Neural Computing and Applications manuscript No.
(will be inserted by the editor)

Prediction of convective clouds formation using evolutionary neural computation techniques

David Guijo-Rubio[1] ✉ · **Pedro A. Gutiérrez**[1] · **Carlos Casanova-Mateo**[2] · **Juan Carlos Fernández**[1] · **Antonio Manuel Gómez-Orellana**[1] · **Pablo Salvador-González**[3] · **Sancho Salcedo-Sanz**[4] · **César Hervás-Martínez**[1]

Received: date / Accepted: date

Abstract The prediction of convective clouds formation is a very important problem in different areas such as agriculture, natural hazards prevention or transport-related facilities, among others. In this paper we evaluate the capacity of different types of evolutionary artificial neural networks to predict the formation of convective clouds, tackling the problem as a classification task. We use data from Madrid-Barajas airport, including variables and indices derived from the Madrid-Barajas airport radiosonde station. As objective variable, we use the cloud information contained in the METAR and SPECI meteorological reports from the same airport and we consider a prediction time-horizon of 12 hours. The performance of different types of evolutionary artificial neural networks has been discussed and analysed, including three types of basis functions (Sigmoidal Unit, Product Unit and Radial Basis Function), and two types of models, a mono-objective evolutionary algorithm with two objective functions and a multi-objective evolutionary algorithm optimised by the two objective functions simultaneously. We show that some of the developed neuro-evolutionary models obtain high quality solutions to this problem, due to its high unbalance characteristic.

Keywords Convection initialization prediction · machine learning algorithms · neural networks · unbalanced databases

1 Introduction

Convective weather conditions can significantly affect many strategic economical areas such as electric supply, communications, logistic services and transport, especially air traffic. The identification of atmospheric situations that favours the initiation of convection as well as the accurate prediction of its timing and location is still a difficult task for the operational weather forecasters [1]. The basic ingredients for convection to occur stated by Johns and Doswell in [2] (a sufficiently moist and deep layer in the low or mid atmosphere, conditional instability and a triggering mechanism) are still the base of many tools and applications developed to support the forecast of convection initiation. However, several works discussing the results obtained in many field studies have been published in the last years [3,4, 5,6], which shows that there is still room for improving our knowledge about convective initiation.

As it is well known throughout the weather forecasters community, one of the most widely used tools for predicting the occurrence of convection are the stability indices derived, in most cases, from the temperature and humidity data measured by the sounding (upper-air) stations. These indices, such as the lifted index [7],

[1] Department of Computer Sciences and Numerical Analysis, Universidad de Córdoba, Córdoba, Spain

[2] Department of Civil Engineering: Construction, Infrastructure and Transport, Universidad Politécnica de Madrid, Madrid, Spain

[3] LATUV, Remote Sensing Laboratory, Universidad de Valladolid, Valladolid, Spain

[4] Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, Spain

✉ Corresponding author at: Department of Computer Science and Numerical Analysis, University of Córdoba, Rabanales Campus, Albert Einstein Building 3rd Floor, 14071 Córdoba, Spain. E-mail address: E-mail: dguijo@uco.es (D. Guijo-Rubio).

the total of totals index [8] or the Convective Available Potential Energy (CAPE) Index [9] among others, try to characterize, with varying degrees of success, if the current atmospheric conditions favours convection and, eventually, the formation of cumulus congestus and cumulonimbus clouds that can bring adverse weather conditions, like turbulence, wind-shear, shower rain and, under some circumstances, hail and electric activity. Normally, a combination of several stability indices are used for improving the prognosis of the atmospheric conditions that favours convection since poor results are usually obtained when using them individually [10]. In this regard, several research works have explored the capability of these indices. For example, [11] performed an analysis of the pre-convective conditions in Argentina based on the information of 713 days of radiosonde data. Later, [12] used 16,421 proximity soundings taken at 32 central European stations to assess environments of severe and non-severe thunderstorms. More recently, [13] employed more than 45,000 soundings from Central and Western Europe to characterize severe and non-severe environments based on thermodynamic and kinematic parameters. Since there are few sounding stations located all over the world and their data are normally available only twice a day (at 00 and 12 UTC), other studies have been focused on the capacity of different numerical weather prediction models to derive reliable stability indexes based on the model outputs at different vertical levels, both with global-scale models, like [14] with the NCAR/NECP reanalysis and [15] with the ECMWF model, and mesoscale models, like [16] with the COSMO model and [17] with the WRF model.

Another research line is focused on the development of accurate products derived from satellite information to forecast the occurrence of convection. This working line assumes that there is insufficient spatial and temporal resolution of surface-station observations and numerical weather prediction models to accurately forecast the exact convective incitation location and time [18]. For example, in [19] the authors identified the precursor signals of convective initiation within sequences of 1-km-resolution visible and 4–8-km infrared imagery from the GOES instrument. Later, [20] presented a new detection scheme for convective initiation under day and night conditions based on the information provided by different channels of the Spinning Enhanced Visible and Infrared Imager (SEVIRI) on board Meteosat Second Generation (MSG) satellites. More recently, some authors have merged numerical weather prediction inputs with satellite information to produce more accurate forecasts, like [21] or [22].

In any case, a common characteristic of all problems is that convection occurrence largely depends on local topography [23]. Therefore, the development of new techniques adapted to local conditions could improve standard approaches to identify atmospheric conditions that favours convection. In this regard, state-of-the-art Machine Learning techniques could be an adequate complement, since they have largely proved their capacity to accurately predict a wide range of local atmospheric phenomenal [24], like precipitation [25,26,27,28,29], low-visibility events formation [30,31,32] or expected solar radiation [33,34,35]. Nevertheless, the number of research works dealing with convective initiation and the use of soft-computing and data mining techniques to assess atmospheric conditions that favours convection or to predict its formation is very limited [36].

A new proposal for the prediction of convection initialization (convective clouds formation prediction) based on machine learning algorithms is presented in this paper. Specifically, Artificial Neural Network models (ANNs) [37] are considered in this work, since this paradigm has been successfully applied before to several prediction problems in different fields related to atmospheric studies. When ANNs are evolved using a single metric as objective function, these models are known as Evolutionary ANNs (EANNs). In the literature, the accuracy is frequently used as objective function, trying to achieve the best possible global performance. However, note that the accuracy is not a proper measure when dealing with imbalanced datasets (as the case of convection initialization) since the model tends to classify all the patterns in the majority class. In this sense, the minimum sensitivity should be also included to optimise the sensitivities of the minority classes. The problem is that, under certain conditions, accuracy and minimum sensitivity are conflictive objectives [38]. These opposite measures can be used to evolve ANNs, i.e. two opposite measures are used to lead to the best performance satisfying the two following main goals: 1) to achieve a high performance for the global problem, i.e. a good accuracy, and 2) to reach a notable performance for the minority classes, i.e. desirable accuracies for the minority classes; these models are known as Multi-Objective Evolutionary Artificial Neural Network models (MOEANNs). Therefore, a multi-variate highly imbalanced classification problem is tackled by the application of several ANNs with different evolutionary algorithms.

Specifically, in this paper we compare five models of ANNs, which are evolved by the two different evolutionary algorithms (mono-objective, EANNs, and multi-objective, MOEANNs), trying to obtain one model achiev-

ing both goals successfully. The main contributions of this paper are: 1) the use of ANNs to improve the predictability of current convective clouds prediction systems; 2) a consistent comparison between different mono- and multi-objective ANN models, to better understand the benefits of MOEANNs; 3) a comparison of different typologies of ANNs, including pure and hybrid models of projection functions (sigmoidal units, SUs, and product units, PUs) and local functions (radial basis functions, RBFs). Moreover, these comparisons have been performed using several metrics showing their advantages and disadvantages. We finally compare the results obtained by the best evolutionary ANN proposed with those by Terminal Aerodrome Forecasts (TAF), the most used aerodrome weather forecast method. The results obtained indicate that a MOEANN with a hybrid hidden layer (SUs and RBFs) reaches to the best solution to predict the formation of convective clouds. Then, a deep analysis using this best model is presented aiming to increase the interpretability of this sort of models and showing that they are a competitive option for the prediction of convective clouds formation at 12-hours time horizon.

The rest of the paper is organized as follows: In Section 2, the dataset considered is introduced and explained, detailing the process carried out for the inclusion of reanalysis data. The methodology used to tackle the problem under study is shown in Section 3, including a brief introduction to ANNs (Section 3.1) and describing the evolutionary algorithms used to evolve the ANNs (Section 3.2). The experiments and results achieved in the detection of convective situations over Madrid-Barajas Airport are shown and discussed in Sections 4 and 5, respectively. Finally, Section 6 closes the paper with some final remarks.

2 Data description

Our study is carried out at the Adolfo-Suárez Madrid-Barajas International Airport (ICAO indicator: LEMD), a facility specially sensitive to convective clouds and thunderstorms. It covers the period from 2011 to 2015 (5 years). Following a classical approach, we have used upper-air data measured at the Madrid-Barajas radiosonde station, which is located in the vicinity of the airport (40° 50'N, 3° 58'W, altitude 633 m) and is part of the National Upper-Air Network managed by the Meteorological State Agency of Spain (AEMET). Fortunately, daily data from the past several years are available on the Internet [1].

[1] http://weather.uwyo.edu/upperair/sounding.html

The selection of the specific group of sounding-derived variables and stability indices for this work was set following a two-step procedure: first, we eliminated those stability indices with a low rate of availability (threshold considered: at least 90% of the data must be available). Secondly, those days where sounding-derived variables were not available were removed. In this sense, our study can be considered as traversal instead of longitudinal, keeping a total of 577 days out of 610 days. Later, we analysed the degree of correlation between the different sounding-derived variables. The correlation test was applied to each single variable measured at the following standard levels: 925, 850, 700, 500, 400, 300 and 200 hPa. Considering the high linear correlations obtained for these variables (greater than 0.90), we calculated the mean and standard deviation coefficients of each variable measured in the vertical axis (i.e. its vertical mean value and standard deviation), thus obtaining 20 new sounding-derived variables. Statistically-derived variables were shown to be less correlated than each single sounding-derived variable and, consequently, more appropriate for the purpose of our study. As a result, we finally chose the group of variables and stability indices shown in Table 1.

On the other hand, valuable information regarding convective environmental conditions can also be obtained from numerical weather prediction models [12]. Consequently, we have also considered several outputs from the ECMWF ERA-Interim reanalysis [39]. Specifically, we have selected all the surface and vertical variables of the grid point nearest to the locations of the sounding station that were similar to the Madrid-Barajas sounding ones. Regarding the vertical variables, we only extracted those corresponding to the same fixed standard levels (925, 850, 700, 500, 400, 300 and 200 hPa) and we also calculated statistically-derived variables as explained before. Finally, we applied the same correlation test (availability tests are not necessary for reanalysis data). As a result, we obtained the list of reanalysis data shown in Table 2.

Information about the presence of convective clouds has been obtained from the Madrid-Barajas METAR and SPECI reports. These meteorological products, issued on a routine basis (METAR) or when necessary (SPECI) are highly reliable, since they were designed to deliver up-to-date information for in-flight aircraft [40]. Specifically, when Cumulonimbus (coded as CB) or Cumulus Congestus (coded as TCU) clouds are detected around the aerodrome, they are included in the METAR or SPECI report, as appropriate. Additionally, when thunderstorms (including thunderstorms in the vicinity) are detected (coded as TS), they are also reported. Therefore, cloud information contained in METAR

Table 1: Variables and indices derived from the Madrid-Barajas radiosonde station.

Mean value and standard deviation coefficient over the vertical	units
Geopotential Height	[m]
Dewpoint Temperature	[Celsius]
Relative Humidity	[%]
Mixing Ratio	[g/kg]
Wind Direction	[degrees true]
Wind Speed	[Knot]
Standard deviation coefficient over the vertical	**units**
Air Temperature	[Celsius]
Potential Temperature	[Kelvin]
Equivalent Potential Temperature	[Kelvin]
Virtual Potential Temperature	[Kelvin]
Stability Indices	**units**
SWEAT	[K]
K	[K]
Cross Totals	[K]
Vertical Totals	[K]
Totals of Totals	[K]
Temp. at Lifted Condensation Level	[K]
Pressure at Lifted Condensation Level	[hPa]
Mean Mixed Layer Mixing Ratio	[g/kg]

Table 2: Variables derived from the ECMWF ERA-Interim reanalysis.

Mean value and standard deviation coefficient over the vertical	units
Potential Vorticity	[K kg^{-1} m^2 s^{-1}]
U-wind Component	[m/s]
V-wind Component	[m/s]
W-wind Component	[m/s]
Relative Vorticity	[s^{-1}]
Variables at surface level	**units**
Total Cloud Cover	[0-1]
U-wind Component (at 10 meters)	[m/s]
V-wind Component (at 10 meters)	[m/s]
Air Temperature (at 2 meters)	[Celsius]

and SPECI reports can be used as objective variable. For the purpose of this study, we have generated four categories obtained from the METAR and SPECI reports. In this way, we are able to tackle a classification problem. It is important to note that our study aims to classify the convective conditions expected in the period of time between 12:00 and 0:00 UTC each day, i.e., is a 12-hours prediction horizon classification problem. Accordingly, we only consider the 12:00 UTC Madrid-Barajas radiosonde and ECMWF ERA-Interim reanalysis data. The four categories are as follows:

- CLEAR: no convective clouds sighted in the next 12 hours.
- TCU: cumulus congestus clouds sighted in the next 12 hours and no CB nor TS phenomena.
- CB: cumulonimbus clouds sighted at least one out of the next 12 hours; cumulus congestus clouds could be sighted in the next 12 hours, but no thunderstorms.
- TS: thunderstorm sighted at least one out of the next 12 hours; cumulonimbus and cumulus congestus clouds could be sighted in the next 12 hours.

Table 3: Distribution of the different types of events for the dataset considered.

Class of day (C_k)	Training	Test
CLEAR	240 (68.18%)	82 (69.49%)
TCU	58 (16.48%)	20 (16.95%)
CB	28 (7.95%)	9 (7.63%)
TS	26 (7.39%)	7 (5.93%)
Total	352 (74.89%)	118 (25.11%)

In this way, the set of labels of the dataset is $Y = \{C_1, C_2, C_3, C_4\}$, where: C_1 represents CLEAR conditions, C_2 represents TCU conditions, C_3 represents CB conditions, and finally, C_4 represents TS conditions. According to the variables exposed in Tables 1 and 2, a set of 38 input variables are used merging data from the radiosonde station and from reanalysis.

Since the study is traversal, the original dataset is divided by means of a stratified-randomised procedure into two different datasets, obtaining approximately a 80% of the patterns for the training set, whereas the remaining 20% is used for the generalisation set, following the guidelines given by Prechelt in [41]. After that, over the training set, an undersampling procedure has been performed removing a 30% of the patterns belonging to the majority class CLEAR. The main goal for the undersampling is reducing the degree of imbalance of the training set, and thus, to classify the minority classes better. Once this procedure has been performed, the class distribution per set is shown in Table 3, in such a way that the final split considers, approximately, a 75% of patterns for training and the remaining 25% for the test set.

As can be checked in Table 3, the dataset is still imbalanced. According to [42], the Imbalance Ratio (IR)

is defined as:

$$IR = \frac{1}{J}\sum_{j=1}^{J} IR_j, \;\; IR_j = \frac{\sum_{q \neq j} N_q}{J \cdot N_j}, \tag{1}$$

where IR_j is the IR for the class j, N_j is the number of patterns belonging to class C_j, in this paper $C_j \in \{\text{CLEAR}, \text{TCU}, \text{CB}, \text{TS}\}$, and $J = 4$ is the number of classes. This dataset has an IR of 1.902, denoting that it is highly imbalanced. Hence, the classification of the minority classes is challenging in this problem, due to the small number of cases.

3 Methodology

In this section, the ANN models considered are introduced with the different kind of basis functions used in this paper. Furthermore, the evolutionary algorithms used for evolving the ANNs (EANNs and MOEANNs for mono-objective and multi-objective ANNs, respectively) are also described.

3.1 Artificial Neural Networks

ANNs [37] are a widely used kind of non-linear model due to its potential to solve a vast range of problems. Feed-forward neural networks were one of the first proposals made in the literature; they are composed by at least one input layer, one hidden layer and one output layer, using in the hidden layer different sorts of basis functions [37], with diverse goals depending on the problem tackled. Furthermore, ANNs have been proven to be universal approximators. Regardless the hidden layer type, the ANN output can be written as:

$$f_q(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta}) = \beta_{q0} + \sum_{j=1}^{m} \beta_{qj} B_j(\mathbf{x}, \mathbf{w}_j), \quad q = 1, \ldots, J-1, \tag{2}$$

where $B_j(\mathbf{x}, \mathbf{w}_j)$ represents the set of non-linear transformations of the input vector $\mathbf{x^T} = (x_1, x_2, \ldots, x_d)$, with d being its length; bias is considered with the element $B_0(\mathbf{x}, \mathbf{w}_j) = 1$; $\boldsymbol{\beta}_q^T = (\beta_{q1}, \beta_{q2}, \ldots, \beta_{qm})$ are the coefficients from linear combination estimated from the data; $\mathbf{w}_j^T = (w_{j1}, w_{j2}, \ldots, w_{jd})$ are the parameters related to the basis function; m is the number of basis functions required to minimise some definite error function; J is the number of outputs of the problem. The outputs are transformed into probabilities by considering the softmax transformation:

$$p_q(\mathbf{x}) = \frac{\exp f_q(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta})}{\sum_{j=1}^{J} \exp f_j(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta})}, \; q = 1, 2, \cdot, J, \tag{3}$$

where, given that all probabilities have to sum 1, $f_J(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta}) = 0$.

For this problem, as it has been described in Section 2, the input layer is composed of 38 neurons, one for each input variable, whereas the output layer is composed of 4 neurons, one for each sort of day (CLEAR, TCU, CB and TS). Regarding the hidden layer, we have focused on the following three basis functions being the most commonly applied in the state-of-the-art:

- Sigmoidal Unit (SU) [43] is the most common basis function due to its ability to approximate any continuous function accurately. However, they fail in local optimum frequently. Using the notation described above, SU is represented as:

$$B_j(\mathbf{x}, \mathbf{w}_j) = \frac{1}{1 + e^{-\left(w_{j0} + \sum_{i=1}^{d} w_{ji} x_i\right)}}, \qquad j = 1, ..., m. \tag{4}$$

- Product Unit (PU) [44] is a basis function that not only it is able to retain the properties of a universal approximator, but also, it only uses a small number of multiplicative neurons [45]. According to the notation followed so far, the PU is expressed as:

$$B_j(\mathbf{x}, \mathbf{w}_j) = \prod_{i=1}^{d} x_i^{w_{ji}}, \qquad j = 1, \ldots, m. \tag{5}$$

- Radial Basis Function (RBF) [46] behaves as different local elements for each hidden neuron, activating a different set of units for every pattern. In this way, the number of local optimum is decreased, making the training stage easier. Regarding the notation exposed above, RBF is represented as:

$$B_j(\mathbf{x}, \mathbf{w}_j) = e^{-\frac{1}{2}\left(\frac{\sum_{i=1}^{d}(x_i - c_{ji})^2}{r_j}\right)}, \qquad j = 1, \ldots, m \tag{6}$$

  where the vector of weights of the j-th hidden neuron, $\mathbf{w}_j$, includes both a centroid $\mathbf{c}_j$ and a radium r_j for the corresponding Gaussian basis function, in such a way that $\mathbf{w}_j = \{r_j, \mathbf{c}_j\}$.

Moreover, in this work, we have considered a mixture of basis functions for the hidden layer, i. e. a linear combination of two types of basis functions is used. For these models, the general ANN can be written in the following way:

$$f_q(\mathbf{x}, \boldsymbol{\theta}) = \beta_{q0} + \sum_{j=1}^{m_1} \beta_{qj}^1 B_j^1(\mathbf{x}, \mathbf{w}_j^1) + \sum_{j=1}^{m_2} \beta_{qj}^2 B_j^2(\mathbf{x}, \mathbf{w}_j^2), \qquad q = 1, 2, \ldots, J-1, \tag{7}$$

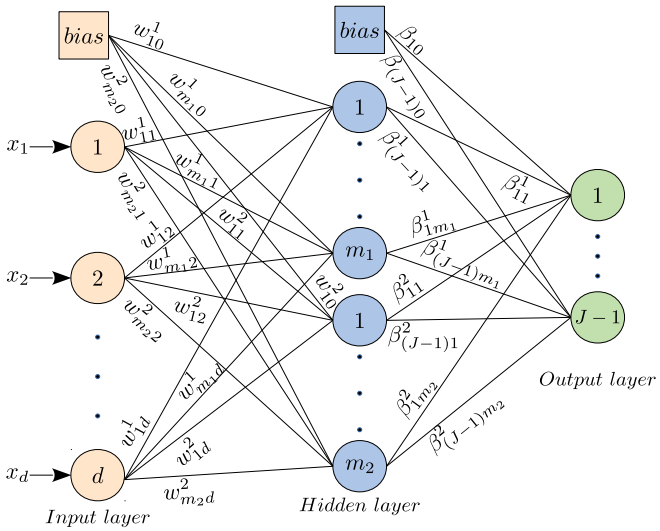


Fig. 1: Hybrid ANN structure. Biases are represented by squares whereas neurons are shown by circles.

where m_1 is the number of hidden neurons of the first type, $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \mathbf{w}_1^1, \ldots, \mathbf{w}_{m_1}^1, \mathbf{w}_1^2, \ldots, \mathbf{w}_{m_2}^2, \}$ is the vector containing all the coefficients of the neural network, $\boldsymbol{\beta} = \{\beta_{q0}, \beta_{q1}^1, \ldots, \beta_{qm_1}^1, \beta_{q1}^2, \ldots, \beta_{qm_2}^2\}$ includes the coefficients between hidden and output layers, and $\mathbf{w}_j^1$ and $\mathbf{w}_j^2$ represents the weights connecting the input layer to the j-th hidden neuron of the first and the second types, respectively. Any of the basis functions previously defined (Equations (4), (5) or (6)) can be used for $B_j^1(\mathbf{x}, \mathbf{w}_j^1)$ and $B_j^2(\mathbf{x}, \mathbf{w}_j^2)$. The general structure of the hybrid ANN is shown in Figure 1.

The combination of different basis functions in the hidden layer has many advantages, such as providing flexible decision rules. Concretely, Donoho *et al.* [47] demonstrated that any continuous function could be decomposed into two different kinds of functions: one associated to SUs or PUs (projection functions) and the other one associated to RBFs (kernel functions). In this way, the hybrid models considered in this paper are two: 1) PU and RBF, and 2) SU and RBF, also known as PURBF and SURBF, respectively.

3.2 Evolutionary Algorithms

Traditionally, ANNs have been evolved by methods that usually fall in local optima, such as backpropagation. Attempting to find a robust search method able to find zones in the search space where good solutions can be found, Evolutionary Artificial Neural Network models (EANNs) were proposed [48]. The main advantage of the EANNs compared to traditional methods is not only their small computational cost but also their robustness to find competitive solutions.

The evolutionary algorithms considered are similar to those described in [49,50], in order to get the structure and estimate the coefficients of the models. These algorithms have proved to be very efficient independently of the initial conditions, reaching good solutions and not requiring the use of gradients. Considering the difficulty of establishing an appropriate architecture and the tuning of the weights, the use of EANNs is justified. In this way, the algorithms presented in this section will be focused to establish a number of hidden neurons for the different neural networks (m for pure networks and m_1 and m_2 for hybrid ones), together with a real value for the synaptic weights of all kind of models (i.e. for the vectors of hidden neuron weights, $\mathbf{w}_j$ or $\mathbf{w}_j^1$ and $\mathbf{w}_j^2$, and for the vector of output weights, $\boldsymbol{\beta}$). Moreover, some of the weights will be deactivated by setting their weights to zero. The error surface associated to such a network optimization is extremely complex, and this justifies the use of a metaheuristic.

EANNs evolve the architecture and weights of the ANN considering a single performance measure (also known as mono-objective optimisation). In this paper, we have considered the following measures:

- The Correct Classification Rate (CCR), which is also known as accuracy, is the sum of all the patterns correctly classified divided by the total number of instances. It is given by the following expression:

$$CCR = \frac{100}{N} \sum_{j=1}^{J} n_{jj}, \tag{8}$$

  where N is the total number of patterns, J is the number of classes and n_{jj} is the number of correctly classified instances for class j. This metric varies between 0 and 100, i.e. from none of the patterns being correctly classified to having a classifier without misclassified patterns.
- The Minimum Sensitivity (MS) [38] measures how the minority classes are classified. It is given by the expression:

$$MS = 100 \cdot min(\frac{n_1}{N_1}, \frac{n_2}{N_2}, \ldots, \frac{n_J}{N_J}) \tag{9}$$

  where the fraction $\frac{n_j}{N_j}$ is the sensitivity, n_j is the number of correctly classified instances and N_j is the total number of instances, all belonging to class j. This metric varies between 0 and 100, where 0 means that at least one class gets totally misclassified, whereas 100 indicates that all the patterns have been correctly classified.

– The Area Under the ROC Curve (AUC) [51] is a measure that represents the ability of the model to distinguish between classes. Since the AUC is a measure used for binary problems, in the case of multi-class problems, it is averaged from all the binary decomposed problems. It varies between 0 and 100, being 0 a model that misclassified all the patterns and 100 a model able to distinguish between classes successfully.

When EANNs are evolved by using CCR as objective function, most of the patterns are classified into the majority class, in this case, in the CLEAR class, maximising the value for this metric. On the other hand, when EANNs are evolved by the MS, the model tries to classify the patterns in the minority classes (CB and TS in this paper), disregarding the classification of the majority class (CLEAR).

In order to solve this problem and maximise both measures, Multi-Objective Evolutionary Artificial Neural Network models (MOEANNs) [52,53] are used, since they allow the evolution of ANN based on the optimisation of two measures, giving the same priority to both. MOEANNs obtain a set of similar optimal solutions, named Pareto front [54], where all the solutions are considered equal (there is a balance between both measures for each solution). The main goal of MOEANNs using these two measures is to obtain a good global performance without misclassifying the minority classes, i.e. obtain the highest accuracy per class possible. For imbalanced datasets, as the one considered in this paper, the MOEANN methodology tries to increase the accuracy of the worst classified class (the one that limits the MS), improving both, the CCR and the MS.

In this paper, five different architectures of ANN are proposed (PU, SU, RBF, PURBF and SURBF), optimised by means of both mono-objective (EANNs) and multi-objective (MOEANNs) evolutionary algorithms, using the CCR and MS as objective functions. Note that instead of using the CCR to evolve the ANN, our algorithms use the cross-entropy function which is related to the maximization of the CCR, but it makes the convergence more robust due to it is a continuous function; the use of the CCR for the results is due to its popularity in the state-of-the-art [55]. The cross-entropy (E_{min}) is an error function defined as:

$$E_{min} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{j=1}^{J} I(y_n = j)\ \log p_j(\mathbf{x}_n), \tag{10}$$

where $I(\cdot)$ is the zero-one loss function, y_n is the label of pattern n and $p_j(\mathbf{x}_n)$ is the probability that the pattern n belongs to class j (Eq. (3)). To perform maximisation, the following transformation is applied:

$$E = \frac{100}{1 + E_{min}}, \tag{11}$$

in this way, the cross-entropy (E) is a maximisation function varying between 0 and 100.

A brief summary of the whole process is shown in Figure 2. For each architecture, the EANN (Pseudocode 1) is run twice, the first one optimising the CCR, and the second one optimising the MS. Then, the best model for each case is applied to the generalisation dataset. On the other hand, the MOEANNs (Pseudocode 2) is run once for each kind of ANN proposed. In this case, the algorithm optimises the ANNs using both the CCR and the MS. Once the algorithm returns the first Pareto front, the two extremes of the front are selected: the upper-most individual of the Pareto front is the best in terms of cross-entropy and CCR, whereas the bottom-most individual of the Pareto front is the best in terms of MS. The best individual in terms of cross-entropy and CCR is associated with the maximisation of the global performance, whereas the best individual regarding MS is associated with the correct classification of the minority class.

Algorithm 1: EANN

```
generate a random population p
while stopping criteria is not satisfied do
    evaluate and rank the individuals using
      cross-entropy or MS
    keep the best individual regarding cross-entropy
      or MS
    for worst 10% of p do
        replace with the best 10% of p
    end
    parametric mutation to the best 10% of p
    structural mutation to the remaining 90% of p
    evaluate and rank the individuals using
      cross-entropy or MS
    the worst individual is replaced by the best
      individual of p stored previously
end
return the best individual using cross-entropy or MS
```

Although CCR and MS seem to be positively correlated (i.e. if one increases, the other one should increase), this only happens for balanced datasets with a small number of classes. However, for imbalanced problems, once the CCR and MS reach to a certain level (trying to achieve the optimum value of 100), both objectives become very competitive, and when an objective improves, it leads to a decrease in the other one. For instance, an improvement in CCR, not necessarily improves the MS (the majority class improves its CCR, decreasing the CCR of the minority classes), as well

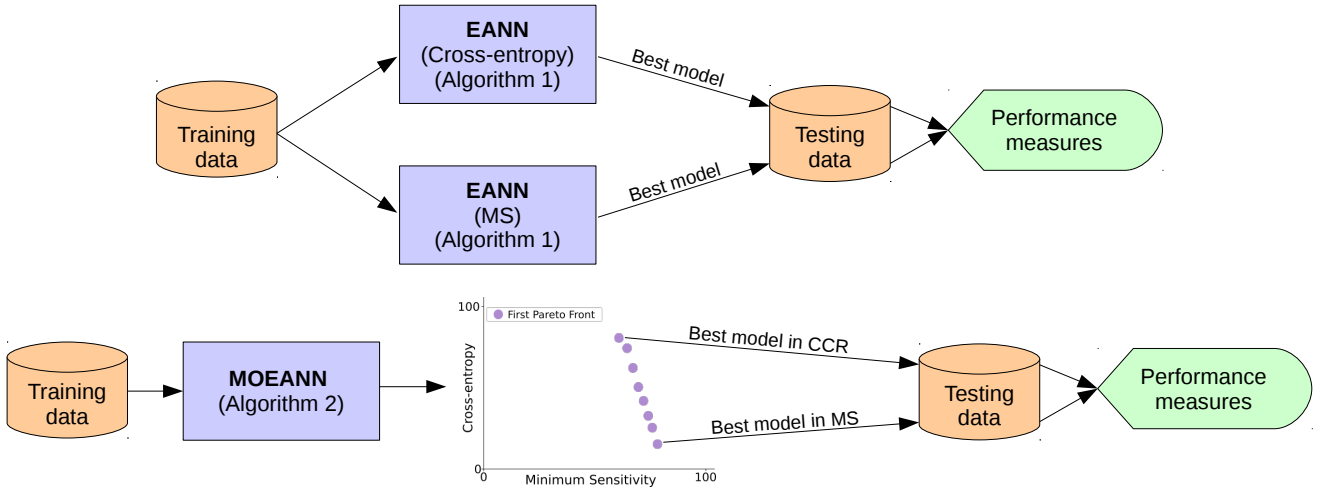


Fig. 2: A brief summary of the whole process. See Algorithms 1 and 2 for more information about the EANNs and MOEANNs, respectively.

Algorithm 2: MOEANN

generate a random population p
evaluate and rank the individuals according to Pareto front strategy
while *stopping criteria is not satisfied* **do**
  tournament selection to choose p individuals for the mutations
  random parametric and structural mutations (offspring of size p: population q)
  parents and mutated offspring $(p+q)$
  evaluate and rank the individuals
  keep the p best individuals
end
return *the first Pareto front*

as, an increase in terms of MS could not correspond with an increase in CCR (the sensitivity of the minority class increases, while the CCR of the majority class decreases); thus, the use of the multi-objective methodology is justified since the primary objective is reaching the highest MS without disregarding the global performance.

4 Experiments and results

This section shows the experimental settings and shows the results obtained using the dataset and methodology described in Sections 2 and 3, respectively.

4.1 Experimental settings

As described in Section 2, the original dataset is shuffled and divided using a 75% for training (352 instances) and the remaining 25% for testing (118 instances), preserving the percentage of patterns belonging to each class.

The mono-objective EANN and the multi-objective MOEANN algorithms have been applied using the following parameters: both algorithms have been run 30 times using 320 generations and a population size of 1000 and 100 for the EANN and the MOEANN, respectively. Furthermore, the number of hidden nodes for the PU is initialised in the range $[1, 3]$, due to its simplicity; in the case of SU, this number is initialised in the range $[1, 4]$; the RBF and the hybrid models need a higher number of hidden nodes, thus, they vary in the range $[2, 4]$[2].

Previously to the application of the ANN models, all the input variables have been scaled to the range $[0.1, 0.9]$ in order to avoid values close to 0 which produces large values in case of negative exponents for PU and saturation issues for SU.

4.2 Results

The results obtained for the experimental settings previously described in Section 4.1 are shown in Table 4. All the results are expressed as the mean and standard deviation of the 30 executions for each algorithm and ANN topology. Table 4 is divided horizontally into four parts: the first two contain the mono-objective methodology (EANN), whereas the last two contain the multi-objective methodology (MOEANN). Specifically, considering the mono-objective approaches, the first row

[2] Further information of the parameters considered can be found in [38,56,57], whereas, more information regarding the ANNs can be obtained from [37].

contains those ANNs optimised using the cross-entropy (note that we use the term CCR for notation purposes, as both measures are highly correlated and CCR is commonest in the state-of-the-art), whereas the second row shows the same models being optimised by the MS. Regarding the multi-objective approaches, the MOEANN-CCR represents the model of the Pareto front with best value in terms of CCR, whereas, MOEANN-MS refers to the model of the Pareto Front with the highest value regarding MS. The different models are named by the activation function used, i.e. SURBF is an ANN with a hybrid layer combining Sigmoidal Unit (SU) and Radial Basis Functions (RBF). To improve its reading, the notation for the mono-objectives and multi-objectives algorithms is $EANN-XX-YY$ and $MOEANN-XX-YY$, respectively, where XX represents the objective function (CCR or MS) to optimise, in the case of mono-objective methods, and the extreme of the first Pareto front, in the case of multi-objective algorithms, and YY is the name of the basis function used (PU, SU, RBF, PURBF, SURBF).

As can be seen, the best results for each metric are obtained by different models: in the case of CCR, the highest value is obtained by EANN-CCR-SU (73.870 ± 1.667), in the case of MS, the best performance is achieved by the hybrid MOEANN-MS-SURBF architecture (20.508 ± 11.072), the best AUC is also reached by EANN-CCR-SURBF (88.938 ± 1.144), and the simplest model (i.e. the model with the lowest number of links) is EANN-CCR-PU (42.900 ± 10.899). Apart from denoting the difficulty of the problem, these results show that the different objectives are conflictive, thus, the use of a multi-objective methodology is straightforward; for instance, if a model is optimised using the CCR, the classification of the minority classes is ignored whereas a higher priority is given the global performance of the model. Otherwise, models optimised by the MS prioritise obtaining a better performance for the minority classes than for the majority ones. As can be checked in Table 4, especially on those methods optimised by MS, the multi-objective methodology improves the performance on the opposite measures, this is, both the CCR and AUC are improved using MOEANN-MS.

Considering the difficulty of the problem, it is not possible to obtain a single EANN model reaching good performances for all the metrics used. As described in Section 2, the dataset is divided into four classes, depending on the convective situation. Furthermore, from this point of view, two different strategies can be followed: 1) the main goal is to get a classifier reaching competitive global performance for the four classes, and 2) the primary purpose is to prioritise the minority classes, due to they are the ones that could cause trouble to the proper operation of the flights. In this way, we can stand out two ANN models from the results shown in Table 4, the EANN-CCR-SURBF, and the MOEANN-MS-SURBF, for the first and second strategy, respectively (to ease its visualisation, both classifiers have been shaded in the Table). The EANN-CCR-SURBF model obtains the best performance in terms of AUC (88.938 ± 1.144) and the second best performance in terms of CCR (73.503 ± 1.723), where the difference with the best one is not high (73.870 ± 1.667). On the other hand, the MOEANN-MS-SURBF model using MS extreme reaches to the highest value in terms of MS (20.508 ± 11.072), as can be seen, not only the CCR and AUC are improved regarding the same model with mono-objective methodology, but also, the MS obtained is higher, making this model very competitive.

Regarding the number of links, it varies in the range $[42.900, 85.266]$, which is highly dependent on the activation function used. Moreover, those algorithms using a multi-objective methodology needs a higher number of links than the mono-objective versions.

In Table 5, the results for the best model of each ANN are shown. The criterion for selecting the best model from the 30 runs is obtaining the highest MS. If there are ties, we check AUC and, there are still ties, we consider CCR. This decision was made given the difficulty of correctly classifying the minority classes.

As can be checked in Table 5, the highest value in terms of accuracy is obtained by MOEANN-CCR-SURBF (76.271). Furthermore, the best MS is reached by EANN-MS-PU model evolved by MS (44.444), whereas the best AUC is achieved by the EANN-CCR-SU model (90.608). Finally, the most simple model is EANN-CCR-PU (40 links).

Regarding the strategies previously introduced, the main interest is to find a single model able to achieve an acceptable performance for both goals. In this way, although the MOEANN-MS-SURBF model is not able to achieve the best result for any of the three metrics considered, we propose it as the best model, due to the balance between the different objectives. Specifically, in terms of CCR, it obtains a value of 74.576, whereas the best value is 76.271 and, in terms of AUC, it achieves 88.676, whereas the highest value is 90.608. Regarding MS, the best value is 44.444, whereas the one obtained by this model is 42.857. It can be concluded that the multi-objective methodology is highly successful for this problem because it can achieve competitive values for all the metrics, considering that the metrics are conflictive. Summarising, the difference in terms of CCR and AUC is small compared with the best value for both metrics, and, in terms of MS, this model can achieve

Table 4: Mean and standard deviation ($mean \pm sd$) values of CCR, MS, AUC and number of links (#links) of the 30 executions of the models proposed for both methodologies, all measured in the generalisation set. The best result is highlighted in **bold**; the second best result is shown in *italics*.

	Name	CCR	MS	AUC	#links
EANN-CCR	PU	72.062 ± 1.759	0.847 ± 3.249	84.396 ± 6.727	$\mathbf{42.900 \pm 10.899}$
	SU	$\mathbf{73.870 \pm 1.667}$	4.233 ± 6.159	88.300 ± 1.876	82.033 ± 13.828
	RBF	72.062 ± 1.612	0.370 ± 2.029	$\mathit{88.857 \pm 0.937}$	76.633 ± 12.979
	PURBF	72.147 ± 1.193	0.000 ± 0.000	87.023 ± 2.774	69.900 ± 14.079
	SURBF	$\mathit{73.503 \pm 1.723}$	0.476 ± 2.608	$\mathbf{88.938 \pm 1.144}$	75.700 ± 13.473
EANN-MS	PU	44.463 ± 5.849	$\mathit{19.262 \pm 10.637}$	62.659 ± 9.057	52.467 ± 9.537
	SU	40.734 ± 7.687	15.405 ± 8.646	60.940 ± 7.417	67.000 ± 17.834
	RBF	42.034 ± 7.085	17.438 ± 12.886	65.142 ± 7.859	$\mathit{46.600 \pm 15.843}$
	PURBF	41.610 ± 5.282	16.068 ± 10.708	62.051 ± 6.524	66.733 ± 18.260
	SURBF	40.876 ± 7.152	17.217 ± 10.496	61.919 ± 6.633	58.633 ± 18.277
MOEANN-CCR	PU	70.056 ± 1.051	0.166 ± 0.912	70.436 ± 9.983	59.200 ± 11.868
	SU	72.062 ± 1.787	0.333 ± 1.826	82.886 ± 4.243	83.800 ± 9.513
	RBF	71.638 ± 1.691	0.704 ± 2.339	86.944 ± 1.299	68.966 ± 17.058
	PURBF	71.780 ± 1.638	0.370 ± 2.029	85.896 ± 3.195	69.066 ± 11.316
	SURBF	72.514 ± 1.748	1.347 ± 3.705	86.599 ± 2.103	78.266 ± 10.913
MOEANN-MS	PU	40.508 ± 13.503	15.753 ± 11.003	61.433 ± 9.718	55.366 ± 12.081
	SU	50.367 ± 10.217	18.272 ± 10.790	70.462 ± 7.880	85.266 ± 8.808
	RBF	59.463 ± 7.850	13.913 ± 9.726	75.928 ± 6.305	70.933 ± 22.190
	PURBF	53.644 ± 11.906	16.958 ± 11.477	71.117 ± 9.321	72.066 ± 17.071
	SURBF	56.780 ± 8.999	$\mathbf{20.508 \pm 11.072}$	74.931 ± 5.020	77.133 ± 13.302

Table 5: Values of CCR, MS, AUC and number of links (#links) of the best neural network models obtained, all measured in the generalisation set. The best result is highlighted in **bold**; the second best result is shown in *italics*.

	Name	CCR	MS	AUC	#links
EANN-CCR	PU	72.881	14.286	89.206	**40**
	SU	*75.424*	14.286	**90.608**	50
	RBF	74.576	11.111	*90.322*	87
	PURBF	72.881	0.000	89.739	79
	SURBF	72.881	14.286	89.043	76
EANN-MS	PU	50.000	**44.444**	70.128	67
	SU	49.153	35.000	65.351	79
	RBF	59.322	*42.857*	75.517	81
	PURBF	43.220	42.683	72.236	70
	SURBF	45.763	33.333	66.078	47
MOEANN-CCR	PU	72.881	0.000	84.302	*45*
	SU	*75.424*	0.000	85.251	81
	RBF	*75.424*	0.000	89.190	76
	PURBF	*75.424*	0.000	88.008	66
	SURBF	**76.271**	0.000	86.275	56
MOEANN-MS	PU	39.830	39.024	69.250	62
	SU	50.000	40.000	73.476	78
	RBF	71.186	33.333	81.214	57
	PURBF	43.220	33.333	76.245	45
	SURBF	74.576	*42.857*	88.676	72

the second best value. For visualisation purposes, this model has been shaded in Table 5.

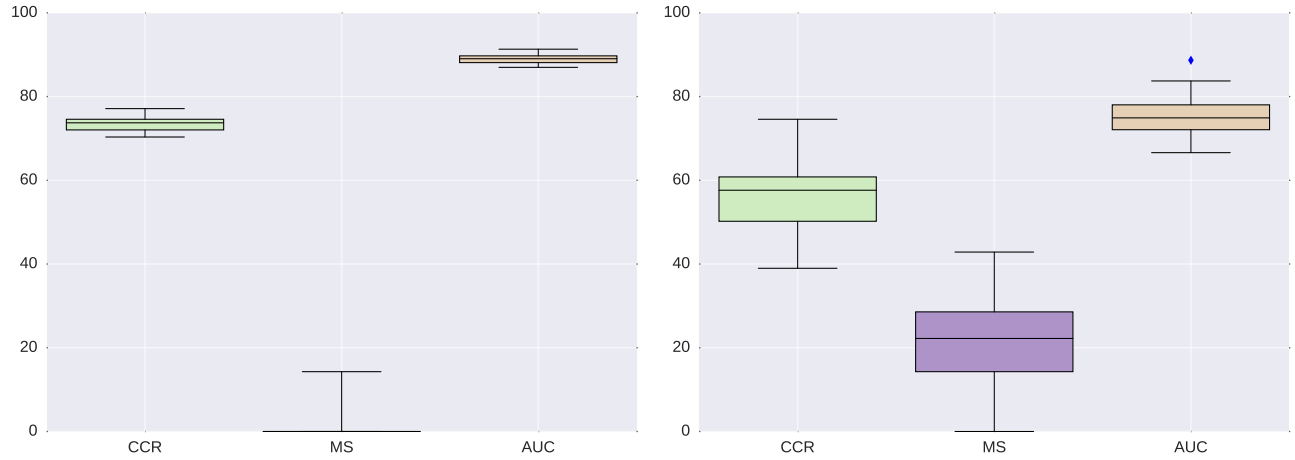
5 Discussion

Regarding the best two algorithms shaded in Table 4, the distribution of the results for the 30 executions can be visually observed in Figure 3(a) for the mono-objective EANN-CCR-SURBF method, and in Figure 3(b) for the multi-objective MOEANN-MS-SURBF method. As observed, mono-objective method seems to be more stable at the cost of obtaining lower values for MS, i.e. the classes corresponding to minority categories are not correctly classified.

In Figure 4, the Pareto front of one of the 30 runs carried out for the MOEANN-MS-SURBF is shown. The purple circles represent the individuals of the first Pareto front obtained in the last generation, whereas, the remaining dominated solutions are represented by green crosses. From the plot, the Pareto front is wide, what is a sign that the multi-objective algorithm is correctly covering many different solutions.

Following the strategies exposed in the previous Section 4.2, the multi-objective MOEANN-MS-SURBF model attending to the MS extreme of the Pareto front is analysed. In this way, its confusion matrix is presented in Table 6.

As can be seen in the confusion matrix of Table 6, the minority classes are classified with acceptable accuracy. For the CB and TS classes, all the patterns are labelled with classes representing convective situations,



(a) EANN-CCR-SURBF model (b) MOEANN-MS-SURBF model

Fig. 3: Box-plot obtained from the 30 runs of each model.

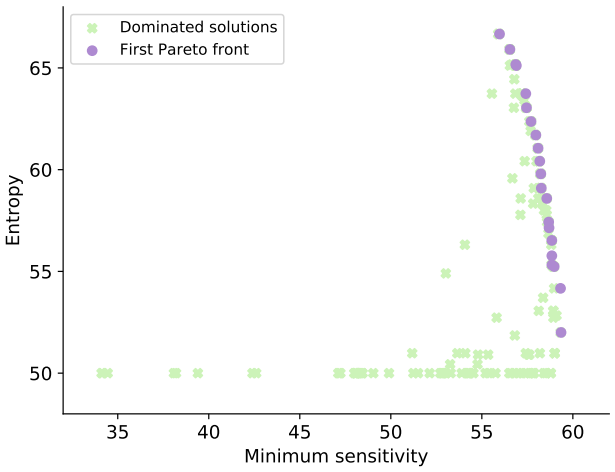


Fig. 4: The Pareto front obtained for the MOEANN-MS-SURBF model. The purple circles represent the individuals of the first Pareto, whereas the remaining dominated solutions are shown by green crosses.

Table 6: Test confusion matrix obtained by MOEANN-MS-SURBF.

	Prediction			
Target	CLEAR	TCU	CB	TS
CLEAR	71	8	2	1
TCU	5	10	4	1
CB	0	4	4	1
TS	0	3	1	3

which means that the classifier can distinguish between days with convective situations and clear ones. Regarding the class TCU, the majority of the instances are correctly classified. Finally, although the model obtained is focused on maximising MS, the CLEAR class is well classified. From this confusion matrix, the balance between achieving a good classification of the minority and majority classes can be seen.

5.1 Comparison with aerodrome forecasts

To test how well the best methodology proposed (MOEANN-MS-SURBF algorithm) is performing, we have compared the results obtained against the most used aerodrome weather forecast: the Terminal Aerodrome Forecasts. This product, known as TAF, is issued at fixed times, and consists of a concise statement of the expected meteorological conditions around an airport for a specified period. Its specifications are regulated in the Annex 3 to the Convention on International Civil Aviation [40] and they are given in a coded format.

To describe significant changes during the forecast period, the TAF is subdivided into one or more smaller time segments preceded by the keyword TEMPO or BECMG. TEMPO is used when temporary fluctuations in weather conditions are expected to occur during the forecast period, whereas BECMG is used when a more or less uniform change is expected. In addition, when there is a significant degree of uncertainty associated with the forecast, the abbreviation “PROB” followed by the probability value (30 or 40 per cent, as appropriate) is used. PROB groups can be used alone or placed before the change indicator “TEMPO”. As can

be seen, TAF products can contains deterministic and probabilistic statements. Therefore, it is necessary to lay down a methodology to compare the quality of our results against TAF reports. In order to make a meaningful comparison we decided to choose the most advantageous situation for TAF forecasts in the following sense: if the forecast considers the "type of day" subsequently verified by the METAR or SPECI reports, even with the lowest degree of probability (PROB30), we consider that the TAF succeeded. On the basis of this premise, we obtained the TAF confusion matrix shown in Table 7.

Table 7: Test confusion matrix obtained by the TAF product.

	Prediction			
Target	CLEAR	TCU	CB	TS
CLEAR	66	15	1	0
TCU	0	18	2	0
CB	1	3	4	1
TS	1	1	0	5

A comparative assessment between Table 6 and Table 7 in terms of the proportion of occurrences that were correctly classified can be carried out.

First, note that for the CLEAR category (the majority class), both methods obtain comparable results, but better in the case of the proposed MOEANN-MS-SURBF classification method (71 over 82) than in the case of the TAF reports (66 over 82). For the CB category, TAF and the MOEANN-MS-SURBF algorithm obtains the same proportion of correct classifications (4 over 9). The proportion of hits in the TS class is similar, but better in the case of the TAF reports (5 over 7) than for the MOEANN-MS-SURBF algorithm (3 over 7). Finally, for the Cumulus Congestus category (TCU), the TAF product (18 over 20) clearly provides better results than the proposed MOEANN-MS-SURBF classification method (10 over 20). Here, the Aeronautical Meteorological Forecaster's skill and the deep knowledge of the local conditions (from an atmospherical and topographical point of view) is critical to the success of the forecast. The reason is that some of these local conditions favour the development of the convective clouds of importance for aviation which are an intermediate stage between Cumulus Mediocris clouds (CLEAR category) and Cumulonimbus clouds (CB category).

6 Conclusions

In this paper, we have analysed the performance of different evolutionary artificial neural networks in a problem of convective cloud prediction. The analysis of the performance has been carried out with real data from the Adolfo-Suárez Madrid-Barajas airport (Madrid), where predictive variables have been obtained from radiosonde and reanalysis data, and the objective values (convective clouds formation at 12-hours time horizon) from METAR and SPECI information at the airport. The final problem has been tackled as a multivariate highly imbalanced classification problem, where evolutionary neural networks have obtained good results in previous studies. Specifically, we have tested mono- and multi-objective evolutionary algorithms, with different types of units in the hidden layer of the network. We have shown that the proposed approaches are able to solve the problem with a high level of accuracy, even for the minority classes, which are also well classified in the majority of cases. We have also carried out a direct comparison of the best results obtained by the evolutionary artificial neural networks with that of the Terminal Aerodrome Forecasts (TAF), obtaining improvements in the classification of the most difficult convective situation prediction. This shows that evolutionary neural networks are a very good option for the prediction of convective situations, at short-term prediction time horizons.

Acknowledgements This research has been partially supported by the Ministerio de Economía, Industria y Competitividad of Spain (Refs. TIN2017-90567-REDT, TIN2017-85887-C2-1-P and TIN2017-85887-C2-2-P). D. Guijo-Rubio's research has been supported by the FPU Predoctoral Program from Spanish Ministry of Education and Science (Grant Ref. FPU16/02128).

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Cao, Z.; H. Cai, 2016. Identification of forcing mechanisms of convective initiation in the mountain areas through high-resolution numerical simulations. Advanced in Atmospheric Sciences 33(10), 1104-1105.
2. Johns, R.H., Doswell III, C.A., 1992. Severe local storms forecasting. Weather & Forecasting 7, 588-612.
3. Browning, K; Blyth, A; Clark, P; Corsmeier, U; Morcrette, C; Agner, J; Bamber, et al., 2007. The Convective Storm Initiation Project. Bulletin of the American Meteorolical Society 88, 1939-1955.

4. Kalthoff, N.; Adler, B.; Barthlott, C.; Corsmeier, U.; Mobbs, S;, Crewell, S; 2009. The impact of convergence zones on the initiation of deep convection: A case-study from COPS. Atmospheric Research 93, 680-690.
5. Wang, H., Y. Luo, and B. J.-D. Jou, 2014: Initiation, maintenance,and properties of convection in an extreme rainfall eventduring SCMREX: Observational analysis. Journal of Geophysical Research: Atmospheres 119, 13206-13232.
6. Bouin, M.N.; Redelsperger, J.L.; Lebeaupin Brossier, C, 2017. Processes leading to deep convection and sensitivity to sea-state representation during HyMeX IOP8 heavy precipitation event, Quarterly Journal of the Royal Meteorological Society 143(707), 2600-2615.
7. Galway, J.G., 1956. The lifted index as a predictor of latent inestability. Bulletin of the Ameican Meteorological Society 37, 528-529
8. Miller, R.C., 1975: Notes on analysis and severe storm forecasting procedures of the military weather warning center, Technical Report No. 200, AWS, USAF.
9. Moncrieff, M.W., Miller, M.J., 1976. A theory of organised steady convection and its transport properties. Q. J. R. Meteorol. Soc. 102, 373-394.
10. Sánchez, J.L., Marcos J.L., Dessens, J., López, L., Bustos, C., García-Ortega, E., 2009. Assessing sounding-derived parameters as storm predictors in different latitudes. Atmospheric Research 93, 446-456.
11. Sánchez, J.L., López, L., Bustos, C., Marcos, J.L., Ortega, E.G., 2007. Short-term forecast of thunderstorms in Argentina. Atmospheric Research 88, 36-45.
12. Púcik, T., Groenemeijer, P., Rýva D., and Kolár, M., 2015. Proximity soundings of severe and non severe thunderstorms in central Europe. Monthly Weather Review 143, 4805-4821.
13. Taszarek, M.; Brooks, H.E.; Czernecki, B. Sounding-Derived Parameters Associated with Convective Hazards in Europe. Mon. Wea. Rev. 2017, 145, 1511-1528.
14. Brooks, H.E., Lee, J.W., Graven, J.P., 2003. The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. Atmospheric Research 67-68, 73-94.
15. Kaltenböck R., Diendorfer G., Dotzek N., 2009. Evaluation of thunderstorm indices from ECWMF analyses, lightning data and severe storm reports. Atmospheric Research 93, 381-396.
16. Bertolotto, P. A., Roggero, G. 2016. Comparison of instability indices from COSMO-I7 and ECMWF-IFS analyses over the Piedmont Region, Italy, and new modifications to the K Index. Meteorological Applications, 23, 605-613.
17. Gascon, E., Merino, A., Sanchez, J.L., Fernández-González, S., García-Ortega, E., López, L., Hermida, L., 2015. Spatial distribution of thermodynamic conditions of severe storms in southwestern Europe. Atmospheric Research 164-165, 194-209.
18. Zhuge, X., Zou, X. 2018. Summertime convective initiation nowcasting over Southeastern China based on Advanced Himawari Imager observations. Journal of the Meteorological Society of Japan, 96, 337-353.
19. Mecikalski, J., and Bedka, K., 2006. Forecasting convective initiation by monitoring the evolution of moving cumulus in daytime GOES imagery. Monthly Weather Review, 134, 49-78
20. Merk, D., and Zinner, T., 2013. Detection of convective initiation using Meteosat SEVIRI implementation in and verification with the tracking and nowcasting algorithm Cb-TRAM. Atmospheric Measurement Techniques, 6, 1903-1918.
21. Mecikalski, J., Williams, J. K., Jewett, C. P., Ahijevych, D., LeRoy, A., and Walker, J. R., 2015. Probabilistic 0-1-h convective initiation nowcasts that combine geostationary satellite observations and numerical weather prediction model data. Journal of Applied Meteorology and Climatology 54, 1039-1059.
22. Neto, C. P., Barbosa, H. A., Beneti, C.A., 2016. A method for convective storm detection using satellite data. Atmósfera 29, 343-358.
23. Doswell, C. A. III, 2001. Severe convective storms - An overview. Severe Convective Storms Meteor. Monogr., No. 50, American Meteorological Society, 1-26
24. Maqsood, I., Khan, M. R., and Abraham, A. 2004. An ensemble of neural networks for weather forecasting. Neural Computing and Applications, 13(2), 112-122.
25. Sánchez-Monedero, J., Salcedo-Sanz, S., Gutiérrez, P., Casanova-Mateo, C.,Hervás-Martínez, C., 2014. Simultaneous modelling of rainfall occurrence and amount using a hierarchical nominal-ordinal support vector classifier. Engineering Applications of Artificial Intelligence 34, 199-207.
26. Ortiz-García, E. G., Salcedo-Sanz, S. and Casanova-Mateo C., 2014. Accurate precipitation prediction with support vector classifiers: A study including novel predictive variables and observational data. Atmospheric Research 139, 128-136.
27. Sachindra, D. A., Ahmed, K., Rashid, M., Shahid, Perera, S., 2018. Statistical downscaling of precipitation using machine learning techniques. Atmospheric Research 212, 240-258.
28. Ghada, W., Estrella, N., Menzel, 2019. A Machine Learning approach to classify rain type based on thies disdrometers and cloud observations Atmosphere 10(5), 251.
29. Seo, Y., Kim, S., Singh, V. P., 2018. Machine Learning Models Coupled with Variational Mode Decomposition: A New Approach for Modeling Daily Rainfall-Runoff. Atmosphere 9(7), 251.
30. Cornejo-Bueno, L., Casanova-Mateo, C., Sanz-Justo, J., Cerro-Prada, E., S. Salcedo-Sanz, 2017. Efficient prediction of low-visibility events at airports using machine-learning regression. Boundary Layer Meteorology 165, 349-370.
31. Guijo-Rubio, D.; Gutiérrez, P.A.; Casanova-Mateo C.; Sanz-Justo, J., Salcedo-Sanz, S., Hervás-Martínez, C. 2018. Prediction of Low-visibility Events due to Fog using Ordinal Classification. Atmospheric Research 214, 64-73.
32. Durán-Rosal, A. M., Fernández, J. C., Casanova-Mateo, C., Sanz-Justo, J., Salcedo-Sanz, S., and Hervás-Martínez, C. 2018. Efficient fog prediction with multi-objective evolutionary neural networks. Applied Soft Computing 70, 347-358.
33. Salcedo-Sanz, S., Casanova-Mateo, C., Pastor-Sánchez, A., Sánchez-Girón, M., 2014. Daily Global Solar Radiation Prediction based on a Hybrid Coral Reefs Optimization – Extreme Learning Machine Approach. Solar Energy 105, 91-98.
34. Aybar-Ruiz, A., Jiménez-Fernández, S., Cornejo-Bueno, L., Casanova-Mateo, C., Sanz-Justo, J., Salvador-González P., S. Salcedo-Sanz, S., 2016. A novel Grouping Genetic Algorithm – Extreme Learning Machine approach for global solar radiation prediction from numerical weather models inputs. Solar Energy 132, 129-142.
35. Cornejo-Bueno, L., Casanova-Mateo, C., Sanz-Justo, J., Salcedo-Sanz, S. 2019. Machine learning regressors for solar radiation estimation from satellite data. Solar Energy 183, 768-775.

36. Bala, K.; Choubey, D.K.; Paul, S., 2017. Soft computing and data mining techniques for thunderstorms and lightning prediction: a survey. In: International Conference of Electronics, Communication and Aerospace Technology (ICECA 2017), RVS Technical Campus, Coimbatore, Tamilnadu, India, 1, 42-46.
37. Bishop, C. M., 1995. Neural networks for pattern recognition. Oxford university press.
38. Fernández, J. C., Martínez-Estudillo, F. J., Hervás-Martínez, C., and Gutiérrez, P. A. 2010. Sensitivity versus accuracy in multiclass problems using memetic pareto evolutionary neural networks. IEEE Transactions on Neural Networks, 21(5), 750-770.
39. Dee D. P., Uppala S. M., Simmons A. J., Berrisford P., Poli P., Kobayashi S., et al. 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society 137, 553-597.
40. Annex 3 to the Convention on International Civil Aviation: Meteorological Service for International Air Navigation, 2016.
41. Prechelt, L. 1994. Proben1: A set of neural network benchmark problems and benchmarking rules.
42. Pérez-Ortiz, M., Gutierrez, P. A., Hervás-Martínez, C., and Yao, X. 2014. Graph-based approaches for oversampling in the context of ordinal regression. IEEE Transactions on Knowledge and Data Engineering 27(5), 1233-1245.
43. Lippmann, R. P. 1989. Pattern classification using neural networks. IEEE communications magazine, 27(11), 47-50.
44. Martínez-Estudillo, F. J., Hervás-Martínez, C., Gutiérrez, P. A., and Martínez-Estudillo, A. C. 2008. Evolutionary product-unit neural networks classifiers. Neurocomputing, 72(1-3), 548-561.
45. Schmitt, M. 2002. On the complexity of computing and learning with multiplicative neural networks. Neural Computation, 14(2), 241-301.
46. Billings, S. A., and Zheng, G. L. 1995. Radial basis function network configuration using genetic algorithms. Neural Networks, 8(6), 877-890.
47. Donoho, D. L., and Johnstone, I. M. 1989. Projection-based approximation and a duality with kernel methods. The Annals of Statistics, 58-106.
48. Del Ser, J., Osaba, E., Molina, D., Yang, X. S., Salcedo-Sanz, S., Camacho, D., et al., 2019. Bio-inspired Computation: Where We Stand and What's Next. Swarm and Evolutionary Computation 48, 220-250.
49. Angeline, P. J., Saunders, G. M., and Pollack, J. B. 1994. An evolutionary algorithm that constructs recurrent neural networks. IEEE transactions on Neural Networks 5(1), 54-65.
50. García-Pedrajas, N., Hervás-Martínez, C., and Muñoz-Pérez, J. 2002. Multi-objective cooperative coevolution of artificial neural networks (multi-objective cooperative networks). Neural networks: the official journal of the International Neural Network Society, 15(10), 1259-1278.
51. Fawcett, T. 2006. An introduction to ROC analysis. Pattern recognition letters 27(8), 861-874.
52. Baldominos, A., Saez, Y., and Isasi, P. 2019.On the automated, evolutionary design of neural networks: past, present, and future. Neural Computing and Applications, 1-27.
53. Konak, A., Coit, D. W., and Smith, A. E. 2006. Multi-objective optimization using genetic algorithms: A tutorial. Reliability Engineering and System Safety 91(9), 992-1007.
54. Zitzler, E., and Thiele, L. 1999. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE transactions on Evolutionary Computation 3(4), 257-271.
55. Fernández, J. C., Cruz-Ramírez, M., and Hervás-Martínez, C. 2018. Sensitivity versus accuracy in ensemble models of Artificial Neural Networks from Multi-objective Evolutionary Algorithms. Neural Computing and Applications, 30(1), 289-305.
56. Martínez-Estudillo, A., Martínez-Estudillo, F., Hervás-Martínez, C., and García-Pedrajas, N. 2006. Evolutionary product unit based neural networks for regression. Neural Networks 19(4), 477-486.
57. Gutiérrez, P. A., Hervás, C., Carbonero, M., Fernández, J. C. 2009. Combined projection and kernel basis functions for classification in evolutionary neural networks. Neurocomputing, 72(13-15), 2731-2742.