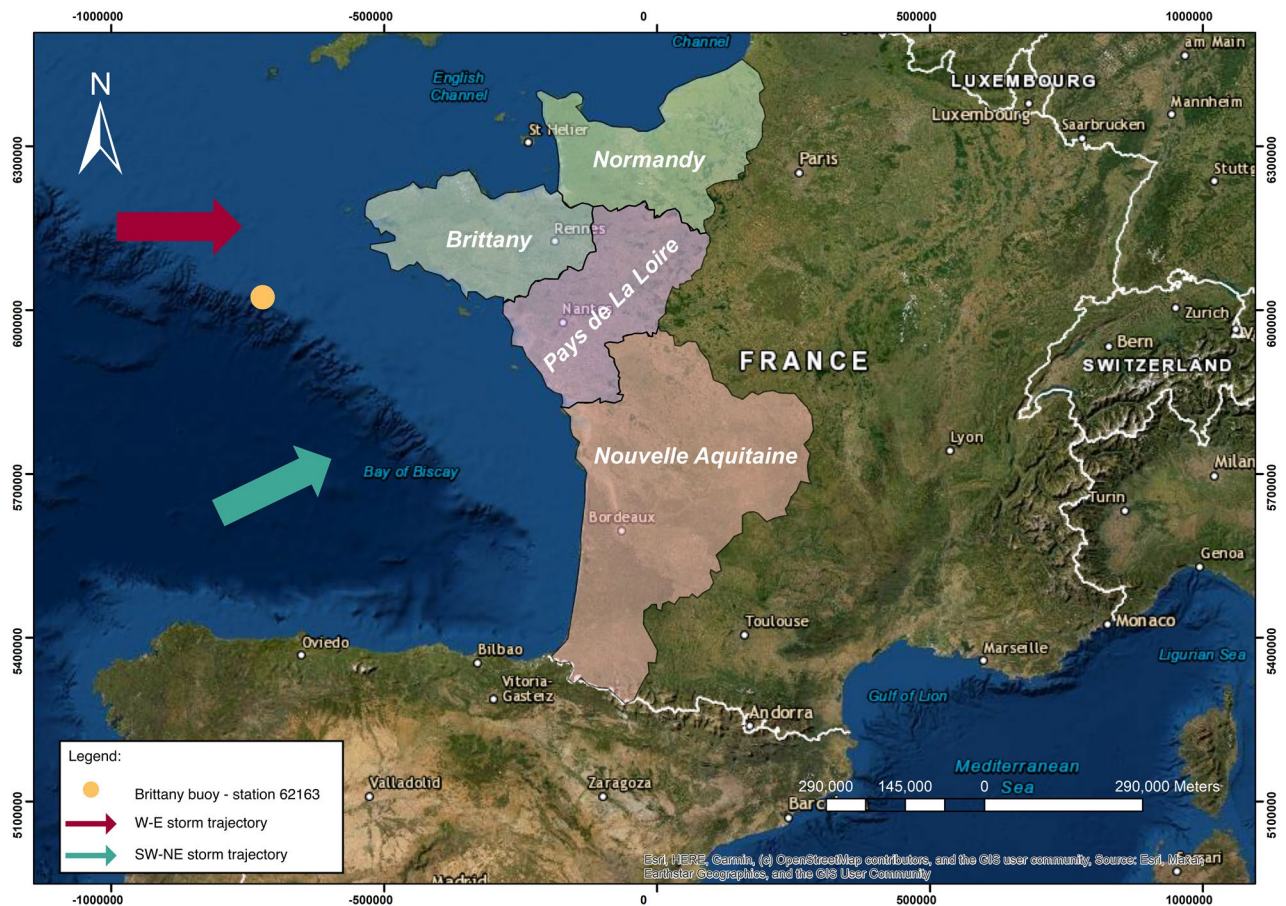


Figure 1. Location of the study area.

such as wave height, duration, and angle of incidence, and the consequent patterns of erosion⁴⁵. In addition, Castelle et al.⁴⁶ developed a new climate index based on sea level pressure, termed the Western Europe Pressure Anomaly (WEPA) index, which accurately accounts for the variations in winter wave height throughout the Western European coast. The WEPA index outperforms other prominent atmospheric modes in explaining winter wave variability and excels at capturing extreme wave height events, as demonstrated by its performance during the severe 2013–2014 winter storms that affected the study area⁴⁶. Moreover, Castelle et al.⁴⁷ conducted a study on the changes in winter-mean wave height, variability, and periodicity in the northeast Atlantic from 1949 to 2017, and examined their connections with climate indicators. Their research highlighted the growing patterns in the height, variability, and regularity of winter waves. These patterns are mainly influenced by climate indices like the North Atlantic Oscillation (NAO) and WEPA. As the positivity and variability of the WEPA and NAO increase, the occurrence of extreme winter-mean wave heights becomes more common⁴⁷, which makes these indices valuable predictors in forecasting coastal hazards.

This study combines data from a storm database and offshore buoy data from 1996 to 2020. The Meteo France website provided the storm database, containing all the storm events in the study area (<http://tempetes.meteo.fr/spip.php?rubrique6>), and the historical measurements from the Brittany Buoy—station 62,163, with hourly records of wave height, wave period, wind speed, temperature, pressure, and humidity from January 1, 1996 to December 31, 2020 (https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=95&id_rubrique=32). These hourly records were transformed into daily values by calculating the average value for each day and adding this to the storm names and occurrences extracted from the storm database. Storm occurrence is defined as days with wind speed values ranking above the 98th percentile of historical wind speed measurements from buoy data and coinciding with the days identified in the storm database as having storm events. Using this definition of storm occurrence, we were able to identify all the storm events in the study area since 1996. The occurrence of storms followed a binomial probability distribution, where the value 1 represented the occurrence of a storm, and the value 0 represented no storm. The combination of buoy data and storm database provided multivariate time series data (Table 1) representing a day-by-day record of historical weather and marine data and the precise days on which storms occurred, which could be used to predict the occurrence of storms and their different characteristics.

Prediction of storm characteristics

The dataset was fed into a multivariate LSTM model to predict the values of the six storm characteristics at once. The following steps were taken to prepare the data for implementing the algorithm: first, any missing values

Date	Storm name	Storm occurrence	Wave height (m)	Wave period (s)	Wind speed (m/s)	Temperature (K)	Pressure (Pa)	Humidity (%)
01-01-1996	-	0	4.336	9.494	6.931	285.249	99,436.485	87.471
02-01-1996	-	0	3.452	8.236	7.457	286.219	101,116.521	88.826
...
26-12-1999	Lothar	1	7.795	9.757	23.95	286.046	97,000	81.588
27-12-1999	Martin	1	6.852	9.828	22.853	297.942	96,480	83.533
28-12-1999	-	0	5.404	8.875	9.315	285.420	101,280	70.210
...
30-12-2020	-	0	3.45	7	9.808	283.770	101,017.1	79.291
31-12-2020	-	0	2.654	6.041	9.704	282.508	101,144.6	67.833

Table 1. Part of the dataset used for the prediction at Brittany Buoy—station 62,163.

for each attribute were replaced with the median value for all the known values of the attribute. This approach is more reliable than mean imputation when the data includes outliers or is skewed since the median is less impacted by extreme values. The LSTM model was then constructed in the Keras framework, the input to the LSTM consisting of a time series vector x for T time steps in the series.

$$x = [x(1), x(2), \dots, x(T)]$$

Each input vector $x(t)$ contained the six characteristics:

$$x(t) = [Temperature(t), Humidity(t), Windspeed(t), Pressure(t), Waveheight(t), Waveperiod(t)]$$

These input sequences were generated by moving a time window through the dataset, each time capturing a sequence of T input vectors. An input of 30-time steps was favored during hyperparameter optimization. Consequently, the characteristics from time $(T-30)$ to $(T-1)$ were used as features to predict the corresponding characteristics y at time T .

$$y(T) = [Temperature(T), Humidity(T), Windspeed(T), Pressure(T), Waveheight(T), Waveperiod(T)].$$

To train and test the LSTM model, the time series data were split into training and test sets at a ratio of 80% to 20%. Each division was given a continuous period of the time series to preserve the serial correlation between successive observations. This yielded a training dataset with an input size of $[nx, T, v] = [7281, 30, 6]$, where nx is the number of training sequences of length T and v is the number of variables, and a test dataset with an input size of $[1821, 30, 6]$. The training dataset only was used to select the optimal hyperparameters, and the test partition, used to test how well the model generalizes to unseen conditions, was set aside to avoid information leakage. After splitting the data, the standard score normalization (Z-Score) method was applied to normalize the input variables. The values for an attribute were normalized by subtracting the mean for each value in the distribution and dividing the result by the standard deviation⁴⁸. Normalization helps speed up the learning phase and prevents initially large-range attributes from outweighing initially smaller ones⁴⁸.

Optimization of the LSTM model required decisions on a combination of large hyperparameters, such as the number of neurons, number of layers, batch size, and learning rate. A scalable hyperparameter optimization framework called Keras Tuner, based on the Bayesian optimization algorithm, was therefore applied to find the best hyperparameter values. To optimize the parameters, this neural network was trained using cross-validation and by setting mean absolute error (MAE) as a performance evaluator. The TimeSeriesSplit technique was used to apply the cross-validation. This strategy is especially appropriate for our investigation since it preserves the temporal order of the data, which is critical for the predictive accuracy of the LSTM model. The model was therefore trained and validated in three splits, which allowed us to systematically analyze and enhance the model's performance by modifying hyperparameters using the Bayesian optimization strategy. Table 2 lists the details of hyperparameter value ranges and choices, and the following paragraph covers each hyperparameter tuned using the Keras Tuner.

Twenty different model configurations were tested to find the optimum set of hyperparameters. For each hyperparameter combination, the LSTM model was trained for a maximum of 100 epochs, and the epoch corresponding to the minimum validation set MAE was recorded. The Google Colab A100 GPU system was used to run the entire model training. The best-proposed model by the Keras Tuner applied three layers. The minimum and maximum numbers of neurons at each hidden layer were set as 32 and 512 respectively. The dropout layer, used to prevent overfitting by eliminating certain connections between neurons in each iteration⁴⁹, was defined within the range of 0 to 0.5. The output layer was then linked to a dense layer with six output neurons and the activation function was set to Linear. The network was compiled with a MAE loss function and Adam optimizer to update the weight of each layer after each iteration. The learning rate, which was set using Keras Tuner to one of three values (0.01, 0.001, or 0.0001), determined the size of the step at each iteration of the optimization method⁴⁹. Four other value choices (16, 32, 64, 128) were set for the batch size, representing the number of sub-samples the network used to update the weights⁴⁹. Proceeding through the search space, the combination with the lowest validation set MAE was chosen as the optimum set of hyperparameters (Table 3).

	Hyperparameters	Value Ranges			Choices	Value
		Min	Max	Step		
LSTM	Number of layers	1	4	–	–	–
	LSTM layers units	32	512	32	–	–
	Dropout	0	0.5	0.1	–	–
	Learning rate	–	–	–	0.01, 0.001, 0.0001	–
	Batch size	–	–	–	16, 32, 64, 128	–
	Optimizer	–	–	–	–	Adam
	Objective	–	–	–	–	val_mean_absolute_error
	Max trials	–	–	–	–	20
	Executions per trial	–	–	–	–	2
XGBoost	learning_rate	–	–	–	0.0001, 0.001, 0.01, 0.1	–
	max_depth	–	–	–	3, 4, 5, 6, 7, 8, 9, 10, 11, 12	–
	gamma	–	–	–	0, 0.1, 0.2, 0.3, 0.4	–
	colsample_bytree	–	–	–	0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9	–
	reg_alpha	–	–	–	0.00001, 0.01, 0.1, 1, 10, 100	–
	reg_lambda	–	–	–	0.00001, 0.01, 0.1, 1, 10, 100	–

Table 2. Hyperparameters optimization of the LSTM and XGBoost models.

	Hyperparameters	Best hyperparameter values
LSTM	Number of layers	3
	LSTM layer 1 unit	192
	LSTM layer 2 units	288
	LSTM layer 3 units	32
	Dropout	0.2
	Activation function for output layer	linear
	Learning rate	0.01
	Batch size	32
	Optimizer	Adam
	Time steps	30
XGBoost	learning_rate	0.1
	max_depth	12
	gamma	0
	colsample_bytree	0.4
	reg_alpha	0.00001
	reg_lambda	0.01

Table 3. Optimal set of hyperparameters found for LSTM and XGBoost.

A comprehensive analysis of autocorrelation functions (ACF), together with a trial-and-error process, was used to identify the ideal input time steps for improving the model's performance. The use of the ACF on the daily dataset demonstrates how each attribute corresponds with its historical observations. As seen in Fig. 2, the autocorrelation plots for pressure and wind speed indicate a shorter memory effect, with substantial correlations vanishing rapidly. In contrast, humidity has longer-term meaningful associations that stay generally steady over time. Temperature, wave period, and wave height show more gradual decay, indicating a longer memory effect. Based on the findings from the ACF analysis, a sequence length that captures the memory of the most persistent variable while still giving useful information for variables with shorter memories should be selected. As a result, a sequence length of 20-time steps was chosen as an ideal compromise. This input sequence was thought sufficient to capture the longer-term dependency found in temperature and wave-related variables, while also catching essential shorter-term information for pressure and wind speed. The model's performance was then examined by training and validating the LSTM model using cross-validation with the initial sequence length selected based on the ACF results. The sequence length was further optimized by gradually decreasing and increasing the number of lags used to anticipate future values until the optimal sequence length was determined. This iterative procedure resulted in the discovery of an optimal input sequence of 30-time steps, allowing the LSTM model to achieve higher performance.

MAE, mean absolute percentage error (MAPE), root mean square error (RMSE), and coefficient of determination (R^2) were used to evaluate the performance of the LSTM model. These evaluation metrics are defined as follows:

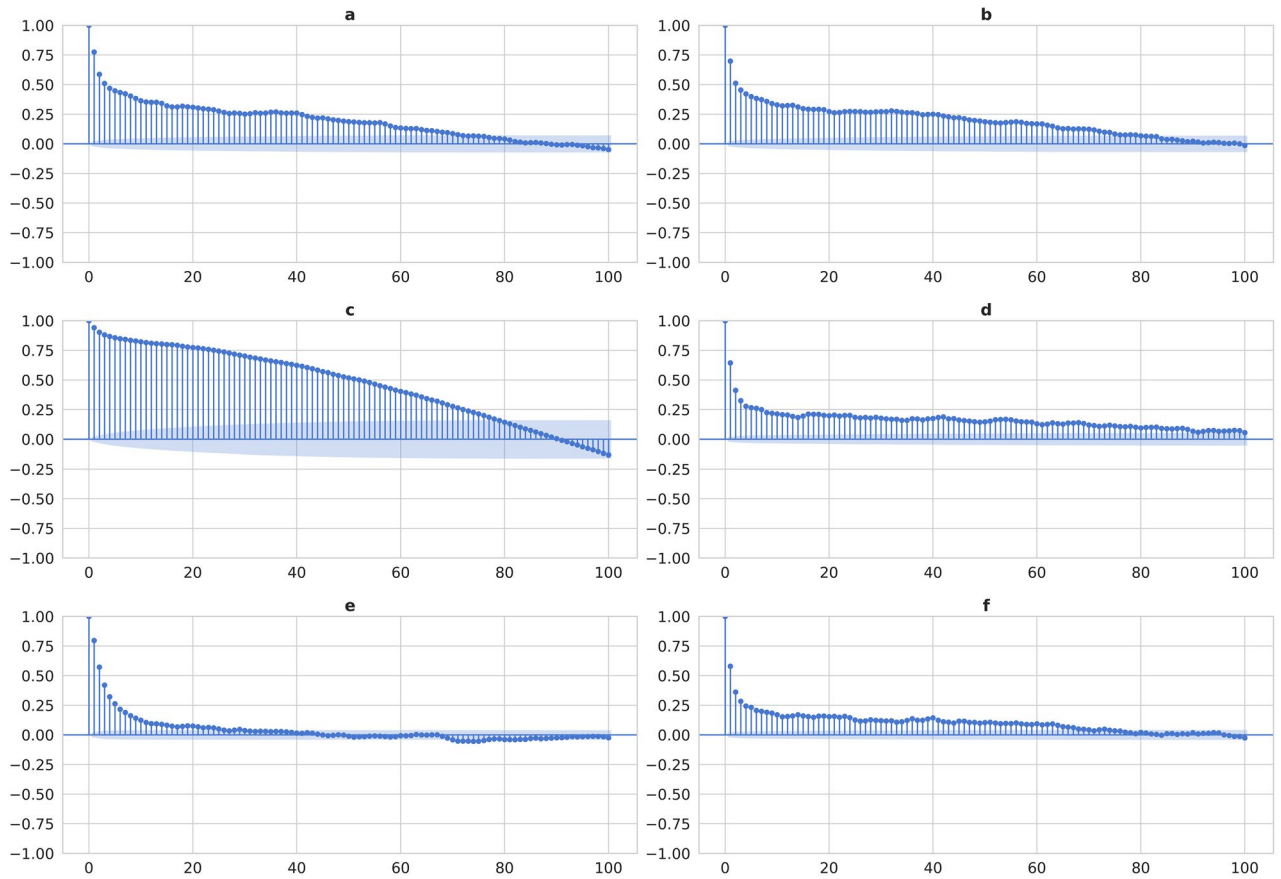


Figure 2. Autocorrelation Analysis (a) Autocorrelation of Wave Height (b) Autocorrelation of Wave Period (c) Autocorrelation of Temperature (d) Autocorrelation of Humidity (e) Autocorrelation of Pressure (f) Autocorrelation of Wind Speed.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{1}$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{Y})^2} \tag{4}$$

where n is the number of samples, y_i the real value, \hat{y}_i the predicted value, and \bar{Y} is the mean of the observed values. The best performance is achieved when the values of MAE, MAPE, and RMSE are close to 0 and the value of R^2 is close to 1.

The Pearson correlation coefficient (r) was also used to assess the performance of the LSTM model, specifically the strength of the linear association between the observed and predicted values. The definition of this measure is as follows:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \tag{5}$$

where x_i are the observed values, y_i are the predicted values, \bar{x} and \bar{y} are the means of the observed and predicted values, respectively, and n is the number of samples.

Prediction of storms

The objective of the second part was to predict the occurrence of storms based on their characteristics. To achieve this, an XGBoost binary classifier was developed using the XGBoost library. The XGBoost model used the six characteristics as input X_i , and storm occurrence was used as model output y_i , given a binary class label $y_i \in \{0, 1\}$, indicating the absence or occurrence of a storm, respectively (Table 1). The median value of each attribute was used to fill in the missing values of the independent variables X_i . The dataset was split into training and testing sets at the ratio of 80 to 20. The 20% testing data (unseen data) corresponding to the period from January 2016 to December 2020 was set aside to avoid information leakage. The training dataset from January 1996 to December 2015 only was used for tuning the hyperparameters. The preprocessing phase of normalizing independent variables X_i was carried out using the method described in the section 'Prediction of storm characteristics'. This section applied Bayesian optimization using the Hyperopt library in Python to find the best combination of hyperparameters. Bayesian optimization using Hyperopt is an effective method for Hyperparameter optimization of the XGBoost algorithm. It performs better than other widely-used approaches such as grid search and random Search⁵⁰.

The XGBoost model was trained using stratified k-fold cross-validation with $K = 3$. Stratified k-fold is highly appropriate for imbalanced data, as in our study, as it helps keep the class ratio in the folds the same as the training dataset⁵¹. The training dataset was divided into three subintervals—two for training and one for evaluating. Six hyperparameters were optimized, namely Learning rate (eta), Maximum Tree Depth, Gamma, Column samples by a tree, Alpha, and Lambda, and the search space defined is shown in Table 2. Fifty different combinations were tested to find the optimum set of hyperparameters. For each combination, recall was set as the evaluation metric for cross-validation. The combination at which the highest recall value was found is identified as the best set of hyperparameters (Table 3). After identifying the best hyperparameters, XGBoost was trained with the entire training dataset and tested using the unseen dataset. The Google Colab A100 GPU system was used to run the hyperparameter optimization and model training.

Due to the low frequency of storm occurrence, the following performance measures were used to evaluate the accuracy of the model: Recall, specificity, false positive rate (FPR), and false negative rate (FNR). These assessment metrics are not sensitive to imbalanced data⁵² and can be calculated using the following equations:

$$\text{Recall (Sensitivity)} = \text{Truepositiverate} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{Specificity} = \text{Truenegativerate} = \frac{TN}{TN + FP} \quad (7)$$

$$\text{FPR} = \frac{FP}{FP + TN} \quad (8)$$

$$\text{FNR} = \frac{FN}{FN + TP} \quad (9)$$

where:

- TP : true positives, where the model predicted samples correctly as positive. In this case, the storms were classified as 'storm'.
- TN : true negatives, where the model predicted samples correctly as negative (no storms predicted as 'no storm').
- FP : false positives, where the model wrongly predicted samples as positive (no storms predicted as 'storm').
- FN : false negatives, where the model wrongly predicted samples as negative (storms predicted as 'no storm').

The performance of the model was also measured using ROC (receiver operating characteristics) curve. ROC curve is a two-dimensional graph where the x-axis is the FPR and the y-axis the TPR (true positive rate), and it is generated by changing the threshold on the confidence score. The ROC curve is not sensitive to imbalanced data and can illustrate the diagnostic capability of a binary classifier⁵². The area under the ROC curve (AUC) metric is used to assess the performance of the XGBoost model since there is no scalar value representing the expected performance in the ROC curve. The AUC metric ranges from 0 to 1, and a perfect model will have an AUC of 1.

Results

LSTM results

The trained LSTM model was used to predict storm characteristics for the unseen test period from January 2016 to December 2020. The prediction errors are summarized in Table 4. Figure 3 shows the prediction outcomes of the LSTM model. The blue curve is the actual values and the orange curve is the forecast results. The figure shows that the model provides a good representation of the observed values for most of the time period as the actual and predicted curves roughly correspond.

The LSTM model demonstrated impressive accuracy in temperature predicting, as shown by a low MAE of 0.7574 K, an RMSE of 0.9881 K, and a MAPE of 0.26%. In addition, the model's R^2 score of 0.8753 and Pearson Correlation Coefficient of 0.9385 indicate that it effectively captures a substantial amount of the temperature data's variability and exhibits a robust linear association with the actual values.

		MAE	RMSE	MAPE	R ²	r
LSTM	Wave height	0.6271	0.8617	24.27%	0.6083	0.7914
	Wave period	0.6439	0.8189	9.69%	0.6522	0.8141
	Temperature	0.7574	0.9881	0.26%	0.8753	0.9385
	Humidity	6.0851	7.6008	7.61%	0.3272	0.5843
	Pressure	467.2703	630.4106	0.46%	0.5536	0.7495
	Wind speed	2.0061	2.5666	32.76%	0.2337	0.5482
XGBoost	Recall	Specificity	FPR	FNR	–	–
	1	1	0	0	–	–

Table 4. Evaluation metrics of the LSTM and XGBoost models.

The performance of the model for humidity forecasts was less accurate, with MAE and RMSE values of 6.0851% and 7.6008%, respectively. The moderate value of the R² score (0.3272) and the comparatively high MAPE value (7.61%) show the model's limited skill in capturing the variation in humidity. The Pearson Correlation Coefficient of 0.5843 suggests a moderate linear connection, however with much opportunity for improvement.

The LSTM model faced a larger difficulty in predicting wind speed, as evidenced by the high MAPE of 32.76%, MAE of 2.0061 m/s and RMSE of 2.5666 m/s. In addition, the low R² score (0.2337) and modest Pearson Correlation Coefficient (0.5482) demonstrate the model's limitations in predicting wind speed. The low performance of the LSTM for wind speed prediction might be due to the complex and varied nature of wind behavior, and the presence of very extreme values (outliers) at storm events.

The model's performance for pressure prediction was relatively good, with MAPE of 0.46% indicating that the model's predictions closely match the actual pressure levels. The R² score of 0.5536 and the Pearson Correlation Coefficient of 0.7495 indicate that the model is capable of capturing the variability in pressure.

The predictions of wave conditions (wave height and period) by the LSTM model were reasonable, with MAE values of 0.6271 m and 0.6439 s, respectively, and RMSE values of 0.8617 m and 0.8189 s, respectively. However, the high MAPE of 24.27% in wave height shows a considerable relative error, demonstrating the underestimation of the model in wave height prediction during storm events. The R² scores of 0.6083 and 0.6522 for wave height and period, together with Pearson Correlation Coefficients of 0.7914 and 0.8141, imply a high predictive association with the actual values.

The model's forecasts at storm events for temperature and pressure were particularly accurate, whereas humidity revealed inconsistencies when compared to the observed values. Predictions of wave-related variables were consistent but underlined the necessity for refinement, specifically for wave height, which showed a tendency to be underpredicted during storm events. An even bigger underprediction can be seen towards the extreme values of wind speed during storm events.

The low frequency of storm occurrence can explain these findings; storm events represent only 0.7% of the total data used in this study. Most of the available data for training therefore covers lower wave height and wind speed values, which correspond to non-storm events. More storm observations are required in the training dataset to improve further the accuracy of the model in predicting storm characteristics.

In conclusion, the LSTM model exhibits proficiency in predicting temperature and pressure characteristics of storms with high accuracy. The model provides a robust foundation for temperature and pressure forecasts, while the prediction of humidity, wind speed, and wave characteristics shows potential but requires further optimization.

XGBoost results

The trained XGBoost model was used to predict storm occurrence for the unseen test period of January 2016–December 2020. There were five storms in this period: Zeus on March 6, 2017, Eleanor on January 3, 2018, storm on December 13, 2019, Ciara on February 10, 2020, and Bella, a two-day storm on December 27 and 28, 2020. There are two classes: Storm and No-storm. The Storm class is positive, and the No-storm class is negative. Figure 4a shows the prediction of storm occurrence; predicted storms are represented by light orange points and actual storms by light blue points. As shown in the figure, the predicted storms concur precisely with the actual storms; the model predicted the occurrence of all the storms. The confusion matrix of the XGBoost model is given in Fig. 4b. Six samples are correctly classified as Storm and 1821 correctly classified as No-storm; the respective TN, TP, FN, and FP values are therefore 1821, 6, 0, and 0. The classification metrics are given in Table 4. The proposed XGBoost classifier performed very well in predicting storm occurrence. Recall and specificity values, which represent the correctly predicted samples, are equal to 1, while the FPR and FNR, which represent the incorrectly predicted samples, are equal to 0. Figure 4c shows the AUC value of the model. As shown, the model has an AUC of 1, which implies higher accuracy in predicting storm occurrence.

Discussion

This study applied a new data-driven approach based on LSTM and XGBoost to forecast storm characteristics and occurrence in Western France. There were two main elements to the study. First, comprehensive research into storm prediction using six different storm attributes and the days on which they occurred. Second, exploring

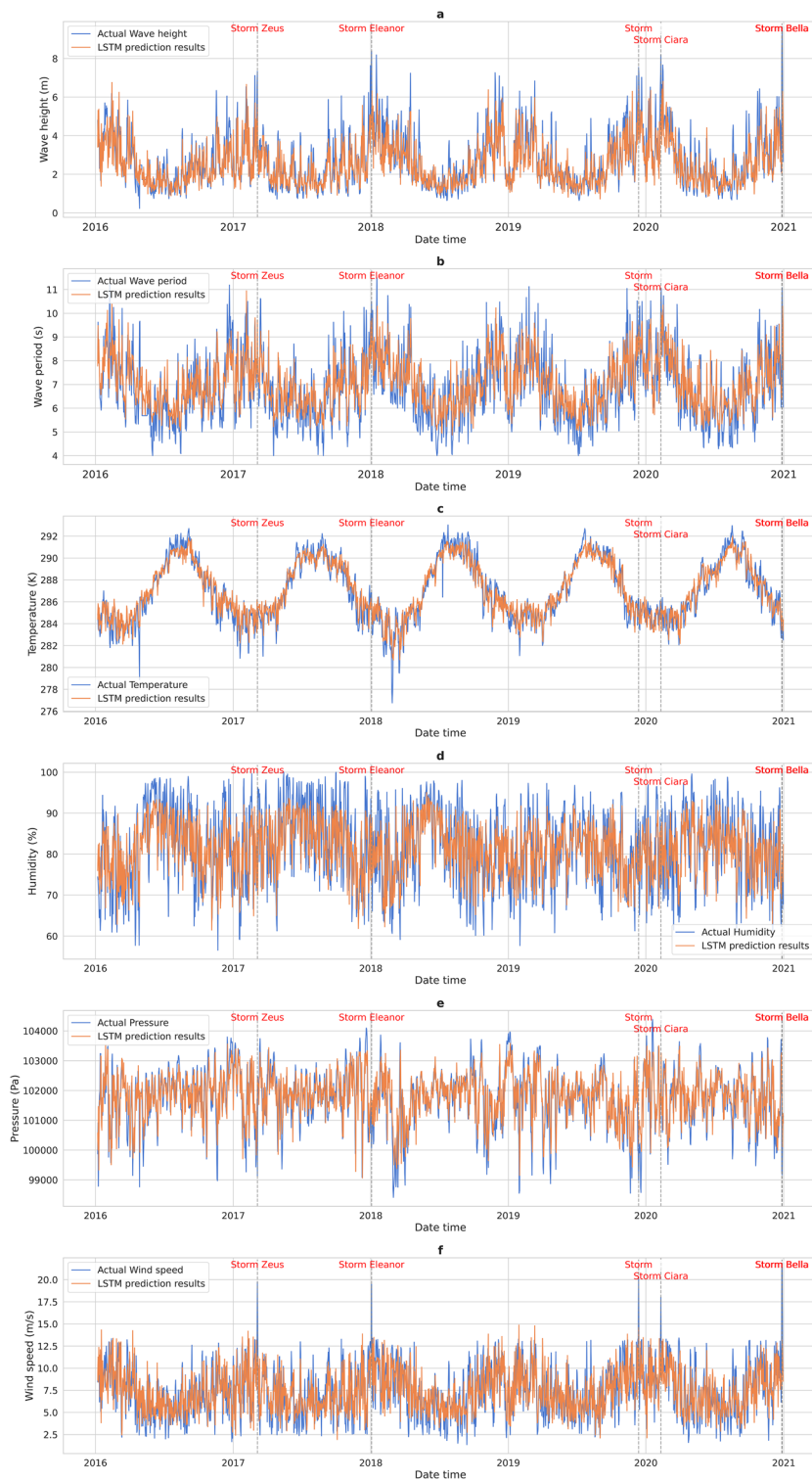


Figure 3. Prediction outcomes of the LSTM model for the unseen test partition (a) Wave height prediction results (b) Wave period prediction results (c) Temperature prediction results (d) Humidity prediction results (e) Pressure prediction results (f) Wind speed prediction results..

the efficiency of the ML and DL methods in capturing rare events such as storms. Experiments were conducted to build a multivariate LSTM model to predict storm characteristics. The LSTM model was evaluated using the unseen test period data from January 2016 to December 2020. In a second step, an XGBoost binary classifier

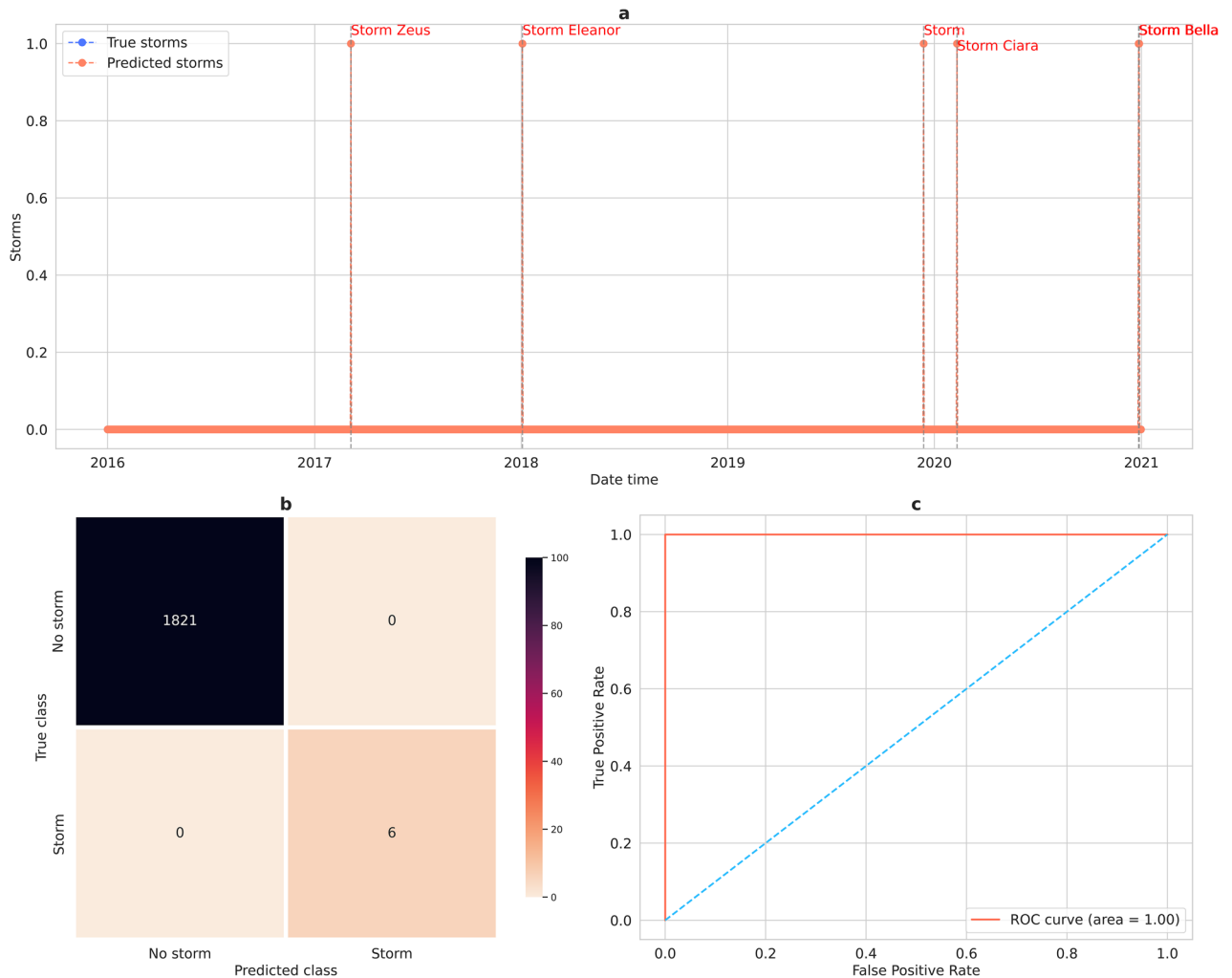


Figure 4. Prediction of storms using XGBoost model (a) Storm occurrence prediction results (b) Confusion matrix of the XGBoost model (c) ROC curve of the XGBoost model.

was developed to predict the occurrence of storms based on their characteristics, and data from January 2016 to December 2020 was used for model evaluation. The following key points elaborate on the outcome of this study.

The LSTM model was found to perform well in predicting the six variables (wave height, wave period, wind speed, temperature, humidity, and pressure) over all the unseen test period. The visual comparison of forecasted and observed values demonstrated that the model typically produced a decent representation of actual values. During storm events, the accuracy of the LSTM model in forecasting storm features fluctuated significantly. It excelled in anticipating temperature and pressure, displaying high accuracy. However, it proved less effective in forecasting humidity and notably underpredicted wave-related variables and wind speed during storms. This tendency to underestimate wave height and wind speed can be explained by the limited number of extreme wave heights and wind speeds available for model training. Comparable results were found by Hu, Haoguo, et al. for predicting wave height during storm events²⁴. The models tended to underestimate extreme wave height and perform extremely well for regular events²⁴. To overcome this issue, Dixit & Londhe²³, Prahlada & Deka⁵³ applied a discrete wavelet transform to decompose time series data into low and high-frequency components. Separate neural network models were then trained for both parts, potentially increasing the accuracy of extreme event prediction^{23,53}.

Despite the limited availability of storm data for model training, the XGBoost model performed well and predicted all storms in the unseen test period. Most accuracy-driven 'vanilla' machine learning methods suffer a decline in performance when they encounter a label-imbalanced classification situation⁵⁴. In this experiment, the number of samples in the Storm and No-storm classes was 68 (0.7%) and 9064 (99.3%), respectively. If the model were to predict all the samples as 'no storm,' then the accuracy would be 99%, which is remarkably high. However, failure to predict any storm can lead to severe consequences. The most commonly used evaluation metrics, such as accuracy and precision, were therefore not considered in this study due to their sensitivity to imbalanced data⁵².

Consequently, the model was evaluated using recall, specificity, FPR, FNR, and AUC metrics and provided values of 1, 1, 0, 0, and 1, respectively. These assessment metrics are not sensitive to changes in data distribution

and can be used to evaluate classification performance with imbalanced data⁵². Kabir & Ludwig also suggest addressing the problem of imbalanced data by adopting several data resampling techniques before applying XGBoost for classification⁵⁵. More recently, Wang et al. introduced imbalance-XGBoost. The aim of this XGBoost-based Python package is to deal with binary label-imbalanced classification issues by implementing weighted cross-entropy and focal loss functions on XGBoost⁵⁴.

As part of our future work, we plan to implement a multilevel decomposition of data using a discrete wavelet transform. A separate model will then be trained on the low-frequency components (extreme wave height, wave period and wind speed partition) to improve prediction accuracy and predict the characteristics of storm events precisely. To overcome the data imbalance problem, we must also investigate the sampling-based approach of modifying the dataset to balance the class distribution before using it to train the XGBoost classifier. Future research may extend the outcomes of this study to other regions, deep learning model architectures, and hyper-parameter tunings.

Conclusions

This paper considerably expands on past storm prediction research by adopting a unique, integrated strategy that uses both LSTM and XGBoost models to anticipate a wide range of storm characteristics as well as storm occurrence along France's western coast. The detailed investigation reported in this research led to the following conclusions:

- **Relevance to Previous Studies:** Building on the existing literature that studied the usage of LSTM and XGBoost models in isolation, this study proposes a unique technique that leverages both models in unison. While past work focused largely on specific features of storm prediction, such as storm surges, and approached them as regression issues, the present method blends regression and classification approaches to give a more holistic understanding of storm dynamics. This dual-model technique provides for the thorough prediction of varied storm parameters (temperature, pressure, humidity, wind speed, wave height, and wave period) as well as the prediction of storm occurrence days.
- **Performance of the LSTM Model:** The LSTM model has shown great accuracy in forecasting temperature and pressure, key elements for understanding and anticipating storm conditions. However, it revealed limits in reliably forecasting factors linked with high variability and extreme circumstances (extreme values), such as wind speed and wave height, particularly during severe storm events. This shows that although the LSTM model is resilient under stable conditions, more refinement and enhancement are required to capture the entire spectrum of storm-induced variabilities.
- **Efficiency of the XGBoost Model:** The XGBoost model successfully predicted storm occurrences with extraordinary accuracy, indicated by flawless recall and specificity scores. This accuracy is crucial for practical applications in storm forecasting, where the cost of FN (failing to anticipate a storm) can be extraordinarily significant. The performance of the XGBoost model in this situation underlines its promise as a trustworthy tool in operational meteorology.

In essence, this study indicates a big step forward in storm prediction. It not only expands the scope of forecasted storm features but also combines the capabilities of LSTM and XGBoost, opening up new paths for increasing storm preparation and risk reduction measures. The method adopted in this study is not only relevant to the western coast of France but also has possibilities for adaption and usage in other places prone to extreme weather events.

Data availability

The final dataset analyzed in this study is available on GitFront: <https://gitfront.io/r/user-9374464/JTvFdsL1Q3C5/Storm-Prediction/>.

Code availability

Scripts reproducing the experiments of this paper are available at GitFront <https://gitfront.io/r/user-9374464/JTvFdsL1Q3C5/Storm-Prediction/>.

Received: 23 November 2022; Accepted: 13 May 2024

Published online: 18 May 2024

References

1. Usbeck, T. *et al.* Increasing storm damage to forests in Switzerland from 1858 to 2007. *Agric. For. Meteorol.* **150**, 47–55 (2010).
2. Browning, K. A. The sting at the end of the tail: Damaging winds associated with extratropical cyclones. *Q. J. R. Meteorol. Soc.* **130**, 375–399 (2004).
3. Fink, A. H., Brücher, T., Ermert, V., Krüger, A. & Pinto, J. G. The European storm Kyrill in January 2007: Synoptic evolution, meteorological impacts and some considerations with respect to climate change. *Nat. Hazards Earth Syst. Sci.* **9**, 405–423 (2009).
4. Liberato, M. L. R., Pinto, J. G., Trigo, I. F. & Trigo, R. M. Klaus—An exceptional winter storm over northern Iberia and southern France. *Weather* **66**, 330–334 (2011).
5. Peiris, N. & Hill, M. Modeling wind vulnerability of French houses to European extra-tropical cyclones using empirical methods. *J. Wind Eng. Ind. Aerodyn.* **104**, 293–301 (2012).
6. Schwierz, C. *et al.* Modelling European winter wind storm losses in current and future climate. *Clim. Change* **101**, 777–780 (2010).
7. Haylock, M. R. European extra-tropical storm damage risk from a multi-model ensemble of dynamically-downscaled global climate models. *Nat. Hazards Earth Syst. Sci.* **11**, 2847 (2011).
8. Genovese, E. & Przyłuski, V. Storm surge disaster risk management: The Xynthia case study in France. *J Risk Res* **16**, 825 (2013).

9. Dorland, C., Tol, R. S. J. & Palutikof, J. P. Vulnerability of the Netherlands and Northwest Europe to storm damage under climate change: A model approach based on storm damage in the Netherlands. *Clim. Change* **43**, 513 (1999).
10. Leckebusch, G. C., Ulbrich, U., Fröhlich, L. & Pinto, J. G. Property loss potentials for European midlatitude storms in a changing climate. *Geophys. Res. Lett.* **34**, 5 (2007).
11. Hawcroft, M. K., Shaffrey, L. C., Hodges, K. I. & Dacre, H. F. How much Northern Hemisphere precipitation is associated with extratropical cyclones?. *Geophys. Res. Lett.* **39**, 24 (2012).
12. Dedieu, F. Alerts and catastrophes: The case of the 1999 storm in France, a treacherous risk. *Sociol. Trav.* **52**, e1–e21 (2010).
13. Sacré, C. Extreme wind speed in France: The '99 storms and their consequences. *J. Wind Eng. Ind. Aerodyn.* **90**, 1163–1171 (2002).
14. Bertin, X. *et al.* A modeling-based analysis of the flooding associated with Xynthia, central Bay of Biscay. *Coast. Eng.* **94**, 80–89 (2014).
15. Kron, W., Löw, P. & Kundzewicz, Z. W. Changes in risk of extreme weather events in Europe. *Environ. Sci. Policy* **100**, 74–83 (2019).
16. Feser, F. *et al.* Storminess over the North Atlantic and northwestern Europe—A review. *Q. J. R. Meteorol. Soc.* **141**, 1. <https://doi.org/10.1002/qj.2364> (2015).
17. Della-Marta, P. M. *et al.* The return period of wind storms over Europe. *Int. J. Climatol.* **29**, 1 (2009).
18. Ren, F. M. *et al.* A research progress review on regional extreme events. *Adv. Clim. Chang. Res.* **9**, 161–169 (2018).
19. Buizza, R. & Hollingsworth, A. Storm prediction over Europe using the ECMWF Ensemble Prediction System. *Meteorol. Appl.* **9**, 289–305 (2002).
20. Renggli, D., Leckebusch, G. C., Ulbrich, U., Gleixner, S. N. & Faust, E. The skill of seasonal ensemble prediction systems to forecast wintertime windstorm frequency over the North Atlantic and Europe. *Mon. Weather Rev.* **139**, 3052–3068 (2011).
21. Befort, D. J. *et al.* Seasonal forecast skill for extratropical cyclones and windstorms. *Q. J. R. Meteorol. Soc.* **145**, 92–104 (2019).
22. Behrens, A. & Günther, H. Operational wave prediction of extreme storms in Northern Europe. *Nat. Hazards* **49**, 387–399 (2009).
23. Dixit, P. & Londhe, S. Prediction of extreme wave heights using neuro wavelet technique. *Appl. Ocean Res.* **58**, 241 (2016).
24. Hu, H., van der Westhuysen, A. J., Chu, P. & Fujisaki-Manome, A. Predicting Lake Erie wave heights and periods using XGBoost and LSTM. *Ocean Model (Oxf)* **164**, 101832 (2021).
25. Wei, Z. Forecasting wind waves in the US Atlantic Coast using an artificial neural network model: Towards an AI-based storm forecast system. *Ocean Eng.* **237**, 109646 (2021).
26. Elman, J. L. Finding structure in time. *Cogn. Sci.* **14**, 179–211 (1990).
27. Hüskens, M. & Stagge, P. Recurrent neural networks for time series classification. *Neurocomputing* **50**, 223 (2003).
28. Bengio, Y., Simard, P. & Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **5**, 157–166 (1994).
29. Pascanu, R., Mikolov, T. & Bengio, Y. On the difficulty of training recurrent neural networks. 1310–1318 (2013).
30. Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **9**, 1735–1780 (1997).
31. Sak, H., Senior, A. & Beaufays, F. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. (2014).
32. Sutskever, I., Vinyals, O. & Le, Q. V. Sequence to sequence learning with neural networks. *Adv. Neural Inf. Process. Syst.* **4**, 3104–3112 (2014).
33. Elsaraiti, M. & Merabet, A. A comparative analysis of the ARIMA and LSTM predictive models and their effectiveness for predicting wind speed. *Energies (Basel)* **14**, 1 (2021).
34. Geng, D., Zhang, H. & Wu, H. Short-term wind speed prediction based on principal component analysis and LSTM. *Appl. Sci. Switzerland* **10**, 4416 (2020).
35. Ji, C. *et al.* XG-SF: An XGBoost classifier based on shapelet features for time series classification. *Procedia Comput. Sci.* **147**, 24–28 (2019).
36. Friedman, J. H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **38**, 367–378 (2002).
37. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016). <https://doi.org/10.1145/2939672>.
38. Ranjan, C., Reddy, M., Mustonen, M., Paynabar, K. & Pourak, K. Dataset: Rare Event Classification in Multivariate Time Series. 1–7 (2018).
39. Fang, Z. G., Yang, S. Q., Lv, C. X., An, S. Y., & Wu, W. Application of a data-driven XGBoost model for the prediction of COVID-19 in the USA: a time-series study. *BMJ Open* 1–8 (2022). <https://doi.org/10.1136/bmjopen-2021-056685>.
40. Abbasi, R. A. *et al.* Short Term Load Forecasting Using XGBoost. in *Advances in Intelligent Systems and Computing* vol. 927 (2019).
41. Jan, V. K., Tse, R., Tang, S. K. & Pau, G. Bridging the Gap: Enhancing storm surge prediction and decision support with bidirectional attention-based LSTM. *Atmos. Basel* **14**, 1082 (2023).
42. Pouzet, P. & Maanan, M. Climatological influences on major storm events during the last millennium along the Atlantic coast of France. *Sci. Rep.* **10**, 1 (2020).
43. Athimon, E. & Maanan, M. Vulnerability, resilience and adaptation of societies during major extreme storms during the Little Ice Age. *Clim. Past* **14**, 1487–1497 (2018).
44. Drevet, C., Benech, B. & Jourdain, S. Classification des tempêtes sur la France à l'usage des assureurs. *La Météorologie* **8**, 23 (1997).
45. Castelle, B. *et al.* Impact of the winter 2013–2014 series of severe Western Europe storms on a double-barred sandy coast: Beach and dune erosion and megacusp embayments. *Geomorphology* **238**, 135 (2015).
46. Castelle, B., Dodet, G., Masselink, G. & Scott, T. A new climate index controlling winter wave activity along the Atlantic coast of Europe: The West Europe Pressure Anomaly. *Geophys. Res. Lett.* **44**, 1384 (2017).
47. Castelle, B., Dodet, G., Masselink, G. & Scott, T. Increased winter-mean wave height, variability, and periodicity in the Northeast Atlantic Over 1949–2017. *Geophys. Res. Lett.* **45**, 3586 (2018).
48. Kumar, S. Efficient K-mean clustering algorithm for large datasets using data mining standard score normalization. *Int. J. Recent Innov. Trends Comput. Commun.* **2**, 3161 (2014).
49. Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F. & Troncoso, A. Deep learning for time series forecasting: A Survey. *Big Data* **9**, 3. <https://doi.org/10.1089/big.2020.0159> (2021).
50. Putatunda, S. & Rama, K. A comparative analysis of hyperopt as against other approaches for hyper-parameter optimization of XGBoost. *ACM Int. Conf. Proc. Ser.* <https://doi.org/10.1145/3297067.3297080> (2018).
51. Berrar, D. Cross-validation. in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* vols 1–3 (2018).
52. Tharwat, A. Classification assessment methods. *Appl. Comput. Inf.* **17**, 168 (2018).
53. Prahada, R. & Deka, P. C. Forecasting of time series significant wave height using wavelet decomposed neural network. *Aquat Procedia* **4**, 540 (2015).
54. Wang, C., Deng, C. & Wang, S. Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. *Pattern Recognit. Lett.* **136**, 1 (2020).
55. Kabir, M. F. & Ludwig, S. Classification of Breast Cancer Risk Factors Using Several Resampling Approaches. in *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018* (2019). doi:<https://doi.org/10.1109/ICMLA.2018.00202>.

Acknowledgements

This work was supported by UMR 6554 CNRS LETG-Nantes laboratory, Geosciences Laboratory, and the CNRST (Centre National pour la Recherche Scientifique et Technique) as part of the research excellence scholarship program.

Author contributions

M.M.: Conceptualization and methodology. M.M. and R.H.: Data acquisition and interpretation. F.A.: Custom code creation and analysis of algorithms performance. R.H. and M.M.: Results analysis and evaluation of data analysis. All the authors drafted, read, and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024