Expert Systems With Applications 200 (2022) 117091

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Is the vessel fishing? Discrimination of fishing activity with low-cost intelligent mobile devices through traditional and heuristic approaches

M.M. Galotto-Tébar [a,*], A. Pomares-Padilla [a], I.A. Czerwinski [b], J.C. Gutiérrez-Estrada [c]

[a] Dpto. de Ingeniería de Computadores, Universidad Miguel Hernández. Campus de Elche, Avenida de la Universidad s/n, 03202 Elche, Alicante, Spain
[b] Instituto Español de Oceanografía, Centro Oceanográfico de Cádiz, Puerto Pesquero, Muelle de Levante, s/n, 11006 Cádiz, Spain
[c] Dpto. de Ciencias Agroforestales, Escuela Técnica Superior de Ingeniería, Campus de El Carmen, Universidad de Huelva, 21007 Huelva, Spain

ARTICLE INFO

Keywords:
Vessel monitoring system
Mobile device
Sensor
Trawl fishing
Ship's behavior
Machine learning

ABSTRACT

Knowing the activity of fishing vessels accurately and in real time means a leap in quality in the management of fishing activity. This paper presents the development of a new fishing activity monitoring integral system (FAMIS) that can complement and overcome the limitations of current fishing vessel monitoring systems (VMS). FAMIS is developed on the basis of a low-cost mobile device with GPS sensors, accelerometer, gyroscope and magnetic field and integrates different statistical methods (discriminant functions) and heuristics (artificial neural networks and vectorial support machines) as techniques to classify the information recorded by the sensors of a mobile device during fishing activity. The results obtained with FAMIS indicate that, in general, heuristics have a high degree of discrimination of each of the phases of fishing operation and that, in particular, multilayer perceptrons (MLPs) are capable of correctly identifying 96.3% of towing phases using only GPS and gyro sensors.

1. Introduction

For the exploitation of living aquatic resources to be sustainable, there is a need to limit the impact of fishing fleets on marine ecosystems. To assess the impact of a fishing vessel during its activity, we use the concept of fishing effort defined as fishing capacity multiplied by the duration of fishing activity (EC, 2007). Fishing capacity can be quantified from the vessel's technical characteristics and the fishing gear used. On the other hand, assessment of the duration of fishing activity requires knowledge of the time during which the fishing capacity of a vessel is effectively operated, a parameter which is highly variable and specific to each type of fisheries and fishing operation. Therefore, to reliably identify the duration of activity of a vessel, we need to have information, beyond just the time spent at sea, that allow us to track the activity of a fishing vessel.

In 1995, the European Commission undertook a pilot project to assess the functionality and costs of various satellite systems that might make it possible to continuously track the position and activity of fishing vessels. The result of this pilot was the implementation, in 2000, of a Vessel Monitoring System (VMS) with the goal of improving the management and monitoring of fishing activity. Specifically, all fishing vessels $\geq$ 24 m long were required to use the VMS, this requirement being extended to vessels $\geq$ 15 m in 2003 and those $\geq$ 12 m in 2012 (EC, 2009). Data provided by the VMS (the vessel's geographical position, course and speed) are transmitted every hour to the Spanish Fisheries Monitoring Centre where all this information is stored and shared internationally in real time with parties in whose waters European vessels operate.

Based on the information provided by the VMS, several authors have proposed actions and protocols to strengthen the monitoring and planning of fishing activity (Gerritsen and Lordan, 2011; Gerritsen et al., 2012; Jennings and Lee, 2012), a large proportion of published studies having focused on estimating fishing effort (Rijnsdorp, 1998; Deng et al., 2005; Murawski et al., 2005; Salthaug & Johannessen, 2006; Walter et all., 2007; Fock, 2008). Nonetheless, despite the large amount of information provided by the VMS and the indisputable advantages of its analysis for monitoring the fishing fleet, the long sampling period, the limited information provided in each sample and the exclusion of the coastal fleet limit the reliability and accuracy of the estimations that can be obtained from this information (Russo et al, 2016).

To overcome these limitations, there is a need for new devices and/or monitoring procedures. For example, the Location and Track System for Andalusian Fishing Vessels (SLSEPA), also known as the "green box" system (Junta de Andalucía, 2004) developed by the regional

* Corresponding author.

https://doi.org/10.1016/j.eswa.2022.117091
Received 12 November 2020; Received in revised form 20 September 2021; Accepted 28 March 2022
Available online 1 April 2022

government of Andalusia (Spain) uses the GPRS mobile network to transmit data on location, course and speed of the artisanal fisheries to a control centre at 3-min intervals. The higher sampling rate improves the accuracy of the estimation of fishing effort and spatial distribution of the fleet and the low data transmission costs make the system feasible and operative for vessels < 12 m long of the artesanal fleets (Cojan and Burgos, 2015). Despite SLSEPA representing a significant improvement, however, this type of system still relies on assigning the type of activity based on the recorded speed alone (Lee et al., 2010; Burgos et al., 2013).

We intend to develop a new fisheries monitoring system that will provide better quality information and at a low cost to facilitate their deployment across the entire fishing fleet. Considering that fishing gear in operation has an effect on the dynamic behaviour of a vessel (Sun et al., 2011; Russo et al, 2011) and that the manoeuvres of a vessel may be related to its fishing activity, a device equipped with positioning and movement sensors could potentially provide the information necessary to accurately determine when a vessel is fishing.

Previous research has found that the sensors that are usually used in smart mobile devices such as mobile phones and tablets (GPS, accelerometer, gyroscope and compass) are capable of accurately detecting likely changes in the dynamic behaviour of a vessel during each trawl phase (Galotto-Tebar et al., 2020). On the other hand, the processing capacity of these devices enables us to include machine learning algorithms (artificial intelligence) that are able to learn about the dynamic behaviour of a vessel during the different phases of operation and identify the activity of a fishing vessel in real time, at any time.

Along these lines, some authors have succeeded in using various learning machines to classify a situation or recognise movements of objects or people based on data provided by movement sensors from smart mobile devices (Rodriguez-Martin et al., 2013; Tian et al., 2019; Cust et al., 2019). In this paper we have selected four supervised machine learning classification techniques: linear discriminant analysis (LDA) used to classify linearly separable samples, multilayer perceptron (MLP) and support vector machine (SVM) used to classify non-linearly separable samples and Bayesian classifier or probabilistic neural network (PNN). The selected classifiers are commonly used in various fields such as livestock (Rodero et al., 2012), fisheries (Bertrand et al., 2008; Czerwinski et al., 2007; Gutiérrez-Estrada et al., 2000, 2007, 2008, 2010; Queirolo et al., 2012; Robotham et al., 2010), hydraulic management (Pulido-Calvo and Portela, 2007), economics (Pérez-Ramírez and Fernández-Castaño, 2007), quality control (Gutierrez and Vázquez, 2013), etc.

Given this, the aims of this study were first to assess the viability of a range of statistical and heuristic methods as tools to classify the data retrieved by sensors on a mobile device during fishing activity, and secondly, to test their ability to identify which trawl phase a fishing vessel is in at a given time.

2. Material and methods

The models developed in this paper works with the Fishing Activity Monitoring Integral System (FAMIS) (Galotto-Tebar et al., 2020). FAMIS is a mobile application (APP) developed with Android Studio that records the vessel's movement during its fishing activity. During the recording, the data provide by the sensors of devices like smart phones and tablets are temporarily stored in a local database (SQLite) linked the fishing phase, ship's informations and fishing gear. When the mobile device connects to a server, it sends the information to a MySql database. The server allows access to information through the REST API service which allows to analyse the data using RStudio application (Fig. 1).

2.1. Mobile device and sensors

In this study, we used a Samsung SM-P600 tablet with movement sensors that include as standard: 6-axis inertial measurement unit (Bosch Sensortec BMI055) composed of a digital triaxial 12-bit

Fig. 1. FAMIS user interface installed in a smart phone and information flow.

acceleration sensor and a digital triaxial 16-bit gyroscope, a 16-bit 3-axis magnetometer (Asahi Kasei Microdevices AK8963C) and a GPS receptor. Using these sensors and sampling periods of 50 ms and 1 s, data were collected on 11 different variables (Table 1).

With the acceleration sensor, we measured the acceleration ($\overrightarrow{Ace}$) of the mobile device while it is moving, this movement being produced when there are external forces on the device, the value recorded being the force acting on the mass of the sensor. Once the mobile device is attached to the vessel's structure, we can say that the movement of the device reflects that of the vessel. One of the forces always involved in the movement and hence in the acceleration recorded by the sensor is gravity (G) (Eq. (1)). In this equation, the mass is the sum of the mass of

Table 1
Sensors used: main characteristics, data generation interval of each sensor (sampling period) and variables extracted from each sensor).

Sensor	*Range*	*Sensitivity*	Sampling period	Variable
Acceleration	*±19.6133 m/s²*	*0,009570 m/s²*	50 ms	AceX AceY AceZ
Gyroscope	*±8.7266 rad/s*	*0.0002661 rad/s*	50 ms	GirX GirY GirZ
Magnetometer	±4900 μT	0.6 μT	50 ms	MagX MagY MagZ
GPS	–	–	1 s	GpsR GpsV

the device and that of the vessel; this term can be ignored and we can consider that the accelerometer will record the components of the gravity vector ($\overrightarrow{G}$) at each point.

$$\overrightarrow{Ace} = -\overrightarrow{G} - \overrightarrow{\sum F}/Mass \approx -\overrightarrow{G} \tag{1}$$

With the gyroscope sensor, we measure the speed of rotation around the three main coordinate axes: OX, OY and OZ in radians per second (rad/s). If an observer placed on the positive side of an axis detects that the rotation around the axis is anticlockwise, the speed of rotation will be positive and when the rotation is clockwise, the speed of rotation will be negative.

With the magnetometer, we measure the Earth's magnetic field and the magnetic field around the device, in microTesla (μT). This sensor obtains the vector sum of the Earth's magnetic field and the magnetic fields generated by objects around the device such as engines, cables, etc., and provides the vector components resulting from the sum along OX, OY and OZ axes (Equation (2)):

$$\overrightarrow{Mag} = \overrightarrow{Mag_{Earth}} + \overrightarrow{Mag_{Environ}} \tag{2}$$

Using a network of satellites, the global positioning system sensor measures the 3-D geographical position of the device, providing values for latitude, longitude and altitude. The sensor also records the time-point of the measurement and is able to process these data and generate new information indicating the course, instantaneous and mean velocity, and accuracy. For this study, we have only considered it relevant to record the course (GpsR) and instantaneous speed (GpsV) every second during the activity of the vessel.

2.2. Data collection

Data were collected from the fishing and oceanographic research vessel Miguel Oliver of the Spanish General Secretariat for Fisheries, between 28 May and 1 June 1, 2016 in the waters off Castellón, Tarragona, Barcelona and Girona, during the Mediterranean International Trawl Survey (MEDITS)-Spain bottom trawl survey. This survey is undertaken annually by the Spanish National Institute of Oceanography (IEO), to assess demersal fish stock along the continental shelf and slope in the Spanish Mediterranean Sea (Bertrand et al., 2002).

Seeking to record the position of a vessel as accurately as possible, the mobile device was placed in the navigation bridge at around 18 m above the centre of gravity of the boat at the point where the pitch and roll axes intersect (Ibrahim and Grace, 2010), aligning the device with the OY axis parallel to the longitudinal axis (running from stern to prow), the OX axis parallel to the transverse axis (running from port to starboard) and the OZ axis parallel to the vertical axis. The vector information from the sensors is recorded with respect to these coordinate axes. In relation to this, we previously developed a mobile application (FAMIS) to allow users to activate sensors, select a vessel's phase of hauling at each point in time and store the information recorded in a database (Galotto-Tebar et al., 2020).

We recorded the vessel's movements during 22 trawl hauls using a GOC 73 sampling gear with Morgère trawl doors at a speed of 3 knots, of which 20 were 30-minute hauls at depths of 50 to 200 m and 2 were 60-minute hauls at depths greater than 200 m. We divided the vessel's fishing activity into four phases: steaming, setting, towing and hauling, and recorded the start and end time of each phase of operation.

2.3. Automated learning techniques

Automated machine learning refers to a set of techniques that allow computer systems themselves to create algorithms that are able to analyse data and acquire the knowledge necessary to carry out tasks such as predicting, classifying, ranking and decision making, without the need for defining initial rules to facilitate these tasks. For this study, we selected four supervised automated learning classification techniques with very different strategies: (i) linear discriminant analysis (LDA), a statistical technique that classifies samples that are linearly separable and estimates the probability of being in each group; (ii) support vector machines (SVMs) that classify nonlinearly separable samples using a geometric approach; (iii) multilayer perceptrons (MLPs) that are universal proxies of nonlinear functions; and (iv) Bayesian classifiers or probabilistic neural networks (PNNs).

2.4. Data processing

The first step in data processing was filtering, to eliminate zeros and outliers in the data recorded by the sensors. Next, we created 22 independent continuous quantitative variables (attributes) with a mean (Xm) and standard deviation (Xs) of the data in 10-s periods, allowing sufficient time for the boat to complete half a pitch/roll cycle (Barrass and Derrett, 2012): For the Accelerometer, the mean and standard deviation were recorded for the three axes, resulting in variables *AceXm, AceXs, AceYm, AceYs, AceZm,* and *AceZs*. Similar variables were recorded for the gyroscope (*GirXm, GirXs, GirYm, GirYs, GirZm, GirZs*) and the Magnometer (*MagXm, MagXs, MagYm, MagYs, MagZm, MagZs*). For the GPS the mean and standard deviation for the course and instanteneous speed were recorded: *GpsRm, GpsRs, GpsVm* and *GpsVs*.

We then created an independent nominal qualitative variable (label) called Activity to identify the launch phase of each sample from the official MEDITS-2016 survey data. The labelling of this variable follows a window-type criterion similar to that used by O'Farrell with VMS samples (O'Farrell et al., 2017), assigning the most repeated value during the 10 s prior to the sample.

The table of labelled data was created by merging the 22 attributes and the label, removing the records with at least one null, normalising attributes and balancing the samples by activity. As a result, we obtained a total of 6184 samples (1546 samples per activity).

To evaluate the classification performance of each model, we divided the table of data into two blocks: a training data set consisting of 3844 samples (62.16%) from the first 3 days of the survey, covering a total of 15 hauls, and a test data set consisting of 2340 samples (37.83%) from the last 2 days of the survey, during which 7 hauls were completed. This procedure allows models to be trained and tested under different weather conditions.

2.5. Assessment metrics

The classification performance of each model is related to the number of hits and misses. In this study, we used five specific metrics to assess the predictive power of the classifiers:

Accuracy: Percentage of positive and negative predictions that are correct.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100 \tag{3}$$

Precision: Percentage of positive predictions that are correct.

$$Precision = \frac{TP}{(TP + FP)} \times 100 \tag{4}$$

Recall: Percentage of true positives retrieved.

$$Recall = \frac{TP}{(TP + FN)} \times 100 \tag{5}$$

F1 Score: Harmonic mean of precision and recall.

$$F1Score = \frac{2*Precision*Recall}{(Precision + Recall)} \tag{6}$$

where *TP* stands for *True Positive*, the number of fishing sets correctly predicted to be a member of a class, *TN* for *True Negative*, the number of fishing sets correctly predicted to not be a member of a class, *FP* for *False*

Positive, the number of fishing set incorrectly predicted to be a member of a class, and *FN* for *False negative*, the number of fishing sets incorrectly predicted to not be a member of a class.

Kappa index: Measure of the degree of agreement between the classifier's prediction and the true classification (Cohen, 1960).

$$Kappa = \frac{p_o - p_e}{1 - p_e} \tag{7}$$

where p_o is the probability of success of the classifier and p_e is the probability of success of a random classifier.

2.6. Model hyperparameter tuning

Automated learning models are parameterised in order that their behaviour can be adjusted to fit a given problem. These models can have multiple parameters and finding the optimal combination of values sometimes requires in-depth analysis that is beyond the scope of this paper. In this study, we selected the values for each parameter using a grid search procedure for hyperparameters, establishing a discrete range of values for each parameter and creating a grid with all the possible combinations of hyperparameters to methodically assess all the resulting models. The best model was selected based on using the F1 score to compare the model's performance.

To avoid choosing hyperparameters of the model that best fit the test data set, the validation process uses different data to train and validate each configuration, saving the test data set for the final assessment of the model. The validation of each configuration of the model was carried out using k-fold cross-validation resampling (Burman, 1989) (Fig. 2). To avoid overly small data sets, we used K = 5. Having selected the parameters that provided the best results in the validation process, the model was trained with all the training data and then the classification performance is assessed with the test data.

2.7. Simplification of the model

The samples recorded during the vessel's fishing activity reflect the influence of trawling on the dynamic behaviour of the vessel. Galotto et al. (2020) demonstrated that the sensors of a mobile device were capable of identifying four distinct periods labelled: steaming, setting, towing and hauling. The task of the models used in this study is to classify the samples into one of these four categories. We should recall that the main objective of this study is to identify the start and end of trawling to obtain real-time information on fishing, and hence, the efficacy in recognising the samples taken during the towing phase is key to achieving this objective. The remaining phases could potentially be merged, thereby reducing the number of phases, if, by doing so, we improve the performance of the model when classifying the towing category. In relation to this, we validated the models with 4, 3, and 2 categories, always keeping the towing category.

The inclusion of too many attributes of samples may lead to learning machines performing the classification task less well. Given that each attribute is linked to a sensor, and in some cases, to the orientation of the mobile device with respect to the vessel, we grouped the attributes by sensor, and analysed the response of the four learning machines to all of the potential reductions in the group of attributes.

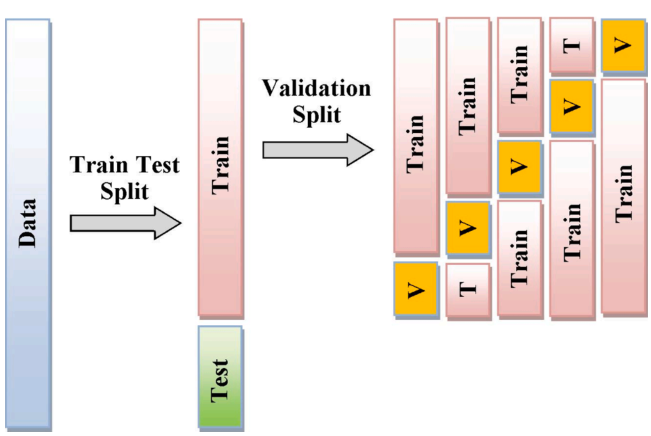

Fig. 2. K-fold cross-validation process. The samples from the table of data (Data) are split into training data (Train - T) to be used in the cross-validation process and test data (Test) to be used in the final model assessment. The 5-fold cross-validation process calculates the mean of the results from repeating the model training and validation process five times. Each process uses different data, ensuring that all the data have been used once in the validation phase.

2.8. Machine learning

2.8.1. Linear discriminant analysis (LDA)

Linear discriminant analysis (LDA) is a supervised learning algorithm used for data classification and dimension reduction. Based on a dependent qualitative variable and a set of independent quantitative variables, discriminant analysis allows samples to be classified into one of the groups established by the dependent variable.

This type of analysis provides classification procedures for new observations with an unknown origin into one of the groups analysed, by providing discriminant scores from which we can estimate the probability of being in each group. For this, the algorithm uses new variables known as discriminant variables that are able to characterise and differentiate between groups, described using discriminant functions which are linear combinations of the original variables. More detailed descriptions of the method can be found in Kim et al. (2007), Bernstein et al. (2019) and Li et al. (2020).

2.8.2. Artificial neural networks: Multilayer perceptrons (MLPs) and probabilistic neural networks (PNNs)

Artificial neural networks are mathematical models inspired by the neural architecture of the human brain (Rumelhart et al, 1986). To create an artificial neural system, we use artificial neurons functionally organised in layers. The information flows through the layers of the neural network via one-way connections that simulate the synapses between neurons. Each connection has an associated weight (w_i) equivalent to the synapse strength, and each input (x_i) through this connection is multiplied by this weight. The receptor neurone calculates the weighted sum of all the inputs, adds a numerical value known as bias (b) and applies an activation function to the result (f) to determine the output (Fig. 3).

In particular, multilayer perceptrons (MLPs) are one-way neural networks formed by an input layer with as many neurons as there are input attributes, one or several hidden layers with a variable number of neurons and output layer with the number of neurons required to show the output (Fig. 4). The data flow in one direction from the input layer towards the output layer. The outputs from each neuron of the input layer and hidden layers are connected to the input neurons of the following layer. MPLs are able to analyse complex data sets and perform nonlinear classification into two or more groups, and given this, have been widely used in a range of technical applications (Lek & Guégan, 1999; Gutiérrez-Estrada et al., 2000; Dedecker et al., 2005; Goethals et al., 2007; Pulido-Calvo and Portela, 2007; Gutiérrez-Estrada et al., 2008).

On the other hand, probabilistic neural networks (PNNs) are Bayes-Parzen classifiers composed of four layers: an input layer with the same number of neurons as input attributes; a first hidden layer, which is a pattern layer with the same number of neurons as training or calibration samples; a second hidden layer, which is a summation layer with the same number of neurons as classes; and an output layer, with a neuron that provides the result of the classification (Fig. 5) (Specht, 1990; Hajmeer and Basheer, 2002; Rodero et al., 2012).

When we supply an input to the PNN, the pattern layer assesses the

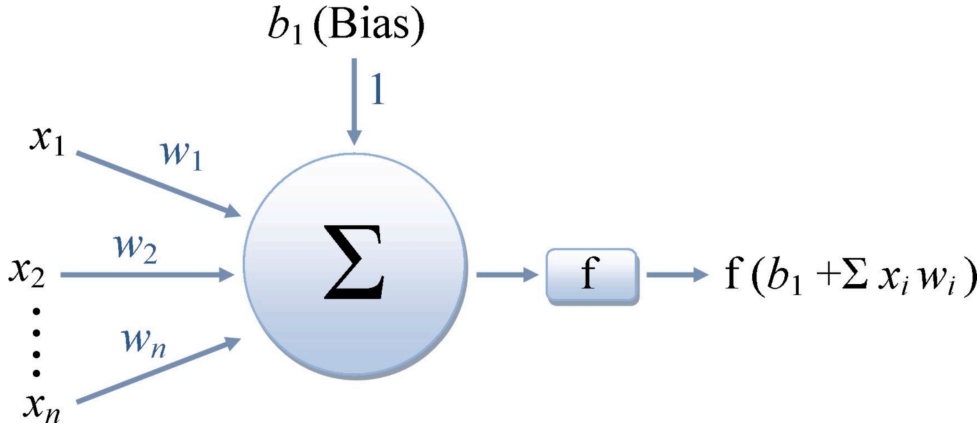

Fig. 3. Artificial neuron. It sums the inputs (x_i) multiplied by the weight associated with each one (w_i), adds the bias (b) and applies the activation function (f).

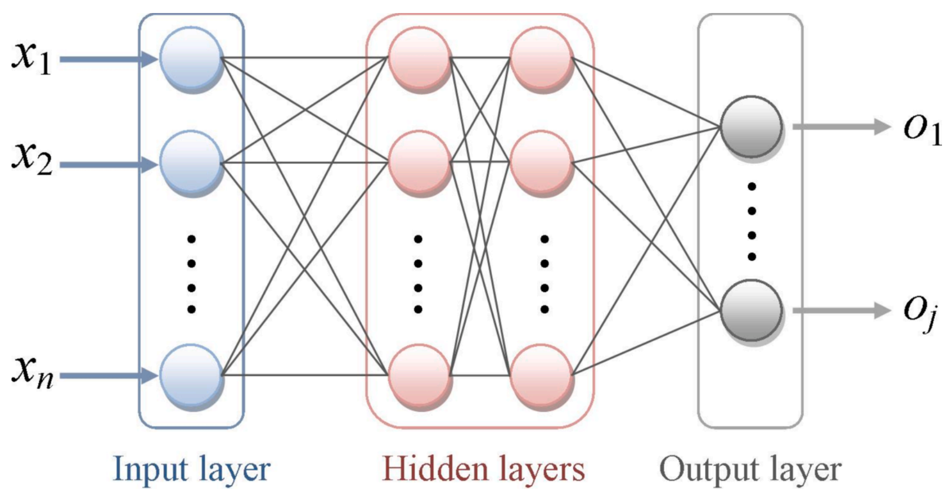

Fig. 4. Architecture of the multi-layer perceptron neural network. Input layer with n neurons that receive input attributes (x_n). Hidden layers composed of a variable number of layers and neurons. Output layer that provides the response of the neural network (o_j). The lines represent the connections that transmit the data between layers.

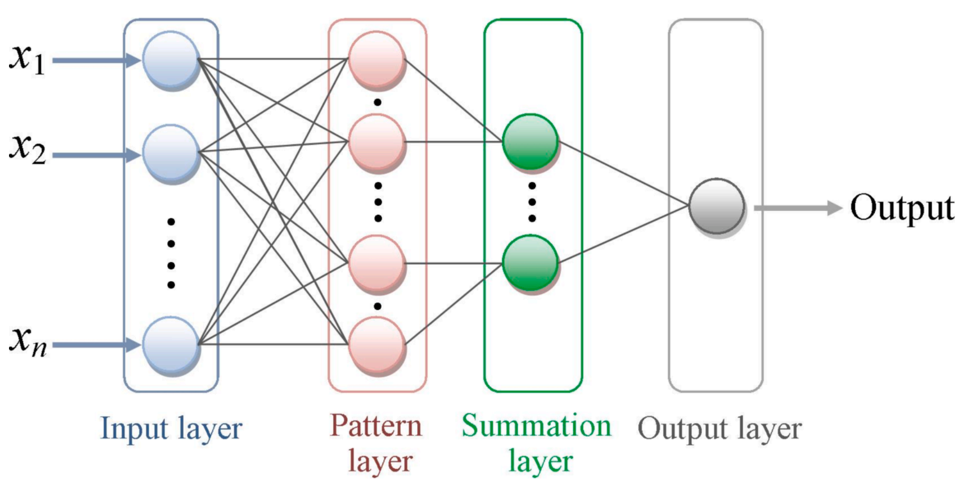

Fig. 5. Architecture of the probabilistic neural network. Input layer with n neurons that receive input attributes (x_n). Pattern layer with one neuron per sample. Summation layer with one neuron per class. Output layer with one neuron providing the output of the neural network.

distances from the input vector to the training vectors using a vector of distances, the elements of which indicate how close the input is to each training input. The summation layer (the second layer) sums the patterns of each class generating a vector of probabilities which is then used by a competitive transfer function at the output layer to select the most probable class (Pérez-Ramírez and Fernández-Castaño, 2007).

2.8.3. Support vector machine

The support vector machine (SVM) are statistical classifiers proposed by Vapnik (1995) that belong to a family of linear classifiers. SVMs seek to identify hyperplanes that divide the input feature space. At the algorithmic level, SVM learning is modelled as a quadratic optimisation problem with linear constraints, the size of the problem depending on the dimension of the feature space (Fig. 6). While markedly less popular than the aforementioned models, they have been used in a range of applications and shown a discriminatory power similar to that of MLPs (Robotham et al., 2010; Rodero et al., 2012).

2.9. Calibration and validation procedures

The construction of learning machine models starts with the processes of validation (k-fold cross-validation) and parameter section allowing adaption of their behaviour to the problem to be solved. In the case of the LDA modelling, given the use of balanced samples, the initial probability of class membership is the same for all the classes. In MLP modelling, the following were selected: a hidden layer with 8 neurons, a maximum of 30 iterations of learning, Randomize_Weights(0,1) as the initialization function, Std_Backpropagation(0.2,0) as the learning function, Topological_Order(0) as the update function, and Act_Logistic as the activation function of all hidden units and all output units. As the initialisation function assigns random initial weights, producing small differences between the models, we calculated the mean of 10 training and testing runs of the MLP. In the PNN modelling, a value of 1.4 was set as the smoothing parameter for the pattern-layer activation function. Lastly, in the SVM modelling, we chose a linear kernel and set the constraint violation cost to 1.

For data processing, simulating the learning machines and obtaining the results, we used the R programming language (https://www.r-project.org) with specific libraries to work with learning machines such as Modern Applied Statistics with S (Ripley et al., 2020), R Neural Networks using the Stuttgart Neural Network Simulator (Neural, 2022), probabilistic neural networks PNN (Chasset, 2016) and Misc Functions of the Department of Statistics Probability Theory Group (E1071; Meyer, 2019).

3. Results

Table 2 shows the results concerning the performance in the final test of the four learning machines analysed in this study.

In general, MLPs and SVMs provided better results than LDA or PNNs. The best harmonic mean between precision and recall (F1 scores) was obtained with the SVM model, with a mean among the categories of 68.21% compared to the 65.41% observed in the MLP model. Further, the SVM model was better than the MLP model in precision, with a mean of 76.04% compared to 67.52%. In contrast, the MLP model was better in terms of accuracy and sensitivity, with mean percentages of 81.75% and 63.50% compared to 80.92% and 61.84% in the SVM model. On the other hand, the MLP model provided more balanced results than the SVM one, with regards to classifying the samples recorded during the trawling phase, with an accuracy of 87.24%, a precision of 71.46% and an F1 score value of 76.26%. The SVM model only performed better than the MLP model in terms of sensitivity, being able to identify 97.61% of the trawling activity samples, at the expense of a very high number of false positives. We should highlight the difficulties of the LDA, MLP and PNN models in identifying the setting category, these only being able to recognise 31.11%, 32.43% and 17.95%, respectively, of the samples recorded in this phase.

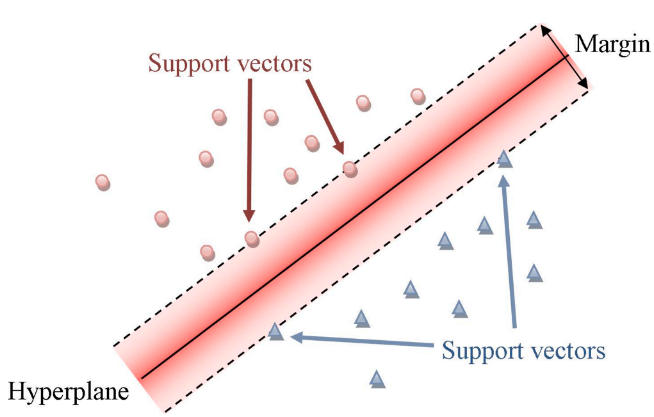

Fig. 6. Graphical representation of the hyperplane used by the support vector machine to discriminate samples of different classes. Support vectors show which samples are closest to the hyperplane and define the edge of the margin.

Table 2
Accuracy, *precision*, *recall* and *F1 score* (expressed as percentages) in the final test of the LDA, MLP, PNN and SVM learning machines. This analysis included the samples recorded during all categories of activity with all their attributes.

Activity	*Accuracy*	*Precision*	*Recall*	*F1 Score*
LDA				
Steaming	73.63	47.66	55.73	51.38
Setting	77.82	61.07	31.11	41.22
Towing	80.30	62.76	52.14	56.96
Hauling	78.76	55.05	82.05	65.89
MLP				
Steaming	72.15	46.96	70.17	55.93
Setting	80.00	73.00	32.43	43.66
Towing	87.24	71.46	81.97	76.26
Hauling	87.60	78.66	69.42	73.51
PNN				
Steaming	76.88	54.26	47.86	50.86
Setting	71.88	37.10	17.95	24.19
Towing	68.08	40.24	57.09	47.21
Hauling	71.45	44.16	53.68	48.46
SVM				
Steaming	83.80	99.05	35.56	52.33
Setting	86.45	76.69	65.81	70.84
Towing	68.59	44.20	97.61	60.84
Hauling	84.83	84.23	48.38	61.45

Table 3 presents the ability of learning machines to distinguish samples from two fishing phases. The analysis from using samples from two trawl phases (1 versus 1) is reported in the top half of the table and from using all the samples, three phases being merged in a new class called “the rest” (1 versus the rest) in the bottom half. This table provides F1 scores as a percentage in the final test of the four models. The results obtained reveal the weaknesses of the LDA, MLP and PNN models in different trawl phases. The LDA and MLP models struggle to differentiate samples from the steaming and setting phases, being able to identify <33% of the samples recorded during setting. On the other hand, the PNN model did not differentiate well between steaming and setting, but in this case, the poor results correspond to the steaming phase, this model identifying only 22.56% of the samples obtained during steaming. Similarly, the PNN provided poor results for differentiating between setting and towing, being able to identify only 14.87% of the samples from the setting phase.

Next, we explored the possibility of improving performance in the classification process by reducing the number of categories, testing all four models, maintaining the towing and steaming activities and removing the setting or hauling categories or both. Table 4 shows that all the models performed better in classifying towing (F1 score) when setting and hauling were eliminated and assigning the samples taken in these phases to steaming and towing, respectively.

We also explored the response of the four models to a reduction in the number of input attributes. For this purpose, we grouped the

Table 3
F1 score (expressed as a percentage) in the final test of the LDA, MLP, PNN and SVM learning machines. The samples are included with all their attributes. *1*-vs-*1* includes samples from two categories. *1*-vs-*Rest* includes all the samples, maintaining the category on the left and grouping the other categories in the *Rest* category.

Type of partition	LDA	MLP	PNN	SVM
1-vs-1				
Steaming-vs-Setting	59.37	69.48	34.46	57.65
	35.58	44.82	68.11	73.77
Steaming-vs-Towing	74.34	84.94	51.52	63.56
	77.12	85.04	75.39	78.87
Steaming-vs-Hauling	79.77	85.12	58.25	66.59
	84.01	86.42	76.27	79.97
Setting-vs-Towing	90.66	92.49	25.89	87.24
	91.05	91.84	70.14	88.58
Setting-vs-Hauling	95.21	94.14	81.59	93.81
	95.22	94.47	83.44	93.17
Towing-vs-Hauling	81.23	66.37	75.88	84.83
	80.28	75.58	54.79	79.27
1-vs-Rest				
Steaming-vs-Rest	69.43	75.85	48.51	61.26
	63.15	67.77	73.05	71.79
Setting-vs-Rest	46.09	53.16	73.12	62.08
	65.88	72.90	60.32	69.30
Towing-vs-Rest	60.54	88.72	69.60	71.03
	61.67	88.54	22.46	34.95
Hauling-vs-Rest	80.09	81.84	74.73	77.48
	70.55	79.71	64.62	61.32

Table 4
F1 score (expressed as a percentage) for the towing category in the final test of the LDA, MLP, PNN and SVM learning machines with all the attributes and after reducing the number of categories. The results without a reduction and with 3 and 2 categories. The symbol ">" written between two categories indicates that the category on the left is merged with that on the right, the label of the samples of the category on the left being changed to that of the one on the right.

Activity	LDA	MLP	PNN	SVM
Steaming-Setting-Towing-Hauling				
Unreduced	56.96	76.26	47.21	60.84
Steaming-Towing-Hauling				
Setting > Steaming	57.14	73.77	55.70	64.81
Setting > Towing	55.16	68.13	58.70	67.06
Steaming-Setting-Towing				
Hauling > Steaming	41.64	71.15	55.06	18.88
Hauling > Towing	77.75	84.48	64.95	71.16
Steaming-Towing				
Setting > Steaming/Hauling > Steaming	55.47	77.51	70.72	71.05
Setting > Steaming/Hauling > Towing	85.38	87.34	76.18	79.42
Setting > Towing/Hauling > Steaming	44.98	73.97	69.41	53.75
Setting > Towing/Hauling > Towing	76.42	77.81	76.02	79.35

Table 5
Accuracy, precision, recall and *F1 score* (expressed as percentages) for the towing category in the final tests of the LDA, MLP, PNN and SVM learning machines, with all the categories and after reducing attributes grouped by sensors. The results are shown before reduction (Ace-Gry-Mag-GPS) and after reduction using the set of sensors that provided the best *F1 score* for the *Towing* phase. The attributes were grouped by sensors as follows: Ace = (AceXm, AceXs, AceYm, AceYs, AceZm, AceZs), Gir = (GirXm, GirXs, GirYm, GirYs, GirZm, GirZs), Mag = (MagXm, MagXs, MagYm, MagYs, MagZm, MagZs) and Gps = (GpsRm, GpsRs, GpsVm, GpsVs).

Sensors	Accuracy	Precision	Recall	F1 Score
LDA				
Ace-Gir-Mag-Gps	80.30	62.76	52.14	56.96
Ace-Gir-Gps	82.35	60.29	86.15	70.94
MLP				
Ace-Gir-Mag-Gps	87.24	71.46	81.97	76.26
Gir-Gps	88.62	74.39	83.37	78.57
PNN				
Ace-Gir-Mag-Gps	68.08	40.24	57.09	47.21
Gps	66.67	42.55	95.21	58.82
SVM				
Ace-Gir-Mag-Gps	68.59	44.20	97.61	60.84
–	–	–	–	–

Table 6
Accuracy, *precision*, *recall*, *F1 score* and *Kappa index* for the towing category in the final test of the LDA, MPL, PNN and SVM learning machines. The results are shown before reduction (all activities/all sensors) and after the reduction in categories and attributes grouped by sensors that provide the best *F1 scores* in the *towing* phase.

Activities/Sensors	Accuracy	Precision	Recall	F1 Score	Kappa
LDA					
All activities/All sensors	80.30	62.76	52.14	56.96	0.60
Steaming-Towing/Ace-Gir-Gps	84.02	76.78	97.52	85.92	0.73
MLP					
All activities/All sensors	87.24	71.46	81.97	76.26	0.70
Steaming-Towing/Gir-Gps	89.69	85.11	96.26	90.33	0.83
PNN					
All activities/All sensors	68.08	40.24	57.09	47.21	0.34
Steaming-Towing/Gps	74.19	66.45	97.69	79.10	0.54
SVM					
All activities/All sensors	68.59	44.20	97.61	60.84	0.56
Steaming-Towing/All sensors	70.04	62.66	99.23	76.81	0.52

attributes by sensor and tested all the possible combinations. Table 5 shows the F1 score for towing with each of the four models with all the attributes and the best response to a potential reduction in attributes grouped by sensor. We can observe the substantial improvement in the LDA model after removing the variables based on data from the magnetometer while the PNN model provided better results using only the GPS-based attributes. Similarly, the MLP model improved 3% using only the gyroscope- and GPS-based attributes, while the performance of the SVM model did not improve by reducing attributes.

Table 6 shows the results in the final test of the four models considering the reduction of attributes and categories. The MLP model performs markedly better than the others, with a hit rate of 89.69% in their positive and negative predictions concerning the towing phase, a positive predictive rate of 85.11%, a towing phase sample identification rate of 96.26%, a harmonic mean precision and recall of 90.33%, and a Kappa index with a high degree of agreement of 0.83 ($p < 0.0001$). Fig. 7 shows graphically a part of the final test of the models.

4. Discussion

The information provided by the VMS (position, course and speed) to identify fishing grounds and the activity of a fishing fleet can be improved using additional sensors as is suggested in this study. In trawl fishing, precision is key to assessing haul duration and track, the spatial and time distribution of the effort, and subsequently the catch per unit

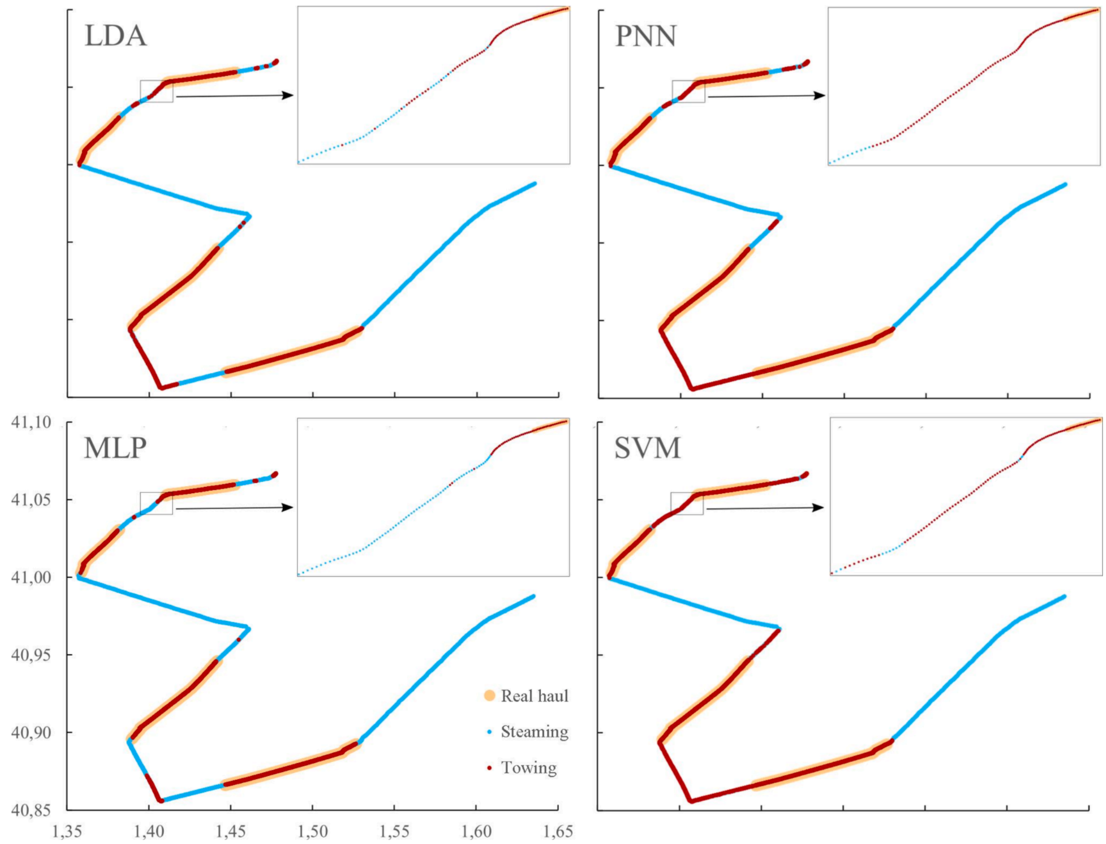

Fig. 7. Graphical representation of the prediction of 4 hauls of the LDA, MLP, PNN and SVM models. The orange area represents the real hauls, the blue dots the *steaming* prediction and the red dots the *towing* prediction.

effort in each fishing area. Deng (2005) stated that the errors in the prediction of trawl fishing increase with the time interval between VMS records and that the loss in precision is relatively high with sampling intervals longer than 30 min. On the other hand, Mills et al. (2007) indicated that a vessel's speed alone is not a suitable criterion for identifying trawling activity and confirmed that the sampling frequency of the VMS is too low to allow trawl tracks to be characterised, most trawls being represented by just one record and some not being detected at all, the haul occurring between two records. In addition, for Peruvian anchoveta, Bertrand et al. (2008) reported that speed provided by VMS data leads to the number of fishing events being overestimated by nearly 182%. These findings warrant the development and deployment of new low-cost monitoring systems, complementary to VMS, that are flexible and easily upgradeable and that allow accurate estimation of the activities of fishing vessels.

Since the deployment of the VMS in 2000, many researchers have studied how to process and complement the data provided by this system to improve the management and monitoring of fishing activity. For example, Szostek et al. (2017) included information extracted from interviews of fishermen about their experience fishing king scallops in the English Channel, while Bastardie et al. (2010) combined the data provided by the VMS with information extracted from logbooks through a linking process considering the degree of mismatch. In these ways, these authors were able to obtain disaggregated fishing effort data at a fine geographical scale. Overall, they indicate that the procedures they used significantly improve the delimitation of the catchment area as well as the assessment of fishing effort in time and space, but to achieve this, there was a need for offline processing of external data and its subsequent integration with data generated by the VMS. In relation to this, the data provided by FAMIS (Galotto-Tebar et al., 2020), the system we propose, is at the same level as that from VMS, and hence, it is fully complementary and therefore easy to integrate.

The use of the FAMIS in fishing vessels would significantly improve on the use of the VMS alone, as it adds movement sensors on the vessel and processing capacity to allow the use of artificial intelligence algorithms to identify the activity of a vessel in situ. Galotto et al. (2020) showed that standard low-cost devices such as mobile phones and tablets do record different data in each trawl phase, and building on that, in this study, we demonstrate that certain models, such as MLPs or SVMs, yield classifiers that are good tools for automated classification.

The four types of models analysed in this study use very different classification strategies, and in line with that, their classification performance varied. The results obtained with samples recorded in the four trawl phases and attributes associated with the four sensors indicated that SVMs and MLPs provide better results than LDA and PNNs. These results are similar to those obtained by other authors. Robotham et al. (2010) found that MLPs and SVMs performed significantly better than PNNs or LDA, when attempting to distinguish between four pelagic fish species caught off the northern coast of Chile based on echograms of fish shoals. Further, Rodero et al. (2012) compared the same four types of models for distinguishing between four Andalusian cattle breeds based on morphometric variables and found that MLPs and SVMs had better classification performance. This is attributable to the fact that both MLPs and SVMs associate highly nonlinear discriminant functions with the input patterns. This is a result of the large number of variables in each sample, the physical magnitudes that they represent and the complexity of the identification task, in our case, the identification of movement patterns of a vessel exposed to forces from the waves, wind, engine, rudder and fishing gear.

The high nonlinearity of the input data is clearly reflected in the inability of LDAs to find a discriminant function that distinguishes the samples from different classes. In relation to this, the PNN-based model, which takes a radial approach, identifying new samples based on similarities with training samples, provided the poorest results with a mean

F1 score of just 44%. In contrast, the MLP as a universal classifier, with a single hidden layer and a small number of neurons, provided an F1 score of 65.41%, being able to identify nearly 82% of samples in the trawl phase. Finally, the most efficient configuration of the SVM model with a linear kernel finds spaces with new dimensions where it is possible to separate different classes with hyperplanes, providing an acceptable mean F1 score of 68.21% and excellent recall of the trawl phase of over 97%.

These results significantly improve on the classification performance found by previous authors based on VMS data. Gerritsen and Lordan (2011) reported that the use of vessel speed as a classification criterion for fishing activity provided a low rate of false negatives but a very high rate of false positives in the classification (32%), only 68% of cases being correctly classified. These authors highlighted that the potential reasons for these results are related to the fact that a boat may travel at a speed similar to that when trawling when sail at a slow speed while waiting for the right tide or because of poor weather conditions, among other factors.

In relation to this, the four trawl phases identify periods in time when the fishing vessel is involved in different activities. In each phase, the dynamic behaviour of the vessel is influenced by various elements involved such as speed, course, waves, fishing gear, catch, etc. These elements have a different impact at each phase. For example, the steaming phase ends when the boat is manoeuvring and slowing down until it positions itself where it is going to start trawling. After that, the setting phase starts by shooting the net; then the doors are set and the towing cable paid out. During the towing phase, the drag on the vessel grows as the growing volume of the catch is added to the resistance of the fishing gear itself. Next, in the hauling phase, the towing cable is winched in, the doors are raised, and finally, the fishing gear and the catch are brought on board. At the end of the haul, the vessel starts a new steaming phase, manoeuvring and increasing its speed to move to the next fishing spot. Given all this, the effect of the different manoeuvres within each phase on the behaviour of the vessel cannot be detected by the speed data provided by the VMS unless this information is complemented with movement sensor data such as those recorded through FAMIS.

On the one hand, homogenous behaviour of the vessel within a phase and marked changes between phases should facilitate the task of classifying the samples and identifying the start and end of the trawl phases. The start of setting is the most problematic time point. This is because until the doors are in the water, the resistance of the nets is low and the behaviour of the vessel is similar to that in the steaming phase. In contrast, the towing phase starts when the fishing gear, moving at the towing speed, reaches the correct depth, the cable stops being paid out and the speed of the vessel is reduced to the towing speed. This change of speed enables the models to identify the first samples of the towing phase and therefore the start of the real phase of fishing. Similarly, the hauling phase is easy to identify, given that it starts with a reduction in the vessel's speed to make it easier to bring the fishing gear on board and finishes with the fishing gear inactive and an increase in the vessels speed.

The results of the models analysed are significantly better when we simplify the description of the activity of the fishing vessel to two phases, a common approach for discriminating the activity in a fishing boat when onboard observers log whether the boat is “fishing” or “not fishing” as a function of whether the fishing gear is deployed in the water (Chang and Yuan, 2014).

By merging the phases steaming + setting and towing + hauling, we obtained F1 scores of over 76% in all the models. These proposed mergers avoid the difficulty of identifying the samples at the start of the setting phase and considers that the vessel is fishing from when it starts towing until the fishing gear is taken out of the water. This partially consistent with the results of Joo et al. (2011) who used MLPs to estimate fishing events in Peruvian anchoveta fisheries, based on VMS parameters and validated in-situ by onboard observers. In that study, observers validated two segments of activity in the speed data series provided by VMS, identified as acceleration between the previous and current pace (which would coincide in part with the merging of steaming + setting) and the acceleration between the current and the following pace (approximately towing + hauling).

In relation to the sensor systems used in this study, we should indicate that all four sensors provided relevant information about changes in the vessels behaviour in the four trawl phases (Galotto-Tebar et al., 2020). The reduction in computational costs associated with removing variables had a relatively small impact on the behaviour of the models. The sensitivity analysis indicated that the GPS sensor is undoubtedly the device that provides the most important information, given that using this device alone, the models assessed achieved a mean F1 score of over 61%. However, the use of information from the other sensors improved the behaviour of some models. For example, the MLP model was able to correctly identify 96.26% of the samples recorded by the GPS + gyroscope in the towing phase. Specifically, for a standard 60-minute haul, the model identified the act of fishing during 57 min and 45 s, in the worst-case scenario with the false negatives concentrated at the start or end of trawling. The prediction error of 2 min and 45 s translates to the identification of a shorter-duration hauling operation and an advance or delay in the start or end of the hauling. On the other hand, we should bear in mind that characteristics of the boat where we performed our experiment (Miguel Oliver) mean that it had a higher vessel size to fishing gear ratio than that in fishing boats. This suggests that the dynamics in a fishing boat may be even more affected by the fishing gear, and hence, the difference between the trawl phases will be easier to identify using this system.

5. Conclusions

The results of this work indicate that the use of the proposed device (FAMIS) on the basis of a mobile device with GPS, accelerometer, gyroscope and magnetic field sensors significantly improves the accuracy of current VMS systems.

The 4 models analysed have shown different capacities to classify fishing activity. The results obtained with samples recorded in the 4 phases of the set (steaming, setting, towing and hauling) and attributes associated with the 4 sensors (Gps, accelerometer, gyroscope and magnetic field) indicate that the SVM and MLP models offer better results than the LDA and PNN models.

On the other hand, the reduction to 2 phases (steaming and towing) considering that the towing phase starts when the fishing gear reaches the correct depth and trawling speed, and ends when the fishing gear is on the deck of the vessel, facilitates the task of classification and substantially improves the results of the 4 models. Furthermore, the reduction of the sensors involved in each model also improves the quality of their predictions.

In summary, the results of this work indicate that the proposed system is capable of classifying the information recorded by the sensors of a mobile device during fishing activity and identifying the phase of the set in which the vessel is at any given time.

CRediT authorship contribution statement

M.M. Galotto-Tebar: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Visualization. **A. Pomares-Padilla:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – review & editing. **I.A. Czerwinski:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing – review & editing, Visualization, Supervision. **J.C. Gutiérrez-Estrada:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing – review & editing, Visualization, Supervision.

Declaration of Competing Interest

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.

References

Barrass, C. B., & Derrett, D. R. (2012). *Ship Stability for Masters and Mates*. Elsevier Ltd.

Bastardie, F., Nielsen, J. R., Ulrich, C., Egekvist, J., & Degel, H. (2010). Detailed mapping of fishing effort and landings by coupling fishing logbooks with satellite-recorded vessel geo-location. *Fisheries Research, 106*(1), 41–53.

Bernstein, R., Osadchy, M., Keren, D., & Schuster, A. (2019). LDA classifier monitoring in distributed streaming systems. *Journal of Parallel and Distributed Computing, 123*, 156–167.

Bertrand, J. A., Gil de Sola, L., Papaconstantinou, C., Relini, G., & Souplet, A. (2002). The general specifications of the MEDITS surveys. *Scientia Marina, 66*(2), 9–17.

Bertrand, S., Díaz, E., & Lengaigne, M. (2008). Patterns in the spatial distribution of Peruvian anchovy (*Engraulis ringens*) revealed by spatially explicit fishing data. *Progress in Oceanography, 79*(2–4), 379–389.

Burgos, C., Gil, J., & del Olmo, L. A. (2013). The Spanish blackspot seabream (Pagellus bogaraveo) fishery in the Strait of Gibraltar: spatial distribution and fishing effort derived from a small-scale GPRS/GSM based fisheries vessel monitoring system. *Aquatic Living Resources, 26*(4), 399–407. https://doi.org/10.1051/alr/2013068

Burman, P. (1989). A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika, 76*(3), 503–514.

Chang, S. K., & Yuan, T. L. (2014). Deriving high-resolution spatiotemporal fishing effort of large-scale longline fishery from vessel monitoring system (VMS) data and validated by observer data. *Canadian Journal of Fisheries and Aquatic Sciences, 71*, 1363–1370.

Walter, J. F., Hoenig, J. M., & Gedamke, T. (2007). Correcting for effective area fished in fishery-dependent depletion estimates of abundance and capture efficiency. *ICES Journal of Marine Science, 64*(9), 1760–1771. https://doi.org/10.1093/icesjms/fsm147

Chasset, P.O. (2016). Package pnn. https://cran.r-project.org/web/packages/pnn/pnn.pdf.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement, 20*, 1.

Cojan, M., & Burgos, C. (2015). Análisis de la información proporcionada por los sistemas de localización vía satélite de la flota que explota la chirla (Chamelea gallina) en el Golfo de Cádiz. In *Teledetección: Humedales y Espacios Protegidos. XVI Congreso de la Asociación Española de Teledetección, 550-553* http://ocs.ebd.csic.es/index.php/AET/2015/schedConf/presentations.

Cust, E. E., Sweeting, A. J., Ball, K., & Robertson, S. (2019). Machine and deep learning for sport-specific movement recognition: A systematic review of model development and performance. *Journal of Sports Sciences, 37*(5), 568–600.

Czerwinski, I. A., Gutiérrez-Estrada, J. C., & Hernando-Casal, J. A. (2007). Short-term forecasting of halibut CPUE: Linear and non-linear univariate approaches. *Fisheries Research, 86*(2–3), 120–128.

Dedecker, A. P., Goethals, P. L. M., D'heygere, T., Gevrey, M., Lek, S., & De Pauw, N. (2005). Application Of Artificial Neural Network Models To Analyse The Relationships Between Gammarus pulex L. (Crustacea, Amphipoda) And River Characteristics. *Environmental Monitoring and Assessment, 111*(1), 223–241. https://doi.org/10.1007/s10661-005-8221-6

Deng, R., Dichmont, C., Milton, D., Haywood, M., Vance, D., Hall, N., & Die, D. (2005). Can vessel monitoring system data also be used to study trawling intensity and population depletion? The example of Australia's northern prawn fishery. *Canadian Journal of Fisheries and Aquatic Sciences, 62*(3), 611–622.

European Commission. (2007). On improving fishing capacity and effort indicators under the common fisheries policy. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52007DC0039&from=EN.

Fock, H. O (2008). Fisheries in the context of marine spatial planning: Defining principal areas for fisheries in the German EEZ. *Marine Policy*. https://doi.org/10.1016/j.marpol.2007.12.010

Galotto-Tebar, M. M., Pomares-Padilla, A., Czerwinski, I. A., & Gutiérrez-Estrada, J. C. (2020). Using mobile device's sensors to identify fishing activity. *Journal of Marine Science and Technology, 25*, 978–989.

Gerritsen, H., & Lordan, C. (2011). Integrating vessel monitoring systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. *ICES Journal of Marine Science, 68*(1), 245–252.

Gerritsen, H. D., Lordan, C., Minto, C., & Kraak, S. B. M. (2012). Spatial patterns in the retained catch composition of Irish demersal otter trawlers: High-resolution fisheries data as a management tool. *Fisheries Research, 129–130*, 127–136.

Goethals, P. L. M., Dedecker, A. P, Gabriels, W., Lek, S., & De Pauw, N. (2007). Applications of artificial neural networks predicting macroinvertebrates in freshwaters. *Aquatic Ecology, 41*(3), 491–508. https://doi.org/10.1007/s10452-007-9093-3

Gutierrez, P. T., & Vázquez, J. A. (2013). Aplicación de la red neuronal probabilística para la clasificación de productos conforme a sus especifícaciones. *Innovation in Engineering, Technology and Education for Competitiveness and Prosperity*, 1–10.

Gutiérrez-Estrada, J. C., Pulido-Calvo, I., & Prenda, J. (2000). Gonadosomatic index estimates of an introduced pumpkinseed (*Lepomis gibbosus*) population in a Mediterranean stream, using computational neural networks. *Aquatic Sciences, 62*(4), 350–363.

Gutiérrez-Estrada, J. C., Silva, C., Yáñez, E., Rodríguez, N., & Pulido-Calvo, I. (2007). Monthly catch forecasting of anchovy (*Engraulis ringens*) in the north area of Chile: Non-linear univariate approach. *Fisheries Research, 86*(2–3), 188–200.

Gutiérrez-Estrada, J. C., Vasconcelos, R., & Costa, M. J. (2008). Estimating fish community diversity from environmental features in the Tagus estuary (Portugal): Multiple Linear Regression and Artificial Neural Network approaches. *Journal of Applied Ichthyology, 24*(2), 150–162.

Gutiérrez-Estrada, J. C., & Bilton, D. T. (2010). A heuristic approach to predicting water beetle diversity in temporary and fluctuating waters. *Ecological Modelling, 221*(11), 1451–1462.

Hajmeer, M., & Basheer, I. (2002). A probabilistic neural network approach for modeling and classification of bacterial growth/no-growth data. *Journal of Microbiological Methods, 51*(2), 217–226.

Ibrahim, R. A., & Grace, I. M. (2010). Modeling of Ship Roll Dynamics and Its Coupling with Heave and Pitch. *Mathematical Problems in Engineering, 2010*, 1–32.

Jennings, S., & Lee, J. (2012). Defining fishing grounds with vessel monitoring system data. *ICES Journal of Marine Science, 69*(1), 51–63.

Joo, R., Salcedo, O., Gutierrez, M., Fablet, R., & Bertrand, S. (2015). Defining fishing spatial strategies from VMS data: Insights from the world's largest monospecific fishery. *Fisheries Research, 164*, 223–230.

Junta de Andalucía (2004). Localización y seguimiento de embarcaciones pesqueras (SLSEPA). http://www.juntadeandalucia.es/organismos/agriculturapescaydesarrollorural/areas/pesca-acuicultura/slsepa.html.

Kim, H., Drake, B. L., & Park, H. (2007). Multiclass classifiers based on dimension reduction with generalized LDA. *Pattern Recognition, 40*(11), 2939–2945.

Lee, J., South, A. B., & Jennings, S. (2010). Developing reliable, repeatable, and accessible methods to provide high-resolution estimates of fishing-effort distributions from vessel monitoring system (VMS) data. *ICES Journal of Marine Science, 67*(6), 1260–1271.

Lek, S., & Guégan, J. F (1999). Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling, 120*(2), 65–73. https://doi.org/10.1016/S0304-3800(99)00092-7

Li, C. N., Shao, Y. H., Yin, W., & Liu, M. Z. (2020). Robust and Sparse Linear Discriminant Analysis via an Alternating Direction Method of Multipliers. *IEEE Transactions on Neural Networks and Learning Systems, 31*(3), 915–926.

Meyer, D. (2019). Package e1071. https://cran.r-project.org/web/packages/e1071/e1071.pdf.

Mills, C. M., Townsend, S. E., Jennings, S., Eastwood, P. D., & Houghton, C. A. (2007). Estimating high resolution trawl fishing effort from satellite-based vessel monitoring system data. *ICES Journal of Marine Science, 64*(2), 248–255.

Neural Networks in R using the Stuttgart Neural Network Simulator. https://github.com/cbergmeir/RSNNS.

Murawski, S. A, Wigley, S. E, Fogarty, M. J, Rago, P. J, & Mountain, D. G (2005). Effort distribution and catch patterns adjacent to temperate MPAs. *ICES Journal of Marine Science, 62*(6), 1150–1167. https://doi.org/10.1016/j.icesjms.2005.04.005

O'Farrell, S., Sanchirico, J. N., Chollett, I., Cockrell, M., Murawski, S. A., Watson, J. T., … Perruso, L. (2017). Improving detection of short-duration fishing behaviour in vessel tracks by feature engineering of training data. *ICES Journal of Marine Science, 74*(5), 1428–1436.

Pérez-Ramírez, F. O., & Fernández-Castaño, H. (2007). Las redes neuronales y la evaluación del riesgo de crédito. *Revista Ingenierías Universidad de Medellín, 6*(10), 77–91.

Pulido-Calvo, I., & Portela, M. M. (2007). Application of neural approaches to one-step daily flow forecasting in Portuguese watersheds. *Journal of Hydrology, 332*(1–2), 1–15.

Queirolo, D., Hurtado, C. F., Gaete, E., Soriguer, M. C., Erzini, K., & Gutiérrez-Estrada, J. C. (2012). Effects of environmental conditions and fishing operations on the performance of a bottom trawl. *ICES Journal of Marine Science, 69*(2), 293–302.

Ripley, B., Venables, B., Bates, D.M, Hornik, K., Gebhardt, A., & Firth, D. (2020). Package mass. http://www.stats.ox.ac.uk/pub/MASS4/.

Rijnsdorp, A. (1998). Micro-scale distribution of beam trawl effort in the southern North Sea between 1993 and 1996 in relation to the trawling frequency of the sea bed and the impact on benthic organisms. *ICES Journal of Marine Science, 55*(3), 403–419. https://doi.org/10.1006/jmsc.1997.0326

Robotham, H., Bosch, P., Gutiérrez-Estrada, J. C., Castillo, J., & Pulido-Calvo, I. (2010). Acoustic identification of small pelagic fish species in Chile using support vector machines and neural networks. *Fisheries Research, 102*(1–2), 115–122.

Rodero, E., González, A., Luque, M., Herrera, M., & Gutiérrez-Estrada, J. C. (2012). Classification of Spanish autochthonous bovine breeds. Morphometric study using classical and heuristic techniques. *Livestock Science, 143*(2–3), 226–232.

Rodriguez-Martin, D., Samà, A., Perez-Lopez, C., Català, A., Cabestany, J., & Rodriguez-Molinero, A. (2013). SVM-based posture identification with a single waist-located triaxial accelerometer. *Expert Systems with Applications, 40*(18), 7203–7211.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by backpropagation errors. *Nature, 323*, 533–536.

Russo, T., Parisi, A., Prorgi, M., Boccoli, F., Cignini, I., Tordoni, M., & Cataudella, S. (2011). When behaviour reveals activity: Assigning fishing effort to métiers based on VMS data using artificial neural networks. *Fisheries Research, 111*, 53–64.

Russo, T., D'Andrea, L., Parisi, A., Martinelli, M., Belardinelli, A., Boccoli, F., … Cataudella, S. (2016). Assessing the fishing footprint using data integrated from different tracking devices: Issues and opportunities. *Ecological Indicators, 69*, 818–827.

Salthaug, A., & Johannessen, T. (2006). The Norwegian in-year monitoring fishery for sandeel in the North Sea using satellite-based VMS data and landings information. *ICES CM documents* (pp. 1949–2011). ICES.

Specht, D. F. (1990). Probabilistic Neural Networks. *Neural Networks, 3*, 109–118.

Sun, X., Yin, Y., Jin, Y., Zhang, X., & Zhang, X. (2011). The modeling of single-boat, mid-water trawl systems for fishing simulation. *Fisheries Research, 109*(1), 7–15.

Szostek, C. L., Murray, L. G., Bell, E., & Kaiser, M. J. (2017). Filling the gap: Using fishers' knowledge to map the extent and intensity of fishing activity. *Marine Environmental Research, 129*, 329–346.

Tian, Y., Wang, X., Chen, W., Liu, Z., & Li, L. (2019). Adaptive multiple classifiers fusion for inertial sensor based human activity recognition. *Cluster Computing, 22*(4), 8141–8154.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. New York: Springer.