Ecological Informatics 75 (2023) 102014

Contents lists available at ScienceDirect

Ecological Informatics

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

Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification

Joan Gómez-Gómez, Ester Vidaña-Vila *, Xavier Sevillano

HER - Human-Environment Research, La Salle-Universitat Ramon Llull, Quatre Camins, 30, 08022 Barcelona, Spain

ARTICLE INFO

Keywords: Bird song Audio dataset Species identification Deep learning Neural network Spectrogram

ABSTRACT

The deployment of an expert system running over a wireless acoustic sensors network made up of bioacoustic monitoring devices that recognize bird species from their sounds would enable the automation of many tasks of ecological value, including the analysis of bird population composition or the detection of endangered species in areas of environmental interest. Endowing these devices with accurate audio classification capabilities is possible thanks to the latest advances in artificial intelligence, among which deep learning techniques stand out. To train such algorithms, data from the sources to be classified is required. For this reason, this paper presents the Western Mediterranean Wetland Birds (WMWB) dataset, consisting of 201.6 min and 5795 annotated audio excerpts of 20 endemic bird species of the Aiguamolls de l'Empordà Natural Park. The main objective of this work is to describe and analyze this new dataset. Moreover, this work presents the results of bird species classification experiments using four well- known deep neural networks fine-tuned on our dataset, whose models are also made public along with the dataset. These results are aimed to serve as a performance baseline reference for the community when using the WMWB dataset for their experiments.

1. Introduction

Bioacoustic avian life monitoring is a valuable means for obtaining relevant information regarding birdlife in a specific environment (Shonfield and Bayne, 2017).

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *E*cological Informatics *2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

In this context, the deployment of automatic systems running over bioacoustic monitoring devices in areas of environmental interest paves the way for remote bioacoustic sensing (Wijers et al., 2021).
This enables the study of birdlife in little accessible locations (e.g. reedbeds, or in strictly protected areas), or when access becomes more complicated (e.g., at night, during wintertime in highlands, or during lockdown periods). For instance, bioacoustic avian data can be used for the quantitative and qualitative analysis of bird population composition (Klingbeil and Willig, 2015; Rosenstock et al., 2002), the study of different ecological moments of a species (migration or reproductive seasons), long-term monitoring of bird individuals (Frommolt and Tauchert, 2014) or the detection of the presence of endangered species in a particular site (Garnett et al., 2011). However, the analysis of the recordings captured by acoustic monitoring devices is still often made by human experts (for instance, by listening to the recordings and/or through the visual inspection of sonograms), which is a complex and time-consuming task. Fortunately, the advances in audio signal processing and artificial intelligence (especially, deep learning) have enabled the development of algorithms capable of accurately detecting bird sounds from the recordings (Stowell et al., 2019; Tseng et al., 2020) and classifying bird species upon their sounds (see e.g. (Kahl et al., 2021; Knight et al., 2020)), which eases the automation of these relevant birdlife monitoring tasks. However, ecologists willing to apply deep learning classification techniques for bioacoustic monitoring should take into account that properly training a deep neural network from scratch requires huge amounts of annotated audio recordings. In fact, their performance increases logarithmically with the volume of properly annotated training data (Sun et al., 2017). Whereas there exist several excellent and publicly available repositories of bird sounds recordings (e.g. Xeno-Canto (Xeno-Canto Foundation, 2022), Avibase (Lepage, 2021) or eBird (Sullivan et al., 2009)), these often suffer from label reliability and the presence of environmental noise (Vidaña-Vila et al., 2017), requiring further annotation efforts to provide training data for the classifiers to perform properly. Thus, the lack of annotated bird song audio data collections constitutes a bottleneck in the development of this kind of approaches. For this reason, in this work we introduce the Western Mediterranean Wetland Birds (WMWB) dataset, a brand new annotated bird song audio dataset of species belonging to the authors closest environment, the Western Mediterranean coast. We took as a reference scenario the Aiguamolls de l'Empordà Natural Park, located in Catalonia, in north eastern Spain (42◦ 13′ 28.09" N, 3◦ 05′ 34.92″ E) (Rosell et al., 2013). This Mediterranean coastal wetland covers 4729 ha, and its privileged location between the Muga and Fluvià rivers makes it a natural shelter and rest area for over 320 nesting and migratory bird species, 82 of which have been reported to be regular nesters (Fatoric and Morén- Alegret, 2013). Advised by biologists of the park, 20 of its endemic species were selected. Using recordings of these species available in the Xeno-Canto (XC) portal, we collected and annotated a total 5795 clips totalizing over 200 min of bird songs.

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

* Corresponding author.
E-mail addresses: juan.g@salle.url.edu (J. Gómez-Gómez), ester.vidana@salle.url.edu (E. Vidaña-Vila), xavier.sevillano@salle.url.edu (X. Sevillano).
https://doi.org/10.1016/j.ecoinf.2023.102014
Received 23 September 2022; Received in revised form 1 February 2023; Accepted 1 February 2023

Available online 4 February 2023

Moreover, as a companion to the WMWB dataset, we also present the results of several bird species classification experiments based on fine tuning already-trained deep neural networks using spectrogram images of audio clips of this new labelled data collection. The resulting fine-tuned models are also published, so that they can serve as a reference baseline for researchers developing new classification algorithms with the proposed dataset. Thus, our main goal is to offer the research community a novel annotated dataset of birds that typically inhabit wetlands. Compared to other works, the main novelty of this work is threefold. First, and contrarily to other datasets such as (Stowell and Plumbley, 2014), the WMWB dataset is offered with hard labels (i.e., the labels specify the specific starting and ending point where a bird is vocalizing), which permits researchers to deeply analyze the audio samples if desired (e.g., compute time span between repetitions of vocalizations). Second, the WMWB dataset is made available to the community not only as wave recordings (like other datasets, e.g. (Lostanlen et al., 2018; Pamula, 2022)), but also as already-calculated features (i.e., spectrograms). This is intended to ease the development of new classification techniques, as, if desired, users could skip the feature engineering process and focus on classification only. And third, we accompany the WMWB dataset with fine-tuned deep neural network models that can serve as a baseline for classifier performance benchmarking. The remainder of this paper is organized as follows: first, Section 2 describes the main relevant related work in the field. Next, Section 3 explains the conception of the newly annotated WMWB dataset, describes its contents from different perspectives including the extraction of audio features, and describes a testbed of deep learning models that have been fine-tuned on our dataset to provide a baseline performance benchmark for future developments. Then, Section 4 presents the results and discusses the classification performance of the algorithms, comparing them in terms of accuracy, stability, and network footprint size. Finally, Section 5 concludes the paper and exposes the future research directions derived from this work.

2. Related work

This section aims at illustrating how the availability of annotated bird sound data collections has helped the development of automated bioacoustic avian life monitoring algorithms. On the one hand, we first describe several publicly available

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

bird vocalization datasets designed to train artificial intelligence algorithms for detection or classification tasks. And on the other hand, we also briefly review the evolution of acoustic feature extraction and pre-processing, including an analysis

of how increasingly complex machine learning approaches have been applied to solve these tasks.

2.1. Annotated bird vocalization datasets

The difficulty of gathering good quality and representative bird vocalization audio data makes currently available datasets often limited in size and diversity. As a matter of fact, many of them consist of recordings from a small number of individuals or species, usually ranging from a couple of species to a few dozens. In other cases, the limitation comes from the reduced number of recordings available for each species.

In this sense, one of the earliest works on bird species classification that provided its own annotated dataset was the work by (Sprengel et al., 2016). In that work, the authors compiled a large amount of bird categories (999 in the training set) but, in global, few instances per each category were collected (average of 25 files per class).

On the other end of this size vs diversity spectrum, we find the recent contribution of (Nicholson et al., 2022), a dataset containing a single male bird species named Bengalese finches (*Lonchura striata* var. *domestica*) recorded in laboratory conditions but with 50 classes corresponding to syllables.

Another example of how the classification of birds can become a complicated task due to the variation of vocalizations within the same family is the dataset of (Vidaña-Vila et al., 2017). In that work, the authors introduced a dataset with vocalizations of seven *Picidae* species inhabiting the Iberian Peninsula emitting up to three different sounds: call, drumming or song. The dataset is formed from 161 recordings lasting 4984 s which were obtained from the Xeno-Canto website.

Other datasets provide vocalizations that aim to obtain a classifier of multiple and diverse bird species. This is the case of (Pamula, 2022) which offers more than 56.5 h of recordings with annotations of nocturnal flight calls of six different passerine bird species migrating along the Baltic Sea coast, Poland. A larger number of species names are found is the dataset presented in (Salamon et al., 2016) with over 5000 flight calls from 43 different species from different locations of the USA. Another example can be found in (Morfi et al., 2019) which offers recordings from 7 regions from Spain and France and consists of 61 bird species and 87 different labels contained in 30 h of recordings.

While the aforementioned datasets have been created with the purpose of training bird species classification algorithms, others are designed with the detection of bioacoustic events in mind. This is the case of (Lostanlen et al., 2018) where a new dataset called BirdVox-full-night was presented. It consists of 62 h of audio recordings of nocturnally migrating birds, collected through a network of 6 acoustic sensors that include 35,402 flight calls annotated for use in the automatic detection of these vocalizations.

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

Regardless of the specific target application, to expand the number of species that a classifier can identify or to improve the accuracy of existing species (from the same or different family), it is crucial to have new and more diverse datasets of bird vocalizations that are collected in a variety of natural habitats and under different conditions.

Last but not least, it is important to highlight that there are several competitions or online challenges that encourage bioacoustics practitioners to research on new automatic detection and classification techniques to push the state-of-the- art forward. Often, in these competitions, new datasets are made available to pose new challenges to the community. In this field, the two main competitions are: (1) BirdCLEF and (2) DCASE (Challenge on Detection and Classification of Acoustic Scenes and Events). Specifically, DCASE is divided in several tasks from different domains, and usually one of the tasks is related to bioacoustics. In the 2022 edition, the Western Mediterranean Wetlands Bird dataset presented in this paper has been used as one of the training sets of the Few-shot Bioacoustic Event Detection Task (Nolasco et al., 2022).

2.2. Acoustic feature extraction and machine learning for bird species classification

In the last two decades, the literature has reflected a shift in the type of techniques employed for sound-based bird species classification. Indeed, early efforts in this area typically applied one-dimensional features already employed for human speech recognition. For example, in (Franzen and Gu, 2003), authors developed a bird-song production model able to generate synthetic bird songs and a hierarchical bird-song classifier based on a speech recognition approach.

The most commonly one-dimension speech features used for bioacoustic classification purposes were: (1) Wavelets, as in the case of the work proposed in (Selin et al., 2006), where authors stated that it is possible to recognize bird sounds using four features extracted from wavelets and a neural network as a classifier. (2) Mel-Frequency Cepstral Coefficients (MFCC), used for example in (Rai et al., 2016), where authors used the coefficients as an input to a Support Vector Machine. However, the dataset that they used for the experimental evaluation contained only four bird species commonly found in India (blackbird, duck, house crow and parrot) and few instances per species. (3) A combination, also referred to as fusion, of different one-dimensional features. This is the case of the work presented in (Vidana-Vila et al., 2020). In that work, authors used a two-layers hierarchical system that first detects bird vocal- izations using MFCC and Zero Crossing Rate (ZCR) and then classifies the detected vocalizations using a fusion of MFCC, Linear Predictive Cepstral Coefficients (LPCC) and Perceptual Linear Predictive Coefficients (PLPC).

However, recent literature in this field shows the increasing use of two-dimensional audio representations based on spectrograms as sound descriptors, like (de Oliveira et al., 2015). Moreover, this trend has been reinforced by the emergence of deep learning. Indeed, trying to take advantage of the excellent performance of

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

deep neural networks in image classification tasks (Krizhevsky et al., 2017), audio classification is increasingly being performed using spectrograms –which represent the temporal evolution of audio frequency contents– as image representations of sounds. Nonetheless, the accuracy and effectiveness of these deep learning algorithms is largely dependent on the quality and diversity of the training data.

One of the earliest works in this area is the work by (Sprengel et al., 2016), where authors trained from scratch a convolutional neural network (CNN) consisting in five convolutional layers and one dense layer using pre-processed spectrograms of bird vocalizations as inputs. For the spectrogram pre-processing, authors first separated the part of the spectrogram that contains the bird sound from the background. This system won the BirdCLEF2016 recognition challenge.

In this context, the recent work of (Stowell, 2022) reviews the state of the art in deep learning for computational bioacoustics. According to this work, spectrograms outperform MFCC when using CNNs. Moreover, this approach has the advantage of automatically learning features directly from less pre-processed data representations.

One of the latest and most prominent advances in the field is BirdNET (Kahl et al., 2021) a task-specific ResNet- based deep neural network architecture that was massively trained on audios of 984 bird species, including audios from Xeno-Canto (Xeno-Canto Foundation, 2022) and the Macaulay Library of Natural Sounds (Macaulay, 2022). Based on BirdNET, there exists the BirdNET-Pi solution for real-time acoustic bird classification for the Raspberry Pi (McGuire, 2022).

It must be highlighted, though, that the works from the field of bird sound classification usually focus on the detection or classification of a single vocalization within an audio segment. Nevertheless, given that in the real world multiple species can overlap, the classification of multiple birds vocalizing at the same time remains an open challenge (Briggs et al., 2012; Denton et al., 2022; Parrilla and Stowell, 2022).

Finally, to conclude this section, it is worth mentioning the work of (Priyadarshani et al., 2018), which reviews different approaches of birdsong recognition in complex acoustic environments. Also, this work highlights the importance of making benchmark datasets available so different researchers can compare their classification methodologies. This is one of the main reasons for making the Western Mediterranean Wetlands Bird dataset presented in this paper available for public use.

3. Materials and methods

3.1. Bird song dataset

Bird data acquisition and annotation is typically an exhaustive and time-consuming task, which requires expert birders and birdwatchers to manually annotate audio files. In this regard, several platforms such as Xeno-Canto (Xeno-Canto Foundation, 2022) present an excellent alternative for content sharing, allowing the scientific community to access thousands of bird audio files. However, the Xeno-

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *E*cological Informatics *2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

Canto portal has not been conceived as a database for training machine learning or deep learning algorithms, in the sense that it does not include strong annotations, such as the start and end points of each bird vocalization. For this reason, when using their data for that purpose, the audio-files require a pre-processing stage to remove background noise and sounds that may belong to sources different from the bird of interest, and to annotate the acoustic events that occur in the recording.

3.1.1. Data gathering

The first step of the project was to collect recordings from Xeno- Canto corresponding to the following 20 en- demic bird species of ecological interest suggested by the ornithologists in charge of bird species supervision in the Aiguamolls de l'Empordà Natural Park: *Acrocephalus arundinaceus, Acrocephalus melanopogon, Acrocephalus scir-paceus, Alcedo atthis, Anas strepera, Anas platyrhynchos, Ardea purpurea, Botaurus stellaris, Charadrius alexandri- nus, Ciconia ciconia, Circus aeruginosus, Coracias garrulus, Dryobates (Dendrocopos) minor, Fulica atra, Gallinula chloropus, Himantopus himantopus, Ixobrychus minutus, Motacilla flava, Porphyrio porphyrio* and *Tachybaptus rufi collis* (see Table 1 for the common name of each species).

The quality of the audio files is especially relevant when deciding what audios are included in the dataset to ensure the exclusion of audios with high background noise, audios in which the bird is only heard in the background or audios with presence of other interfering sounds. In Xeno-Canto, audios are labelled by quality from A to E, where A means that the audio quality is excellent and E means that the audio quality is poor. Hence, authors gathered only files that were labelled with categories A and B from the selected species.

3.1.2. Data annotation

The data pre-processing task was manually carried out using the Audacity software (https://www.audacityteam. org/), a free open source program that allows to listen, record and easily label audio files.
An advantage of Audacity is that it enables the visualization of the audio file as a waveform or as a spectrogram. Moreover, when using the spectrogram view, different parameters can be configured to help the audio labeller, such as the frequency range displayed in the screen, the scale (mel, linear, logarithmic, bark, etc.), the gain or even the color scale (color or grey-scale).

Fig. 1 depicts a screenshot of Audacity. As it can be seen, the process of labelling consists of opening a complete audio file in one track, selecting the spectral view with the desired configuration parameters (pink and yellow spectrogram) and then creating an empty labels track. The person dedicated to label must then carefully listen to the audio file and mark in the labels track which region of the file contains relevant information. In this case, the relevant information of the audio file XC436938.mp3 was given the label 'call', as it corresponds to a call of a *Fallinula*

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

chloropus (Common moorhen) individual. After labelling the full audio file, a labels file is exported in .txt format with the same name as the audio file (the Xeno-Canto identifier number, in this case, XC436938.txt).
These text files are organized in three columns: the first one indicates the starting time (in seconds) of the bird vocalization, the second column indicates the ending time (in seconds), and the third column is the label of the event: song, call,

Table 1
Length of the Western Mediterranean Wetland Birds dataset in terms of time (seconds) and number of audio files per species.

Bird species	Common name	Sound type	Total time (sec.)	Number of cuts	Number of XC files
Acrocephalus arundinaceus	Great reed warbler	songs	1982	453	34
Acrocephalus melanopogon	Moustached warbler	songs	2037	221	50
Acrocephalus scirpaceus	Common reed warbler	songs	2260	121	37
Alcedo atthis	Common kingfisher	songs / calls	351	418	64
Anas crecca	Gadwall	songs	292	96	9
Anas platyrhynchos	Mallard	songs	229	70	19
Ardea purpurea	Purple heron	calls	120	207	43
Botaurus stellaris	Eurasian bittern	songs	414	436	56
Charadrius alexandrinus	Kentish plover	songs / calls	109	275	58
Ciconia ciconia	White stork	bill clapping	479	121	40
Circus aeruginosus	Western marsh harrier	calls	165	207	38
Coracias garrulus	European roller	calls	179	267	34
Dendrocopos minor	Lesser spotted woodpecker	drumming	563	494	39
Fulica atra	Eurasian coot	calls	123	272	55
Gallinula chloropus	Common moorhen	calls	107	262	54
Himantopus himantopus	Black-winged stilt	calls	1212	277	70
Ixobrychus minutus	Little bittern	songs / calls	148	559	38
Motacilla flava	Western yellow wagtail	songs	292	400	42
Porphyrio porphyrio	Western swamphen	songs / calls	363	186	53
Tachybaptus ruficollis	Little grebe	songs	543	153	56
Total	–	–	12,096	5796	879

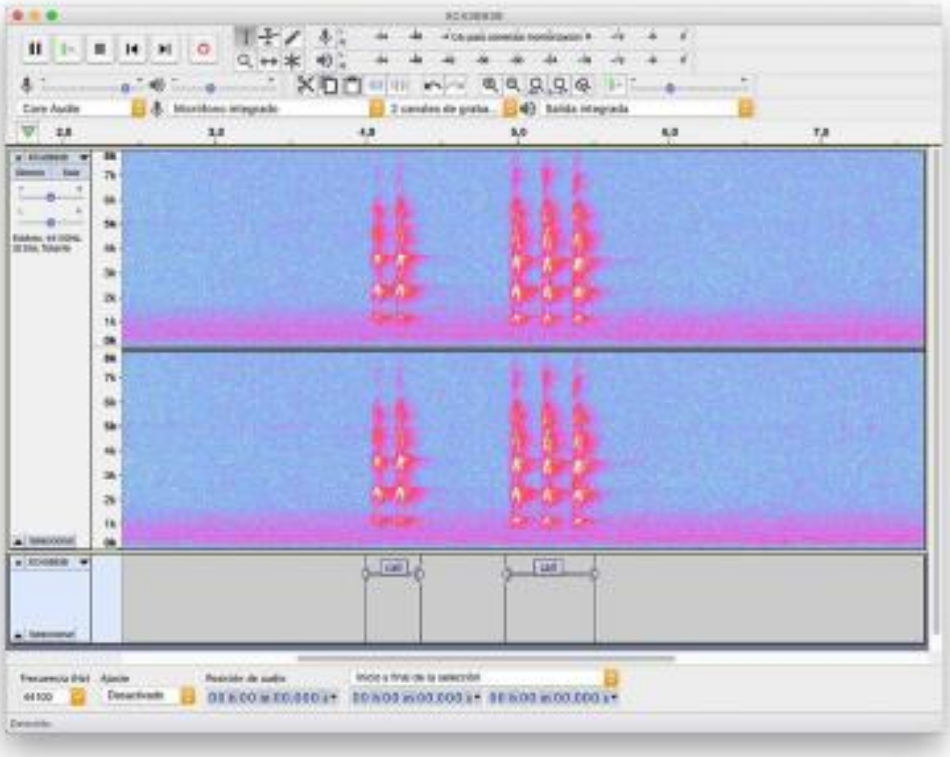

Fig. 1. Screenshot of Audacity, the framework used for manually labelling the audio files.

drumming, clapping, etc. The seconds are always relative to the start of the original Xeno-Canto audio file.

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

As for the example of Fig. 1, as only two vocalizations have been labelled in the audio file, the output file would have the following information:

3.1.3. Dataset analysis

After the manual labelling process, the authors obtained a dataset of 201.6 min (12,096 s) and 5795 audio excerpts from 879 original Xeno-Canto audio files. Table 1 shows in detail the amount of acoustic information obtained per each species. In most of the species, only call vocalizations or song vocalizations have been considered. However, for some species, both calls and songs have been obtained. This decision was taken depending on the amount of available samples on the Xeno-Canto portal. Moreover, for the *Dendrocopos minor* species, the *drumming* effect has been selected as the identifier of the presence of the bird, despite of being a sound that is characteristic for all woodpeckers in general, and not only the *Dendrocopos minor*. The reasoning of this decision is that the environmental area of reference is typically inhabited by the *Dendrocopos minor* species, and not other woodpeckers. A similar case occurs for the species *Ciconia ciconia*, that produces a characteristic sound produced by a repetitive clap with their bills that has been labelled as *bill clapping*. It is acoustically similar to the drumming of woodpeckers, but the spectral distribution is different as woodpeckers drum on trees using their bill and *Ciconia ciconia* specimens use exclusively their bills to clap.

As shown in Table 1, the dataset is not completely balanced. The major limitation for obtaining clean samples was the availability of XC data at the time of the dataset creation. The class that has more audio cuts (clean vocalizations) is *Ixobrychus minutus*, with 559 cuts. However, this class is not the class with more seconds of vocalizations. Actually, this class presents one of the fewest seconds of vocalizations, with only 148 s of clean data. On the contrary, the class that has more seconds of clean vocalizations is *Acrocephalus melanopogon*, with 2360 s of audio coming from only 221 audio cuts. This exemplifies one of the most complex aspects that bioacoustic automatic monitoring systems must face: the duration variability among vocalizations from different species.

To better illustrate this problematic, in Fig. 2 we show the boxplot representation of the duration of each audio fragment labelled in the dataset (i.e., a clean vocalization) grouped by species. As the variability of the fragments duration is so high, the boxplots are represented on a logarithmic axis, to be able to show with precision both the species with long vocalizations (such as *Acrocephalus scirpaceus* —which has a mean value between 10 and 20 s), and the species with short vocalizations (such as *Charadius alexandrius* —which has a mean value shorter around 20 milliseconds). In the boxplot, the width of the bars is proportional to the number of fragments of that species in the WMWB dataset. The audios in the dataset are provided in *.mp3* format, and with the same sampling frequency they originally had in Xeno-Canto. By doing so, we aim to give the user the freedom to either process the audios with their original sampling rate or resample them to the most convenient frequency.1

To provide the reader with a perspective on the sampling frequency diversity in the dataset, Fig. 3 presents the number of files as a function of the sampling rate. As it

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

can be observed, most of the audio files (580) were recorded using a standard frequency of 44,100 Hz. The second most common sampling rate is 48,000 Hz for 271 files, and then there are 22 files with a frequency of 32,000 Hz and 6 files with a sampling frequency lower than 30,000 Hz. These 6 files belong to the following species: Moreover, the dataset contains both single-channel (mono) and two-channels (stereo) recordings. Table 2 shows the number of audio files recorded in mono/stereo. As it can be observed, most audios (around 75% out of the total) are stereo. Again, we are providing the audio files with the original configuration so users can decide how to operate with them. A common practice would be to use a single channel of the stereo files, but if desired, mono audios can be converted to stereo by duplicating the content of one channel. Finally, Fig. 4 shows a geographic map of the original recordings location. The origin of the bird vocalizations is specially relevant considering that birds have dialects (Baker and Cunningham, 1985). As it can be observed, even though this dataset contains data from birds that typically inhabit the Western Mediterranean wetlands, the samples that compose the dataset are varied in their origin. Again, this is due to the sample availability in Xeno-Canto. Even though most of the samples were obtained in Europe, there are also samples from Asia, Africa and three samples of *Anas platyrhynchos* from North America. An interactive version of the map, where the user can zoom in and out and see to which audio file corresponds each dot in the map, can be downloaded together with the dataset.

1 It should be taken into account that if the user choses to resample all the files to the most common sampling rate of 44,100 Hz and then compute spectral parameters such as spectrogram, those audio files with a lower sampling rate will obtain part of the spectrogram in black and filled with zeros.

3.2. Audio feature extraction

Besides the audio files, the Western Mediterranean Wetland Birds dataset also comprises audio features that can be readily used to train artificial intelligence models. In particular, to characterise the audio events so they can be automatically classified, the spectrograms of the audio files have been generated.
A spectrogram is a two dimensional representation of an acoustic signal that plots the variation of frequency of a signal in function of the variation of intensity on time: the variation on intensity is represented as a change on the brightness or color of the signal (like in a heatmap), with the variation on frequency plotted vertically and the time evolution, horizontally. This way, the spectrogram of an audio fragment can be regarded as a ‘picture’ of it.
In this work, the spectrograms have been calculated using 1-s windows, meaning that the manually cleaned audio files of the dataset were split in fragments of one second each. For those audio files shorter than one second, and as usually done in these type of problems (Singh et al., 2019), the vocalization was repeated to fill the window size. The reasoning for choosing a window size of one second is that the window should contain—at least—one complete bird vocalization so a pattern can be obtained from the spectrogram. Also, the window should contain the

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *E*cological Informatics *2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

smallest possible amount of noise. However, the duration of the audio files on the dataset for each species are very different (see Fig. 2), which makes it hard to choose a window size that satisfies both requirements. For this reason, the selected window length offers a fair trade-off between both restrictions. Whereas in some species, such as *Acrocephalus arundinaceus, Acrocephalus melanopogon* or *Acrocephalus scirpaceus* a single audio file will be divided in several windows, in some others such as *Ixobrychus minutus* it is very likely that the audio has to be repeated in order to fill the window. Also, the mel-scale has been used to emphasise the frequencies of interest, obtaining at the end what is typically referred to as the mel-spectrogam of the audio file.

Fig. 5 shows an example spectrogram for each bird species of the Western Mediterranean Wetland Birds dataset. As it can be seen, spectrograms are typically different for each bird species and follow different patterns, making these representations suitable to train machine learning or deep learning models. All the spectrograms of Fig. 5 represent a 1-s window, and the displayed frequencies range from 0 Hz to 11,025 Hz (all the audios have been resampled to a sampling frequency of 22,050 Hz).

3.3. Fine-tuning deep learning models on the WMWB dataset

As a companion to the Western Mediterranean Wetland Birds dataset, we also provide a set of deep learning models that have been fine-tuned on our dataset, so that they can be used as a reference baseline by researchers willing to use it to develop new bird species classification models. These models follow the rationale of tackling the recognition of bird species from sound as an image classification problem using CNN (Chandu et al., 2020; Florentin et al., 2020).
In the following paragraphs, we describe the models employed, and the process followed to fine-tune them on our dataset.
The network models included in the testbed are: *i)* VGG16 (Simonyan and Zisserman, 2015), a seminal and relatively shallow deep network that is included in the analysis as a classic baseline reference, *ii)*

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

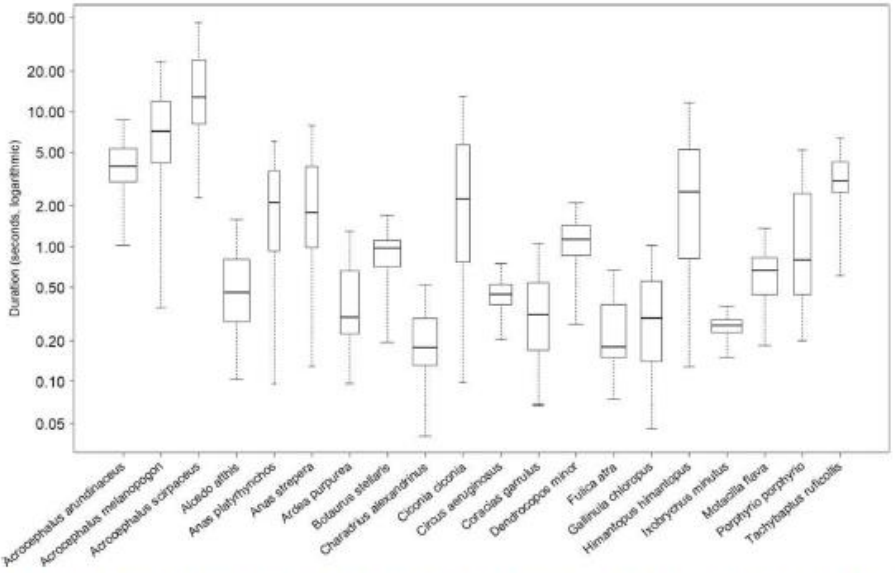

Fig. 2. Boxplot representing the duration in seconds of each audio fragment of the Western Mediterranean Wetland Birds dataset, divided per species. The width of each bar represents the amount of audio fragments from that species in the dataset. Notice that the vertical axis is in logarithmic scale.

ResNet50 (He et al., 2016), a very deep and complex network that is well-known for obtaining high accuracy in diverse classification problems, and two small-footprint networks that can be of interest for developing low cost bioacoustic monitoring devices: *iii)* MobileNetV2 (Howard et al., 2017), a network especially designed for being embedded in mobile devices and *iv)* EfficientNet-B0, a mobile-sized network of the EfficientNet CNN family (Tan and Le, 2019), a new generation of networks that obtain higher accuracy values (compared to older CNNs) in several classification problems, such as ImageNet.

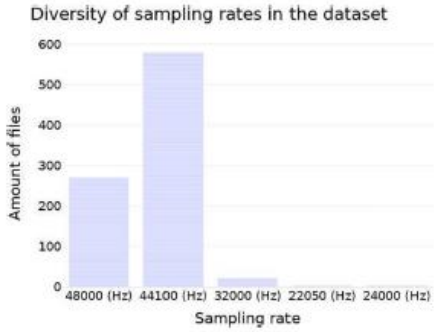

Fig. 3. Diversity of sampling rates on the dataset. *Coracias garrulus, Himantopus himantopus, Tachybaptus ruficollis* (24,000 Hz) and *Acrocephalus arundinaceus, Acrocephalus scirpaceus, Botaurus stellaris* (22,050 Hz).

Table 2
Diversity of the number of channels of the audio files of the dataset.

Number of channels	Mono	Stereo
Amount of files	245	634

A summary of the most important characteristics of the four models is presented in Table 3, including the model footprint size in Megabytes, the number of tunable parameters and the topological depth of their architecture.

At this point, it is worth justifying that the BirdNET network –a reference deep learning model in the area of bird species audio classification– is excluded from the comparison because it has been trained on audios from Xeno-Canto (Kahl et al., 2021), which is the same source used for building the Western Mediterranean Wetland Birds dataset. Thus, there is a non-negligible risk of data leakage that would inevitably flaw any comparison between the models fined-tuned on the WMWB dataset and BirdNET.

As the starting point of the CNN model fine-tuning on the WMWB dataset, we took the models already pre-trained on the ImageNet data set (Deng et al., 2009) available in the Applications API of Keras (Chollet et al., 2015)).

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

As for the input data fed into the CNNs, it must have a size of (X, Y, 3), meaning it needs three input channels. As spectrograms are only 2D, and in our case of size (224, 224), we need to triplicate the layers into the third dimension, thus obtaining an input size of (224, 224, 3) for each spectrogram image.
As the convolutional layers of the pre-trained models have already learned and tuned filters, fine-tuning is per- formed by freezing all layers in the body of the network and then training only the new fully- connected head as a warm-up phase using part of the spectrogram dataset described in Section 3.1. After this first phase, the original layers are unfrozen and another training phase with a smaller learning rate is carried out to increase the overall accu- racy. This second phase was only carried out for models ResNet50 and MobileNetV2, as we did not see a significant improvement in the evaluation metrics of the rest of the models when designing the experiments pipeline.

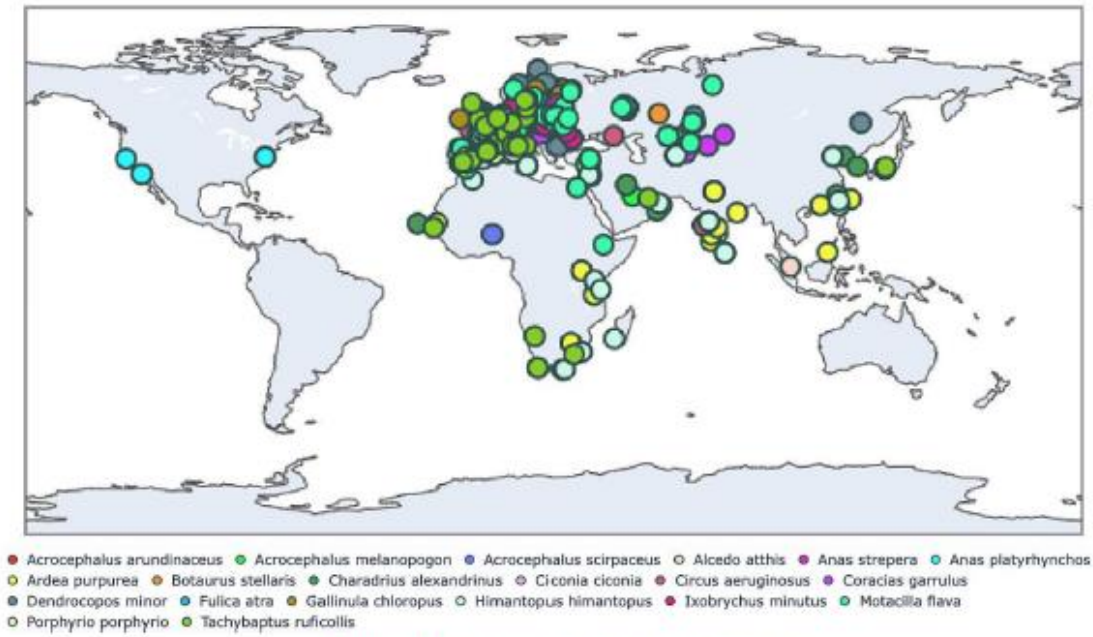

Fig. 4. Map with the original geographic locations of the recordings.

As for the tuning of the network hyperparameters, the final selected values are presented in Table 4 for reproducibility. These values are the result of multiple tuning experiments leading to the maximum accuracy while avoiding overfitting. In additional experiments not reported here, we also applied the ADAM optimizer, obtaining very similar results to those presented in Table 4. Notice that the number of training epochs for the warm-up phase ranges between 20 and 400 to give enough time to train the parameters of the new top layers of the network. It is also worth mention ing that between the two fully-connected top layers of the architectures of all the models, a dropout (Srivastava et al., 2014) with a rate of 0.5 has been employed to provide regularisation.
Also, early stopping was applied to stop the training phase when the validation loss was not decreasing anymore (patience of 20 epochs in the first phase and patience of 3 epochs in the *sec*ond one). This parameter was especially important in the training phase of the models that were set up with more epochs, which are VGG16 and EfficientNet- B0.

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *E*cological Informatics *2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

The evaluation of the performance of the deep neural network models has been made in terms of the following four scalar metrics: precision, recall, F1-measure and accuracy (Tharwat, 2021). To obtain global scores for each classifier, the macroaveraging strategy is used, which consists in computing the aforementioned metrics for each class, and then averaging over all the classes.
To ensure that the results are statistically significant, a stratified 5- fold cross-validation strategy has been applied. All spectrograms coming from the same Xeno-Canto audio file will only belong to a specific subset (fold) on each iteration to avoid giving our neural network model an advantage when classifying spectrograms outside the training stage, which would result in a falsely improved accuracy score. This means that a single Xeno-Canto file, after it has been divided in different sub-files when creating the Western Mediterranean Wetland Birds dataset, will be placed only on a training fold or testing fold, but never in both of them. Finally, Table 4 shows the average training time of the 5 folds for each model. All tests have been carried out with a computer with an AMD Ryzen Threadripper 3970X32-Core Processor, 64 GB of RAM, a Samsung 980 EVO SSD with 1 TB of capacity, and a GeForce RTX 3090 GPU with 24 GB of RAM.

4. Results and discussion

4.1. The Western Mediterranean Wetland birds dataset

The first and main outcome of this work is the annotated Western Mediterranean Wetland Birds dataset, which is available to the scientific community in two forms. First, we share the total of 5795 annotated audio clips (in *.mp3* format) generated from a source of 879 recordings, adding up to a total of 201.6 min (12,096 s) of vocalizations of different lengths, alongside with their corresponding annotations (in *.txt* format).
And second, we also share the Mel spectrogram version of the dataset, where each image corresponds to 1-s window of the original audio, resulting in a total of 17,536 spectrogram images stored as NumPy arrays.
These two versions of the brand new annotated WMWB dataset are available for download at https://doi. org/https://doi.org/10. 5281/zenodo.7505820.

4.2. Fine-tuned deep learning models performance analysis

The second outcome of this work is the set of four deep learning models that have been fine-tuned on the WMWB dataset. They are publicly available at https://github.com/jogomez97/bird-ml-classification.

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

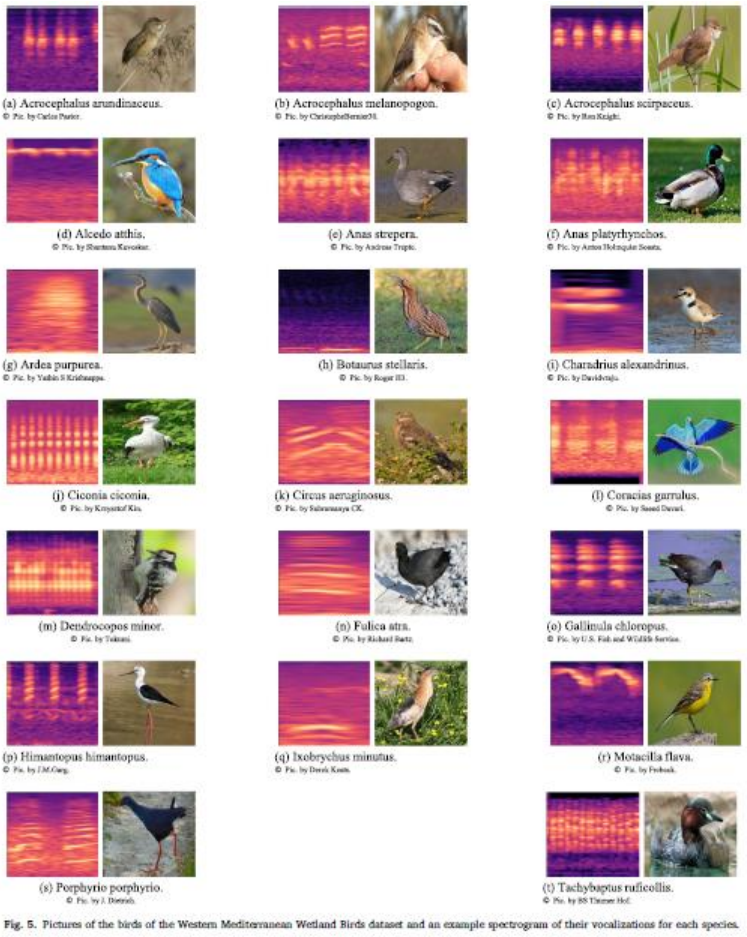

Fig. 5. Pictures of the birds of the Western Mediterranean Wetland Birds dataset and an example spectrogram of their vocalizations for each species.

This section presents a comparative analysis of each individual model performance that aims at serving as a baseline for further developments. To provide the reader with an at-a-glance comparison between all the evaluated models, Table 5 presents their macroaveraged recall, precision, F1-score and accuracy values averaged across the 5 folds of the stratified cross- validation experiments. Moreover, to show how each model performs compared to the others, the rightmost column of Table 5 presents the average ranking of each model averaged across all the stratified 5-fold cross-validation experiments (the lower the rank value, the better the performance). One first issue worth noting is that the EfficientNet-B0 model achieves the best average ranking and outperforms older networks like VGG16 and ResNet50 as expected. This result corroborates previous studies (Tan and Le, 2019), and illustrates how the evolution of network architectures allows to obtain increased accuracies in complex classification tasks. Focusing on the performance of the two networks with the smallest footprint (EfficientNet-B0 and MobileNetV2), it can be observed that they perform really well. When compared between them, their performances are very similar: in none of the evaluation metrics employed the difference between both small-footprint models exceeds 0.3%, and their average ranking is also very similar. When compared to the larger models, it can be observed that both small-footprint networks *i)* outperform by a large margin the heaviest model in the testbed (VGG16, 528 MB in size)), and *ii)* obtain better results than the heavier ResNet-50 network, despite their footprints are 3 to 7 times smaller. A deeper analysis of the

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

Table 3
Summary of the architecture features of the VGG16, ResNet50, MobileNetV2 and EfficientNet-B0 CNNs.

Model	Size	Parameters	Depth
VGG16	528 MB	138,357,544	23
ResNet50	98 MB	25,636,712	50
MobileNetV2	14 MB	3,538,984	88
EfficientNet-B0	29 MB	4,049,571	132

Table 4
Training hyperparameters and average time that it took to complete the training for each fold for all the models.

Hyperparameter	VGG16	ResNet-50	MobileNetV2	EfficientNet-B0
Batch size	16	32	32	32
# Epochs 1 (warm-up phase)	400	20	20	100
Optimizer 1 (warm-up phase)	SGD	RMSprop	RMSprop	SGD
Learning rate 1 (warm-up phase)	0.001	0.001	0.001	0.001
# Epochs 2	–	50	50	–
Optimizer 2	–	SGD	SGD	–
Learning rate 2	–	0.0001	0.0001	–
Training time (sec.)	5368.4	1224.02	1588.40	1110.18

Table 5
Comparison of the performance of the models after the 5-fold cross-validation in terms of macroaveraged recall, precision, F1-score, accuracy and average ranking. The results are the average of the 5-folds, ± the standard deviation.

Model	Recall	Precision	F1-score	Accuracy	Avg ranking
VGG16	69.7% ± 5%	78.7% ± 3%	72.7% ± 4%	77.7% ± 3%	5 ± 0
ResNet50	92.6% ± 4%	93.8% ± 2%	92.4% ± 3%	93.7% ± 3%	2.9 ± 0.1
MobileNetV2	94.4% ± 2%	95.3% ± 2%	94.7% ± 2%	94.8% ± 2%	2.6 ± 0.2
EfficientNet-B0	94.7% ± 2%	95% ± 2%	94.6% ± 2%	94.5% ± 2%	2.4 ± 0.3

classification results obtained by each network confirms the observed trend in perfor- mance. In particular, we have analyzed how accurately each bird species is classified by each model in the testbed. In the case of VGG16, 10 out of the 20 classes are classified with an accuracy greater than the average accuracy of this model (77.7%). From the remaining 10 classes, there are five that achieve an accuracy greater than 50%, while the other five are classified with even lower accuracies: *Ciconia ciconia* (19.7%), *Himantopus himantopus* (20.4%), *Acrocephalus melanopogon* (47.4%), *Acrocephalus arundinaceus* (42.6%) and *Charadrius alexandrinus* (27.5%). A detailed analysis reveals that VGG16 tends to classify the classes that have the poorest results as *Ardea purpurea* (as a matter of fact, the model classifies more samples of *Ciconia ciconia* as *Ardea purpurea* than in the correct category). As regards *Himantopus himantopus* and *Charadrius alexandrinus*, we observe more diversity in its errors. As for the former, it is one of the classes with more data (in seconds) but it has an amount of cuts similar to other categories, meaning that the vocalizations of this bird are usually longer. Therefore, we conjecture that in this case the model would have performed better using a longer window size. As for *Charadrius alexandrinus*, it is one of the classes with less samples in the dataset. Moreover, it comprises a mix of calls and songs, so the low accuracy is probably due to fact that the VGG16 model fails to generalize due to the relatively small amount of data available in this category. The classification performance of ResNet50 is far superior than VGG16, achieving macroaveraged recall and precision scores of 92.6% and 93.8% respectively, and a F1-score of 92.4%. With the ResNet-50 model, 9 out of 20 classes have an accuracy higher than the average (93.7%). Moreover, all the classes score above the 70% mark, which represents a great improvement with respect to the VGG16 model. Focusing on the two bird species that obtained the worst scores in the VGG16, we observe that both of them signifi- cantly increase their accuracy (for *Ciconia Ciconia*, the accuracy increases from 19.7% to 70.2% and for *Himantopus Himantopus*, from 20.4% to 83%). This reinforces the notion that the VGG16 model was unable to generalize these classes, while the more complex ResNet50 architecture is able to do so. However, for the ResNet50 model, the class with the lowest accuracy score is still *Ciconia Ciconia*, which is mainly confused with *Himantopus Himantopus*. Despite being the smallest footprint network in our testbed, the MobileNetV2 model achieves better results than the two previous networks. A per-class analysis shows that 9 classes are classified better than its average accuracy, and actually all the

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

classes achieve an accuracy greater than 80%. The species with the lowest accuracy is again *Ciconia Ciconia*, with an 83.4% accuracy (note that it was 19.7% and 70.2% for VGG16 and ResNet50 models, respectively), confusing it with *Acrocephalus arundinaceus* (which is the same confusion that occurred in model VGG16). Finally, as regards the top-performing network in our testbed, EfficientNet-B0, it is especially important to high- light that, as regards the individual class performance, only *Himantopus himantopus* has an accuracy below 90% (84.7%) with this model. That is, EfficientNet-Bo achieves the most stable classification results across categories among the networks in our experimental testbed. To complement this detailed performance study of the four tested models, we also analyze their accuracy stability. The reason is that the higher the accuracy across classes, the more likely is that they can be used to build robust and reliable bioacoustic monitoring systems. To this end, Fig. 6 depicts the macro-averaged accuracy histograms of the 20 classes for the four models, showing their distribution. The less scattered the histogram, the greater the stability of the model. It can be observed that MobileNetV2 and EfficientNet-B0 have the most compact histograms. This indicates that their performance is the most stable and balanced of the networks in the testbed, which is an interesting feature of a classifier in terms of robustness and reliability. Also, all the models seem to have struggled with more or less the same species (*Ciconia ciconia*, *Himantopus himantopus*, *Acrocephalus arundinaceus*). Considering that these three classes are among the categories with more seconds of samples in the dataset, these might be a warning or an indicator of potential class-imbalance problems for the researchers using the dataset. Despite the main focus of this work is not on the performance of the deep neural networks fine-tuned on the WMWB dataset, it is worth noticing that the classification performance of the best models in our testbed –despite they have not been optimized– is in line with those

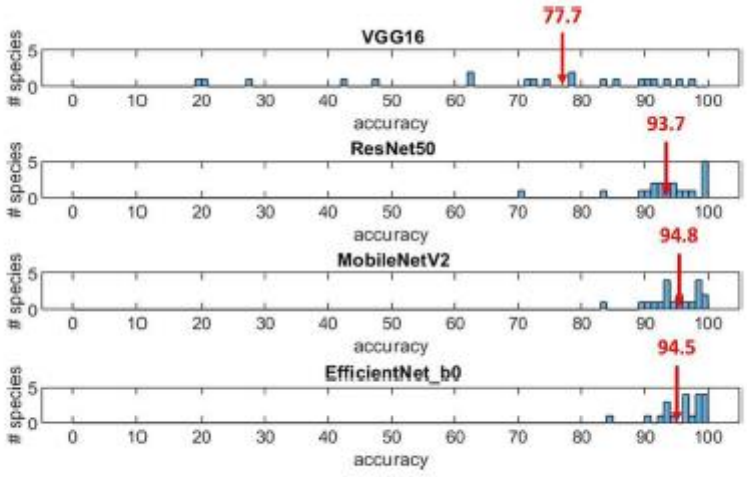

Fig. 6. Accuracy score histograms for the four models. The average accuracy of each model is indicated in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

presented in recent works in this area that operate on datasets of comparable complexity. Examples include the classification of 25 species of the Furnariidae family using multilayer perceptrons, support vector machines and random forest, achieving accuracies around 90% of (Albornoz et al., 2017), the 96% accuracy obtained by a CNN on a dataset of 21 species in (Kucuktopcu et al., 2019), the classification of 16 birds species using multi-view feature fusion in (Xie et al., 2022) which reported

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

accuracies around 96%, the ResNet50 model evaluated in (LeBien et al., 2020) on 13 species that achieved 82.5% average precision, or the modified ResNet50 model of (Morales et al., 2022) that classified 16 species with an average precision of 79%. These results suggest that the WMWB dataset is a good starting point for building audio-based bird species classification systems.
Another interesting conclusion of our work is that even non-optimized small-footprint networks can achieve excellent classification performance. This fact, which had been previously confirmed in other classification domains (Tan and Le, 2019), paves the way for the development of low cost and 'intelligent' bioacoustic monitoring devices.
And finally, it is important to highlight that, in most of those works, a strongly annotated version of the dataset is not directly available (at best, the data is available upon request). This is a distinctive feature of our work, which reflects our commitment to reproducible open science.

5. Conclusion

As the limited availability of annotated data is one of the main hurdles in the field of bioacoustic event classification for biodiversity monitoring (Vidaña-Vila et al., 2017), in this work we have introduced the annotated Western.
Mediterranean Wetland Birds dataset containing sounds of 20 endemic bird species of the Aiguamolls de l'Empordà Natural Park, encouraging the community to use it to foster research on this topic.
Besides the dataset itself, another relevant contribution of this work is the critical comparative analysis of the performance of four deep learning models pre-trained on image classification tasks and fine-tuned on the brand new WMWB dataset.
These results demonstrate that i) the current advances in the field of deep learning allow obtaining high accuracies with lightweight architectures, and ii) thanks to these advances, it is possible to design low cost bioacoustic monitoring devices (i.e. resource constrained) that perform reliably thanks to embedded audio classification capabilities provided by small footprint deep neural networks. We believe this conclusion will be useful for ecologists willing to develop bird species classification systems powered by deep learning. In this sense, our work reinforces the evidences that pave the way for the automation of bioacoustic avian life monitoring in scenarios where wireless access to cloud computing facilities is difficult or not possible. As mentioned earlier, we are making the Western Mediterranean Wetland Birds dataset available to the public to encourage the community to use it to benchmark new algorithms for bird species classification, using the deep learning models fine-tuned in this work as a baseline for performance analysis. In this context, we would like to suggest several lines of future research. First, by publishing not only the spectrograms but also the original audio files, we encourage the use and combination of different spectrogram configurations, as they have been recently proven to better classification performance (Knight et al., 2020). Second, we suggest the use of data augmentation techniques on the Western Mediterranean Wetland Birds dataset, such as noise mixing, time shifting, or mixup (Stowell, 2022), as it could be a way for obtaining even more accurate models. Also, as our dataset is suitable for laboratory experiments, taking steps towards a real-world deployment would require denoising and source separation preprocessing steps prior

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

to classification, as it is highly likely that bird vocalizations of different individuals or species are mixed. Finally, we consider that another interesting continuation of this work would be the creation of ensembles of networks as a means to improve classification and reduce the confusion between the bird species that obtained the poorest scores.

Declaration of Competing Interest

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

Data availability

The dataset can be downloaded from the following repository https://doi.org/10.5281/zenodo.7505820. The deep learning models can be downloaded from the following repository https://github.com/ jogomez97/ bird-ml-classification.

Acknowledgements

Authors would like to thank the Departament de Recerca i Universitats (Generalitat de Catalunya) under Grant Ref. 2021 SGR 01396. They would also like to acknowledge Albert Brugas Riera and Sergio Romero de Tejada Mart'ınez from the Aiguamolls de l'Empord`a Natural Park for their valuable help when defining the species of interest to be classified. Also, authors would like to acknowledge all the Xeno-Canto community and their contributors for making the creation of the dataset possible. Specially, authors would like to thank the following contributors for giving us special permission to use their recordings in this work despite having uploaded them on the Xeno-Canto portal under the terms BY-NC-ND: Anh¨auser, Arnold Meijer, Bodo Sonnenburg, Chie-Jen Jerome Ko, Ding Li Yong, Eveny Luis, Fernand Deroussen (Sonoth`eque du MNHN), Hans Matheve, Herman van der Meer, Itziar Guti´errez, Jacques Prevost, Jarek Matusiak, J´er´emy Simar, Joost van Bruggen, Krzysztof Deoniziak, Lars Lachmann, Mandar Bhagat, Marc Anderson, Marco Dragonetti (www.birdsongs.it), Matthias, Feuersenger, Maudoc, Niels Krabbe, Patrick Franke, Peter Boesman, Piotr Szczypinski, Ruud van Beusekom.

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

References

Albornoz, E.M., Vignolo, L.D., Sarquis, J.A., Leon, E., 2017. Automatic classification of Furnariidae species from the Paranaense Littoral region using speech-related features and machine learning. Ecol. Inform. 38, 39–49.
Baker, M.C., Cunningham, M.A., 1985. The biology of bird-song dialects. Behav. Brain Sci. 8 (1), 85–100.
Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X.Z., Raich, R., Hadley, S.J., Hadley, A. S., Betts, M.G., 2012. Acoustic classification of multiple simultaneous bird species: a multi-instance multi-label approach. J. Acoust. Soc. Am. 131 (6), 4640–4650.
Chandu, B., Munikoti, A., Murthy, K.S., Murthy, G., Nagaraj, C., 2020. Automated bird species identification using audio signal processing and neural networks. In: Proc. International Conference on Artificial Intelligence and Signal Processing (AISP), pp. 1–5.
Chollet, Francois, et al., 2015. Keras. Github repository. [online]. [accessed on November 2022]. https://github.com/keras-team/keras. de Oliveira, A.G., Ventura, T.M., Ganchev, T.D., de Figueiredo, J.M., Jahn, O., Marques, M.I., Schuchmann, K.L., 2015. Bird acoustic activity detection based on morphological filtering of the spectrogram. Appl. Acoust. 98, 34–42.
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A Large-scale Hierarchical Image Database. http://image-net.org/.
Denton, T., Wisdom, S., Hershey, J.R., 2022. Improving bird classification with unsupervised sound separation. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 636–640. Fatoric, S., Morén-Alegret, R., 2013. Integrating local knowledge and perception for assessing vulnerability to climate change in economically dynamic coastal areas: the case of natural protected area Aiguamolls de l'Emporda, Spain. Ocean Coast. Manag. 85, 90–102.
Florentin, J., Dutoit, T., Verlinden, O., 2020. Detection and identification of European woodpeckers with deep convolutional neural networks. Ecol. Inform. 55, 101023.
Franzen, A., Gu, I.Y.H., 2003. Classification of bird species by using key song searching: a comparative study. In: Proc. IEEE International Conference on Systems, Man and Cybernetics, 1, pp. 880–887. Frommolt, K.-H., Tauchert, K.-H., 2014. Applying bioacoustic methods for long-term monitoring of a nocturnal wetland bird. Ecol. Inform. 21, 4–12. https://doi.org/ 10.1016/j.ecoinf.2013.12.009.
Garnett, S., Szabo, J., Dutson, G., 2011. The action plan for Australian birds 2010. In: Victoria (Australia). CSIRO Publishing.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.
Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861.
Kahl, S., Wood, C.M., Eibl, M., Klinck, H., 2021. BirdNET: A deep learning solution for avian diversity monitoring. Ecol. Inform. 61, 101236.

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

Klingbeil, B.T., Willig, M.R., 2015. Bird biodiversity assessments in temperate forest: the value of point count versus acoustic monitoring protocols. PeerJ. 3, e973. Knight, E.C., Hernandez, S.P., Bayne, E.M., Bulitko, V., Tucker, B.V., 2020. Pre-processing

spectrogram parameters improve the accuracy of bioacoustic classification using convolutional neural networks. Bioacoustics 29 (3), 337–355.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60 (6), 84–90.
Kucuktopcu, O., Masazade, E., Ü nsalan C, Varshney PK., 2019. A real-time bird sound recognition system using a low-cost microcontroller. Appl. Acoust. 148, 194–201.
LeBien, J., Zhong, M., Campos-Cerqueira, M., Velev, J.P., Dodhia, R., Lavista-Ferres, J., Aide, T.M., 2020. A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. Ecol. Inform. 59, 101113.
Lepage, D., 2021. Avibase — The World Bird Database. Available online: http://avibase. bsc-eoc.org/avibase.jsp (accessed on November 2022).
Lostanlen, V., Salamon, J., Farnsworth, A., Kelling, S., Bello, J.P., 2018. Birdvox-full-night: a dataset and benchmark for avian flight call detection. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 266–270.
Macaulay, 2022. The World's Premier Scientific Archive of Natural History Audio, Video, and Photographs. https://www.macaulaylibrary.org/about/history.
McGuire, P., 2022. BirdNET-Pi. https://github.com/mcguirepr89/BirdNET-Pi.
Morales, G., Vargas, V., Espejo, D., Poblete, V., Tomasevic, J.A., Otondo, F., Navedo, J. G., 2022. Method for passive acoustic monitoring of bird communities using UMAP and a deep neural network. Ecol. Inform. 72, 101909.
Morfi, V., Bas, Y., Pamula, H., Glotin, H., Stowell, D., 2019. NIPS4BPLUS: a richly annotated birdsong audio dataset. PeerJ Comp. Sci. 5, e223.
Nicholson, D., Queen, J.E., Sober, S.J., 2022. Bengalese finch song repository. Figshare 7 (18 Oct. 2022). https://doi.org/10.6084/m9.figshare.4805749.v7.
Nolasco, I., Singh, S., Vidana-Vila, E., et al., 2022. Few-shot bioacoustic event detection at the DCASE 2022 challenge. ArXiv. https://doi.org/10.48550/ARXIV.2207.07911.
Pamula, H., 2022. Nocturnal flight calls dataset: long-term acoustic monitoring of birds migrating at night (1.0) [data set]. Zenodo. https://doi.org/10.5281/ zenodo.6359955.
Parrilla, A.G.A., Stowell, D., 2022. Polyphonic sound event detection for highly dense birdsong scenes. DCASE 2022, 146–150.
Priyadarshani, N., Marsland, S., Castro, I., 2018. Automated birdsong recognition in complex acoustic environments: a review. J. Avian Biol. 49 (5), jav–01447.
Rai, P., Golchha, V., Srivastava, A., Vyas, G., Mishra, S., 2016. An automatic classification of bird species using audio feature extraction and support vector machines. In: 2016 International Conference on Inventive Computation Technologies (ICICT), vol. 1. IEEE, pp. 1–5.
Rosell, C., Nav`as, F., Romero, F., 2013. Reproduction of wild boar in a cropland and coastal wetland area: implications for management. Anim. Biodivers. Conserv. 35 (2), 209–217.

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

Rosenstock, S.S., Anderson, D.R., Giesen, K.M., Leukering, T., Carter, M.F., Thompson III, F., 2002. Landbird counting techniques: current practices and an alternative. Auk. 119 (1), 46–53.
Salamon, J., Bello, J.P., Farnsworth, A., Robbins, M., Keen, S., Klinck, H., Kelling, S., 2016. Towards the automatic classification of avian flight calls for bioacoustic monitoring. PLoS One 11 (11), e0166866.

Selin, A., Turunen, J., Tanttu, J.T., 2006. Wavelets in recognition of bird sounds. EURASIP J. Adv. Sign. Process. 2007, 051806.
Shonfield, J., Bayne, E.M., 2017. Autonomous recording units in avian ecological research: current use and future applications. Avian Conserv. Ecol. 12 (1), 14.
Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations (ICLR 2015). 1–14.
Singh, S., Pankajakshan, A., Benetos, E., 2019. Audio tagging using linear noise modelling layer. In: Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019).
Sprengel, E., Jaggi, M., Kilcher, Y., Hofmann, T., 2016. Audio based bird species identification using deep learning techniques. In: CLEF (Working Notes), pp. 547–559.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Rec. 15, 1929–1958.
Stowell, D., 2022. Computational bioacoustics with deep learning: a review and roadmap. PeerJ 10, e13152.
Stowell, D., Plumbley, M.D., 2014. freefield1010 - an open dataset for research on audio field recording archives. In: Proceedings of the Audio Engineering Society 53rd Conference on Semantic Audio (AES53). Audio Engineering Society.
Stowell, D., Wood, M.D., Pamuła, H., Stylianou, Y., Glotin, H., 2019. Automatic acoustic detection of birds through deep learning: the first bird audio detection challenge. Methods Ecol. Evol. 10 (3), 368–380.
Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., Kelling, S., 2009. eBird: a citizen-based bird observation network in the biological sciences. Biol. Conserv. 142, 2282–2292.
Sun, C., Shrivastava, A., Singh, S., Gupta, A., 2017. Revisiting unreasonable effectiveness of data in deep learning era. In: Proc. 2017 IEEE International Conference on Computer Vision (ICCV), pp. 843–852.
Tan, M., Le, Q.V., 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. ArXiv. DOI: 10.48550/ARXIV.1905.11946.
Tharwat, A., 2021. Classification assessment methods. Appl. Comp. Inform. 17 (1), 168–192.
Tseng, Y.C., Eskelson, B., Martin, K., LeMay, V., 2020. Automatic bird sound detection: logistic regression based acoustic occupancy model. Bioacoustics. https://doi.org/10.1080/09524622.2020.1730241.
Vidana-Vila, E., Navarro, J., Alsina-Pag`es, R.M., 2017. Towards automatic bird detection: an annotated and segmented acoustic dataset of seven Picidae species. Data 2 (2), 18.

This is a post-print (final draft post-refeering)
Published in final edited form as
Western Mediterranean wetland birds dataset: A new annotated dataset for acoustic bird species classification *Ecological Informatics 2023. Vol. 75.*
https://doi.org/10.1016/j.ecoinf.2023.102014

Vidaña-Vila, E., Navarro, J., Alsina-Pagès, R.M., Ramírez, A., 2020. A two-stage approach to automatically detect and classify woodpecker (Fam. Picidae) sounds. Appl. Acoust. 166, 107312.
Wijers, M., Loveridge, A., Macdonald, D.W., Markham, A., 2021. CARACAL: a versatile passive acoustic monitoring tool for wildlife research and conservation. Bioacoustics 30 (1), 41–57.
Xeno-Canto Foundation, 2022. Xeno-Canto: Sharing Bird Sounds from around the World. Available online. http://www.xeno-canto.org/ (accessed on November 2022).
Xie, S., Lu, J., Liu, J., Zhang, Y., Lv, D., Chen, X., Zhao, Y., 2022. Multi-view features fusion for birdsong classification. Ecol. Inform. 72, 101893.