Received: 19 October 2023 | Accepted: 8 July 2024

DOI: 10.1111/1365-2664.14734

Journal of Applied Ecology BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Using artificial neural networks and citizen science data to assess jellyfish presence along coastal areas

J. Castro-Gutiérrez[1,2] | J. C. Gutiérrez-Estrada[1] | J. C. Báez[3,4]

[1]Departamento de Ciencias Agroforestales, Escuela Técnica Superior de Ingeniería, Campus El Carmen, Universidad de Huelva, Huelva, Spain

[2]Departamento de Biología, Facultad de Ciencias del Mar y Ambientales, Campus Río San Pedro, Universidad de Cádiz, Puerto Real, Spain

[3]Centro Oceanográfico de Málaga, Instituto Español de Oceanografía (CSIC), Fuengirola, Spain

[4]Instituto Iberoamericano de Desarrollo Sostenible (IIDS), Universidad Autónoma de Chile, Temuco, Región de la Araucanía, Chile

Correspondence
J. Castro-Gutiérrez
Email: jairo.castro@dcaf.uhu.es

J. C. Gutiérrez-Estrada
Email: juanc@dcaf.uhu.es

Handling Editor: Tadeu Siqueira

Abstract

1. Jellyfish blooms along coastal areas can pose significant challenges for beach users and local authorities. Understanding the factors influencing jellyfish presence is crucial for effective management and mitigation strategies.
2. In this study, citizen science data from the Andalusian coast (232 beaches, in 40 different localities) and machine learning techniques are used to investigate if the presence and absence of jellyfish along coastal areas can be predicted. A multi-layer perceptron (MLP) neural network was employed to classify user comments regarding jellyfish presence or absence, achieving an accuracy of approximately 96%.
3. The MLP model demonstrated robustness in handling non-linear classification problems and noise, although it showed lower precision for predicting jellyfish presence, likely due to an imbalance in the dataset. Environmental data were also incorporated to characterise the influence of sea surface temperature, wind direction and wind speed on jellyfish distribution. The results align with previous studies, suggesting these environmental factors significantly impact jellyfish presence.
4. *Synthesis and applications*. This research provides actionable recommendations for beach management. The implementation of continuous monitoring of sea surface temperature and wind conditions will enable more accurate predictions of jellyfish distribution. Adaptive management strategies that respond dynamically to environmental data will help mitigate the impact of jellyfish blooms on coastal tourism and public health.

KEYWORDS
beach management, citizen science, Infomedusa APP, jellyfish, machine learning, multi-layer perceptron, sea surface temperature, wind conditions

1 | INTRODUCTION

Jellyfish populations have recently attracted significant attention due to the potential impact that the proliferation of these stinging organisms has on regional tourism, bathers and local ecosystems (Canepa et al., 2014; Pauly et al., 2009; Purcell et al., 2007; Richardson et al., 2009). Their stings can cause a risk to human health (Brotz et al., 2012; Mariottini et al., 2008; Mariottini & Pane, 2010).

Furthermore, the presence of jellyfish on the coast may deter tourists from visiting beach destinations, negatively affecting local economies (Brotz et al., 2012; De Donno et al., 2014; Galil, 2008; Ruiz-Frau, 2022).

Given these impacts, monitoring jellyfish occurrences for the public safety and management is crucial (Bellido et al., 2020; Ghermandi et al., 2015). Nevertheless, monitoring programmes covering large tracts of land over a relatively long period involve a financial and personnel investment that is unaffordable for any administration or research project. In this sense, data collected through citizen participation offer an alternative solution for monitoring environmental phenomena by leveraging user-generated information (Goodchild, 2007).

Citizen science, which involves public participation in scientific research, provides an effective solution for monitoring environmental phenomena by utilising user-generated information (Bonney et al., 2009; Silvertown, 2009). This approach leverages the contributions of non-scientists to achieve extensive spatial and temporal data coverage, making it especially valuable for large-scale environmental monitoring (Dickinson et al., 2010). By engaging the community, citizen science not only reduces the costs associated with traditional data collection methods but also enhances public awareness and involvement in scientific endeavours (Fritz et al., 2019). Various projects employing diverse approaches have been developed to mitigate the impact of jellyfish blooms in coastal areas using citizen science (Canepa et al., 2016; Marambio et al., 2021). This approach has the potential to be cost-effective and accurate, as it relies on observations from many users who are directly experiencing the coastal environment.

In light of this situation, the Provincial Diputación of Malaga, in partnership with Aula del Mar of Malaga developed in 2013 the Infomedusa APP. A smartphone application that allows citizens to monitor the presence and abundance of jellyfish swarms along the coast of Andalucía (south of Spain). Infomedusa APP functions as a forum where users can engage in discussions beyond the topic of jellyfish. Users can freely discuss other subjects, such as sea and beach conditions, and other topics related to their beach experiences. This open forum approach fosters a sense of community and encourages broader engagement among APP users while still providing valuable data on jellyfish occurrences, but it requires a great effort of data pre-processing to be able to perform statistical analyses on the jellyfish information.

In this regard, techniques such as those included in the machine learning field could help automate the extraction of jellyfish-related data from user comments, making the data analysis process more efficient and accurate (Tang et al., 2015; Young et al., 2018). In recent years, there has been a growing interest in text classification using artificial neural networks for information extraction (Hovy, 2022; Mukhamediev et al., 2022; Zhang, 2000). This is due to the increasing volume of data generated on online platforms and the subsequent need to manage this information effectively. Artificial neural networks can offer significant advantages in this area, including the identification and classification of jellyfish species from images and the ability to detect complex patterns and handle large-scale data sets (Do & Tran, 2022; Kim et al., 2016; Minaee et al., 2021).

Additionally, understanding the complex relationship between jellyfish and the environment is crucial for early predicting and managing potential jellyfish blooms in coastal areas. It has been shown that the presence of jellyfish on coasts is influenced by various environmental factors, such as sea surface temperature (SST), wind direction and wind speed. This way, rising SSTs have been linked to increased jellyfish abundance, as higher temperatures can accelerate their growth rates and reproduction, favouring the expansion of their populations (Purcell, 2005). For example, Avian et al. (1991) established that an elevated SST influences the population dynamics of *Pelagia noctiluca* by promoting ovule development and, consequently, their proliferation. Moreover, SST has been proven as one of the main drivers of cnidarian communities composition in the Mediterranean Sea (Guerrero et al., 2018). Wind direction and speed also play a significant role in the distribution and accumulation of jellyfish along coastlines. The wind can transport jellyfish to the shore through the process of wind-driven surface currents as previously studies highlighted (Castro-Gutiérrez et al., 2022; Graham et al., 2014; Gutiérrez-Estrada et al., 2021). Therefore, a better understanding of the combined effects of SST, wind direction and wind speed on jellyfish populations can contribute to more effective management and mitigation strategies to address their impact on coastal ecosystems and human activities.

The objective of this study was to investigate the potential of citizen science data and machine learning techniques to track jellyfish occurrences. To achieve this, we used a multi-layer perceptron (MLP) classifier to process and classify comments from the Infomedusa APP (version 6.7, developed by Desarrollos Digitales Malva2 2020 SL and Aula del Mar de Málaga). Additionally, data sets of SST, wind speed and wind direction on the presence of jellyfish were analysed to increase the understanding of the occurrence of these organisms on beaches. Our research indicates that the application of advanced technologies in the study of jellyfish can provide valuable insights for researching and managing these marine organisms on the beaches, thus contributing to public safety.

2 | MATERIALS AND METHODS

2.1 | Study area

The study area is located along the coast of Andalusia, the southernmost region of Spain (approximately between latitudes 36.0°N to 37.4°N and longitudes 1.5°W to 7.5°W, Figure S1). The hydrographic setting of Andalusia encompasses both the Atlantic Ocean and Mediterranean Sea coasts. This unique location exposes the region to different oceanographic phenomena, such as the influx of Atlantic waters into the Mediterranean through the Strait of Gibraltar, upwelling systems along the Atlantic coastline and the oligotrophic nature of the Mediterranean Sea (García-Lafuente et al., 2021; Peliz et al., 2002). Collectively, these factors give rise to the rich biodiversity and productivity of the Andalusian marine ecosystems (Bianchi & Morri, 2000; Perret et al., 2023; Rueda et al., 2021).

A total of 232 beaches belonging to 40 different localities were analysed in the present study (Table S1; Figure S1). Following Gutiérrez-Estrada et al. (2021) and Castro-Gutiérrez et al. (2022), the beaches were grouped in 11 different sets of beaches to make it easier to analyse the jellyfish presence according to their geographical position. Specifically, this grouping was based on the proximity of each beach to each other and to the nearest weather station, thus allowing for a more precise analysis with local meteorological data that are relevant to understanding the geographical distribution of jellyfish.

2.2 | Data collection

The information about the presence or absence of jellyfish was acquired from the Infomedusa application, which can be accessed through this Google link: https://play.google.com/store/apps/details?id=es.infomedusa&hl=es. A total of 9433 comments were collected in 2019 (from January 5 to December 8). During the conduct of this study, we had access only to data from the year 2019 as provided by the Infomedusa App. Additionally, it is noteworthy that the majority of these comments were sourced during the summer months, aligning with the peak period of beach usage. This seasonal skew in data availability is a reflection of both the operational capacity of the Infomedusa App and the natural increase in beachgoers during summer, which in turn influences the volume and frequency of jellyfish sightings reported.

Furthermore, three different environmental variables were obtained in order to characterise which are the main environmental that could drive jellyfish to the Andalusian coast: (1) sea surface temperature, SST; (2) wind direction; and (3) wind speed. SST was obtained from the Copernicus Marine Environment Monitoring Service (CMEMS, https://marine.copernicus.eu/) satellite observations with an hourly temporal resolution and 0.0625°×0.0625° spatial resolution. Daily average wind direction and speed were obtained from the weather stations closest to the location of each of the beach sets. The weather stations used in this study belong to the Spanish Meteorological Agency (AEMET, http://www.aemet.es).

2.3 | Data pre-processing and one-hot encoding

The database was carefully filtered only to include comments (only in Spanish) where the occurrence of jellyfish was either 'Presence' (1) or 'Absence' (0), thus excluding any instances with missing data (NA).

To illustrate, here are translated examples of comments used in our analysis:

- '*They have been bringing out jellyfish of considerable size all morning. There are not many but they are very big*' was classified as Presence (1) directly reflecting the presence of jellyfish.
- '*Right now, the water temperature is very good, no jellyfish*' was classified as Absence (0) indicating their absence in the comment.
- '*How is the beach today?*' was classified as NA due to the absence of information on the presence of jellyfish.

Additionally, a simple lexical pre-processing of each comment was performed to delete special characters (e.g., @, $ and &), emoticons, words with spelling errors and prepositions, as they did not contribute valuable information to the analysis.

A dictionary of unique words in alphabetical order was generated from the refined set of comments. This dictionary was further filtered to only include Spanish words with a frequency greater than 9 in the main data set because low-frequency words tend to be useless in the learning process (Cunha et al., 2019). This threshold was selected to focus on the comments' most relevant and frequently used terms.

Performing a one-hot encoding method, each comment was subsequently transformed into a binary code system using the dictionary generated (Figure 1). The size of the binary code is equal to the number of words in the dictionary. The code assigned to each comment is a sequence of 0s and 1s, where each position in the sequence corresponds to a word in the dictionary. A value of 1 indicates that the word is present in the comment, while a value of 0 indicates that it is not present.

Finally, each digit in the code was split into individual columns. Hence, through the one-hot encoding process the information of each word in each comment is represented in a compact and efficient binary representation of the information contained in each comment *j*. This entire procedure was carried out through a customised code using the R version 4.1.2 (R Core Team, 2021).

2.4 | Multi-layer perceptron

In this study, MLP classifiers developed in Python (version 3.9.16) by means of the *scikit-learn* library (Pedregosa et al., 2011) were employed to analyse the citizen comments (Figure 1). The primary purpose of employing the MLP in our study was to harness the advanced analytical capabilities of neural networks to effectively classify user comments on the Infomedusa application, distinguishing between those reporting the presence or absence of jellyfish. The choice of MLP was based on its recognised robustness and effectiveness in text classification (Goldberg, 2022), particularly for binary classification tasks (Shridhar et al., 2020). MLPs are able to capture and model complexities in large textual datasets, making them ideal for analysing and classifying user comments.

The inputs to the MLP model were user comments previously transformed into binary representations by the one-hot encoding process. Each comment was converted into a vector of 0s and 1s during data pre-processing. The output of the model was the binary

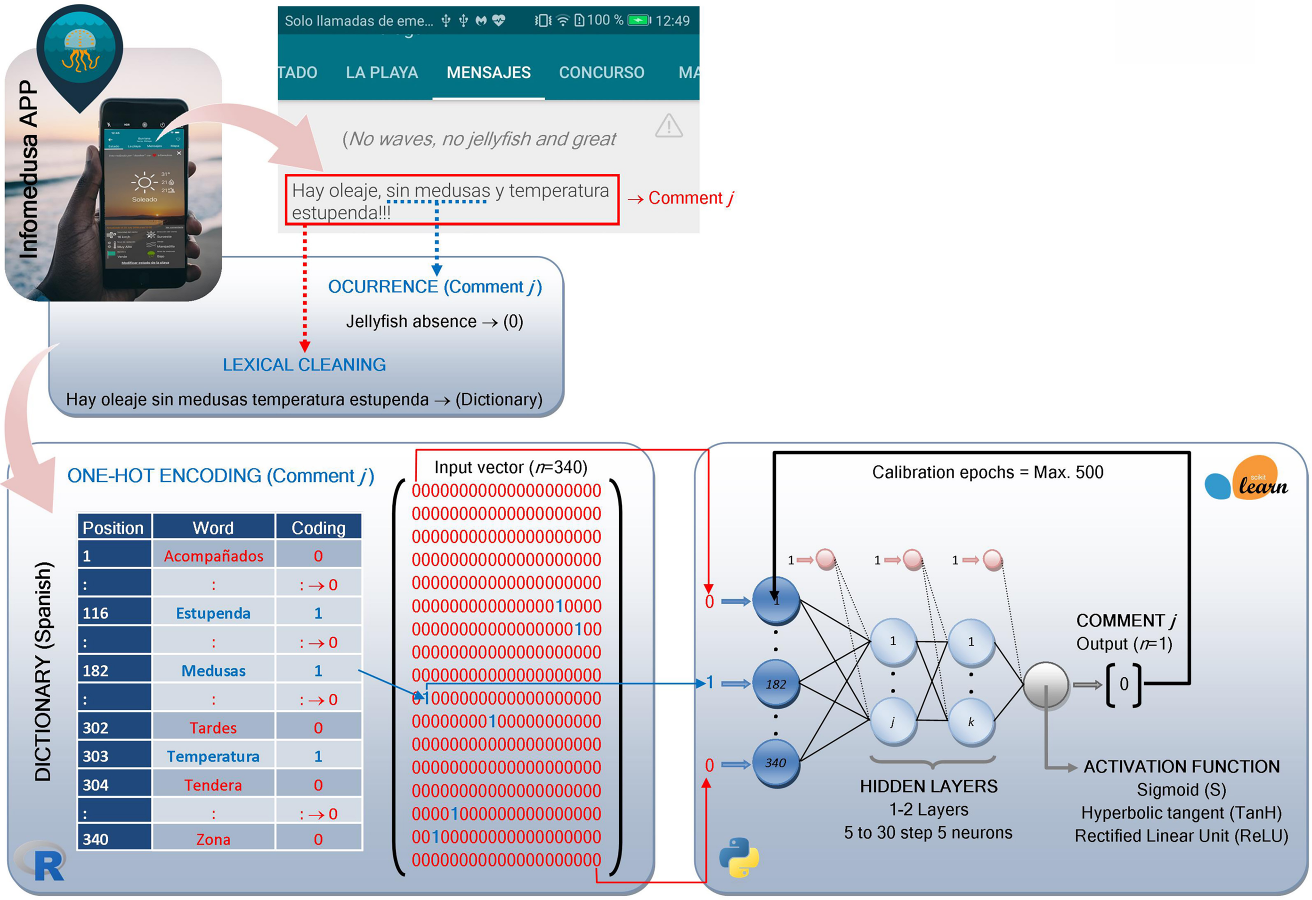

FIGURE 1 Process of one-hot encoding of the comment *j* introduced in Infomedusa APP by a user and the multi-layer perceptron (MLP) schema used in the analysis of user comments from the Infomedusa APP. The one-hot encoding procedure transforms user comments into a binary vector format, where each word is represented as either present (1) or absent (0) in the comment. This encoded data serves as the input for the MLP, which consists of 341 inputs, 1–2 hidden layers and one output neuron. The MLP tested three different activation functions: sigmoid (S), hyperbolic tangent (TanH) and rectified linear unit (ReLU). The model calibration involved a maximum of 500 epochs. The colon (':') symbols in the one-hot encoding table represent a series of omitted entries for brevity. The full table includes a comprehensive list of terms used in the comments. These omissions are made solely for visual simplification in the figure.

classification of each comment, assigning '1' to indicate the presence of jellyfish and '0' to indicate their absence.

The data set was split into training and testing sets using a 70/30 ratio, with 70% of the data allocated for training and 30% for testing. The hyperparameters were tuned using *GridSearchCV*, an exhaustive search method included in *scikit-learn* that evaluates all possible combinations of hyperparameters provided in a parameter grid. It performs cross-validation for each combination to estimate the model performance and selects the best hyperparameter neuron values based on the specified scoring metric (Pedregosa et al., 2011).

The parameter grid for the search included various combinations of single and double hidden layers, comprising all possible combinations of 5, 10, 15, 20, 25 and 30 neurons. Three activation functions (sigmoid—*S*-; hyperbolic tangent—*TanH*-; and rectified linear unit—*ReLU*-) were considered, and the value of maximum iterations or calibration epochs was 500 (Figure 1).

A repeated stratified *k*-fold cross-validation strategy with 5 folds and 30 repetitions was employed to evaluate the performance of the different combinations of hyperparameters. Stratified *k*-fold cross-validation ensures that the proportion of each class in the dataset is preserved in each fold, resulting in a more representative evaluation of the model performance (Kelleher et al., 2015; Kohavi, 1995). The repeated nature of the cross-validation process allows for multiple iterations of the stratification, further enhancing the reliability of the performance estimates. This level of repetitions implies that the chosen model is among the best 14% of the distribution of all possible models at the 99% confidence level (Iyer & Rhinehart, 1999). The models were evaluated and ranked based on their performance using the Recall scoring metric, which quantifies the proportion of true positive predictions among all actual positive instances. This metric is particularly useful for imbalanced data sets, where it is crucial to correctly identify the minority class (Branco et al., 2016). Recall (R), also known as sensitivity, is the ratio of true positive predictions to the sum of true positive and false negative (FN) predictions (Equation 1).

$$R = \frac{TP}{TP + FN}. \tag{1}$$

Other metrics are often employed for a comprehensive evaluation: precision (P) and the F1 score (Sokolova & Lapalme, 2009).

Precision (P) is defined as the ratio of true-positive predictions (TP) to the sum of true positive and false-positive (FP) predictions (Equation 2).

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{2}$$

The F1 score (F1) is the harmonic mean of precision and recall, providing a balanced measure of a model's performance (Equation 3).

$$\mathrm{F1} = \frac{2 \times (\mathrm{P} \times \mathrm{R})}{\mathrm{P} + \mathrm{R}}. \tag{3}$$

The grid search was performed using the training data, and the best model, along with its hyperparameters, was identified. The best model performance, as well as the top 10 models based on their rank, was evaluated using the test data. The models were fitted to the training data, and their predictions were compared to the true values of the test data to calculate the accuracy. Additionally, a classification report was generated for each of the top 10 models, providing further insights into their performance. Finally, the best model in the grid was evaluated, and the process was repeated if necessary, until the finding of the best MLP architecture.

2.5 | Environmental factor analysis

To characterise the environmental variables on jellyfish presence within each beach set, it was necessary to assess the representation of jellyfish data in the comments across all beach sets. This involved the elimination of sets that had minimal data representation to ensure the reliability of our analysis. Visualisation and comparison of the distribution of jellyfish presence (1) and absence (0) relative to the three environmental variables (SST, and wind direction and speed) were conducted. Each variable was examined for each data set, grouped according to the different sets of beaches. To visualise the distribution of data and explore the relationship between the presence/absence of jellyfish and other environmental factors, kernel density estimation (KDE) plots were generated using the *Seaborn* library in Python (Waskom, 2021). KDE is a non-parametric method used to estimate the probability density function of a random variable. The KDE approach functions by placing a kernel, which is a smooth, symmetric function (such as a Gaussian function), around each data point, which refers to an individual observation or measurement in the data set. These kernels are then summed to form a smooth density curve that represents the distribution of the data. In this study, a Gaussian function kernel was utilised, which by default creates a smooth bell-shaped curve around each data point. The KDE plots were generated separately for each set of beaches and each environmental variable under consideration. The distribution of these variables was compared between comments indicating the jellyfish presence (Class 1) and absence (Class 0).

The environmental variables were further statistically analysed for the magnitude of their variation as a function of their effect on the different sets of beaches. The non-parametric Kruskal–Wallis test was implemented to analyse whether there was a statistically significant difference in the variable distribution (SST, and wind direction and speed) between the different sets of beaches. For each variable, a Kruskal–Wallis test was conducted, and the resulting p-values were corrected for multiple comparisons using the Benjamini–Hochberg procedure. This approach controls the false discovery rate, enhancing the reliability of our findings when multiple tests are performed (García, 2003). Whenever the corrected p-value from the Kruskal–Wallis test was less than 0.05, a post hoc Dunn's test was conducted for pairwise comparisons between the groups. In these pairwise comparisons, we adjusted the p-values using the Holm method to account for multiple comparisons.

To characterise the impact of environmental variables on jellyfish presence (or absence) within each beach set, generalised additive mixed models (GAMM) were employed (Wood, 2017). This method models non-linear relationships between the response variable (jellyfish presence) and predictor variables while incorporating random effects to handle data dependency.

The GAMM was structured to include smoothing terms for SST, wind direction, wind speed and a random effect for beach sets. Implementation was carried out using the *mgcv* library in R (Wood, 2021). Parameters were selected using restricted maximum likelihood (REML) to optimise model fit without overfitting. The fitted model was evaluated by visual inspection of smooth effects and statistical summaries.

This comprehensive approach allowed for a thorough and systematic analysis of the impact of each predictive variable on the presence of jellyfish.

3 | RESULTS

3.1 | Data pre-processing and one-hot encoding

A total of 9433 user's comments from the Infomedusa app forum were analysed. Only 1234 (13.08%) contained explicit information about jellyfish presence or absence. Among these, 148 observations reported the presence of jellyfish and 1086 reported their absence. The remaining records did not include information about jellyfish and were thus classified as not applicable (NA). The frequency of the comments in each set of beaches is shown in Figure S2.

The dictionary of unique words comprised a total of 340 words. Subsequently, following the one-hot encoding process, each comment was represented as a data point in a 340-dimensional feature space. In addition, a further dimension was included to represent the response variable, resulting in a matrix of $n \times 340$, where n is the total number of comments with jellyfish information.

3.2 | Multi-layer perceptron

A MLP neural network algorithm was implemented for classification Infomedusa pre-processed users comments based on whether they mentioned the presence (1) or absence (0) of jellyfish.

In the process of hyperparameter optimisation, a comprehensive search over the specified parameter grid was undertaken. The grid comprised 21 different configurations of hidden layer sizes and neurons, and 3 different activation functions for those hidden layers. This leads to a total of 63 unique combinations. Coupled with the adopted repeated stratified *k*-fold cross-validation strategy, which includes 5 folds repeated 30 times, the complete procedure resulted in the fitting and evaluation of 9450 different models trained and evaluated. The top 10 models in terms of minority class (Class 1) recall performance are presented in Table 1.

The models featured various configurations of hidden layers and neurons, utilising ReLU and Sigmoid activation functions. Overall, all models performed similarly on the test data, and accuracy was high across all of them, ranging from 94.9% to 96.2%. For Class 0 (jellyfish absence), precision and recall remained consistently high, indicating reliable performance. However, there was variability in recall for Class 1 (jellyfish presence), with the fourth-ranked model achieving the highest recall of 0.85, making it the best at classifying both classes. This model, with one hidden layer of 10 neurons and ReLU activation, strikes a balance with an accuracy of 96.2%. The confusion matrix of this model showed that it successfully predicted the absence of jellyfish (class 0) 324 times, while it incorrectly predicted absence eight times when jellyfish were actually present (false negatives). Conversely, the model correctly predicted the presence of jellyfish (class 1) 33 times and incorrectly predicted presence six times when jellyfish were actually absent (false positives).

3.3 | Environmental effect on jellyfish presence

In the data collection for this study, a significant concentration of observations was noted during the summer months (June to September), totalling 77, 524, 502 and 104 observations in June, July, August and September, respectively. In contrast, the months of March, April and May presented a very limited number of data (1, 9 and 17 observations, respectively). Due to this uneven distribution, it was decided to exclude the months of March, April and May from the environmental analyses. Beach sets 6 and 7 were also excluded from the subsequent stages of the study due to their insufficient number of user's comments from those locations. The distribution of jellyfish presence (1) and absence (0) for each variable using kernel density plots are shown in Figure 2.

In this study, the influence of various environmental variables (SST, wind direction and wind speed) on the presence of jellyfish across different beach sets was assessed. The Kruskal–Wallis test was applied to each variable, followed by an adjustment of *p*-values using the Benjamini–Hochberg procedure to control for the false discovery rate due to multiple comparisons (García, 2003). Application of the Kruskal–Wallis test to each environmental variable revealed statistically significant differences between sets of beaches, indicating variations in environmental conditions that could influence the presence of jellyfish.

Even after the Benjamini–Hochberg adjustment, the results remained significant, reinforcing the impact of these environmental variables on jellyfish presence across different beach sets.

As shown in Table S2, the mean values for SST, wind direction and wind speed were compared across different beach sets using Dunn's test.

For SST, the sets 1, 2, 3, 9, 10 and 11 had similar means ($p > 0.05$), with values ranging from 23.59°C to 24.41°C. However, sets 4, 5 and

TABLE 1 Summary of the 10 best models generated by the hyperparameter tunning optimization algorithm using *GridSearchCV* with the 63 unique combinations.

Model ranking	Hidden layers	Neurons	Activation function	Accuracy (%)	Class 0 Precision	Class 0 Recall	Class 0 F1 score	Class 1 Precision	Class 1 Recall	Class 1 F1 score
#1	1	15	ReLU	96.0	0.98	0.98	0.98	0.82	0.79	0.81
#2	1	30	ReLU	96.2	0.98	0.98	0.98	0.82	0.82	0.82
#3	1	20	ReLU	95.7	0.98	0.98	0.98	0.79	0.79	0.79
#4	1	10	ReLU	96.2	0.98	0.98	0.98	0.80	0.85	0.83
#5	2	20, 15	S	94.9	0.98	0.96	0.97	0.73	0.82	0.77
#6	2	15, 10	S	94.6	0.98	0.96	0.97	0.71	0.82	0.76
#7	2	10, 10	S	94.6	0.98	0.96	0.97	0.71	0.82	0.76
#8	2	10, 5	ReLU	95.7	0.98	0.97	0.98	0.78	0.82	0.80
#9	2	25, 25	ReLU	95.4	0.97	0.97	0.97	0.79	0.77	0.78
#10	1	5	ReLU	94.9	0.98	0.97	0.97	0.74	0.79	0.77

Note: The model ranking is sorted from the best model (#1) in terms of minority class (Class 1) recall. Absence of jellyfish (Class 0); presence of jellyfish (Class 1). Rectified Linear Unit (ReLU); sigmoid (S).

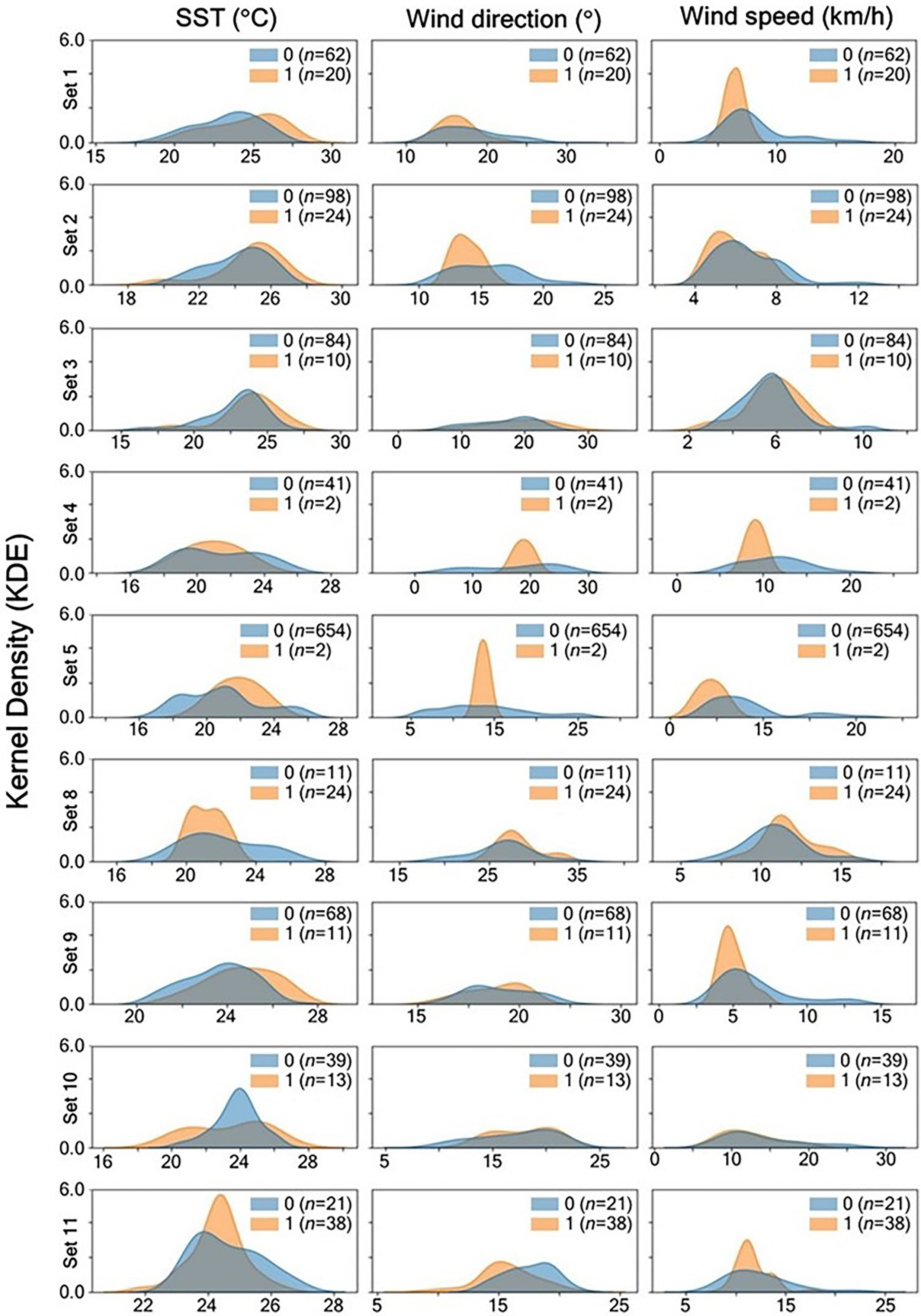

FIGURE 2 Kernel density plots for the variables sea surface temperature (SST), wind direction and wind speed across each beach set. Each row of subplots corresponds to the beach sets, while each column represents one of the variables. For each subplot, two distributions are plotted: One for the presence of jellyfish (in orange indicated as 1) and another for the absence of jellyfish (in blue indicated as 0).

8 showed significantly lower SST, with means of 21.37°C, 20.95°C and 21.47°C, respectively.

In relation to wind direction, sets 1, 3, 4, 9, 10 and 11 had similar values ($p > 0.05$), with means ranging from 16.35 to 17.98 (° ×10). In contrast, Set 2 and Set 5 had significantly lower wind directions, 15.26 and 14.02 (° ×10), respectively. Notably, Set 8 had a significantly higher mean wind direction of 37.72 (° ×10).

As for wind speed, Sets 1, 5, 10 and 11 showed similar averages ($p > 0.05$), with values ranging from 7.69 to 13.50 km/h. Sets 2 and 9 demonstrated significantly lower wind speeds, with 6.45 and 6.29 km/h, respectively. Set 3 exhibited the lowest mean wind speed of 5.72 km/h. Conversely, Set 8 had a significantly higher mean wind speed of 11.51 km/h.

The generalised additive mixed model (GAMM) revealed significant influence of environmental variables on the presence of jellyfish. The fitted model demonstrated a deviance explained of 40.3% and an adjusted *R*-squared of 0.34, indicating a moderate fit to the data. Among the smooth terms, wind speed (edf = 1.000, $p = 0.0032$) and beach sets (edf = 7.570, $p < 2e\text{-}16$) were highly significant predictors, and wind direction showed marginal significance (edf = 4.130, $p = 0.07$). SST (edf = 2.011, $p = 0.13$) was not statistically significant at the 0.05 level but showed trends towards significance.

The smooth effect plots (Figure 3) illustrate the non-linear relationships between the environmental variables and jellyfish presence. Wind speed had a notably strong effect, with higher speeds being associated with a lower probability of jellyfish presence. SST showed a slight increasing trend in jellyfish presence with higher temperatures, though not statistically significant. Wind direction exhibited a more complex relationship, with certain directions showing a higher likelihood of jellyfish presence, indicating potential onshore winds facilitating jellyfish accumulation near the shore. The random effect for beach sets captured substantial variability among the different beach locations.

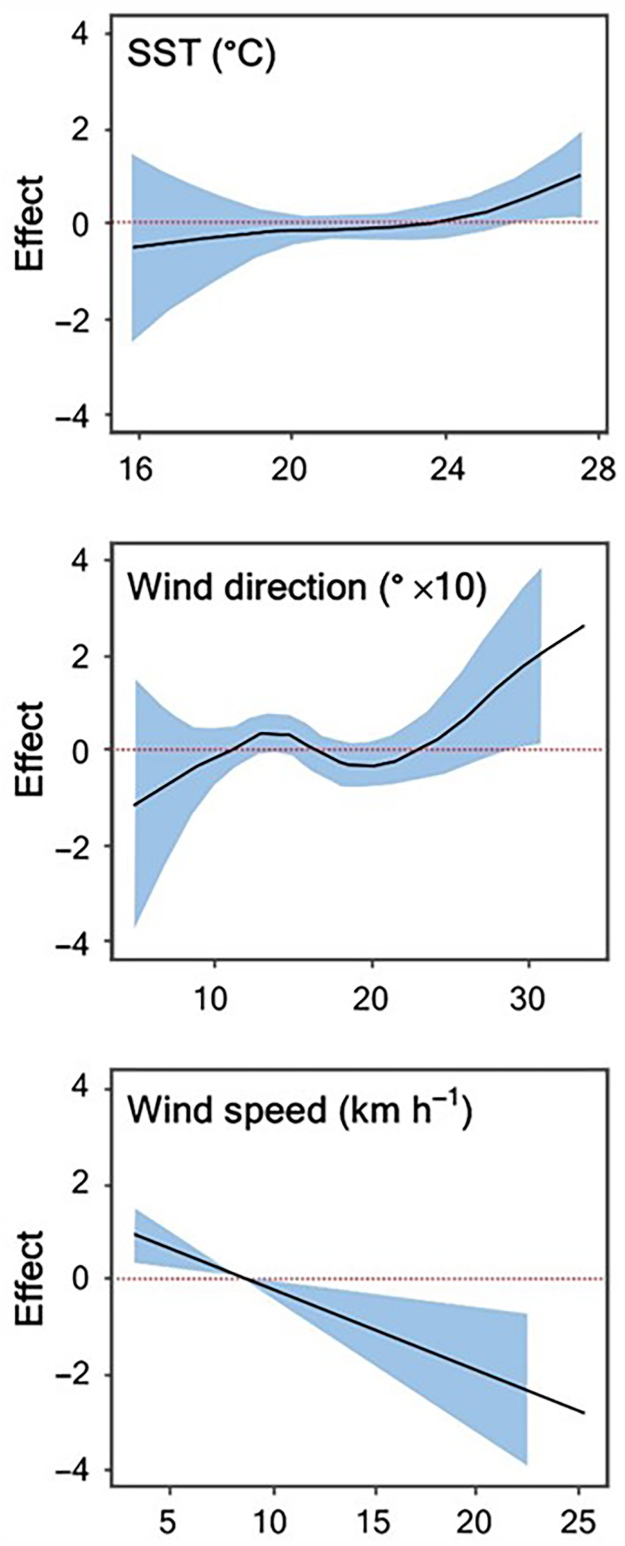

FIGURE 3 Smooth effects of environmental variables on jellyfish presence. The shaded areas represent the 95% confidence interval. SST, sea surface temperature.

4 | DISCUSSION

Our research successfully employed citizen science data from Infomedusa APP and machine learning methods to examine the presence of jellyfish on the Andalusian coast. A MLP classifier was effectively used to classify citizen comments, and environmental data were combined with user comments to characterise the jellyfish presence along the entire coast of southern Spain, thus providing valuable information for both beach users and authorities.

4.1 | Neural networks to elicit the information

MLPs model efficiently handles non-linear classification problems with robustness to noise and outliers (Haykin, 2009). These neural networks demonstrated good performance specially providing information about jellyfish absence. However, it had lower precision when informing about jellyfish presence. This might be due to the imbalance of the data set, with a lower number of comments indicating the presence of jellyfish, which is a common challenge in binary classification tasks when using real-world data (He & Garcia, 2009; Szeghalmy & Fazekas, 2023).

Analogously to our study, Gutiérrez-Estrada et al. (2021) implemented an Augmented Transition Network (ATN) to automatically process the Infomedusa comments from 2018. Through a syntax analysis procedure, this algorithm achieved an accuracy of 89.2% (which is approximately a 7% less precise than our MLP). For instance, the results from the present study imply a significant improvement concerning previous uses of this source of information, implying an increase in precision for future studies carried out with citizen science from Infomedusa. The same methodology was applied by Gutiérrez-Estrada et al. (2005) who developed and tested an expert system as a tool in the diagnosis of eel pathologies. The ATN had a success rate of 69.9% on strictly correct answers, which increased to 96.6% when approximate answers were considered correct. In comparison with the ATN success rate with strictly correct answers, our best artificial neural network classifier achieved a precision of 96.2% (Table 1, Model #4), suggesting a higher level of accuracy and reliability.

Deep learning techniques, such as convolutional neural networks have also been used in jellyfish management for jellyfish image recognition and the development of monitoring tools (Gauci et al., 2020; Han et al., 2021; Martin-Abadal et al., 2020; Ruiz-Frau et al., 2022). These deep learning techniques have proven to be efficient in identifying and classifying jellyfish species in images, allowing for faster and more accurate analysis by enhancing the quality and quantity of data collected, thus contributing to a better understanding of jellyfish populations and their distribution patterns. However, to fully exploit the potential of these technological advances, a comprehensive analysis of the environmental factors driving jellyfish occurrence is essential.

4.2 | Environmental factors

About the possible impact of environmental factors on jellyfish presence, our results are in line with previous studies suggesting that sea surface temperature (SST), and wind direction and speed significantly influence jellyfish distribution (Canepa et al., 2014; Graham et al., 2001; Messié & Chavez, 2017; Purcell, 2005; Zavodnik, 1987). SST varied significantly in certain areas of the Andalusian coast, which could be influenced by specific geographical or environmental characteristics of these areas, such as the orientation of the coastline, marine currents or local weather patterns (Canepa et al., 2014). We observed that the presence of jellyfish tends to be more frequent under elevated temperature conditions, as stated by other authors (Brotz et al., 2012; Heim-Ballew & Olsen, 2019; Purcell, 2005). This aligns with studies that have shown a direct effect of SSTs on jellyfish reproduction by favouring egg development (Avian et al., 1991). In particular, in the Mediterranean region, such increases in temperature have been notably associated with the prevalence of the mauve stinger, *Pelagia noctiluca* (Forsskål, 1775), a species known for its notoriety and frequency of outbreaks (Brotz & Pauly, 2012). The Infomedusa app has an option where users can attach an image along with their comments. Although it is rarely used, it allowed us to verify that the most frequent species was indeed *P. noctiluca*. Furthermore, research indicates that the peak spawning period for this species occurs during spring, with a peak at temperatures around 17°C (Augustine et al., 2014), suggesting a seasonal variation in jellyfish populations. This highlights the importance of considering specific jellyfish species and their unique reproductive cycles when examining the impact of SSTs on jellyfish presence. However, it is important to consider that these findings may not fully represent the patterns of jellyfish presence in other months and seasons, and the responses could vary among different jellyfish species.

Nevertheless, high-density jellyfish blooms are not only a consequence of high proliferation rates which mainly dependent on SST but also the result of changes on aggregation and distribution patterns of the species lead by wind direction changes (Bellido et al., 2020; Castro-Gutiérrez et al., 2022).

Following our results, wind conditions seem to favour the presence of jellyfish on the beaches when the wind is blowing onshore, as seen in the frequency of occurrence of jellyfish on beach sets 1, 2, 9 and 11. This pattern is consistent given the planktonic nature of these organisms that can be passively transported and accumulated near shore by ocean currents, tides and winds (Graham et al., 2001; MacAli et al., 2018; Mills, 2001; Zavodnik, 1987). It should be noted that in some areas (as the beach set 3, 4, 8 and 10) the influence of wind direction and speed on the presence of jellyfish is unclear due to the low frequency of jellyfish occurrence or the overlap of environmental conditions where jellyfish are sometimes present and sometimes not present. The findings offer valuable insights for future research aimed at predicting jellyfish occurrence based on specific environmental conditions. The complexity of temperature and wind patterns, together with the variability in jellyfish occurrence, makes it difficult to determine precisely the relationship between these factors, and the overlap indicates that, although certain conditions may be conducive to the occurrence of jellyfish, their presence is not always guaranteed. This indicates that other factors at play are also influencing the presence of jellyfish which has not been considered in this study.

Furthermore, our findings indicate that wind speed, although it may not have as obvious an effect as wind direction, could influence the distribution of jellyfish. The effect of wind speed differs with each set of beaches, which is in line with Gutiérrez-Estrada et al. (2021), who found that regional jellyfish distribution was significantly affected by local factors. This may be related to the interaction of the wind with the ocean gyres along the southern Spanish coast, where jellyfish may be aggregated inside or released towards the coast (Bellido et al., 2020; Castro-Gutiérrez et al., 2022; Franks, 1992). In our study area, wind speed tends to decrease during the summer (Vargas-Yáñez et al., 2021), which is when our data were collected. Due to the seasonal decrease in wind speed during the summer, which coincides with our data collection period, our ability to observe and fully understand its effect on jellyfish distribution may be limited in some locations. This is because if wind speeds are generally low during this period, we may not capture a full range of wind conditions, potentially underrepresenting the impacts of varying wind speeds on jellyfish distribution.

Modelling jellyfish population dynamics through the use of environmental variables can be very useful when predicting the appearance of this type of organisms, which can sometimes be dangerous to health due to their stings (Gershwin et al., 2014). The models we developed could potentially be used to forecast jellyfish presence on Andalusian beaches, thereby alerting beachgoers and potentially preventing jellyfish stings, enhancing beach user safety and enjoyment (CIESM, 2001; Lucas et al., 2014; Martin-Abadal et al., 2020).

4.3 | Limitations of our study

Despite the significant findings obtained in this study, it is important to recognise its limitations. Firstly, the APP data are predominantly collected in the summer months, which limit our ability to generalise these findings to other seasons. Secondly, the imbalanced nature of the data (fewer reports of jellyfish presence compared to absence) may have affected the accuracy of our results, especially in classifying jellyfish presence. Given the context of our study, where the presence of jellyfish can pose a risk to beachgoers, maximising the sensitivity (recall) of the model is particularly valuable. The F-beta score, which balances recall and precision, could be considered as an alternative metric for this purpose once resolved the balance of the classes.

In the same way, jellyfish absences, as reported by beachgoers through the Infomedusa App, may not always accurately represent the true absence of jellyfish. Consequently, while the reported absences provide valuable information, they come with inherent limitations that must be considered when interpreting the data.

The open nature of the app allows users to comment on various topics, many of which are unrelated to beach conditions. For this study, only comments with explicit information about jellyfish presence or absence were included, resulting in the utilisation of less than 20% of the total comments. Our team is currently collaborating with Aula del Mar de Málaga to promote the app and implement changes that encourage users to provide explicit information about jellyfish presence. These efforts are expected to significantly increase the quality and quantity of data in future studies.

4.4 | Citizen science for monitoring jellyfish

The importance of continued local monitoring of jellyfish cannot be overstated. To enhance our knowledge of jellyfish distribution patterns, it is essential to engage in continuous monitoring at the local level. This monitoring allows us to capture the various fluctuations and improve our understanding of these patterns.

According to Danielsen et al. (2005), local monitoring is not only cost-effective but also yields results that are locally relevant and can be just as reliable as those obtained through professional monitoring. This is particularly important given the localised nature of environmental conditions affecting jellyfish populations. When properly designed and implemented, local monitoring schemes can yield valuable data to inform timely management decisions and actions (Fraisl et al., 2022). Moreover, local monitoring can reinforce existing community-based resource management systems and lead to a shift in local attitudes towards more environmentally sustainable resource management. This aligns with the findings of other studies that have underscored the potential of community-based monitoring in biodiversity conservation (Berkes et al., 2007; Conrad & Hilchey, 2011).

In the case of Infomedusa app, although the app is not explicitly designed for scientific purposes, the data it gathers can still be valuable for scientific analysis when properly validated and processed. This approach leverages daily observations from citizens to enhance environmental monitoring and management. The Infomedusa app collects user comments about beach conditions and jellyfish presence. While the app has administrators to manage inappropriate use and inform users about updates, the citizens contributing data are unaware that their inputs are used for scientific research. Throughout the study, the confidentiality of user data was maintained at all times, ensuring that individual identities and personal information were protected. This presents both challenges and opportunities for improving the accuracy and utility of the collected data. Our findings have a potential real-world impact, particularly within the tourism industry, a vital sector of Spanish economy (Moreno-Luna et al., 2021). According to the Office of the National Statistics Institute (INE), Andalusia received 10 million international tourists in 2022 with an average daily expenditure of 133.33 euros. Beach tourism in Andalusia and other Spanish coastal regions can be significantly influenced by the presence of jellyfish, as their stings can lead to both minor discomfort and serious health issues, posing a deterrent to tourists (Bordehore et al., 2016; Cantarero Prados & Moreno Portillo, 2021; Crowley-Cyr et al., 2022; Rubio Gómez & Gutiérrez-Hernández, 2020; Ruiz-Frau, 2022). Consequently, predicting jellyfish blooms and issuing early warnings based on our results could substantially improve the safety and satisfaction of beachgoers, fostering more positive beach tourism experiences (De Donno et al., 2014).

Similarly, various projects employing diverse approaches have been developed to mitigate the impact of jellyfish blooms in coastal areas using citizen science. For instance, the MED-JELLYRISK project focused on creating a forecasting platform for jellyfish blooms in the western and central Mediterranean Basin (Canepa et al., 2016). It utilised an integrated citizen science approach and species distribution models to predict jellyfish occurrences, while offering up-to-date information to the public through the mobile IMedjelly APP. Besides that, the worldwide citizen science initiative Jellywatch relies on user-reported jellyfish sightings to track and enhance the understanding of jellyfish populations and distribution patterns. Meanwhile, JellyMonitor aims to establish a portable imaging platform capable of detecting jellyfish near water intakes, providing an early warning system for jellyfish ingress (French et al., 2018).

Similarly, Marambio et al. (2021) used citizen science to investigate the dynamics of jellyfish blooms in the Mediterranean Sea relying on volunteer observers to report sightings from four different countries. In addition, the MedusApp enables users to geolocate jellyfish, identify species and their abundance, and provides information on species identification and the toxicity of jellyfish stings, along with first aid recommendations. The app also allows users to send images of stings for evaluation by a specialised medical service for scientific purposes (Blasco Talavan et al., 2016).

In this context, citizen science has proven to be instrumental in our understanding of jellyfish ecology, allowing for the collection of extensive data across broad geographical ranges (Marambio et al., 2021). This inclusive approach not only democratises scientific research by engaging the public in data collection but also circumvents the logistical and financial limitations typically faced by research teams. The substantial data volume provided by citizen scientists enhances our knowledge of jellyfish distribution and their temporal dynamics.

Moreover, these initiatives significantly boost public awareness and involvement with biodiversity, fostering a well-informed community poised to contribute positively to conservation efforts. The engagement with citizen science leads to more informed and responsible behaviour regarding environmental stewardship, crucial for the protection and conservation of marine species.

The long-term data garnered through such initiatives are indispensable for monitoring environmental changes and their impact on jellyfish populations. The insights gained are vital for shaping effective conservation strategies and informing policy decisions, underscoring the irreplaceable value of citizen science in ecological research.

This approach has the potential to be cost-effective and accurate, as it relies on observations from many users who are directly experiencing the coastal environment. Nevertheless, it is important to consider observer-based biases in such citizen science biodiversity monitoring projects since users' decisions about what and when to report can influence observed data patterns (Arazy & Malkinson, 2021).

5 | CONCLUSIONS

The results of this study underscore the efficacy of using citizen science data and machine learning techniques to monitor jellyfish presence along coastal areas. The MLP neural network demonstrated high accuracy in classifying jellyfish presence and absence, despite challenges posed by data imbalance. Our findings highlight the significant impact of environmental variables, particularly SST and wind conditions, on jellyfish distribution.

To enhance beach management based on these insights, it is essential to continuously monitor environmental variables such as SST, wind direction and wind speed. Developing and implementing seasonal preparedness plans based on these environmental factors can help mitigate risks. For instance, heightened monitoring and readiness for jellyfish blooms during periods of elevated SSTs and specific wind conditions can be particularly effective.

Integrating real-time SST and wind data into the Infomedusa app and other monitoring tools will provide up-to-date information, enabling more accurate and timely predictions of jellyfish presence. This integration aids in proactive beach management. Developing public information systems that notify beachgoers about current environmental conditions and the likelihood of jellyfish presence is also crucial. Real-time updates through mobile apps and beachside displays can inform the public and enhance safety.

Investing in research to explore the influence of other potential environmental factors such as nutrient levels, ocean currents and anthropogenic influences on jellyfish populations is recommended. Understanding these factors can further refine predictive models and management strategies. Employing adaptive management strategies that are responsive to real-time environmental data and predictive model outputs allows for dynamic adjustments to beach management practices, reducing the impact of jellyfish blooms on beachgoers.

By adopting these recommendations, beach management authorities can improve the safety and enjoyment of beachgoers, enhance the accuracy of jellyfish monitoring and contribute to the broader understanding of marine ecosystems. This targeted approach, based on environmental analysis, can lead to more effective and efficient management of coastal areas affected by jellyfish blooms.

AUTHOR CONTRIBUTIONS

Jairo Castro-Gutiérrez: Data curation; formal analysis; investigation; software; validation; visualisation; writing—original draft. Juan Carlos Gutiérrez-Estrada: Conceptualization; investigation; resources; supervision; validation; writing—review and editing. José Carlos Báez: Conceptualization; resources; supervision; writing—review and editing.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the team at the Aula del Mar in Malaga for their continuous support and assistance throughout this research and to the technical team behind the Infomedusa APP. We also extend our thanks to all the users who have contributed their feedback on Infomedusa. The authors would like to express our gratitude to the anonymous reviewers for their useful comments, which helped us to improve the paper's quality. Funding for open access charge: Universidad de Huelva / CBUA.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

Data are available via the Zenodo Repository https://doi.org/10.5281/zenodo.12732089 (Souviron-Priego et al., 2024).

ORCID

J. Castro-Gutiérrez https://orcid.org/0000-0002-4466-3645

J. C. Gutiérrez-Estrada https://orcid.org/0000-0002-2647-6002

J. C. Báez https://orcid.org/0000-0003-2049-0409

REFERENCES

Arazy, O., & Malkinson, D. (2021). A framework of observer-based biases in citizen science biodiversity monitoring: Semi-structuring unstructured biodiversity monitoring protocols. *Frontiers in Ecology and Evolution*, *9*, 1–13. https://doi.org/10.3389/fevo.2021.693602

Augustine, S., Rosa, S., Kooijman, S. A., Carlotti, F., & Poggiale, J. C. (2014). Modeling the eco-physiology of the purple mauve stinger, *Pelagia noctiluca* using Dynamic Energy Budget theory. *Journal of Sea Research*, *94*, 52–64.

Avian, M., Rottini Sandrini, L., & Stravisi, F. (1991). The effect of seawater temperature on the swimming activity of *Pelagia noctiluca* (Forsskål). *The Italian Journal of Zoology*, *58*(2), 135–143.

Bellido, J. J., Baez, J. C., Souviron-Priego, L., Ferri-Yáñez, F., Salas, C., López, J. A., & Real, R. (2020). Atmospheric indices allow anticipating the incidence of jellyfish coastal swarms. *Mediterranean Marine Science*, *21*(2), 289–297.

Berkes, F., Berkes, M. K., & Fast, H. (2007). Collaborative integrated management in Canada's north: The role of local and traditional knowledge and community-based monitoring. *Coastal Management*, *35*(1), 143–162.

Bianchi, C. N., & Morri, C. (2000). Marine biodiversity of the Mediterranean Sea: Situation, problems and prospects for future research. *Marine Pollution Bulletin*, *40*(5), 367–376.

Blasco Talavan, E., Palacios Sáez, R., Fondría, E., & Bordehore, C. (2016). *MEDUSAPP: Mobile citizen science App for quantitative geolocation of jellyfish sightings and stings registration for educational, scientific and medical purposes*.

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen science: A developing tool for expanding science knowledge and scientific literacy. *BioScience*, *59*(11), 977–984.

Bordehore, C., Alonso, C., Sánchez-Fernández, L., Canepa, A., Acevedo, M., Nogué, S., & Fuentes, V. L. (2016). Lifeguard assistance at Spanish Mediterranean beaches: Jellyfish prevail and proposals for improving risk management. *Ocean and Coastal Management*, *131*, 45–52. https://doi.org/10.1016/j.ocecoaman.2016.08.008

Branco, P., Torgo, L., & Ribeiro, R. P. (2016). A survey of predictive modeling on imbalanced domains. *ACM Computing Surveys*, *49*(2), 1–50. https://doi.org/10.1145/2907070

Brotz, L., Cheung, W. W. L., Kleisner, K., Pakhomov, E., & Pauly, D. (2012). Increasing jellyfish populations: Trends in large marine ecosystems. *Hydrobiologia*, *690*(1), 3–20.

Brotz, L., & Pauly, D. (2012). Jellyfish populations in the Mediterranean Sea. *Acta Adriatica*, *53*(2), 213–231.

Canepa, A., Fuentes, M., Marambio, M., Lopez, L., Deidun, A., Yahia, K. D. O., Yahia, M. N. D., Piraino, S., & Fuentes, V. (2016). Forecasting jellyfish blooms in the Mediterranean Sea: The med-JellyRisk project. In *5th International Jellyfish Bloom Symposium Barcelona Spain*. Institut de Ciències del Mar—CSIC.

Canepa, A., Fuentes, V., Sabatés, A., Piraino, S., Boero, F., & Gili, J. M. (2014). *Pelagia noctiluca* in the Mediterranean Sea. In *Jellyfish blooms* (pp. 237–266). Springer.

Cantarero Prados, F. J., & Moreno Portillo, A. M. (2021). Jellyfish swarms and degree of exposure and vulnerability of recreational and tourist activities on beaches. Methodological approach to their assessment in the Lagos-Ferrara sector (Málaga, Spain). In *Advances in tourism, technology and systems: Selected papers from ICOTTS20* (Vol. *1*, pp. 331–340). Springer.

Castro-Gutiérrez, J., Gutiérrez-Estrada, J. C., Aroba, J., Pulido-Calvo, I., Peregrín, A., Báez, J. C., Bellido, J. J., & Souviron-Priego, L. (2022). Estimation of jellyfish abundance in the south-eastern Spanish coastline by using an explainable artificial intelligence model based on fuzzy logic. *Estuarine, Coastal and Shelf Science*, *277*, 108062.

CIESM. (2001). Gelatinous zooplankton outbreaks: Theory and practice. *CIESM Workshop Series*, no. 14.

Conrad, C. C., & Hilchey, K. G. (2011). A review of citizen science and community-based environmental monitoring: Issues and opportunities. *Environmental Monitoring and Assessment*, *176*, 273–291.

Crowley-Cyr, L., Gershwin, L., Bremser, K., Abraham, V., Moreno Martin, M. J., Carreño, M., & Wüst, K. (2022). Jellyfish risk communications: The effect on risk perception, travel intentions and behaviour, and beach tourism destinations. *Journal of Hospitality and Tourism Management*, *51*, 196–206. https://doi.org/10.1016/j.jhtm.2022.03.002

Cunha, A. A. L., Costa, M. C., & Pacheco, M. A. C. (2019). Sentiment analysis of youtube video comments using deep neural networks. In *Artificial Intelligence and Soft Computing: 18th International Conference, ICAISC 2019, Zakopane, Poland, June 16–20, 2019, Proceedings, Part I* (Vol. *18*, pp. 561–570). Springer International Publishing.

Danielsen, F., Burgess, N. D., & Balmford, A. (2005). Monitoring matters: Examining the potential of locally-based approaches. *Biodiversity and Conservation*, *14*, 2507–2542.

De Donno, A., Idolo, A., Bagordo, F., Grassi, T., Leomanni, A., Serio, F., Guido, M., Canitano, M., Zampardi, S., Boero, F., & Piraino, S. (2014). Impact of stinging jellyfish proliferations along south Italian coasts: Human health hazards, treatment and social costs. *International Journal of Environmental Research and Public Health*, *11*(3), 2488–2503.

Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. (2010). Citizen science as an ecological research tool: Challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, *41*, 149–172.

Do, A. N. T., & Tran, H. D. (2022). Potential application of artificial neural networks for analyzing the occurrences of fish larvae and juveniles in an estuary in northern Vietnam. *Aquatic Ecology*, *57*, 813–831.

Fraisl, D., Hager, G., Bedessem, B., Gold, M., Hsing, P. Y., Danielsen, F., Hitchcock, C. B., Hulbert, J. M., Piera, J., Spiers, H., Thiel, M., & Haklay, M. (2022). Citizen science in environmental and ecological sciences. *Nature Reviews Methods Primers*, *2*(1), 64.

Franks, P. J. (1992). Sink or swim: Accumulation of biomass at fronts. *Marine Ecology Progress Series*, *82*(1), 1–12.

French, G., Mackiewicz, M., Fisher, M., Challiss, M., Knight, P., Robinson, B., & Bloomfield, A. (2018). JellyMonitor: Automated detection of jellyfish in sonar images using neural networks. In *2018 14th IEEE International Conference on Signal Processing (ICSP)* (pp. 406–412). IEEE.

Fritz, S., See, L., Carlson, T., Haklay, M., Oliver, J., Fraisl, D., Mondardini, R., Brocklehurst, M., Shanley, L., Schade, S., Wehn, U., Abrate, T., Anstee, J., Arnold, S., Billot, M., Campbell, J., Parker, A., Gold, M., Hager, G., … West, S. (2019). Citizen science and the United Nations sustainable development goals. *Nature Sustainability*, *2*(10), 922–930.

Galil, B. S. (2008). The price of change: The economic impacts of alien species and jellyfish outbreaks in the Mediterranean Sea. In *Economic valuation of natural coastal and marine ecosystems*. CIESM Workshop Monographs, No. 37. CIESM.

García, L. V. (2003). Controlling the false discovery rate in ecological research. *Trends in Ecology & Evolution*, *18*(11), 553–554.

García-Lafuente, J., Sánchez-Garrido, J. C., Garcia, A., Hidalgo, M., Sammartino, S., & Laíz, R. (2021). Biophysical processes determining the connectivity of the Alboran Sea fish populations. In J. C. Báez, J. T. Vázquez, J. A. Camiñas, & M. M. Idrissi (Eds.), *Alboran sea—Ecosystem and marine resources* (pp. 459–487). Springer Nature.

Gauci, A., Deidun, A., & Abela, J. (2020). Automating jellyfish species recognition through faster region- based convolution neural networks. *Applied Sciences*, *10*, 8257.
Gershwin, L., Condie, S. A., Mansbridge, J. V., & Richardson, A. J. (2014). Dangerous jellyfish blooms are predictable. *Journal of the Royal Society Interface*, *11*(96), 20131168.
Ghermandi, A., Galil, B., Gowdy, J., & Nunes, P. A. (2015). Jellyfish outbreak impacts on recreation in the Mediterranean Sea: Welfare estimates from a socioeconomic pilot survey in Israel. *Ecosystem Services*, *11*, 140–147.
Goldberg, Y. (2022). *Neural network methods for natural language processing*. Springer Nature.
Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, *69*(4), 211–221.
Graham, W. M., Gelcich, S., Robinson, K. L., Duarte, C. M., Brotz, L., Purcell, J. E., Madin, L. P., Mianzan, H., Sutherland, K. R., Uye, S., Pitt, K. A., Lucas, C. H., Bøgeberg, M., Brodeur, R. D., & Condon, R. H. (2014). Linking human well-being and jellyfish: Ecosystem services, impacts, and societal responses. *Frontiers in Ecology and the Environment*, *12*(9), 515–523.
Graham, W. M., Pagès, F., & Hamner, W. M. (2001). A physical context for gelatinous zooplankton aggregations: A review. *Hydrobiologia*, *451*(1–3), 199–212.
Guerrero, E., Gili, J. M., Grinyó, J., Raya, V., & Sabatés, A. (2018). Long-term changes in the planktonic cnidarian community in a mesoscale area of the NW Mediterranean. *PLoS One*, *13*(5), e0196431.
Gutiérrez-Estrada, J. C., Pulido-Calvo, I., Peregrín, A., García-Gálvez, A., Báez, J. C., Bellido, J. J., Souviron-Priego, L., Sánchez-Laulhé, J. M., & López, J. A. (2021). Integrating local environmental data and information from non-driven citizen science to estimate jellyfish abundance in Costa del Sol (southern Spain). *Estuarine, Coastal and Shelf Science*, *249*, 107112.
Gutiérrez-Estrada, J. C., Sanz, E. D. P., López-Luque, R., & Pulido-Calvo, I. (2005). SEDPA, an expert system for disease diagnosis in eel rearing systems. *Aquacultural Engineering*, *33*(2), 110–125.
Han, Y., Chang, Q., Ding, S., Gao, M., Zhang, B., & Li, S. (2021). Research on multiple jellyfish classification and detection based on deep learning. *Multimedia Tools and Applications*, *81*, 19429–19444.
Haykin, S. S. (2009). *Neural networks and learning machines* (Vol. *3*). Pearson Upper.
He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284.
Heim-Ballew, H., & Olsen, Z. (2019). Salinity and temperature influence on scyphozoan jellyfish abundance in the Western Gulf of Mexico. *Hydrobiologia*, *827*, 247–262.
Hovy, D. (2022). *Text analysis in python for social scientists: Prediction and classification*. Cambridge University Press.
Iyer, M. S., & Rhinehart, R. R. (1999). A method to determine the required number of neural-network training repetitions. *IEEE Transactions on Neural Networks*, *10*(2), 427–432.
Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2015). *Fundamentals of machine learning for predictive data analytics: Algorithms, worked examples, and case studies*. MIT Press.
Kim, H., Koo, J., Kim, D., Jung, S., Shin, J. U., Lee, S., & Myung, H. (2016). Image-based monitoring of jellyfish using deep learning architecture. *IEEE Sensors Journal*, *16*(8), 2215–2216.
Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence—Volume 2 (IJCAI'95)* (pp. 1137–1143). Morgan Kaufmann Publishers Inc.
Lucas, C. H., Gelcich, S., & Uye, S. I. (2014). *Living with jellyfish: Management and adaptation strategies*. Springer Netherlands.
MacAli, A., Semenov, A., Venuti, V., Crupi, V., D'Amico, F., Rossi, B., Corsi, I., & Bergami, E. (2018). Episodic records of jellyfish ingestion of plastic items reveal a novel pathway for trophic transference of marine litter. *Scientific Reports*, *8*(1), 6105.
Marambio, M., Canepa, A., Lòpez, L., Gauci, A. A., Gueroun, S. K., Zampardi, S., Boero, F., Yahia, O. K. D., Yahia, M. N. D., Fuentes, V., Piraino, S., & Deidun, A. (2021). Unfolding jellyfish bloom dynamics along the Mediterranean basin by transnational citizen science initiatives. *Diversity*, *13*(6), 274.
Mariottini, G. L., Giacco, E., & Pane, L. (2008). The mauve stinger *Pelagia noctiluca* (Forsskål, 1775). Distribution, ecology, toxicity and epidemiology of stings. A review. *Marine Drugs*, *6*, 496–513.
Mariottini, G. L., & Pane, L. (2010). Mediterranean jellyfish venoms: A review on Scyphomedusae. *Marine Drugs*, *8*(4), 1122–1152.
Martin-Abadal, M., Ruiz-Frau, A., Hinz, H., & Gonzalez-Cid, Y. (2020). Jellytoring: Real-time jellyfish monitoring based on deep learning object detection. *Sensors*, *20*, 1708.
Messié, M., & Chavez, F. P. (2017). Nutrient supply, surface currents, and plankton dynamics predict zooplankton hotspots in coastal upwelling systems. *Geophysical Research Letters*, *44*(17), 8979–8986.
Mills, C. E. (2001). Jellyfish blooms: Are populations increasing globally in response to changing ocean conditions? *Hydrobiologia*, *451*(1–3), 55–68.
Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep learning-based text classification: A comprehensive review. *ACM Computing Surveys (CSUR)*, *54*(3), 1–40.
Moreno-Luna, L., Robina-Ramírez, R., Sánchez, M. S. O., & Castro-Serrano, J. (2021). Tourism and sustainability in times of COVID-19: The case of Spain. *International Journal of Environmental Research and Public Health*, *18*(4), 1859.
Mukhamediev, R. I., Popova, Y., Kuchin, Y., Zaitseva, E., Kalimoldayev, A., Symagulov, A., Levashenko, V., Abdoldina, F., Gopejenko, V., Yakunin, K., Muhamedijeva, E., & Yelis, M. (2022). Review of artificial intelligence and machine learning technologies: Classification, restrictions, opportunities and challenges. *Mathematics*, *10*(15), 2552.
Pauly, D., Graham, W. M., Libralato, S., Morissette, L., & Palomares, M. L. D. (2009). Jellyfish in ecosystems, online databases, and ecosystem models. *Hydrobiologia*, *616*, 67–85.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Duborg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.
Peliz, Á., Rosa, T. L., Santos, A. M. P., & Pissarra, J. L. (2002). Fronts, jets, and counter-flows in the Western Iberian upwelling system. *Journal of Marine Systems*, *35*(1–2), 61–77.
Perret, J., Cobelli, O., Taudière, A., Andrieu, J., Aumeeruddy-Thomas, Y., Souissi, J. B., Bernard, G., Casazza, G., Crochet, P.-A., Decaëns, T., Denis, F., Geniez, P., Loizides, M., Médail, F., Pasqualini, V., Speciale, C., Battesti, V., Chevaldonné, P., Lejeusne, C., & Richard, F. (2023). Time to refine the geography of biodiversity hotspots by integrating molecular data: The Mediterranean Basin as a case study. *Biological Conservation*, *284*, 110162.
Purcell, J. E. (2005). Climate effects on formation of jellyfish and ctenophore blooms: A review. *Journal of the Marine Biological Association of the United Kingdom*, *85*(3), 461–476.
Purcell, J. E., Uye, S. I., & Lo, W. T. (2007). Anthropogenic causes of jellyfish blooms and their direct consequences for humans: A review. *Marine Ecology Progress Series*, *350*, 153–174.
R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/
Richardson, A. J., Bakun, A., Hays, G. C., & Gibbons, M. J. (2009). The jellyfish joyride: Causes, consequences and management responses to a more gelatinous future. *Trends in Ecology & Evolution*, *24*(6), 312–322.
Rubio Gómez, A., & Gutiérrez-Hernández, O. (2020). Impact of jellyfish and other gelatinous organisms on the Andalusian coast.

Implications for sun and beach tourism. *Estudios Geográficos*, *81*(288), e038. https://doi.org/10.3989/estgeogr.202053.033

Rueda, J. L., Gofas, S., Aguilar, R., de la Torriente, A., García Raso, J. E., Lo Iacono, C., Luque, A. A., Marina, P., Mateo-Ramírez, A., Moya-Urbano, E., Moreno, D., Navarro-Barranco, C., Salas, C., Sánchez-Tocino, L., Templado, J., & Urra, J. (2021). Benthic fauna of littoral and deep-sea habitats of the Alboran Sea: A hotspot of biodiversity. In *Alboran sea-ecosystems and marine resources* (pp. 285–358). Springer International Publishing.

Ruiz-Frau, A. (2022). Impacts of jellyfish presence on tourists' holiday destination choices and their willingness to pay for mitigation measures. *Journal of Environmental Planning and Management*, *66*, 2107–2125. https://doi.org/10.1080/09640568.2022.2061926

Ruiz-Frau, A., Martin-Abadal, M., Jennings, C. L., Gonzalez- Cid, Y., & Hinz, H. (2022). The potential of Jellytoring 2.0 smart tool as a global jellyfish monitoring platform. *Ecology and Evolution*, *12*, e9472.

Shridhar, K., Jain, H., Agarwal, A., & Kleyko, D. (2020). End to end binarized neural networks for text classification. In *Proceedings of SustaiNLP: Workshop on simple and efficient natural language processing* (pp. 29–34). Association for Computational Linguistics.

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution*, *24*(9), 467–471.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*(4), 427–437.

Souviron-Priego, L., Bellido-López, J. J., López-Jaime, J. A., Castro-Gutiérrez, J., Gutiérrez-Estrada, J. C., & Báez-Barrionuevo, J. C. (2024). Data from: Using artificial neural networks and citizen science data to assess jellyfish presence along coastal areas. *Zenodo*, https://doi.org/10.5281/zenodo.12732089

Szeghalmy, S., & Fazekas, A. (2023). A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning. *Sensors*, *23*(4), 2333.

Tang, D., Qin, B., & Liu, T. (2015). Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1422–1432). Association for Computational Linguistics.

Vargas-Yáñez, M., García-Martínez, M. C., Moya, F., Balbín, R., & López-Jurado, J. L. (2021). The oceanographic and climatic context. In J. C. Báez, J. T. Vázquez, J. A. Camiñas, & M. M. Idrissi (Eds.), (pp. 85–109). Alboran sea—Ecosystem and marine resources.

Waskom, M. L. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021.

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). CRC Press.

Wood, S. N. (2021). *mgcv: Mixed GAM computation vehicle with automatic smoothness estimation*. R package version 1.8-36. https://cran.r-project.org/package=mgcv

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, *13*(3), 55–75.

Zavodnik, D. (1987). Spatial aggregations of the swarming jellyfish *Pelagia noctiluca* (Scyphozoa). *Marine Biology*, *94*, 265–269.

Zhang, G. P. (2000). Neural networks for classification: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *30*(4), 451–462.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Table S1: Summary of the localities and number of beaches covered by each beach set.

Table S2: Mean values for the environmental variables sea surface temperature (SST, °C), wind direction (° ×10), and wind speed (km/h) in the different beach sets.

Figure S1: Study area.

Figure S2: Frequency of user comments from the Infomedusa APP explicitly mentioning the jellyfish presence (orange bar, labelled as 1) or absence (blue bar, labelled as 0) for each beach set.

How to cite this article: Castro-Gutiérrez, J., Gutiérrez-Estrada, J. C., & Báez, J. C. (2024). Using artificial neural networks and citizen science data to assess jellyfish presence along coastal areas. *Journal of Applied Ecology*, *61*, 2244–2257. https://doi.org/10.1111/1365-2664.14734