

Article

# Deep Learning-Based Wave Overtopping Prediction

Alberto Alvarellos <sup>1,\*</sup> , Andrés Figuero <sup>2</sup> , Santiago Rodríguez-Yáñez <sup>1</sup> , José Sande <sup>2</sup> , Enrique Peña <sup>2</sup>, Paulo Rosa-Santos <sup>3</sup>  and Juan Rabuñal <sup>4</sup> 

<sup>1</sup> Software Engineering Laboratory (ISLA), Centro de Innovación Tecnológica en Edificación e Enxeñaría Civil (CITEEC), Universidade da Coruña, Campus Elviña s/n, 15071 A Coruña, Spain; santiago.rodriguez@udc.es

<sup>2</sup> Water and Environmental Engineering Group (GEAMA), Centro de Innovación Tecnológica en Edificación e Enxeñaría Civil (CITEEC), Universidade da Coruña, Campus Elviña s/n, 15071 A Coruña, Spain; andres.figuero@udc.es (A.F.); jose.sande@udc.es (J.S.); enrique.penag@udc.es (E.P.)

<sup>3</sup> Department of Civil Engineering, Faculty of Engineering, University of Porto (FEUP), Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal; pjsantos@fe.up.pt

<sup>4</sup> RNASA Group, Centro de Innovación Tecnológica en Edificación e Enxeñaría Civil (CITEEC), Universidade da Coruña, Campus Elviña s/n, 15071 A Coruña, Spain; juan.rabunal@udc.es

\* Correspondence: alberto.alvarellos@udc.es

**Featured Application:** Port managers can use predictions of the wave overtopping predictors created in this work to take preventative measures and optimize operations, ultimately improving safety and helping to minimize the economic impact that overtopping events have on the port's activities.

**Abstract:** This paper analyses the application of deep learning techniques for predicting wave overtopping events in port environments using sea state and weather forecasts as inputs. The study was conducted in the outer port of Punta Langosteira, A Coruña, Spain. A video-recording infrastructure was installed to monitor overtopping events from 2015 to 2022, identifying 3709 overtopping events. The data collected were merged with actual and predicted data for the sea state and weather conditions during the overtopping events, creating three datasets. We used these datasets to create several machine learning models to predict whether an overtopping event would occur based on sea state and weather conditions. The final models achieved a high accuracy level during the training and testing stages: 0.81, 0.73, and 0.84 average accuracy during training and 0.67, 0.48, and 0.86 average accuracy during testing, respectively. The results of this study have significant implications for port safety and efficiency, as wave overtopping events can cause disruptions and potential damage. Using deep learning techniques for overtopping prediction can help port managers take preventative measures and optimize operations, ultimately improving safety and helping to minimize the economic impact that overtopping events have on the port's activities.



**Citation:** Alvarellos, A.; Figuero, A.; Rodríguez-Yáñez, S.; Sande, J.; Peña, E.; Rosa-Santos, P.; Rabuñal, J. Deep Learning-Based Wave Overtopping Prediction. *Appl. Sci.* **2024**, *14*, 2611. <https://doi.org/10.3390/app14062611>

Academic Editor: Gianluigi Ferrari

Received: 24 February 2024

Revised: 12 March 2024

Accepted: 14 March 2024

Published: 20 March 2024

**Keywords:** machine learning; neural networks; deep learning; wave overtopping prediction; port management; port security



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Around 80% of the goods we consume are carried by ships due to shipping's ability to offer economical and efficient long-distance transport. These goods include raw materials that need to be processed and products like food, medical supplies, fuel, clothes, and other essential goods that are ready to be consumed [1].

Maritime trade has increased almost yearly since 1970, except for 2020, due to the global COVID-19 pandemic [1]. We can also observe this growing trend in maritime trade in Europe, where many countries have invested in maritime port infrastructures [2], including Spain, where this work took place [3].

The existence of a worldwide transport network and the globalization of production and consumption have fostered competition between ports to attract the largest number of customers. In order to attract more customers, port terminals must be as competitive as possible [4]. A port must be safe and efficient to be competitive. These two characteristics are not mutually exclusive, and measures to improve safety would also impact efficiency and vice versa.

Besides having good shelter conditions, connections, and large surfaces for product handling and storage, the ship's stay in port must be safe, allowing it to carry out loading and unloading operations securely. Safety also includes the physical safeguard of port personnel working on port operations (besides loading and unloading cargo).

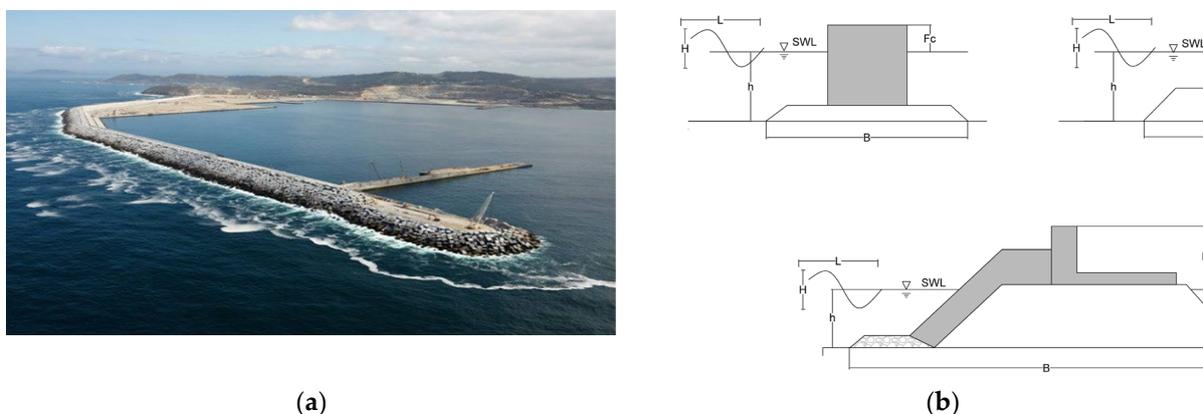
The overall importance of maritime trade in the world economy due to its weight in the consumption economy, and the particular importance in the case of Spain, reveals that events that could disrupt the normal operations of a port, impacting the operation safety or performance, can have a significant negative economic impact. This importance also reveals that improvements in safety or optimization of the operations in a port can have a significant positive economic impact.

A commercial port aims to harbour ships and provide safe loading areas where ships can load and unload cargo (or passengers). A loading area is only safe if the port has some large structure that shelters and protects the ships moored there from the force of powerful waves.

One element a port uses to protect the loading areas is a breakwater (also known as a jetty). A breakwater is a permanent protective structure built in a coastal area to protect anchorages against waves, tides, currents, and storm surges. A breakwater achieves this protection through the reflection and dissipation of the incident wave energy. Although breakwaters may also be used to reduce coastal erosion, to protect beaches, for instance, this work focused on studying breakwaters as a port-protection structure.

A breakwater may be connected to land or be freestanding and may contain a walkway or road for vehicle access. A standard breakwater section is usually composed of a core made of granular material, protected with armour layers of rockfill (or riprap, or different pieces like cubes, tetrapods, or similar elements), and covered with a superstructure of mass or reinforced concrete.

The most conventional types of breakwaters are vertical, composite, and sloping (or a rubble mound). We can see these three types of breakwaters in Figure 1b (source [5]), where  $h$  is the depth at the site, and  $H$ ,  $L$ ,  $F_c$ , and  $B$  are, respectively, the characteristic wave height, characteristic wavelength, freeboard, and the representative magnitude of the breakwater width.



**Figure 1.** (a) Breakwater at the outer port of Punta Langosteira, A Coruña, Spain. (b) Breakwater sections and parameters representing vertical, composite, and sloping breakwaters.

Figure 1a (source [6]) shows a sloping breakwater with all the previously explained elements. This image corresponds to the outer port of A Coruña, Spain, where we carried out the primary fieldwork for this work.

Even a properly designed breakwater cannot protect the ships anchored there from all the possible wave-caused phenomena, i.e., a breakwater protects up to a point; it is designed to protect against phenomena up to a given magnitude. However, budget and physical constraints do not allow a breakwater to be built to protect ships against every wave-caused phenomenon.

The main disrupting wave-related event that could hinder the protection offered by a breakwater is wave overtopping. Wave overtopping is a dangerous phenomenon that occurs when waves meet a submerged reef or structure or an emerged reef or structure with a height up to double the approximate wave height. The latter case is the one that affects a port's breakwater and the one studied in this work.

During an overtopping, two processes occur: wave transmission and water passing over the structure. This work focuses on measuring the passing of water over the structure (a breakwater in this case). This process can occur in three different ways, either independently of each other or combined:

- Green Water is the solid (continuous) step of a given volume of water above the breakwater's crown wall due to the wave's rise (wave run-up) above the exposed surface of the said breakwater.
- White Water occurs when the wave breaks against the seaside slope. This event creates so much turbulence that air is entrained into the water body, forming a bubbly or aerated and unstable current and water springs that reach the protected area of the structure either by the wave's impulse or as a result of the wind.
- Aerosol is generated by the wind passing by the crest of the waves near the breakwater. Aerosol is not an especially meaningful event in terms of the damage it can produce, even in the case of storms. This case is the least dangerous of the three, and its impact on the normal development of port activities is negligible.

Regarding the damage they could cause, the more critical overtopping types are green and white water. So, this work focused on these overtopping types. Figure 2 shows an example of White Water overtopping captured in the outer port of Punta Langosteira, A Coruña, as part of this work.



**Figure 2.** Overtopping example. Outer port of Punta Langosteira, A Coruña, Spain.

### 1.1. Recommendations for Overtopping Limits and Mitigation

A breakwater constructed to protect a harbour area should satisfy rigorous conditions if we want to reduce wave overtopping and its impact. Several proposals in the literature define those conditions and propose recommendations to study wave overtopping. The most widely used manual regarding wave overtopping is the EurOtop manual [7], which guides designers and operators of breakwaters, reclamations, inland lakes, or reservoirs on analysing and predicting wave overtopping for flood defences attacked by wave action. The manual describes the main variables involved in an overtopping event and provides formulas and tools for predicting said events based on tests carried out in hydraulic laboratories. However, the manual notes that one must consider the specific characteristics of a port when using these resources due to the sensitivity of overtopping to changes in the geometry of the port, incident waves, and the local bathymetry.

Although it is impossible to establish precise limits to tolerable overtopping for all conditions, this manual includes some guidance on tolerable mean discharges and maximum overtopping volumes for various circumstances or uses. The manual also emphasises that these limits may be adopted or modified depending on the site's circumstances and uses.

Another recommendation for the conditions a breakwater should satisfy to protect a harbour area by reducing wave overtopping and its impact is the Spanish Ports System's Recommendations for Breakwater Construction Projects, ROM 1.1–18 [5]. ROM 1.1–18, for instance, indicates that the conditions a breakwater satisfies regarding wave energy transmission, either from the overtopping of the crown or propagation through the breakwater section, are the following: the breakwater's relative height  $(h + F_C)/H$ , which generally has dimensions on the order of 1,  $O(1)$ , and relative width,  $B/L$ , on the order of  $1/10$ ,  $O(1/10)$ , where  $h$  is the depth at the site, and  $H$ ,  $L$ ,  $F_C$ ,  $B$  are, respectively, the characteristic wave height, characteristic wavelength, freeboard, and the representative magnitude of the breakwater width (see Figure 1).

When a wave overtopping occurs in a commercial port environment, the best-case scenario will be the disruption of activities. Even this scenario has negative financial repercussions. A system that detects overtopping events would provide valuable information to port operators, allowing for the minimisation of the impact of overtopping, including the financial impact, property damage, or even physical harm to port workers. Wave overtopping has traditionally been studied from three possible approaches: small-scale physical modelling, numerical modelling, and in situ measurement.

### 1.2. Overtopping: Small-Scale Physical Modelling

Small-scale physical modelling has traditionally been one of the most used techniques for studying wave overtopping. It involves a small-scale port model, where the overtopping will be studied. These models allow the most significant physical phenomena that intervene in overtopping to be reproduced.

In this type of modelling, the tests are usually conducted using a wave energy spectrum, which is then used to generate a free surface wave time series at the wave paddle. When the water overpasses the model's breakwater due to overtopping, it is channelled and collected using structures built to measure the overtopping water volume.

Small-scale modelling is used, for instance, to determine the influence of certain factors in wave overtopping. For example, in [8], the authors studied the influence of wind waves on overtopping.

The methods involved in this type of modelling continue to evolve. The tests usually use a wave energy spectrum free surface wave time series at the wave paddle. Although this method could generate an infinite number of time series, only one is usually generated due to the expense of running multiple tests in a physical model. Hannah et al. proposed an improvement to this method in [9], where the authors used several time series generated from the same wave energy spectrum to study the variation in the main overtopping

measures. As a result, they showed that using different time series gives different results for some of the overtopping measures, indicating that this is the correct approach.

This method continues to generate new knowledge. For instance, Lashley et al. [10] proposed a new formula for calculating wave overtopping at vertical and sloping structures with shallow foreshores using deep-water wave characteristics. Another example is the work of Orimoloye et al. [11], where the authors conducted an experimental study focusing on overtopping characteristics of coastal seawalls under random bimodal waves and, based on the experimental results, proposed a modification to the EurOtop formula to better capture the overtopping discharge under bimodal conditions.

An example of a different technique for detecting overtopping events is the one used by Formentin and Zanuttigh [12]. In this work, the authors used a semi-automatic and customisable method based on a threshold-down-crossing analysis of the sea surface elevation signals obtained using consecutive gauges. They applied their method to new and past data and compared the results with well-established formulae, obtaining accurate and reliable results.

Although new and clever techniques exist to measure overtopping on a physical model, the downside is that some are effective in a laboratory but not practical in a real environment due to the high number of elements required or the need to construct a gigantic structure to collect the water of an overtopping event, and the high economic cost and time consumption.

### 1.3. Overtopping: Numerical Modelling

The development of computational models for overtopping simulations progressed along with the advances in computing power. For example, some works analysed the wave-infrastructure interaction using 2D or 3D models to study overtopping [13–15]. Alternatively, other works use Machine Learning models to perform water flow estimations during overtopping [16] or predict overtopping events [17,18].

Public institutions and private companies have developed software which includes computational models for overtopping simulations. An example of such a tool is the widely used Computational Fluid Dynamics Toolbox OpenFoam [19]. OpenFOAM has many features to solve complex fluid flows involving chemical reactions, turbulence and heat transfer, acoustics, solid mechanics, and electromagnetics.

Although not technically numerical modelling, it is worth mentioning that, alongside the first EurOtop manual in 2007, they also developed an online Calculation Tool to assist the user through a series of steps to establish overtopping predictions for embankments and dikes, rubble mound structures, and vertical structures. Due to lacking funds, they could not update the web-based Calculation Tool with the new formulae in the manual's latest edition. They removed the old calculation tool from the website but kept PC-Overtopping, a PC version of the tool [20].

Also, in parallel with the EurOtop manual, the authors developed an Artificial Neural Network called the EurOtop ANN. This ANN can predict the mean overtopping discharge using the hydraulic and geometrical parameters of the structure as inputs. The authors created this ANN using the extensive EurOtop database extended from the CLASH database, which contains more than 13,000 tests on wave overtopping. The ANN and both databases are free and available through links on the website [21,22].

Although the evolution and support experienced by tools like OpenFOAM and the EurOtop tools have positioned them as a viable alternative to physical modelling due to their low cost and adaptability, they present the same limitations as physical modelling. These limitations are caused by adopting simplifications of reality in the model to maintain an acceptable computational cost. Another problem with these numerical models is that, given the phenomena's complexity, real data should be used to validate the models whenever possible.

#### 1.4. Overtopping: In Situ Measuring

The third alternative for analysing the wave overtopping phenomenon is in situ measurement, i.e., using devices that record or measure overtopping events in a real scenario.

An example of the in situ measurement of wave overtopping is the work of Briganti et al. [23]. In this work, the authors measured wave overtopping at the rubble mound breakwater of the Rome–Ostia yacht harbour, constructing a full-scale station similar to the ones used in small-scale physical modelling. This structure collects the water from an overtopping event and allows it to be measured. The authors studied wave overtopping during one measurement campaign (2003–2004) and compared the data to the predictions obtained using known formulae based on small-scale model tests. Their results show that these formulae tend to underestimate overtopping events. The same port was also studied in a later work [24]. A similar method was used to study the Zeebrugge port in Belgium [25].

Modern approaches to detecting wave overtopping involve using image devices. For instance, Ishimoto Chiba and Kajiya [26] created a system that used video cameras to detect an overtopping event. The authors set trigger points in different areas of the video cameras' images, and the system could detect overtopping in those areas. When an overtopping was detected, the system sent several images (pre- and post-overtopping) to the servers, where they were analysed.

Similar work was carried out by Seki, Taniguchi, and Hashimoto [27], where the authors created a method for detecting the overtopping wave and other high waves using images from video cameras. In this work, the authors detected overtopping by measuring the wave contour in every video frame using Active Contour Models and tracking the contour. This method has the advantage of being robust to other moving objects. The authors showed the effectiveness of their method by experimenting with actual video sequences of both an approaching typhoon and calm scenes.

More recently, Chi et al. [28] used a shore-based video monitoring system to collect coastal images of a wave overtopping at a 360 m long sea dike during a storm in July 2018 on the Rizhao Coast, China. The system captured images with a sampling frequency of 1 Hz in the first 10 min of each hour during daylight. Using the images, the authors calculated the frequency, location, width, and duration of individual overtopping events, detecting 6252 individual overtopping events during the ten-hour storm. The results of this work indicate the feasibility of a shore-based video monitoring approach to capture the main features of wave overtopping in a safe and labour-saving manner while enabling a detailed analysis of the temporal and spatial variation of wave overtopping.

Despite providing precious information for detecting the problems that affect a given installation, in situ measurement constitutes a less widespread methodology than physical or numerical modelling. The less-general usage of in situ measuring is mainly due to the high economic cost involved, on the one hand, in acquiring monitoring equipment and, on the other, in carrying out extensive field campaigns. Despite the smaller number of scientific studies using in situ measuring and the low amount of data they have collected compared to the other techniques, in situ measuring is a good tool for evaluating operational problems in port facilities due to overtopping events.

Wave overtopping events have an enormous negative impact on a port in terms of the safety of workers and the efficiency of operations. In addressing the escalating risks posed by wave overtopping in the context of climate change, a technical examination reveals the multifaceted impact of global warming on coastal infrastructure resilience, as explained in the Special Report on the Ocean and Cryosphere in a Changing Climate [29] of the Intergovernmental Panel on Climate Change (IPCC). This report elucidates the mechanisms through which climate change can increase the number and severity of overtopping events.

First, the IPCC report articulates a concerning trend of accelerated sea level rise, attributed to the thermal expansion of seawater as it warms and the increased melting of land-based ice. This phenomenon directly elevates the baseline from which storm surges

and wave events originate, reducing the margin of safety provided by coastal defence structures designed based on historical sea level data.

Moreover, the IPCC report forecasts increased frequency and intensity of extreme weather events, including storms that generate significant wave heights. When combined with elevated sea levels, these heightened wave conditions present a compound risk that can surpass the design thresholds of existing coastal defences, leading to an increased number and severity of overtopping events. Current estimations project that, by the end of the 21st century, the globally aggregated annual overtopping hours will be up to 50 times larger than today [30].

The technical challenges climate change poses necessitate advancements in predictive modelling techniques for wave overtopping. The best way to do so is to facilitate in situ measuring overtopping events by developing easy-to-use techniques that use low-cost measuring equipment while providing good results and creating Machine Learning predictive models that capture the complex interplay between sea state, weather conditions, and wave overtopping events to predict the latter.

The potential for these models to enhance predictive accuracy in the face of evolving climate patterns is invaluable in planning and implementing adaptive measures to safeguard coastal infrastructures against the augmented risks associated with a changing climate.

## 2. Materials and Methods

The wave overtopping events data were collected in six field campaigns in the outer port of Punta Langosteira (A Coruña) from 2015 until 2020. During these campaigns, 3098 individual overtopping events were identified. The overtopping events data were integrated with data about the environmental conditions (weather conditions and sea state) the port was subject to during the overtopping events to assemble the datasets used to create several Machine Learning overtopping prediction models.

In the production of their predictions, the models will use the weather conditions and sea state forecasts provided by the Portus platform of the Spanish ports system [31] as inputs and output the probability of an overtopping event for a given input.

The Portus system forecasts the weather and sea state for 72 h. Several variables characterise the sea state, and the sea state forecast is provided as a single hourly data point for that set of variables, i.e., Portus provides a single hourly value for the variables that describe the sea state for the following 72 sea states.

The models created in this work use a single time value of the meteorological conditions and the sea state (the single value for a specific hour, not a time window) as inputs and indicate, as output, whether or not there will be an overtopping event with these conditions. Thus, given that we have a 72 h weather state and sea conditions forecast, we can predict overtopping events for each of the next 72 h.

The overtopping prediction is a classification problem: a model predicts a positive event (overtopping) if the model output exceeds the decision threshold and a negative one (no overtopping) if not.

Besides creating the models, the decision threshold must be studied to evaluate how its value impacts the tradeoff between security and economic cost: a low decision threshold would capture many overtopping events, providing more security, but would have many false positives incurring in higher costs due to stopping port operations, while having a higher decision threshold would capture less overtopping events but with more certainty (less false positives), providing less security, but incurring lower operational costs due to stopping port operations only with a higher certainty of an overtopping event.

In a classification problem, the model's accuracy is usually used as the metric to evaluate its performance, but this is not the best approach if the data are not balanced (when there are several times more samples from one target class (majority class) than from the other (minority class)).

There are several more appropriate metrics (than the accuracy) when dealing with unbalanced datasets that mainly rely on precision and recall, such as the  $F_1$  score (the

harmonic mean of the precision and recall) or the more general F-measure is the weighted harmonic mean of the precision and recall.

No consensus exists on which prevalence values make a dataset imbalanced. In this work, we consider the ranges defined in [32] that indicate the degree of imbalance of a dataset given the percentage of examples belonging to the minority class on the dataset: 20–40% mild, 1–20% moderate, and <1% extreme imbalance.

In order to study how a classifier will behave when changing the decision threshold, more advanced metrics, such as the ROC curve and the associated AUROC [33,34], can be used. However, these metrics are inappropriate when working with imbalanced datasets [35]. In this case, it is better to use the precision–recall curve, which computes a curve from the raw classifier output by varying the decision threshold and calculating the precision and recall at every threshold value. The area under the precision–recall curve can be estimated using the average precision (AP, with values between 0 and 1; the higher, the better). There are multiple ways of calculating the AP. This work uses the definition:

$$AP = \sum_n (R_n - R_{n-1})P_n \quad (1)$$

where  $P_n$  and  $R_n$  are the precision and recall at the  $n$ th threshold. Calculating the area under the curve does not use linear interpolation of points on the precision–recall curve because it provides an overly optimistic measure of the classifier performance [36,37].

Another essential aspect when dealing with an imbalanced dataset is appropriately separating the data into folds when using  $k$ -fold cross-validation to choose among several models (validation). Using regular  $k$ -fold in imbalanced datasets could cause many folds to only have examples of the majority class, negatively affecting the training process. In such cases, it is recommended to use stratified sampling (stratified  $k$ -fold) to ensure that the relative class frequencies are (approximately) preserved in each train and validation fold.

The variables used to create the wave overtopping prediction models can be classified into meteorological conditions (weather conditions and sea state) and overtopping characteristics (location, number of events, and size). The models use the meteorological conditions as inputs and a binary output that indicates whether an overtopping event will happen.

Three dataset variants were created in order to assess which one works better. Two use real historical data for the sea state and weather conditions. The third uses historical data from the Portus forecasting system (the stored predictions instead of real data).

The real data present more variation than the predictions, and the model could better capture the relationship between the actual sea state and weather conditions and the overtopping events a priori.

The historical data from the forecasting system were used for the third dataset variant because the production system will use the forecast data as input to the models. Thus, training the models on data belonging to the same distribution of the production data could provide better results.

For the datasets based on real historical data, the weather conditions and sea state data were gathered using the three measurement systems provided by the port technological infrastructure: a directional wave buoy, a weather station, and a tide gauge.

The directional wave buoy is located 1.8 km off the main breakwater of the port (43°21'00" N 8°33'36" W). This buoy belongs to the Coastal Buoy Network of Puertos del Estado (REDCOS [38]) and has the code 1239 in said network. Due to the sensibility of the buoy data to noise, the system provides its data aggregated in 1 h intervals. This quantisation allows the calculation of some statistical parameters that reflect the sea state while mitigating the noise's effects.

The weather station is located on the main port's breakwater. It belongs to the Port Meteorological Stations Network (REMPOR [39]) and provides data aggregated in 10 min intervals.

The tide gauge is also placed at the end of the main port's breakwater and belongs to the Coastal Tide Gauges Network of Puertos del Estado (REDMAR [40]), with code 3214 on said network. It provides the data in 1 min intervals.

From all the variables these data sources provide, only the ones that were also available in the port's forecast system were chosen to create the first dataset. This work aims to create wave overtopping prediction models that can predict an overtopping event using weather and sea state forecasts as inputs. In order to fulfil this objective, the models' inputs regarding ocean–meteorological variables must be available in the port's forecast system.

The second dataset adds the maximum wave height variable to the previous one. This variable is currently unavailable as a forecast, but Portus is currently modifying its forecasting models to include it, so this dataset was created to assess whether the inclusion of this variable will provide better results once this information is available.

The following are all the meteorological variables available to use as input variables to the models:

- HS (m): the significant wave height, i.e., the mean of the highest third of the waves in a time series representing a specific sea state.
- Hmax (m): the maximum wave height, i.e., the height of the highest wave in a time series representing a specific sea state. This variable is used only in the second dataset and is not available as forecast.
- TP (s): the peak wave period, i.e., the period of the waves with the highest energy, extracted from the spectral analysis of the wave energy.
- $\theta_m$  (deg): the mean wave direction, i.e., the mean of all the individual wave directions in a time series representing a specific sea state.
- WS (km/h): the mean wind speed.
- Wd (deg): the mean wind direction.
- H0 (m): the sea level with respect to the zero of the port.
- H0 state: a calculated variable created to indicate whether the tide is rising, high or low, or falling (1, 0, −1).
- H0 min (m): the minimum sea level with respect to the zero of the port achieved during the current tide.
- H0 max (m): the maximum sea level with respect to the zero of the port achieved during the current tide.

In the datasets based on real historical data for the sea state and weather conditions, the buoy provides  $H_S$ ,  $T_P$ , and  $\theta_m$ , the weather station provides  $W_s$  and  $W_d$ , and the tide gauge provides  $H_0$  and  $H_{max}$ . The rest of the variables are calculated from the raw data.

The Portus system provides all the data for the dataset based on historical forecasted data.

The three datasets use whether an overtopping event occurred or not as the model's output.

During the field campaigns for collecting overtopping data, the buoy and tide gauge in the outer port malfunctioned several times due to storms and other phenomena. Due to this, there are many overtopping events for which there are no associated real meteorological and sea state conditions, only the historical predicted data. Hence, the dataset based on predicted conditions has more examples than the ones based on real conditions (as shown in Table 1).

**Table 1.** Wave overtopping datasets: available data for each dataset (real and predicted data) and purpose (training or testing).

Dataset	Training (h)		Testing (h)	
	Overtopping	No Overtopping	Overtopping	No Overtopping
Real data without hmax	6656	87	2219	27
Real data with hmax	6656	87	2219	27
Predicted data	11,131	168	3710	55

Creating the datasets based on real data involved joining the data provided by the available data sources (buoy, weather station, and tide gauge). These data sources use different data sampling frequencies, so all the data had to be aggregated to have the same frequency as the frequency of the data with the largest period, the sea state data. These data have a period of 1 h, the same as the forecast systems predictions (the data that will be used as the models' inputs in production).

The overtopping events had to be aggregated accordingly. For this, the number of events and their magnitudes were aggregated for every hour, and the output label was set to one if one or more overtopping events happened during the hour and zero otherwise.

After merging, the data were preprocessed to improve the models' performance. Since  $\theta_m$  and  $W_d$  are directional variables in the 0–360 degrees range; 0 and 360 represent the same direction. Using this codification for the directional variables would cause the Machine Learning models to create different internal representations for the same state and could negatively impact the models' performance. This potential problem was avoided using the sine and cosine for each value, so the 0 and 360 directions are close together (sin and cosine vary continuously from 0 to 360).

In the final datasets, the positive examples of overtopping events represent around 1.2% of the total from the two real-data datasets and around 1.5% of the dataset based on predictions. These figures situate these datasets on the lower bound for moderate imbalance datasets, making them almost extremely imbalanced. Due to this, we used the appropriate metrics that consider the data imbalance, as previously explained.

Once merged, aggregated, and preprocessed, the data sources were separated into the corresponding training and testing datasets, used to train and test the different models. The test dataset is usually created by randomly choosing data from the whole dataset. However, having imbalanced datasets would cause the test results to be inaccurate, and the models could be trained with just negative examples, potentially giving the models excellent and unrealistic training results. To avoid these problems, the test sets were created so that they contain a proportion of positive overtopping instances similar to the corresponding training dataset; in each dataset, the testing dataset is around 1/3 of the whole dataset, and the proportions of positive and negative examples are similar among the training and testing datasets, as shown in Table 1.

Hundreds of models based on artificial neural networks were created, trained with the three datasets previously explained, and compared to check which model and dataset combination provided the best results.

The models are based on artificial neural networks (ANNs) [41], given the good results this Machine Learning model can achieve for classification problems. Given that these models require choosing multiple hyperparameters (number of layers, nodes per layer, and training parameters, among others explained later), we used an iterative approach to choose the best hyperparameters while reducing the time required to obtain them:

1. First, several allowed values for each hyperparameter are manually specified, and a hyperparameter grid is created (the outer product of all hyperparameters' values, i.e., all possible combinations).
2. Then, as many individual models as combinations in the grid are created, trained, and compared using cross-validation.
3. The results are analysed, focusing on the grid regions where the best results were obtained to create a new grid to contain that region.
4. Steps 2–3 are repeated several times for each dataset until a satisfactory cross-validation performance is obtained.

The hyperparameters and all the values tested with this approach are the following:

- Network architecture: We tested networks with 1, 2, 3, and 4 layers. They were also tested with 8, 16, 32, 64, 128, and 256 neurons per layer in each case, allowing each layer to use a different number of neurons. All ANNs used fully connected layers. Although the largest networks could seem too complex for the overtopping prediction problem, they were included to use dropout regularisation. Dropout will likely obtain

a better performance on a larger network (it will prune the network, resulting in a smaller one).

- Weights (kernel) initialisation: Several weights initialisations were used, both the uniform and the normal variations, for random, He [42] and Glorot (also known as Xavier) [43] initialisations.
- Optimiser: The Adam [44] optimiser was used.
- Training iterations: The optimiser was tested with 500, 1000, 1500, 2000, and 5000 iterations. In the initial tests, the results did not improve from 2500 iterations up, so the 5000 values were discarded for the subsequent iterations of the hyperparameters search approach.
- Activation function: All the ANNs used the ReLU activation function for the neurons in the hidden layers and the sigmoid function for the output layer.
- Regularisation: The ANNs were regularised using dropout regularisation [45], with one dropout layer for each hidden layer (except the output layer). Several dropout rates were tested: None, 0.0625, 0.125, 0.25, 0.375, and 0.5, where None indicates that no dropout was used.
- Layer weight constraints: A maximum norm constraint on the model parameters was set during training to avoid overfitting, i.e., constraints were put on the weight's incident to each hidden unit to have a norm less than or equal to the desired value. The values None, 1, 3, and 5 were tested, where None indicates no constraint.
- Learning rate (lr): The values 0.00001, 0.0001, 0.001, and 0.01 were tested for the learning rate. Although 0.01 is usually considered too large a learning rate, large values are suggested in the original dropout paper [45].
- Batch size: The datasets have several thousand examples, so several batch size values were tested to speed up the training process (the dataset is divided into smaller batches, and the network weights will be updated after processing each batch). The None, 100, 500, and 1000 values were tested, where None indicates that the whole dataset is to be processed before updating the model's parameters (weights).

As previously explained, the wave overtopping dataset is a moderate–extreme imbalanced dataset. In an imbalanced dataset, regular k-fold cross-validation could potentially give incorrect results [35]. This potential problem was avoided using stratified k-fold, so all the folds used during cross-validation have similar class proportions.

In this problem, we are most interested in obtaining a model that detects as many overtopping events as possible (high recall) to increase the port's safety while having low false positives (high precision) to maximize the port's operativity. Precision and recall are increased and decreased by varying the decision boundary of the ANN's raw output sigmoid function, but usually, increasing one decreases the other. The port's stakeholders are still determining which decision boundary (threshold) is the one that gives the best tradeoff from a practical perspective of the port's safety and performance. Some metrics, such as the balanced accuracy and the F-score, use a 0.5 decision threshold. We used the average precision to choose the best model in cross-validation (although we calculated several metrics to see the models' performance). As previously explained, the average precision calculates the model's overall performance when varying this threshold by obtaining the precision and recall for each threshold and calculating the area under the curve this creates.

Once we trained all the models using cross-validation and the iterative process previously explained, the best model for each dataset was selected and retrained on the whole train dataset (each model in its corresponding dataset). Once the final best three models were obtained, their performance was evaluated using the corresponding testing datasets. The whole process of creating the ANNs overtopping prediction models is summarised in Algorithm 1.

**Algorithm 1.** Iterative Grid-Based Training Process Algorithm Pseudocode

```

do
  manually define grid of model hyperparameters
  for each dataset do
    for each grid cell (model) do
      for each resampling iteration do
        hold-out specific samples using stratified k-fold
        fit model on the remainder
        calculate performance on hold-out samples using metric
      end
      calculate average performance across hold-out predictions
    end
  end
  determine best hyperparameter
  fit final model (best hyperparameter) to all training data
while error >= ε
evaluate final model performance in test set

```

**3. Results**

As explained earlier, several cross-validation metrics were calculated to visualise how the models behave during training (using cross-validation), although the deciding metric was the average precision. Thus, each model's precision, recall,  $F_1$  score, and average precision were calculated.

Table 2 shows the hyperparameters for the best ANN model obtained using the previously described training method. The table shows the best model for each metric used, allowing us to analyse the impact of each metric during cross-validation, i.e., which model would have been chosen had we used the corresponding metric. However, the deciding metric is the average precision for the reasons previously explained.

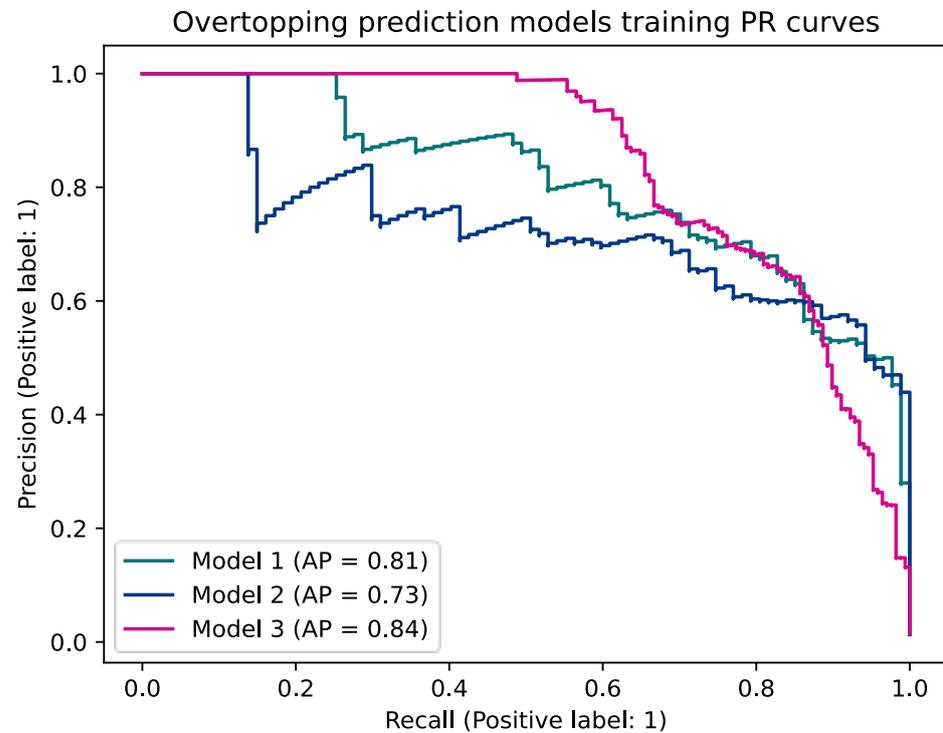
**Table 2.** Best wave overtopping models' hyperparameters for several metrics.

Model	Metric	Metric Value	Neurons Per Layer	Kernel Init.	Hyperparameter			lr	Batch Size
					Iter.	Drop. Ratio	Weight Constraint		
1	Precision	0.77	256, 128	He uniform	1000	0.25	1	0.00001	1000
	Recall	0.81	256, 128, 64	He uniform	1500	0.38	1	0.00001	1000
	$F_1$	0.65	256, 256	He uniform	1500	0.38	None	0.00001	1000
	AP	0.81	256, 128, 64	He uniform	1000	0.5	None	0.0001	1000
2	Precision	0.8	128, 64, 32	He uniform	1000	0.38	3	0.00001	None
	Recall	0.8	256, 128, 64	He uniform	1000	0.38	1	0.00001	500
	$F_1$	0.69	256, 128, 64	He uniform	500	0.38	1	0.00001	500
	AP	0.81	128, 64, 32	He uniform	1500	0.5	3	0.0001	1000
3	Precision	1	128, 64, 32	He uniform	1000	0.5	None	0.0001	1000
	Recall	0.74	256, 128, 64	He uniform	1000	0.38	None	0.00001	1000
	$F_1$	0.68	128, 64, 32	He uniform	1500	0.25	None	0.00001	500
	AP	0.82	128, 64, 32	He uniform	1500	0.25	None	0.00001	500

In this table and all the following tables and figures, models 1 and 2 are the ones trained on historical real data, model 2 being the one that includes the  $h_{\max}$  variable, and model 3 is the one trained on historical predicted data.

Table 2 shows that model 3 obtained the best results during validation, with an average precision of 0.82. Models 1 and 2 performed similarly, with an AP of 0.81. A fairly large Neural Network, with three hidden layers, is needed to solve the problem, resulting in a network of five layers (hidden plus input and output layers). The highest precision, 1, is achieved by a model trained on dataset 3, but the higher recall, 0.81, is obtained by a model trained on dataset 1.

After training the best model (the one with the best AP during validation), the three models' performances in the training dataset for different thresholds were compared using their PR curves, as shown in Figure 3.



**Figure 3.** PR curves for the wave overtopping models using the training datasets.

Figure 3 shows that model 3 obtained the best results, with an AP of 0.84 over the training data, while models 1 and 2 performed worse, achieving lower APs, as we can see with their curves decaying more rapidly than model 3's curve. Surprisingly, using the  $h_{\max}$  variable in model 2 made it perform worse than model 1, indicating that this variable makes it more difficult for the model to find the relationship between inputs and outputs.

The models' performance at specific points in the PR curve was checked using the 0.25, 0.5, and 0.75 threshold values as representatives of the following:

- 0.25: a lax model with lower precision but higher recall.
- 0.5: a regular model with in-between precision and recall.
- 0.75: a conservative model with higher precision but lower recall.

Table 3 shows the confusion matrices for the three models obtained using the above-mentioned thresholds on the training dataset models' predictions.

We can see in Table 3 that model 3 obtains the best overall results on the training data, although models 1 and 2 detect more overtopping using the 0.25 threshold. Model 3 performs better at detecting positive and negative cases for higher thresholds. Models 1 and 2 cannot detect overtopping events when using a 0.5 threshold in model 1 and 0.75 in model 2. This indicates that the positive predictions of models 1 and 2 have low values (the network outputs a low value), i.e., they predict values up to a point (less than 0.5 for model 2). In contrast, model 3 outputs values closer to 1 when it predicts an overtopping.

**Table 3.** Confusion matrices for the three wave overtopping models using a 0.25, 0.5, and 0.75 threshold decision boundary on the training data predictions.

Model	Threshold Real Class	Predicted					
		0.25		0.5		0.75	
		No Overtop.	Overtop.	No Overtop.	Overtop.	No Overtop.	Overtop.
1	No overtop.	6496	160	6,651	5	6656	0
	Overtop.	1	86	56	31	87	0
2	No overtop.	6559	97	6651	0	6656	0
	Overtop.	1	86	87	31	87	0
3	No overtop.	11,027	104	11,120	11	11,131	0
	Overtop.	22	146	63	105	96	72

Table 4 summarises the precision, recall, and F<sub>1</sub> score for each class label (1: overtopping, 0: no overtopping) using 0.25, 0.5, and 0.75 thresholds for the model’s predictions on the training data for the models trained on the three datasets. The table also includes each metric’s macro average (averaging the unweighted mean per label), the weighted average (averaging the support-weighted mean per class), and the precision each model achieves.

**Table 4.** Classification report for the three wave overtopping models using a 0.25, 0.5, and 0.75 threshold decision boundary on the training data predictions.

Model	Threshold	Precision			Recall			F <sub>1</sub>			Accuracy			Support
		0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75	
1	No overtop.	1.00	0.99	0.99	0.98	1.00	1.00	0.99	1.00	0.99	0.98	0.99	0.99	6656
	Overtop.	0.35	0.86	0.00	0.99	0.36	0.00	0.52	0.50	0.00				87
	Macro avg	0.67	0.93	0.49	0.98	0.68	0.50	0.75	0.75	0.50				6743
	Weigh. avg	0.98	0.99	0.97	0.98	0.99	0.99	0.98	0.99	0.98				6743
2	No overtop.	1.00	0.99	0.99	0.99	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.99	6656
	Overtop.	0.47	0.00	0.00	0.99	0.00	0.00	0.64	0.00	0.00				87
	Macro avg	0.73	0.49	0.49	0.99	0.50	0.50	0.81	0.50	0.50				6743
	Weigh. avg	0.99	0.97	0.97	0.99	0.99	0.99	0.99	0.98	0.98				6743
3	No overtop.	1.00	0.99	0.99	0.99	1.00	1.00	0.99	1.00	1.00	0.99	0.99	0.99	11,131
	Overtop.	0.58	0.91	1.00	0.87	0.62	0.43	0.70	0.74	0.60				168
	Macro avg	0.79	0.95	1.00	0.93	0.81	0.71	0.85	0.87	0.80				11,299
	Weigh. avg	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99				11,299

Table 4 shows that, as previously explained, models 1 and 2 detect more overtopping using the 0.25 threshold, i.e., they have a higher recall. For higher thresholds, model 3 outperforms models 1 and 2. Model 3 has a higher precision and F<sub>1</sub> score than models 1 and 2 at any threshold. We can also see that the precision can be misleading in an imbalanced problem. Due to the high prevalence of negative examples, all models achieve high precision (0.99) at every threshold, even when models 1 and 2 cannot detect any overtopping event for higher thresholds.

Now that we have observed how the models performed on the training data and concluded, given the validation performance of the three models shown in Table 2, that model 3 is the best among the three, it can be chosen as the final model.

Now, we will see how the models behave on the test data without using these data as a deciding factor, as we already determined model 3 to be the best at solving the wave overtopping prediction. Figure 4 shows the PR curves for the models obtained in the three testing datasets.

Figure 4 shows that model 3 obtains the best results, with an AP of 0.86 over the testing data, performing even better than on the training data. Models 1 and 2 perform worse than model 3, achieving lower APs (0.67 and 0.48, respectively), as shown in their curves decaying more rapidly than model 3’s curve. These models perform worse in testing than in training. Like in training, surprisingly, using the h<sub>max</sub> variable in model 2 made it perform worse than model 1 in testing.

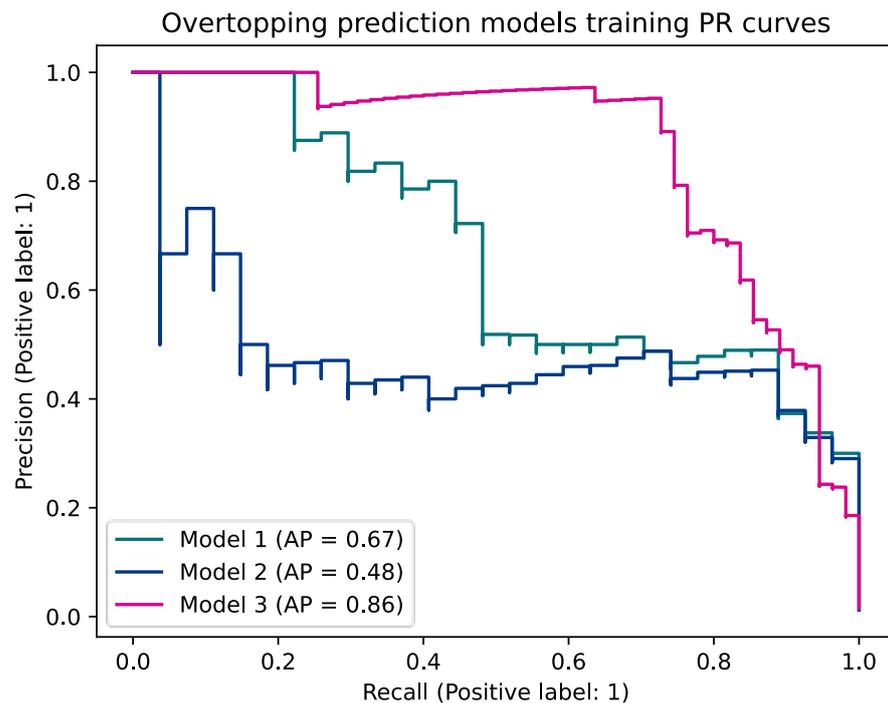


Figure 4. PR curves for the wave overtopping models using the testing dataset.

Table 5 shows the confusion matrices for the three wave overtopping models obtained evaluating the models’ predictions on the testing data at specific points in the PR curve using the 0.25, 0.5, and 0.75 threshold values (lax, regular, and conservative).

Table 5. Confusion matrices for the three wave overtopping models using a 0.25, 0.5, and 0.75 threshold decision boundary on the testing data predictions.

Model	Threshold Real Class	Predicted					
		0.25		0.5		0.75	
		No Overtop.	Overtop.	No Overtop.	Overtop.	No Overtop.	Overtop.
1	No overtop.	2162	57	2217	2	2219	0
	Overtop.	1	26	19	8	27	0
2	No overtop.	2191	28	2219	0	2219	0
	Overtop.	4	23	27	0	27	0
3	No overtop.	3670	40	3708	2	3709	1
	Overtop.	8	47	15	40	29	26

Table 5 shows that model 3 obtains the best overall results over the testing data, although model 1 detects more overtopping using the 0.25 threshold. Model 3 performs better at detecting positive and negative cases for higher thresholds. Both models 1 and 2 cannot detect any overtopping when using a 0.5 threshold in model 1 and 0.75 in model 2.

As seen when using the training data, this fact indicates that the positive predictions of models 1 and 2 have low values, i.e., they predict values up to a point (less than 0.5 for model 1). In contrast, model 3 outputs values closer to 1 when it predicts an overtopping using the testing data.

Table 6 summarises the precision, recall, and F<sub>1</sub> score for each class label (1: overtopping, 0: no overtopping) using 0.25, 0.5, and 0.75 thresholds for the model’s predictions on the testing data for the models trained on the three datasets. The table also includes each metric’s macro average (averaging the unweighted mean per label), the weighted average (averaging the support-weighted mean per class), and the precision each model achieves.

**Table 6.** Classification report for the three wave overtopping models using a 0.25, 0.5, and 0.75 threshold decision boundary on the testing data predictions.

Model	Threshold	Precision			Recall			F <sub>1</sub>			Accuracy			Support
		0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75	
1	No overtop.	1.00	0.99	0.99	0.97	1.00	1.00	0.99	1.00	0.99	0.97	0.99	0.66	2219
	Overtop.	0.31	0.80	0.00	0.96	0.30	0.00	0.47	0.43	0.00				27
	Macro avg	0.66	0.90	0.49	0.97	0.65	0.50	0.73	0.71	0.50				2246
	Weigh. avg	0.99	0.99	0.98	0.97	0.99	0.99	0.98	0.99	0.98				2246
2	No overtop.	1.00	0.99	0.99	0.99	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.72	2219
	Overtop.	0.45	0.00	0.00	0.85	0.00	0.00	0.59	0.00	0.00				27
	Macro avg	0.72	0.49	0.49	0.92	0.50	0.50	0.79	0.50	0.50				2246
	Weigh. avg	0.99	0.98	0.98	0.99	0.99	0.99	0.99	0.98	0.98				2246
3	No overtop.	1.00	1.00	0.99	0.99	1.00	1.00	0.99	1.00	1.00	0.99	1.00	0.77	3710
	Overtop.	0.54	0.95	0.96	0.85	0.73	0.47	0.66	0.82	0.63				55
	Macro avg	0.77	0.97	0.98	0.92	0.86	0.74	0.83	0.91	0.82				3765
	Weigh. avg	0.99	1.00	0.99	0.99	1.00	0.99	0.99	1.00	0.99				3765

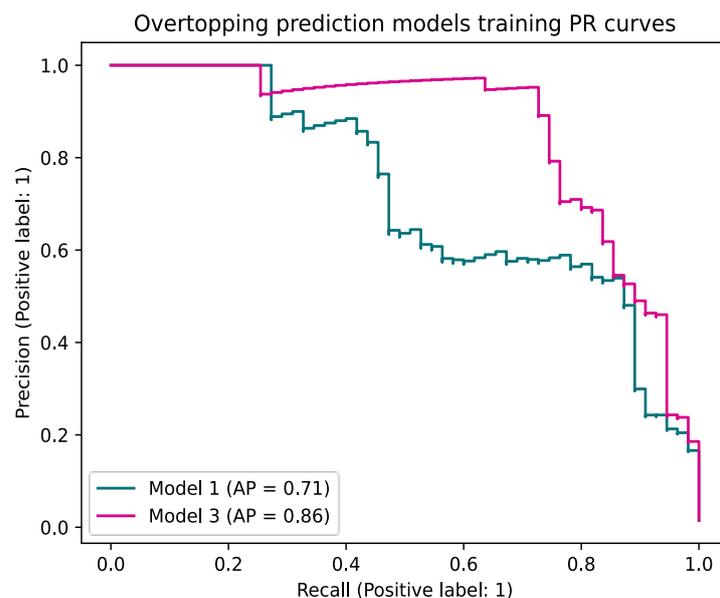
Table 6 confirms the data in Table 5: models 1 and 2 detect more overtopping events on the testing data using the 0.25 threshold (they have a higher recall). For higher thresholds, model 3 outperforms models 1 and 2. Model 3 has a higher precision and F<sub>1</sub> score than models 1 and 2 at any threshold.

As previously explained, once deployed in the production system, the wave overtopping models will input the Portus forecast data to make an overtopping prediction. Model 1, trained on real historical data, uses the same input variables as model 3, trained on historical data of the Portus forecast system.

The previous test results were obtained using the corresponding test dataset created from the original dataset. I.e., model 1 was tested with real historical data and model 3 with historical forecasted data.

Now, we will test both model 1 and model 3 on historical forecasted data and compare them to see how well model 1 will behave using forecasted data like it would in the production system.

Figure 5 shows the PR curves for models 1 and 3 using the testing data of model 3 (i.e., using historical forecasted data).



**Figure 5.** PR curves for the wave overtopping models 1 and 3 using the testing dataset of model 3 (historical forecasted data).

Figure 5 shows that model 3 outperforms model 1 with an AP of 0.86 over its testing data, while model 1 obtains an AP of 0.71 using model 3’s testing data.

Table 7 shows the confusion matrices for models 1 and 3 obtained using the 0.25, 0.5, and 0.75 threshold values as the decision boundary on the predictions over the testing dataset of model 3.

**Table 7.** Confusion matrices for the wave overtopping models 1 and 3 using a 0.25, 0.5, and 0.75 threshold on the predictions over the testing dataset of model 3.

Model	Threshold Real Class	Predicted					
		0.25		0.5		0.75	
		No Overtop.	Overtop.	No Overtop.	Overtop.	No Overtop.	Overtop.
1	No overtop.	3625	85	3709	1	3710	0
	Overtop.	6	49	40	15	55	0
3	No overtop.	3670	40	3708	2	3709	1
	Overtop.	8	47	15	40	29	26

Table 7 shows that model 1 performs slightly better than model 3 regarding recall over model 3’s testing data using the 0.25 and 0.5 thresholds, but model 3 outperforms model 1 overall. This affirmation is confirmed in Table 8, which shows that model 3 generally obtains higher macro precision, recall, and F<sub>1</sub> score, except for recall using the 0.25 threshold.

**Table 8.** Classification report for the wave overtopping models 1 and 3 using a 0.25, 0.5, and 0.75 threshold decision boundary on the testing data predictions.

Mod.	Threshold	Precision			Recall			F <sub>1</sub>			Accuracy			Support
		0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75	
1	No overtop.	1.00	0.99	0.99	0.98	1.00	1.00	0.99	0.99	0.99	0.98	0.99	0.99	3710
	Overtop.	0.37	0.94	0.00	0.89	0.27	0.00	0.52	0.42	0.00				55
	Macro avg	0.68	0.96	0.49	0.93	0.64	0.50	0.75	0.71	0.50				3765
	Weigh. avg	0.99	0.99	0.97	0.98	0.99	0.99	0.98	0.99	0.98				3765
3	No overtop.	1.00	1.00	0.99	0.99	1.00	1.00	0.99	1.00	1.00	0.99	1.00	0.99	3710
	Overtop.	0.54	0.95	0.96	0.85	0.73	0.47	0.66	0.82	0.63				55
	Macro avg	0.77	0.97	0.98	0.92	0.86	0.74	0.83	0.91	0.82				3765
	Weigh. avg	0.99	1.00	0.99	0.99	1.00	0.99	0.99	1.00	0.99				3765

Model latency refers to the delay between when input data are provided to a model and when the model outputs a prediction. In the context of the wave overtopping prediction models, model latency could be a significant concern for several reasons.

In a port environment, timely decisions are crucial for taking preventative measures and optimizing operations when predicting a potential risk. High latency in model predictions could delay the response time, reducing the effectiveness of these measures and potentially compromising port safety and efficiency. Also, port operators and managers expect prompt and accurate predictions to facilitate informed decision making. Any delay in receiving predictions could diminish stakeholder trust in the models, affecting their adoption and continued use.

For the models to be integrated seamlessly into port management systems and decision-making processes, they must operate within the time constraints of those systems. High latency could hinder the integration of these models into existing operational workflows, limiting their practical utility.

To address these concerns, we calculated the latency of model 3, simulating a production environment:

1. We create a dataset of 72 inputs, the same number the model will use in production.
2. We calculated the model latency for the dataset (predicting 72 outputs).
3. We ran step 2 10,000 times and calculated the mean latency. The simulations were run on a laptop (12th Gen Intel® Core™ i7-1260P (Intel, Santa Clara, CA, USA)).

On average, the model takes 0.03 s to make 72 predictions, so we can conclude that model latency is not a concern with the models created in this work.

#### 4. Discussion and Conclusions

The results obtained by these models show that model 3, created using the third dataset variant containing historical forecasted data, achieved the best overall results over the training data, achieving a 0.84 average precision. The third dataset variant also achieved the best overall results during testing, obtaining a 0.86 average precision, while models 1 and 2 obtained a 0.67 and 0.48 average precision, respectively. Comparing models 1 and 3 using the model 3's testing dataset, i.e., forecasted data more similar to that used in production, also shows that model 3 outperforms model 1.

Analysing these results, we can conclude that the historical forecasted data creates the best wave overtopping prediction model (model 3), i.e., creating a model using a dataset with data from the same data distribution as the data the model will use in production provides the best results, as initially hypothesised.

The results obtained by the ANNs overtopping prediction models are satisfactory but can be improved by obtaining more data. As previously explained, the overtopping datasets are imbalanced as there are significantly fewer instances with overtopping events than without them. Gathering more data would help improve the overall results by avoiding overfitting and reducing the need for regularisation. Not needing regularisation could accelerate the training process, which currently takes several days for each model.

However, gathering more data and creating new models is a slow process, as can be observed by the time it took to create the models that include the last processed data from the 2020 field campaign. The extensive time it takes to develop the models can be attributed to several factors, with data collection and preprocessing complexity being the main contributing factors. However, the iterative model development and validation also increase the time.

The overtopping recording devices were installed in the illumination towers of the port, the closest fixed power source. However, to detect more events, some locations should be closer to the breakwater to visualise overtopping events in adverse meteorological conditions, such as heavy rain or dense fog. This situation cannot be easily improved, but we could improve the overtopping detection method to speed up the process and gather data faster. Currently, detecting them involves the following tasks:

1. Periodically collecting the videos stored in the recording devices.
2. Inspecting the videos manually and annotating each overtopping event that can be visually detected.

This method of working has proven to be highly time consuming. The time and resources involved in the first task can be reduced by overcoming the lack of communication infrastructure in the outer port by installing 4G or satellite modems at the recording sites. This infrastructure would allow data to be sent in real time and the videos to be immediately available for processing.

The complexity continues after data collection. Each field campaign gathers tens of terabytes of video (multiple cameras during a year) that must be manually inspected to detect and annotate each overtopping event. This manual process is highly time-consuming (more than a year for each field campaign) and the main bottleneck in the data preparation phase. This task could be sped up by creating a Machine Learning model that automatically processes the videos without human interference and detects the overtopping events. In recent years, deep learning models have dramatically improved the performance of this type of task (image classification). Although creating such models involves initially, and mostly manually, creating large image datasets, we already have the videos of the overtopping events we detected. Also, the time involved in creating such a dataset would far outweigh the time involved in manually inspecting the videos to detect the overtopping events.

Besides improving the throughput of each task with these improvements, allowing us to grow the overtopping dataset faster, the synergy of both would allow new capabilities: the communication infrastructure would allow us to have a real-time video feed of the recording devices, and using this as input of the overtopping classification model developed for the second task would create an overtopping warning system that will inform port operators of an overtopping event in real-time.

As previously indicated, the iterative model development and validation also increase the model creation time. Developing the final models involves the previously explained iterative process of defining hyperparameters, training numerous models, and validating their performance through cross-validation. This iterative process, necessary for optimising model performance, is inherently time consuming, taking several months in our technological infrastructure, particularly in this work, where we trained the same models three times (one per dataset). Now that we know which model architecture and dataset provide the best results, the model training time can be significantly reduced in future works.

Another limitation of these results concerns the ANN-based models used to predict overtopping events, which is model interpretability, i.e., explaining the reason behind a specific model's output. The machine learning models created in this work are intended to replace traditional statistical methods used in port management tools. These traditional models are interpretable, and port operators are accustomed to this characteristic.

Using Machine Learning models, we cannot explain to port operators how and why the model came to the decision it did. The only tool a machine learning practitioner has in this situation is to convince through results, i.e., the models need to prove that they are superior to traditional statistical methods by providing more accurate predictions than traditional models.

During the development of this work, and after obtaining the results, we had several meetings with the A Coruña port authority and the port operators. After reviewing the results obtained in this work, they are convinced that Machine Learning methods provide better results than traditional methods despite the lack of interpretability, so they intend to use the models created in this work in their decision-making pipeline.

In conclusion, analysing the results of the overtopping prediction ANNs and the video recording infrastructure's ability to collect data for years without significant problems makes us conclude that the in situ measuring approach was appropriate, and the models created using the data gathered provided good results. Also, using historical forecasted data to create the models provides the best results, so this is the best approach, which we will focus on in future works.

**Author Contributions:** Conceptualization, A.F., E.P., A.A. and J.S.; methodology, A.A. and A.F.; software, A.A. and S.R.-Y.; validation, A.F., E.P., A.A. and J.S.; formal analysis, A.F., E.P., A.A. and J.S.; investigation, A.A. and A.F.; resources, A.A., A.F. and E.P.; data curation, A.A. and A.F.; writing—original draft preparation, A.A. and P.R.-S.; writing—review and editing, P.R.-S., A.F., J.S. and A.A.; visualization, S.R.-Y., P.R.-S., E.P. and A.F.; supervision, A.F., J.R. and E.P.; project administration, A.F., J.R. and E.P.; funding acquisition, J.R. and E.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Spanish Ministry of Science and Innovation [grant number PID2020-112794RB-I00, funded by MCIN/AEI/10.13039/501100011033].

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets presented in this article are not readily available to protect the confidentiality and security of the Outer Port of A Coruña. Requests to access the datasets should be directed to the corresponding author, and the data will be sent with the permission of the Port Authority.

**Acknowledgments:** The authors would like to thank the Port Authority of A Coruña (Spain) for their availability, collaboration, interest and promotion of research in port engineering.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. UNCTAD. *Review of Maritime Transport 2021*; United Nations: Geneva, Switzerland, 2021.
2. Adam, A.F.; Moldovan, I.A.G.; Nita, S.C.; Hrebenciuc, A. The Importance of Maritime Transport for Economic Growth in the European Union: A Panel Data Analysis. *Sustainability* **2021**, *13*, 7961. [CrossRef]
3. Del Estado, P. Historical Statistics since 1962. Available online: [https://www.puertos.es/en-us/estadisticas/Pages/estadistica\\_Historicas.aspx](https://www.puertos.es/en-us/estadisticas/Pages/estadistica_Historicas.aspx) (accessed on 27 July 2022).
4. Saieva, G. *Port Management and Operations*; S.I. Informa Law from Routledge: London, UK, 2013.
5. Losada Rodríguez, M.Á.; Instituto Interuniversitario de Investigación del Sistema Tierra en Andalucía Grupo de Dinámica de Flujos Ambientales. *ROM 1.1-18: (Articles), Recommendations for Breakwater Construction Projects*; Puertos del Estado: Madrid, Spain, 2019.
6. Port Authority of A Coruña. The Outer Port of A Coruña. Available online: <http://www.puertocoruna.com/en/oportunidades-negocio/puerto-hoy/puertoext.html> (accessed on 3 August 2022).
7. Van der Meer, J.W.; Allsop, N.W.H.; Bruce, T.; De Rouck, J.; Kortenhaus, A.; Pullen, T.; Schüttrumpf, H.; Troch, P.; Zanuttigh, B. *EurOtop: Manual on Wave Overtopping of Sea Defences and Related Structures. An Overtopping Manual Largely Based on European Research, but for Worldwide Application*; EurOtop: 2018. Available online: [www.overtopping-manual.com](http://www.overtopping-manual.com) (accessed on 8 August 2022).
8. Van der Werf, I.M.; Van Gent, M.R. Wave Overtopping over Coastal Structures with Oblique Wind and Swell Waves. *J. Mar. Sci. Eng.* **2018**, *6*, 149. [CrossRef]
9. Williams, H.E.; Briganti, R.; Romano, A.; Dodd, N. Experimental Analysis of Wave Overtopping: A New Small Scale Laboratory Dataset for the Assessment of Uncertainty for Smooth Sloped and Vertical Coastal Structures. *J. Mar. Sci. Eng.* **2019**, *7*, 217. [CrossRef]
10. Lashley, C.H.; van der Meer, J.; Bricker, J.D.; Altomare, C.; Suzuki, T.; Hirayama, K. Formulating Wave Overtopping at Vertical and Sloping Structures with Shallow Foreshores Using Deep-Water Wave Characteristics. *J. Waterw. Port Coastal Ocean Eng.* **2021**, *147*, 04021036. [CrossRef]
11. Orimoloye, S.; Horrillo-Caraballo, J.; Karunarathna, H.; Reeve, D.E. Wave overtopping of smooth impermeable seawalls under unidirectional bimodal sea conditions. *Coast. Eng.* **2021**, *165*, 103792. [CrossRef]
12. Formentin, S.M.; Zanuttigh, B. Semi-automatic detection of the overtopping waves and reconstruction of the overtopping flow characteristics at coastal structures. *Coast. Eng.* **2019**, *152*, 103533. [CrossRef]
13. Altomare, C.; Gironella, X.; Crespo, A.J. Simulation of random wave overtopping by a WCSPH model. *Appl. Ocean Res.* **2021**, *116*, 102888. [CrossRef]
14. Chen, W.; Warmink, J.; van Gent, M.; Hulscher, S. Numerical modelling of wave overtopping at dikes using OpenFOAM®. *Coast. Eng.* **2021**, *166*, 103890. [CrossRef]
15. Neves, M.G.; Didier, E.; Brito, M.; Clavero, M. Numerical and Physical Modelling of Wave Overtopping on a Smooth Impermeable Dike with Promenade under Strong Incident Waves. *J. Mar. Sci. Eng.* **2021**, *9*, 865. [CrossRef]
16. Mares-Nasarre, P.; Molines, J.; Gómez-Martín, M.E.; Medina, J.R. Explicit Neural Network-derived formula for overtopping flow on mound breakwaters in depth-limited breaking wave conditions. *Coast. Eng.* **2021**, *164*, 103810. [CrossRef]
17. Bieman, J.P.D.; van Gent, M.R.; Boogaard, H.F.v.D. Wave overtopping predictions using an advanced machine learning technique. *Coast. Eng.* **2020**, *166*, 103830. [CrossRef]
18. Hosseinzadeh, S.; Etemad-Shahidi, A.; Koosheh, A. Prediction of mean wave overtopping at simple sloped breakwaters using kernel-based methods. *J. Hydroinform.* **2021**, *23*, 1030–1049. [CrossRef]
19. OpenFOAM. Available online: <https://www.openfoam.com/> (accessed on 7 August 2022).
20. Pc-Overtopping-Overtopping Manual. Available online: <http://www.overtopping-manual.com/eurotop/pc-overtopping/> (accessed on 8 August 2022).
21. Neural-Networks-and-Databases-Overtopping Manual. Available online: <http://www.overtopping-manual.com/eurotop/neural-networks-and-databases/> (accessed on 8 August 2022).
22. Steendam, G.J.; Van Der Meer, J.W.; Verhaeghe, H.; Besley, P.; Franco, L.; Van Gent, M.R.A. The international database on wave overtopping. In Proceedings of the 29th International Conference, National Civil Engineering Laboratory, Lisbon, Portugal, 19–24 September 2004; pp. 4301–4313. [CrossRef]
23. Briganti, R.; Bellotti, G.; Franco, L.; De Rouck, J.; Geeraerts, J. Field measurements of wave overtopping at the rubble mound breakwater of Rome–Ostia yacht harbour. *Coast. Eng.* **2005**, *52*, 1155–1174. [CrossRef]
24. Franco, L.; Geeraerts, J.; Briganti, R.; Willems, M.; Bellotti, G.; De Rouck, J. Prototype measurements and small-scale model tests of wave overtopping at shallow rubble-mound breakwaters: The Ostia-Rome yacht harbour case. *Coast. Eng.* **2009**, *56*, 154–165. [CrossRef]
25. Geeraerts, J.; Kortenhaus, A.; González-Escrivá, J.; De Rouck, J.; Troch, P. Effects of new variables on the overtopping discharge at steep rubble mound breakwaters—The Zeebrugge case. *Coast. Eng.* **2009**, *56*, 141–153. [CrossRef]

26. Ishimoto, K.; Chiba, T.; Kajiya, Y. Wave Overtopping Detection by Image Processing. In Proceedings of the Steps Forward. Intelligent Transport Systems World Congress, Yokohama, Japan, 3–11 November 1995; Volume 1, p. 515. Available online: <https://trid.trb.org/view/461709> (accessed on 25 January 2020).
27. Seki, M.; Taniguchi, H.; Hashimoto, M. Overtopping Wave Detection based on Wave Contour Measurement. *IEEE Trans. Electron. Inf. Syst.* **2007**, *127*, 599–604. [[CrossRef](#)]
28. Chi, S.-H.; Zhang, C.; Sui, T.-T.; Cao, Z.-B.; Zheng, J.-H.; Fan, J.-S. Field observation of wave overtopping at sea dike using shore-based video images. *J. Hydrodyn.* **2021**, *33*, 657–672. [[CrossRef](#)]
29. Intergovernmental Panel on Climate Change (IPCC). *The Ocean and Cryosphere in a Changing Climate: Special Report of the Intergovernmental Panel on Climate Change*, 1st ed.; Cambridge University Press: Cambridge, UK, 2022; ISBN 978-1-00-915796-4.
30. Almar, R.; Ranasinghe, R.; Bergsma, E.W.J.; Diaz, H.; Melet, A.; Papa, F.; Vousdoukas, M.; Athanasiou, P.; Dada, O.; Almeida, L.P.; et al. A global analysis of extreme coastal water levels with implications for potential coastal overtopping. *Nat. Commun.* **2021**, *12*, 3775. [[CrossRef](#)]
31. Puertos del Estado Portus: Oceanography Data. Available online: <https://www.puertos.es/en-us/oceanografia/Pages/portus.aspx> (accessed on 9 March 2024).
32. Google. Imbalanced Data | Machine Learning. Google Developers. Available online: <https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data> (accessed on 8 November 2022).
33. Fawcett, T. An Introduction to ROC analysis. *Pattern Recogn. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
34. Powers, D.M.W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv* **2020**. [[CrossRef](#)]
35. Kulkarni, A.; Chong, D.; Batarseh, F.A. 5-Foundations of data imbalance and solutions for a data democracy. In *Data Democracy*; Batarseh, F.A., Yang, R., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 83–106. [[CrossRef](#)]
36. Flach, P.; Kull, M. Precision-Recall-Gain Curves: PR Analysis Done Right. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: New York, NY, USA, 2015; Volume 28, Available online: <https://proceedings.neurips.cc/paper/2015/hash/33e8075e9970de0cfea955afd4644bb2-Abstract.html> (accessed on 9 November 2022).
37. Davis, J.; Goadrich, M. The Relationship Between Precision-Recall and ROC Curves. In Proceedings of the 23rd International Conference on Machine Learning, ACM, Pittsburgh, PA, USA, 25–29 June 2006; pp. 233–240. [[CrossRef](#)]
38. del Estado, P. Red Costera de Boyas de Oleaje de Puertos del Estado (REDCOS). Red Costera de Oleaje de Puertos del Estado. Available online: <https://www.sidmar.es/RedCos.html> (accessed on 25 October 2022).
39. del Estado, P. Red de Estaciones Meteorológicas Portuarias (REMPOR). Red de Estaciones Meteorológicas Portuarias (REMPOR). Available online: [https://bancodatos.puertos.es/BD/informes/INT\\_4.pdf](https://bancodatos.puertos.es/BD/informes/INT_4.pdf) (accessed on 25 October 2022).
40. Del Estado, P. Red de Medida del Nivel del Mar y Agitación de Puertos del Estado (REDMAR). Red de Mareógrafos de Puertos del Estado. Available online: <https://www.sidmar.es/RedMar.html> (accessed on 25 October 2022).
41. Hastie, T.; Tibshirani, R.; Friedman, J. Neural Networks. In *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Hastie, T., Tibshirani, R., Friedman, J., Eds.; Springer: New York, NY, USA, 2009; pp. 389–416. [[CrossRef](#)]
42. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the International Conference on Computer Vision, Las Condes, Chile, 11–18 December 2015; pp. 1026–1034.
43. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, PMLR, Sardinia, Italy, 13–15 May 2010; Volume 9, pp. 249–256. Available online: <http://proceedings.mlr.press/v9/glorot10a.html> (accessed on 20 December 2021).
44. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization, CoRR, vol. *arXiv* **2014**. [[CrossRef](#)]
45. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.