Data Augmentation to Improve the Soundscape Ranking Index Prediction

ROBERTO BENOCCI[1], ANDREA POTENZA[1], GIOVANNI ZAMBON[1],
ANDREA AFIFY[2,3], H. EDUARDO ROMAN[2]

[1]Department of Earth and Environmental Sciences (DISAT)
University of Milano-Bicocca
Piazza della Scienza 1, 20126 Milano
ITALY

[2]Department of Physics
University of Milano-Bicocca
Piazza della Scienza 3, 20126 Milano
ITALY

[3]NEXiD Edge
NEXiD
Via Fabio Filzi 27, 20124 Milano
ITALY

Abstract: Predicting the sound quality of an environment represents an important task especially in urban parks where the coexistence of sources of anthropic and biophonic nature produces complex sound patterns. To this end, an index has been defined by us, denoted as soundscape ranking index (SRI), which assigns a positive weight to natural sounds (biophony) and a negative one to anthropogenic sounds. A numerical strategy to optimize the weight values has been implemented by training two machine learning algorithms, the random forest (RF) and the perceptron (PPN), over an augmented data-set. Due to the availability of a relatively small fraction of labelled recorded sounds, we employed Monte Carlo simulations to mimic the distribution of the original data-set while keeping the original balance among the classes. The results show an increase in the classification performance. We discuss the issues that special care needs to be addressed when the augmented data are based on a too small original data-set.

Key-Words: - soundscape; data augmentation; machine learning; ecoacoustic indices; soundscape ranking index; urban parks;

Received: April 9, 2023. Revised: July 2, 2023. Accepted: September 2, 2023. Published: September 20, 2023.

1 Introduction

Soundscape analysis has grown in importance during the last decades as a tool for a rapid evaluation of environmental sounds. The analysis relies on non-invasive techniques for ecological surveillance thanks to the use of a widespread monitoring technology known as passive acoustic monitoring (PAM) techniques. The latter can provide large amount of data for long periods of time even in hardly accessible geographical locations. In this respect, ecoacoustics has emerged as a new branch of acoustics, which is based on the idea that acoustic ambience measurements can provide a significant information on the sound sources present within a given landscape. Thus, the contributions from different sound sources, identified as geophony (physical), biophony (biological) and anthropogenic (anthrophony or technophony) sources, [1], [2], [3], [4], may provide an estimation of the health of a habitat across space and time. The complexity of the assemblage of sounds in an environment is usually summarized in terms of ecoacoustics indices, which can be evaluated over specific intervals of time, which map the acoustic dynamics of an ecosystem, [5], into time series describing the observed status changes taking place in a given habitat, [6]. The information carried by the calculation of ecoacoustic indices are usually validated by listening to hours of recordings in order to

identify known sound categories. This activity is generally referred to as sound-truthing.

The manual procedure to identify the different sound sources is typically very time consuming, requiring specific technical preparationa and long-time training. For this reason, this methodology is usually applied to small data-sets, [7], [8]. A summative approach has also been proposed by providing qualitative information of the sound characteristics in each audio recording, [9]. In this sense, the consideration of just few or many bird sounds, few or many bird species, and low or high traffic noise, etc., may simplify and speed up the validation process. However, such information alone cannot provide an accurate assessment of the acoustic quality characterizing a complex habitat. In order to estimate a soundscape index, an attempt has been recently made by empirically accounting for the contribution of different sound sources, which assigns a positive weight to natural sounds (biophony) and a negative one to anthropogenic sounds. The results can be summarized in terms of an index denoted as soundscape ranking index (SRI), which has proved to disentangle complex environmental sound blends such as those found in urban parks, [10]. The analysis outlined in this paper refers to the sound recordings taken at Parco Nord (Northern Park) of the city Milan (Italy).

It is well established that machine learning (ML) algorithms can be trained to learn generic features extracted from a database, [11], [12], [13], which can then be used for dealing with many useful and complex tasks. For example, the application of ML techniques for studying the dynamics of a soundscape has provided accurate identifications of different bird species, [14], [15], and also the separation of the different audio sources when they are mixed in a set containing an unknown combination of their components, [16]. In urban areas, ML models have been used to predict long-term acoustic patterns from short-term sound measurements, [17], and for filtering out anomalous noise events before computing the actual traffic noise maps, [18]. ML techniques have been also applied for soundscape classification, [19], [20], species recognition, [21], [22], and the identification of mixed noise sources using a two-stage classifier which is able to discriminate different urban acoustic events in real time, by relying on the normally present redundant information from a network of acoustic sensors in the city of Barcelona, [23].

The implementation of a huge manual labelling (about 60,000 sound recordings) has proved to be satisfactory for the identification of different sound sources, [24]. A huge dataset collected over four years across Sonoma County (California) by citizen scientists was used to predict soundscape components with success, [25]. The automatic recognition of the soundscape quality of urban recordings has been studied by applying four different support vector machine (SVM) regressors to a combination of spectral features, [26]. In [27], a combination of temporal, spectral, and perceptual features was used to classify urban sound events belonging to nine different categories. In [28], different acoustic indicators taken in the city of Barcelona were used to train several clustering algorithms with recognizing the possibility of clustering the city area according to the noise levels. ML has been also used in a preliminary work, to optimize the definition of SRI, [29]. Here, the set of weights that define the SRI were searched by training four machine learning algorithms, Decision Tree (DT), Random Forest (RF), Adaptive Boosting (AdaBoost), Support Vector Machine (SVM), over a relatively small number of labelled sound recorded audio files.

More specifically, it was shown that two classification models, DT and AdaBoost, were able to provide a set of weights characterized by a rather good classification performance. The obtained results were in quantitative agreement with two different statistical approaches: a self-consistent estimation of the mean SRI values at each site, [30], and a cluster analysis performed, over the extracted features at each site, [31].

In this work, we studied the possibility of predicting the SRI at an urban park in the city of Milan (Italy) from the extracted spectral features of audio recordings in the form of seven ecoacoustic indices. Using these features, in [29], we obtained quite poor classification scores. In actual facts, classification performance is strongly affected by unbalanced classes (in our case the number of recording occurring in each sound quality class). To deal with this issue, we aim here at improving the classification scores by using Monte Carlo simulations to create a larger and well-balanced labelled dataset.

2 Materials and Methods

The urban area of study is described, together with the employed instrumentation. The method used is briefly reviewed by reminding the reader of the definition of the SRI and the way it is optimized using machine learning methods.

2.1 Area of Study

Following our previous works, we discuss results on the urban zone known as the Parco Nord (Northern Park) in Milan city. The latter covers an area of about 800 hectares, located within a quite developed and urbanised zone. The park possesses a tree-covered parcel of little more than 20 hectares surrounded by few agricultural fields, lawns, paths and roads (Figure 1). The location of the park is delimited to the north by

trafficked roads, including a highway and a main city street, at about 100 m from the wooded parcel. To the west side of the park there is also a small city airport, which is around 500 m from the tree-line edge.

2.2 Audio recorders

In our study, we employed commercial SMT Security low-cost digital audio recorders. Their working set up was chosen to be able to measure at a sampling rate of 48 kHz in an almost continuous mode. Devices with similar characteristics in terms of frequency response were selected in the measuring campaign. For the analysis, we scheduled the recording to coincide with greatest singing avifauna activity, taken on May 25th 2015, during the morning hours from 06:30 a.m. to 10:00 a.m. (CET). They correspond to 3.5 h for each sensor and recording session. As mentioned in Figure 1, some sensors did not work properly and were excluded. As a result, we were left with 16 sites, while the total recordings data analyzed resulted in 1120 files.

2.3 Aural Survey

In order to quantify the quantify the presence of biophonies, anthrophonies and geophonies, an aural survey was implemented. In particular, an expert listened carefully each one of the recordings according to the following scheme: for every three minutes of continuous recording, only one-minute was listened and interpreted. This yielded a total of 70 min of listening per site. In order to facilitate the task, the expert considered only five sound categories: birds, other animals, road traffic noise, other noise sources (such as airplanes, trains). Then, information about each type of sound source, such as its occurrence or not presence, in addition to the intensity, were determined.

More specifically, four parameters were considered and evaluated: (1) Individual abundance, referred to simply as, no–few–many subjects. (2) Perceived singing activity, represented in terms of the fraction of time (percent) attributed to avian vocalizations, (0 100 %). (3) Species richness, also referred to as, none–one–more than one species. (4) Vocalization intensity, again referred to as, no–low–high intensity. Regarding other animals and/or people, the indicator was considered as an either present or absent event.

It was found that road traffic was the main anthropogenic noise in the zone. For this source, only two parameters were considered: (1) Noise intensity, represented as, no–low–high intensity. (2) Typology of traffic, that is either continuous or intermittent traffic, or no traffic at all.

2.4 The soundscape ranking index, SRI

We briefly review the definition of the SRI, introduced as a simple criterion to estimate the sound quality of a local environment, [10]. Let us consider a single audio recording labeled by the letter ℓ. In this case the SRI is defined as,

$$\mathrm{SRI}_\ell = \sum_{i=0}^{n_c} c_{N_i} N_{i,\ell}, \tag{1}$$

where $n_c + 1$ is the total number of identified categories (birds, other animals, road traffic noise, other noise sources, rain and wind), here $n_c = 4$, $N_{i,r} = 1$ if the ith sound category is present at the recording r, and $N_{i,r} = 0$ otherwise. The coefficients can take one of the following values: $c_{N_i} > 0$ ($c_{N_i} \to c_+$, c_{++}) when the sound category corresponds to a natural sound, in that case we split the values into two possible sub-ranges ($+$, $++$); and $c_{N_i} < 0$ ($c_{N_i} \to c_-$, c_{--}) corresponding to a potential disturbing event, also split into two sub-ranges. The absence of birds vocalization is regarded as a neutral event, $c_{N_0} = 0$, setting the separation between natural and anthropogenic events. In Table 1, we summarize the employed ranges of variability for the coefficients c_{N_i}.

Our definition Eq. (1) is expected to yield SRI values representative of the quality of the environmental sound. For the sake of simplicity, we choose three intervals of SRI to define the environmental sound quality, for a single recording denoted simply as ℓ, given by

$$\begin{aligned} & \mathrm{SRI}_\ell && < 0 && \text{[poor quality]}, \\ 0 \le\ & \mathrm{SRI}_\ell && \le 2 && \text{[medium quality]}, \\ & \mathrm{SRI}_\ell && > 2 && \text{[good quality]}. \end{aligned} \tag{2}$$

2.5 SRI optimization procedure

To each sound category i, we assign a weight, c_{N_i}, according to the attributes extracted from the audio recording (see the column 'Attribute' in Table 2). For instance, the singing activity gets a weight depending on the percentage of singing birds detected in each recording: the interval (0,25]%, corresponds to a weight of $0.25 \times c_{++}$, while the (25,50]% one becomes the weight $0.50 \times c_{++}$, etc. As a general constraint, we assume that c_{N_i} can vary within the intervals mentioned in Table 1.

For the purpose of the using the ML techniques, both the spectral features associated to the ecoacoustic indices, and the corresponding SRIs need to be split into a 'training' set, and a 'test' one. The former is used as input for each classification model, whereas the final performance of the latter is quantified according to the classification metrics used, here employed the F1–score. In the process of classification, the SRIs become the target variable which is expected

Figure 1: Aerial view of the Parco Nord, on which we have indicated the grid of sensors employed in the recordings (red and yellow dots). The red spots correspond to the effectively working sensors, while the fewer yellow spots those which did not function properly and were excluded from the analysis. On the plot, we have highlighted the A4 highway and Padre Turoldo street to the north of the park. The Bresso airport can be seen at the left side of the picture.

Table 1: Intervals of variation for the coefficient c_{N_i} which are assigned to each sound category, $i = 0, \ldots, 4$, to be employed in Eq. (1). In the present calculations, we have taken the total range of variation of the coefficients to be: $-5 \leq c_{N_i} \leq 5$.

c_{N_i}	Range
c_{++}	[2, 5]
c_{+}	[0, 2]
c_{N_0}	0
c_{-}	[-2,0]
c_{--}	[-5,-2]

to be predicted by the classification model. The process is repeated for every combination of weights, c_{N_i}, for $i > 0$, that are varied in the corresponding interval of variability. The optimal set of weights defining our target SRI is taken from the best classification score obtained.

2.6 Classification models

It is convenient to briefly describe the classification models used for predicting the manual labelling sound categories and the SRI index from the ecoacoustic indices. As a matter of fact, ML learning methods are suitable to capture the potentially non-linear relationships among variables, which are not known a priori. The models we have considered here, which have been implemented in Python programming language, [32], are the following,

- Random Forest (RF),
- Single-layer neural network or Perceptron (PPN).

The RF is used to fit a given number of decision tree classifiers on various sub-samples of the data-set, using an averaging procedure to improve its predictive accuracy, and very importantly to also control over-fitting, [33]. For RF, we use default parameters with the exception of *Max-depth* = 3, typically devoted to control the size of the tree to prevent over-fitting.

Table 2: Coefficient c_{N_i} assigned to each sound category to be used in Eq. (1).

Category	Attribute	c_{N_i}
Birds singing	no	c_{N_0}
	few	c_+
	many	c_{++}
Birds species	no	c_{N_0}
	$\lesssim 2$	c_+
	>2	c_{++}
Singing activity (%)	0	$0.00 \times c_{++}$
	(0,25]	$0.25 \times c_{++}$
	(25,50]	$0.50 \times c_{++}$
	(50,75]	$0.75 \times c_{++}$
	(75,100]	$1.00 \times c_{++}$
Traffic type	no traffic	c_+
	continuous	c_-
	intermittent	c_{--}
Traffic intensity	zero	c_+
	low	c_-
	high	c_{--}
Other sound sources	absent	c_{N_0}
	present	c_{--}

The PPN was the first and simplest type of artificial neural network worked out and reported in the Literature, [34]. In a PPN network, the information oscillates back and forth from the initial input nodes, via the (if any) hidden nodes, toward the output nodes. The simplest kind of neural network is clearly just a single-layer Perceptron, consisting of one layer of output nodes. The inputs are then fed directly to the outputs via a series of weights. The sum of all the products of weights and inputs are then calculated at each node. If the resulting value is above a given threshold, the neuron fires and is updated as activated value, otherwise it remains at the deactivated value. For the Perceptron model, we used the following settings,

- seven neurons as input layer with activation function *ReLU*,
- three neurons as output layer with activation function *Softmax*,
- number of epochs 100,
- learning rate 0.001.

In our calculations for the implementation of the SRI optimization procedure, the supervised classification models have been trained on the 80% of the data, and tested on the remaining 20%. The splitting of the data have been performed using a stratified procedure in keeping the proportions between the target variable classes. We chose to implement the RF model and the PPN for essentially two reasons: The former yielded a good performance over other classification models in similar contexts, [29], whereas the latter represents the simplest neural network, both in terms of complexity and computing resources.

2.7 Ecoacoustic indices

In this work, we focused on the following set of ecoacoustic indices: the acoustic entropy index (H), [35], the acoustic complexity index (ACI), [36], the normalized difference soundscape index (NDSI), [37], the bio-acoustic index (BI), [38], the dynamic spectral centroid (DSC), [39], the acoustic diversity index (ADI), [39] and the acoustic evenness index (AEI), [39].

The indices were evaluated using the R statistical package (in particular, the version 3.5.1, [40]). Specifically, the fast Fourier transform (FFT) was computed by the function *spectro* available in the R package "seewave", [41]. The calculations were restricted to the frequency interval (0.1-12) kHz based on 1024 data points, corresponding to a frequency resolution FR = 46.875 Hz and, therefore, to a time resolution TR = 1/FR = 0.0213 s. The indices were calculated using the "soundecology" package, [42]. Finally, a dedicated script running in the "R" environment has been written to calculate the DSC index. For each one-minute recording, obtained as discussed above, we computed seven cumulative indices. Each recording was then represented by seven indices or features.

2.8 Data augmentation by Monte Carlo simulations

We used Monte Carlo simulations to generate a set of random numbers with the same distribution as the original data set, that in our case correspond to the distribution of the seven ecoacoustic indices in each sound quality class. For each combination of weights, c_{N_i}, for $i > 0$, varied in the assigned interval, a distribution for each of the seven ecoacoustics indices was derived. For each combination of weights, a Monte Carlo data augmentation has been applied in order to reproduce the original distribution in each class, and care was taken in order to obtain a balanced distribution of classes. In our case, we managed to obtain 500 instances for each class. As an example, Figure 2 illustrates the density distribution of ACI of the original data-set split into the three classes obtained for a particular combination of weights (c_{++}=2.7, c_{+}= 1.5, c_{-}=-1.0, c_{--}=-2.5).

Operationally, each of the ecoacoustic indices is first normalized, by subtracting its mean value and dividing by its standard deviation. Then, the cumulative probability is computed and, finally, its inverse function is derived (Figure 3). A random number thus corresponds to a value of ACI belonging to the original distribution.

3 Results

We ran RF and PPN models to attempt a prediction of the soundscape ranking indexes calculated assigning a set of weights to each sound category. We varied the different weights according to the intervals reported in Table 2. Then, the best combination of weights is the one which has the highest score provided by each classification model. In our case, we decided to employ the F1score and the accuracy criterion. The F1score is usually suited for dealing with unbalanced classes, whereas the accuracy method for balanced classes. The results are reported in Table 3.

The best F1–score for RF yields 0.63 and 0.60-0.61 for DNN (Table 3). This means that the simpler RF model can provide a slightly higher classification score.

3.1 Implementation of the Monte Carlo simulations

For each combination of weights, the occurrences in each class can substantially vary as it can be observed in Figure 4, showing the counts distribution for each class (Class 1, Class 2, Class 3), produced by different weight combinations whose range of variability is defined in Table 2.

The median values and the interquartile ranges (IQR) for Class 1 and Class 2 show that the majority of counts are unbalanced with respect to Class 3 (Table 4). This result may also be interpreted by considering that the majority of data files show a prevalence of *Good* (Class 3) SRIs.

This unbalanced distribution of classes may introduce a bias in the classification process. One of the methods to augment the data-set is to lean on a Monte Carlo expansion of the data based on the distribution of the original data-set split into each class. Monte Carlo data augmentation can effectively reproduce the original distribution in each class, thus leading to a more balanced distribution of classes. In our case, we managed to obtain 500 instances for each class.

At first, we compared the classification by using the augmented data-set that, as we mentioned, have been balanced to reach 500 elements for each class using the same set of weights that are reported in Table 3. The results are shown in Table 5, suggesting that data augmentation and data balancing provide quite different results than those obtained using the original unbalanced data.

Since the expanded data represent a balanced asset for the classification models, we explored the effect of data augmentation on other combination of weights. Thus, we first expanded the data on the basis of MC calculation, creating new data sets with balanced classes and then we repeated the classification for all the augmented data sets (360,000 files obtained from the combinations of all the possible weights in the range reported in Table 1). One of the major concerns when using MC data augmentation is the numerousness of the original class. Indeed, data expansion based on few data (see data belonging to whiskers of the boxplot of Figure4) would bias the new augmented class. For this reason, we decided to set the numerousness of the original class to be expanded above 100 counts. The results of the distribution for the Accuracy measure, for both models, are reported in Figure5. The two distributions are similar showing an accuracy peak value of 0.50 and 0.53 for PPN and RF, respectively.

In actual situations, we are mainly interested in the highest scores yielded by the two classification models. These results are reported in Table 6.

4 Discussion

The seven spectral features, derived by integrating the ecoacoustic indices over the whole length of the recording (1 minute), contain condensed information about the spectral variability that could be represented by lower integration times. Therefore, it does not appear to be enough to represent the complexity of the soundscape in a single summative index. However, both models yield a significant improvement in classifying the new augmented dataset, in going from 0.60-0.63 (F1–score) to 0.74-0.75 (Accu-

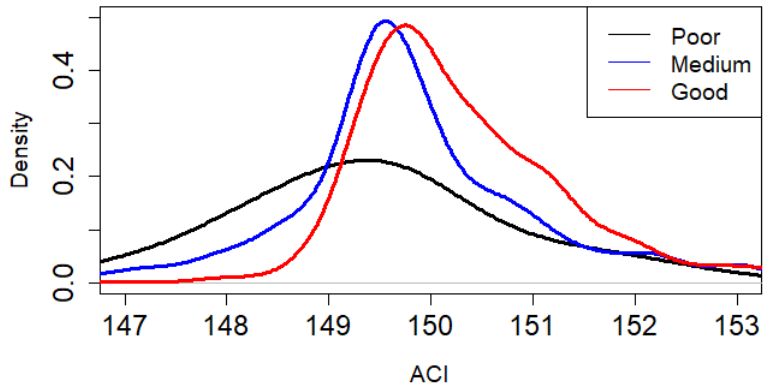

Figure 2: Density distribution of ACI of the original data-set split into the three classes obtained with the following combination of weights: c_{++}=2.7, c_{+}= 1.5, c_{-}=-1.0, c_{--}=-2.5.

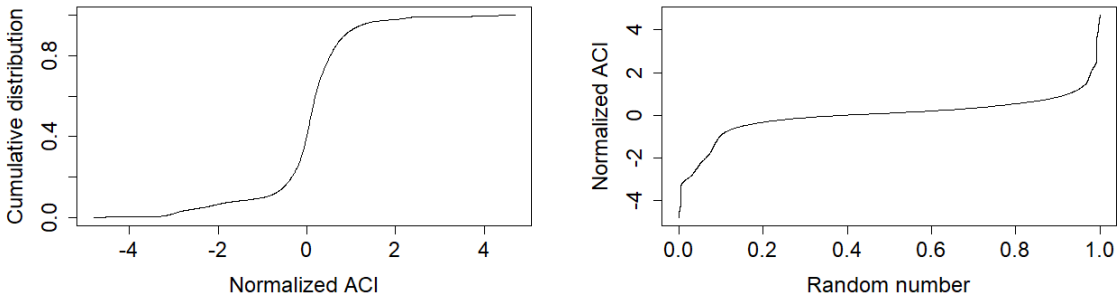

Figure 3: (**Left panel**) Cumulative probability distribution calculated for the normalized ACI, by subtracting the mean value and dividing by the standard deviation. (**Right panel**) Inverse function used for the MC simulations.

Table 3: Results for the RF and PPN models, using eco-acoustic indices as extracted spectral features. Classification measure (F1–Score with its standard deviation), range of weights values and class numerousness are reported. Class 1, Class 2, Class 3, correspond, respectively, to: *Poor, Medium, Good*, soundscape quality.

	F1–score	c_{++}	c_{+}	c_{-}	c_{--}	Class 1	Class 2	Class 3
RF	0.63±0.12	2.0	2.0	[-1.4, -1.5]	[-2.6, -2.7]	228	515	377
DNN	0.61±0.04	2.1	0	0	-4.8	432	364	324
	0.60±0.06	[2.2; 2.3]	[0.9; 1.3]	[-1.9; -1.8]	[-2.2; -2.0]	[245; 276]	[388; 413]	[456; 462]

Table 4: Median values and interquartile ranges (IQR) for Class 1, Class2 and Class 3, corresponding to Figure4.

	Median	IQR
Class1	165	185
Class 2	170	120
Class 2	785	292

racy; F1–score measure is pretty similar). The set of weights that realizes this new classification presents some differences from those obtained in the original unbalanced classification but are similar to each other (RF and PPN with augmented data).

In Figure 6, we compare the SRI maps calculated

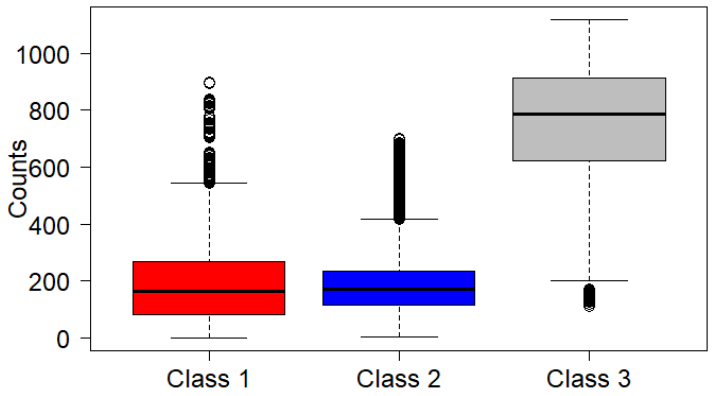

Figure 4: Counts distribution for each class: Class 1, Class 2, Class 3, corresponding to *Poor, Medium, Good* soundscape quality, of the original data-set calculated over all the combination of weights.

Table 5: Accuracy measure calculated for RF and PPN models, using seven eco-acoustic indices as extracted spectral features, 500 augmented data for each class and the same weights reported in Table 3.

	Accuracy
RF	0.48
DNN	0.59

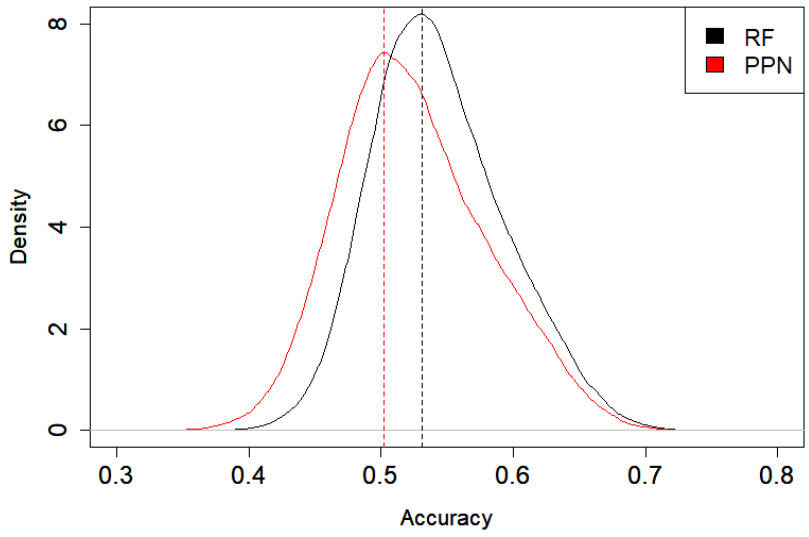

Figure 5: Density distribution of accuracy measure calculated from the augmented data. The peak of the distributions is at 0.51 (dashed red line) and 0.53 (dashed black line) for PPN and RF, respectively.

Table 6: Best accuracy scores obtained for RF and PPN models, using 500 augmented spectral features per class. Weights values and class numerousness of original data are also reported.

	Accuracy max	F1–score	c_{++}	c_{+}	c_{-}	c_{--}	Class 1	Class 2	Class 3
RF	0.751	0.751	2.1	1.9	-0.0	-3.5	123	128	869
PPN	0.746	0.737	2.0	1.5	-0.1	-5.0	371	354	395

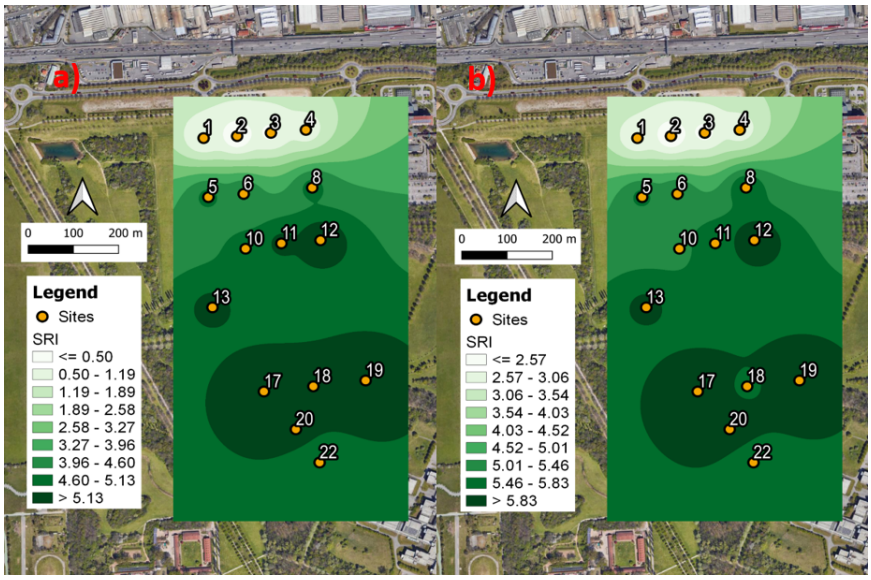

Figure 6: SRI maps obtained for the models: (**a**) PPN and, (**b**) RF.

over the study area based on the weights obtained for the RF and PPN models with the augmented data. For each of the 16 sites, we considered the median value of the SRI computed over all the measurements corresponding to the labeled recordings. As expected, the results are very similar showing a clustering of sites facing the traffic noise sources and those at the park interior. These results are in fairly good agreement with previous analysis obtained using a statistical approach, [30], [31].

5 Conclusions

The soundscape analysis in urban parks can assess the distance between natural habitats and artificial/reconstructed green areas. In this work, seven ecoacoustic indices, calculated over 16 sites of Parco Nord of Milan, Italy, have been used to predict a single index, named the soundscape ranking index, SRI, which has the advantage of yielding a quick overview of the quality of environment sound. SRI is computed through the optimization of the weights which appear in its definition (Eq. 2). Each combination of weights provides a different numerousness of the classes (Class 1, Class 2, Class 3 corresponding to *Poor, Medium, Good* soundscape quality) of the original data-set. These imbalanced classes may represent one of the factors influencing the classification models. In actual situations, the use of two very simple classification models, RF and PPN, with the original data-set, yielded a maximum F1-score of 0.60-0.61. We applied MC calculations to balance the three classes for each combination of weights and recomputed the classification algorithms. From the new computation, we ruled out those combinations of weights with initial numerousness lower than 100. This choice was justified to avoid the introduction of artificial peaked distributions in the original data, which once augmented would have affected the classification performance. In this way, we ended up with the following classification performance expressed in terms of Accuracy (usually employed for balanced classes classification, but we also report the F1–score for comparison with the previous calculations):

- RF : Accuracy = 0.75 (F1–score = 0.751),
- PPN : Accuracy = 0.746 (F1–score = 0.737).

As a future development, we envisage the use of larger labeled datasets, that is by using additional recordings with the corresponding aural survey, and the application of deep neural network to develop more efficient classifications. The definition of SRI and the threshold defining the classes of sound quality will be subject of a further investigation.

References:

[1] Krause, B. The Loss of Natural Soundscapes. Earth Island Journal, Spring 2002. (www.earth-island.org/journal/index.php/magazine/archive)

[2] Pijanowski, B. C., Farina, A., Gage, S. H., Dumyahn, S. L., Krause, B. L. What is soundscape ecology? An introduction and overview of an emerging new science. Landscape Ecology 26, 1213–1232 (2011). (https://doi.org/10.1007/s10980-011-9600-8).

[3] Pavan, G. Fundamentals of Soundscape Conservation. In: Farina A., Gage S. H. (Eds.). Ecoacoustics: The Ecological Role of Sounds, 235–258 (2017). (https://doi.org/10.1002/9781119230724.ch14).

[4] Sethi, S. S., Jones, N. S., Fulcher, B. D., Picinali, L., Clink, D. J., Klinck, H., Orme, C. D. L., Wrege, P. H., Ewers, R. M. Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. Proceedings of the National Academy of Sciences 117(29), 17049–17055 (2020). (https://doi.org/10.1073/pnas.2004702117).

[5] Lellouch, L., Pavoine, S., Jiguet, F., Glotin, H., Sueur, J. Monitoring temporal change of bird communities with dissimilarity acoustic indices. Methods in Ecology and Evolution 5(6), 495–505 (2014). (https://doi.org/10.1111/2041-210X.12178).

[6] Kasten, E. P., Gage, S. H., Fox, J., Joo, W. The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology. Ecological Informatics 12, 50–67 (2012). (https://doi.org/10.1016/j.ecoinf.2012.08.001).

[7] Pérez-Granados, C., Traba, J. Estimating bird density using passive acoustic monitoring: A review of methods and suggestions for further research. Ibis 163, 1–19 (2021). (https://doi.org/10.1111/ibi.12944).

[8] Shonfield, J., Bayne, E. M. Autonomous recording units in avian ecological research: Current use and future applications. Avian Conservation and Ecology 12(1), 14 (2017). (https://doi.org/10.5751/ace-00974-120114).

[9] Benocci, R., Roman, H. E., Bisceglie, A., Angelini, F., Brambilla, G., Zambon, G. Eco-acoustic assessment of an urban park by statistical analysis. Sustainability 13(14), 7857 (2021). (https://doi.org/10.3390/su13147857).

[10] Benocci, R., Roman, H. E., Bisceglie, A., et al. Auto-correlations and long time memory of environment sound: The case of an Urban Park in the city of Milan (Italy). Ecological Indicators 134, 108492 (2022). (https://doi.org/10.1016/j.ecolind.2021.108492).

[11] Cavallari, G. B., Ribeiro, L. S., Ponti, M. A. Unsupervised representation learning using convolutional and stacked auto-encoders: A domain and cross-domain feature space analysis. In: 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE, 440–446 (2018). (https://doi.org/10.1109/SIBGRAPI.2018.00063).

[12] Ponti, M. A., Ribeiro, L. S. F., Nazare, T. S., Bui, T., Collomosse, J. Everything you wanted to know about deep learning for computer vision but were afraid to ask. In: 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T). IEEE, 17–41 (2017). (https://doi.org/10.1109/SIBGRAPI-T.2017.12).

[13] Nunes C., Solteiro Pires E. J., Reis A. Machine Learning and Deep Learning applied to End-of-Line Systems: A review. WSEAS Transactions on Systems 21, 147–156 (2022). (https://doi.org/10.37394/23202.2022.21.16)

[14] Christin, S., Hervet, É., Lecomte, N. Applications for deep learning in ecology. Methods in Ecology and Evolution 10(10), 1632–1644 (2019). (https://doi.org/10.1111/2041-210X.13256).

[15] Fairbrass, A. J., Firman, M., Williams, C., Brostow, G. J., Titheridge, H., Jones, K. E. CityNet–Deep learning tools for urban ecoacoustic assessment. Methods in Ecology and Evolution 10(10), 1632–1644 (2019). Methods in Ecology and Evolution 10(2), 186–197 (2019). (https://doi.org/10.1111/2041-210X.13114).

[16] Lin, T. H., Tsao, Y. Source separation in ecoacoustics: A roadmap towards versatile soundscape information retrieval. Remote Sensing in Ecology and Conservation 6(3), 236–247 (2020). (https://doi.org/10.1002/rse2.141).

[17] Navarro, J. M., Pita, A. Machine Learning Prediction of the Long-Term Environmental Acoustic Pattern of a City Location Using Short-Term Sound Pressure Level Measurements. Applied Sciences 13(3), 1613 (2023). (https://doi.org/10.3390/app13031613).

[18] Orga F., Socoró J. C., Alías F., Alsina-Pagès R. M., Zambon G., Benocci R., Bisceglie A. Anomalous noise events considerations for the computation of road traffic noise levels: The DYNAMAP's Milan case study (2017) 24th International Congress on Sound and Vibration, ICSV 2017. (pdf at: http://hdl.handle.net/2072/376268).

[19] Piczak, K. J. Environmental sound classification with convolutional neural networks. In: IEEE 25th international workshop on machine learning for signal processing (MLSP). IEEE, 1–6 (2015). (https://doi.org/10.1109/MLSP.2015.7324337).

[20] Salamon, J., Bello, J. P. Deep convolutional neural networks and data augmentation for environmental sound classification. IEEE Signal processing letters 24(3), 279–283 (2017). (https://doi.org/10.1109/LSP.2017.2657381).

[21] Ward, J. H. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58(301), 236–244 (1963). (https://doi.org/10.1080/01621459.1963.10500845).

[22] Ruff, Z. J., Lesmeister, D. B., Appel, C. L., Sullivan, C. M. Workflow and convolutional neural network for automated identification of animal sounds. Ecological Indicators 124, 107419 (2021). (https://doi.org/10.1016/j.ecolind.2021.107419).

[23] Vidaña-Vila, E., Navarro, J., Stowell, D., Alsina-Pagès, R. M. Multilabel Acoustic Event Classification Using Real-World Urban Data and Physical Redundancy of Sensors. Sensors 21(22), 7470 (2021). (https://doi.org/10.3390/s21227470).

[24] Mullet, T. C., Gage, S. H., Morton, J. M., Huettmann, F. Temporal and spatial variation of a winter soundscape in south-central Alaska. Landscape Ecology 31(5), 1117–1137 (2016). (https://doi.org/10.1007/s10980-015-0323-0).

[25] Quinn, C. A., Burns, P., Gill, G., Baligar, S., Snyder, R. L., Salas, L., Goetz, S. J., Clark, M. L. Soundscape classification with convolutional neural networks reveals temporal and geographic patterns in ecoacoustic data. Ecological Indicators 138, 108831 (2022). (https://doi.org/10.1016/j.ecolind.2022.108831).

[26] Giannakopoulos, T., Siantikos, G., Perantonis, S., Votsi, N. E. and Pantis, J. Automatic soundscape quality estimation using audio analysis. In: Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments, Corfu, Greece (2015), pp. 1–9. (https://doi.org/10.1145/2769493.2769501).

[27] Tsalera, E., Papadakis, A., Samarakou, M. Monitoring, profiling and classification of urban environmental noise using sound characteristics and the KNN algorithm. Energy Reports 6, 223–230 (2020). (https://doi.org/10.1016/j.egyr.2020.08.045).

[28] Pita, A., Rodriguez, F. J., Navarro, J. M. Cluster analysis of urban acoustic environments on Barcelona sensor network data. International Journal of Environmental Research and Public Health 18(16), 8271 (2021). (https://doi.org/10.3390/ijerph18168271).

[29] Benocci, R., Afify, A., Potenza, A., Roman, H.E., Zambon, G. Toward the definition of a soundscape ranking index (SRI) in an urban park by machine learning techniques. Sensors 23(10), 4797 (2023). (https://doi.org/10.3390/s23104797).

[30] Benocci, R., Afify, A., Potenza, A., Roman, H.E., Zambon, G. Self-Consistent Soundscape Ranking Index: The Case of an Urban Park. Sensors 2023, 23, 3401. (https://doi.org/10.3390/s23073401).

[31] Benocci, R., Potenza, A., Bisceglie, A., Roman, H.E., Zambon, G. Mapping of the Acoustic Environment at an Urban Park in the City Area of Milan, Italy, Using Very Low-Cost Sensors. Sensors 2022, 22, 3528. (https://doi.org/10.3390/s22093528).

[32] Python. Available at https://www.python.org/ (accessed on 05/05/2023).

[33] Tibshirani, R., Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. (Trevor Hastie, Second Edition, 2009).

[34] Zell, A. Simulation Neuronaler Netze (Addison-Wesley, Bonn 1994). (https://doc1.bibliothek.li/aal/FLMF007250.pdf).

[35] Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S. Rapid acoustic survey for biodiversity appraisal. PLoS ONE 3(12), e4065 (2008). (https://doi.org/10.1371/journal.pone.0004065).

[36] Pieretti, N., Farina, A., Morri, D. A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI). Ecological Indicators 11(3), 868–873 (2011). (https://doi.org/10.1016/j.ecolind.2010.11.005).

[37] Grey, J. M., Gordon, J. W. Perceptual effects of spectral modifications on musical timbres. The Journal of the Acoustical Society of America 63(5), 1493–1500 (1978). (https://doi.org/10.1121/1.381843).

[38] Boelman, N. T., Asner, G. P., Hart, P. J., Martin, R. E. Multitrophic invasion resistance in hawaii: Bioacoustics, field surveys, and airborne remote sensing. Ecological Applications 17(8), 2137–2144 (2007). (https://doi.org/10.1890/07-0004.1).

[39] Yang, W., Kang, J. Soundscape and sound preferences in urban squares:

A case study in Sheffield. Journal of Urban Design 10(1), 61–80 (2005). (https://doi.org/10.1080/13574800500062395).

[40] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2018. (Available online: https://www.R-project.org/. (accessed on 05/05/2023).

[41] Seewave: Sound Analysis and Synthesis. Available online: https://cran.r-project.org/web/packages/seewave/index.html. (accessed on 05/05/2023).

[42] Soundecology: Soundscape Ecology. Available online: https://cran.r-project.org/web/packages/soundecology/index.html. (accessed on 05/05/2023).

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

R.B. conceptualized the study, carried out the M.C. simulation and wrote the original draft preparation; A.P. organized the data sets and produced map representations; G.Z. supervised the research; A.A. implemented the M.L. software; H.E.R. reviewed and edited the manuscript.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself:
No funding was received for conducting this study.

Conflicts of Interest
The authors have no conflicts of interest to declare that are relevant to the content of this article.