Testing the Accuracy of a BirdNET, Automatic bird song Classifier

Mehak Arif Richard Hedley Erin Bayne

July 31, 2020

ABSTRACT

In recent years, automated bird song classification programs have been becoming more common among researchers as a way to study, track, and monitor birds. In our research, we tested the accuracy of one such program called BirdNET. We tested 225 recordings by uploading them to BirdNET and manually classifying them to see how often BirdNET was accurate. The overall accuracy of BirdNET was 91.5%, with this number increasing when it came to bird songs that BirdNET was more familiar with, and dropping when it came to other bird songs that BirdNET was unfamiliar with. This paper will explore why such a program is needed, how it can be helpful to biologists, researchers, and anyone else interested in or looking to learn more about bird songs. This study also includes the methods used to test BirdNET, discussion about how automated bird song recognition programs can be improved, limitations when it comes to automated bird song recognition software, and other relevant studies about acoustic monitoring and automatic bird recognition programs.

INTRODUCTION

Biologists and researchers are constantly looking for different ways to study birds. Autonomous recording units (ARUs) are one way of studying and tracking birds. ARUs are small computers that record acoustically active birds in nature (Shonfield & Bayne 2017). ARUs are useful as they allow for the recording of birds with minimal human interference, a permanent record of data that can be referenced anytime, and reduced observer bias (Shonfield & Bayne 2017). Some drawbacks of ARUs include the cost of equipment, storage needed for the recordings, and possible loss of data if units fail (Shonfield & Bayne 2017). The main concern when it comes to studying birds using ARUS is the large amounts of audio recordings they generate. With the use of this technology, our lab has collected recordings that would take more than 2.5 human lifetimes to go through. Manually classifying the bird songs in these recordings is not very practical, therefore, automated bird recognition programs are being implemented where the computer can classify the bird songs instead. Currently, the adoption of these softwares is restricted by a lack of direct comparisons with manual classification (Digby *et al.,* 2013), which is why our research hopes to draw more insight into automated bird song recognition methods.

There are two main reasons why such an automated bird recognition program could be useful: for scientific purposes and for the use of the general public. With recordings from ARUs combined with automatic bird recognition software, researchers can track and monitor acoustically active birds over long periods of time. In addition, such technology also allows researchers to remotely study populations in non-invasive ways, describe the soundscape (complete acoustic environment consisting of biotic and abiotic sounds), study species interactions, and gain insights into social dynamics of acoustically active animals (Blumstein *et al.,* 2011).

Researchers can also determine and track effects of anthropogenic sounds (human-generated noise), climate change, habitat fragmentation, (Blumstein *et al.,* 2011). The second purpose of such a program is for the use of the general public. An easily accessible bird song recognition website or app could be used by the general public to classify the birds they hear in their everyday lives. Such a program could be key for getting the public interested in nature and, therefore, conservation.

The Cornell Lab of Ornithology and the Chemnitz lab University of Technology developed a program “focused on the detection and classification of avian sounds using machine learning” (birdnet.cornell.edu). In simple terms, BirdNET is a tool where users can upload recordings and the program will automatically classify what bird species are making the sounds in the uploaded recording (Fig. 1). BirdNET could be a solution for classifying bird songs in the recordings the ARUs collect in a much more efficient manner. Not only is BirdNET a website, but it is also an app available for anyone to download and could help peak the public’s interest in birds, nature, and conservation. The main objective of this research is to test BirdNet’s algorithms by directly comparing it to manual classification and determining its accuracy.

METHODS

BirdNET uses an artificial neural network to rank the most likely bird species in each recording. It also assigns a probability to each species, indicating the algorithm's level of confidence in that classification (Fig. 2). Currently, BirdNET is capable of classifying 984 of the most common species in North America and Europe.

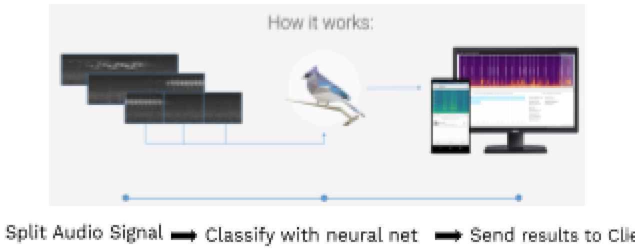

Fig. 1. Basic description of how BirdNET works (Via BirdNET’s home page available at https://birdnet.cornell.edu/).

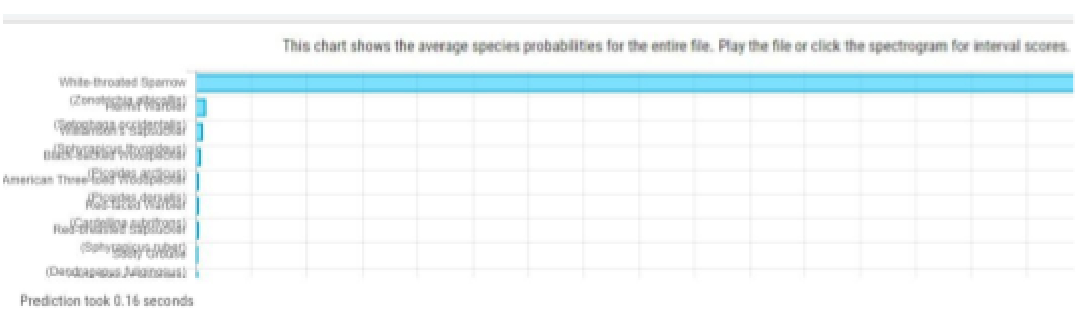

Fig. 2. In this recording, BirdNET has identified the bird sound is most likely a white-throated sparrow (zonotrichia albicollis).

The main purpose of BirdNET is to teach a computer how to recognize birds from sounds and “assist experts and citizen scientists in their work of monitoring and protecting” birds (birdnet.cornell.edu). During the testing phase of this research, we extracted 7-second clips from longer recordings by scanning for sections that met a 70dB sound level threshold. This was to ensure we were only uploading recordings with loud bird sounds as compared to more faint ones. This is because faint sounds are overlapped and more difficult to hear and our objective was to test the classification performance of BirdNET rather than how well it works on faint sounds. Then, we opened the recordings in a program called Praat, to determine if the loudest sound was a bird or something else like traffic or airplane noise (Boersma 2001). In this program, not only could we listen to the recordings, but it also put the recordings on a spectrogram. A spectrogram is a visualization of sounds. On the x-axis, it shows time, and on the y-axis, it shows the frequencies in Hz (Fig. 3). Spectrograms can be used to visualize the frequencies over time. The method we used to determine if the recording was a bird sound or something else was by looking at the frequencies on the spectrogram. If it was a bird sound, the dominant frequencies would be higher, as birds tend to vocalize at relatively high frequencies, as compared to rain, wind, cars, etc which would be at much lower frequencies. The final steps included uploading the recording to BirdNet and taking a screenshot of the output and saving it to a folder for future reference. This process was repeated with a sample of recordings from each recording unit to get a variety of different species of birds and therefore a better picture of how accurate the program is. Another member of the team with experience with bird songs went through the recordings and determined what species was in each recording manually. All of the information was recorded in a spreadsheet to compare results at the end.

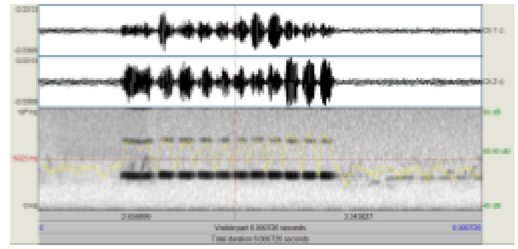

Fig. 3. A screenshot of one of the recordings opened in Praat. The dark lines on the top are called the waveform. The gray picture along the bottom is called the spectrogram. The yellow lines over the spectrogram show the overall loudness of the sound at each moment in the clip. The song in this clip is the whistle song of the white-throated Sparrow.

RESULTS

There were a total of 225 detections made by BirdNET. While there were 205 actual recordings that we tested, for some of the recordings, BirdNET picked up more than one species of bird. Of these 225 detections, 36 (16%) of recordings were unable to be identified manually as they were brief bird calls rather than bird songs that were too short to make a confident assessment. Not taking into account the 16% of recordings we were not able to identify manually, BirdNET had an overall accuracy of 91.5%. Most recordings had a few different bird songs, and in 82.0% of the recordings, BirdNET was able to classify the loudest sounds. Tennessee warblers (Leiothlypis peregrina) were one of the most common detections, with 38.6% of total detections being Tennessee warbler. In total, 100% of the Tennessee Warbler detections were accurate. Another species that was common in the recordings was the white-throated Sparrow, which made up 16.4% of the detections. Only taking into consideration the recordings that we were able to be manually identified, BirdNET was able to detect 100% of white-throated Sparrow bird songs. In conclusion, when it came to the bird songs that had 20+ detections, like the Tennessee warbler and white-throated sparrow, BirdNET still maintained high accuracy levels. One of the species that did not show up as often on the recordings, was the American Redstart (Setophaga ruticilla). The American redstart made up 1.6 % of the total detections, of which 66.7% were accurate. There were a total of 23 different bird species that both BirdNET and manual identification classified. The summary of results for the species detection are found in table 1.

Table 1. Number of detections and accuracy percentage for the bird species.

Species Name	# of Detections	Accuracy %
Tennessee Warbler	73	100%
White-throated Sparrow	31	100%
Ovenbird	17	100%
Winter Wren	10	90%
Bay-breasted warbler	9	100%
Mourning Warbler	6	100%
Northern Waterthrush	5	100%
Hairy Woodpecker	3	66.7%
Chipping Sparrow	2	0%
14 more species combined	33	84.8

*BirdNET identified some species that were not manually identified, therefore contributing to the inaccuracy of the program and percentages in the table may not reflect this.

DISCUSSION

Overall, BirdNET had a 91.5% accuracy rate from the 189 recordings we were able to identify manually. There are a few reasons as to why BirdNET identified 8.5% of the recordings incorrectly. One reason is that the program was developed using recordings mainly made in New York. Birds tend to have dialects (comparable to accents) varying from country to country. Our recordings being from Canada could have been a possible factor that threw off BirdNET's classification system. Another factor could be interfering sounds. While we uploaded recordings that exceed some threshold of loudness, this did not mean they only had one sound. Fainter bird songs, wind, rain, and traffic could have potentially interfered with BirdNET's classification system. Some of our limitations included having a human check the recordings to see if they were a bird song or not. A recommendation would be to have a computer software or program do this instead, which would greatly speed up the entire process and allow for more recordings to be tested. Considering there were a total of 23 species of birds detected in the recordings that we tested, for future research it would be recommended to test on a larger set of recordings. This would give a better idea of how accurate the program is and which species the program is more familiar with and tends to classify correctly compared to others.

Some of the obvious classification errors were geographically implausible as BirdNET would identify species that did not exist in the area the recordings were made. For example, in one of the recordings, BirdNET identified the bird song to belong to the Bay-Breasted Warbler, however, those do not occur in the area where the recordings were made. A recommendation for improvement would be to only allow certain classification to avoid the programs from predicting species that do not occur in the region where the recordings were taken.

Some studies with relevant information will be explored in the rest of this section. According to a study conducted by Schroeder and McRae (2020), managing different factors such as ARU scheduling, improved ability to manipulate parameters within avian detection programs, and with closer attention to quality control, there is potential to locate different bird calls in environments that are acoustically diverse. Another successful study similar to our research was conducted by Evans and Mellinger (1999) relating to monitoring grassland birds in nocturnal migration. Bird sounds were recorded from seven recording stations in New York State and one in Texas, and "bird calls on these audio recordings were later detected by human listening and by automatic sound-detection software" (Evan & Mellinger 1999). They used a signal detection software developed by the Cornell Lab of Ornithology, the same lab that also helped develop BirdNET. Another research paper by Brandes (2008), discusses the need "for increased use and further development" of automated avian sounds recognition, different techniques that can be used softwares to detect and classify avian sounds, ways to improve the performance of such softwarres, and other information relating to acoustic monitoring (Brandes, 2008).

In conclusion, there is a lot of potential for automated bird song recognition programs and such software could drastically change the way researchers study birds. Birds are the very basis of and critical to the maintenance of ecosystems (Brandes, 2008). Efforts towards making avian monitoring more efficient and accurate will significantly improve and advance research on birds and conservation efforts (Brandes, 2008). According to our data, BirdNET has a relatively high accuracy rate and, although it is not a final product yet, BirdNET has a lot of potential to become a tool in the scientific community to better monitor bird population trends and a go-to for people interested in bird songs.

ACKNOWLEDGEMENTS

Thank you to my sponsors, Motorola, the Rotary Club of Edmonton Glenora, and Mr. Prybysh for their generous support. Thank you to Dr. Richard Hedley, the Direct Supervisor, and Dr. Erin Bayne, the Principal Investigator for having me on their team and for their guidance during this project. Finally, thank you to WISEST for organizing this amazing program and giving so many students, including me, the chance to have experiences in STEM fields we love and to have the summer of a lifetime!

UNIVERSITY OF ALBERTA WISEST bioacoustic unit MOTOROLA

REFERENCES

Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., . . . Kirschel, A. N. G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: Applications, technological considerations and prospectus. *The Journal of Applied Ecology, 48*(3), 758-767. doi:10.1111/j.1365-2664.2011.01993.x

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. Glot International 5:9/10, 341-345.

Digby, A., Towsey, M., Bell, B. D., & Teal, P. D. (2013). A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution, 4*(7), 675-683. doi:10.1111/2041-210x.12060

Evans R, W., & David K. Mellinger. (1999). MONITORING GRASSLAND BIRDS IN NOCTURNAL MIGRATION

. *Ecology and Evolution, 9*(1), 1-11. doi:10.1002/ece3.4795

Schroeder, K. M., & McRae, S. B. (2020). Automated auditory detection of a rare, secretive marsh bird with infrequent and acoustically indistinct vocalizations. *Ibis (London, England), 162*(3), 1033-1046. doi:10.1111/ibi.12805

Scott Brandes, T. (2008). Automated sound recording and analysis techniques for bird surveys and conservation. *Bird Conservation International, 18*(S1), S163-S173. doi:10.1017/S0959270908000415

Shonfield, J., & Bayne, E. M. (2017). Autonomous recording units in avian ecological research: Current use and future applications. *Avian Conservation and Ecology, 12*(1), 13. doi:10.5751/ace-00974-120114