



Taneli Voutilainen

# Researching and Testing BirdNET, a Bird Detecting Algorithm

Metropolia University of Applied Sciences

Bachelor of Engineering

Information and Communications Technology

Bachelor's Thesis

6 May 2024

## Abstract

Author: Taneli Voutilainen  
Title: Researching and Testing BirdNET, a Bird Detecting Algorithm  
Number of Pages: 29 pages + 3 appendices  
Date: 6 May 2024

Degree: Bachelor of Engineering  
Degree Programme: Information and Communications Technology  
Professional Major: Smart IoT Systems  
Supervisors: Aarne Klemetti, Senior Instructor

---

This thesis will focus on using BirdNET, a machine learning algorithm designed to detect birds from their vocalisations, testing its efficacy and considering different uses for such a technology, such as a helpful tool for a novice birder.

The process started with getting BirdNET to work on Raspberry Pi's Linux environment and considering possible experiments. The internet has large databases with hundreds of thousands of recordings of different birds, of which a few were chosen. In the experiments, the recordings were played back, captured with a USB microphone on the Raspberry Pi and analysed with BirdNET. The experiments included different volume levels, added environmental noise, and overlapping of bird recordings, to see how the algorithm reacted to varying circumstances. Additional recordings were done in real environments.

The results were highly variable. With near perfect conditions and clearly audible recordings, BirdNET was nearly a hundred percent confident in its predictions. Adding environmental noise, lowering volume levels, and overlapping birds with others made a sizeable negative impact in the algorithm's predictions, sometimes lowering the confidence below the low threshold used. Even with very low prediction rates, the birds are mostly still clearly audible, and a sufficiently competent birder would have no trouble recognising the species.

Even if BirdNET is not perfectly reliable yet, it can be a useful tool. Especially a novice birder, who might not recognise a single bird, can use BirdNET to point them in the right direction. If hours or days long audio files need to be searched for birds, even with low confidence predictions, BirdNET could save large amounts of time by tagging intervals to be checked by a human. BirdNET is in constant development and considering how rapidly the progress of AI has been in the past decade, it will eventually surpass human capabilities.

Keywords: Machine Learning, Artificial Intelligence, Birding

---

The originality of this thesis has been checked using Turnitin Originality Check service.

## Tiivistelmä

Tekijä:	Taneli Voutilainen
Otsikko:	Lintujen tunnistusalgoritmin, BirdNETin, tutkiminen ja testaaminen
Sivumäärä:	29 sivua + 3 liitettä
Aika:	6.5.2024
Tutkinto:	Insinööri (AMK)
Tutkinto-ohjelma:	Tieto- ja viestintätekniikka
Ammatillinen pääaine:	Älykkäät IoT-järjestelmät
Ohjaajat:	Tutkijaopettaja Aarne Klemetti

---

Tässä opinnätetyössä keskityttiin BirdNETin käyttöön, joka on koneoppimisella luotu algoritmi lintujen tunnistamiseen niiden tuottamista lauluista ja äänistä, ja sen tehokkuuteen. Sille pohdittiin myös käyttötarkoituksia, kuten työkaluksi aloittelevalle lintuharrastajalle.

Prosessi aloitettiin panemalla BirdNET toimimaan Raspberry Pi:n Linux-ympäristössä ja mietittiin mahdollisia kokeita algoritmin käytölle. Internetistä löytyy satojatuhansia äänityksiä eri linnuista, joista muutama valittiin. Kokeissa näitä äänityksiä toistettiin, nauhoitettiin Raspberry Pi:hin kiinnitetyllä USB mikrofoniin ja analysoitiin BirdNET:llä. Kokeissa käytettiin eri äänenvoimakkuuksia, lisättiin taustamelua ja toistettiin päällekkäisiä vokalisointeja, jotta näemme miten algoritmi reagoisi erilaisiin tilanteisiin. Ylimääräisiä äänityksiä tehtiin myös oikeissa ympäristöissä.

Tulokset olivat vaihtelevia. Selkeästi kuultavilla nauhoituksilla BirdNET oli melkein täysin varma havainnoistaan. Lisäämällä taustamelua, pienentämällä toistettujen äänityksien äänenvoimakkuutta tai päällekkäisillä lintujen vokalisoinneilla saatiin huomattavasti alennettuja ennustuksia. Joskus BirdNET ei tunnistanut ainuttakaan lintua, vaikka ne olivat selvästi kuultavissa. Kokeneella lintuharrastajalla ei olisi vaikeuksia näiden lintujen tunnistamisessa.

Vaikka BirdNET ei tällä hetkellä ole täydellisen luotettava, löytyy sille käyttötarkoituksia. Aloitteleva lintuharrastaja, joka ei tunnista ainuttakaan lintua, voi käyttää BirdNET:iä ohjaamaan oikeaan suuntaan. Tuntien tai pidempien äänitiedostojen tutkiminen ei ole käytännöllistä ihmiselle. BirdNET voisi analysoida tiedoston, merkitä kohdat, joissa olettaisi linnun vokalisoivan, minkä ihminen myöhemmin tarkistaisi. Koska BirdNET:iä kehitetään jatkuvasti on todennäköistä, että se tulee olemaan ihmisistä kyvykkäämpi lintujen tunnistamisessa.

Avainsanat: Koneoppiminen, Tekoäly, Lintuharrastus

# Contents

## List of Abbreviations

1	Introduction	1
2	BirdNET	4
2.1	Using BirdNET	5
2.1.1	Android	5
2.1.2	Website API	6
2.1.3	Windows Graphical User Interface	7
2.1.4	Command Line	9
2.2	Hardware and Software used	10
3	Experiments	12
3.1	Artificial environments	14
3.2	Real environments	22
4	Conclusions	24
	References	28

## Appendices

Appendix 1: Results for experiment #12

Appendix 2: Results for experiment #13

Appendix 3: Results for experiment #14

## **List of Abbreviations**

ARU: Autonomous Recording Unit

CNN: Convolutional Neural Network

GUI: Graphical User Interface

PAM: Passive Acoustic Monitoring

UI: User Interface

## 1 Introduction

The history of Artificial Intelligence and Machine Learning date back to the 1950s when these terms were coined. From 1957 to 1974 research into AI flourished. With improved algorithms and government funding in the USA, optimism and expectations in this new technology were high. The biggest obstacle, however, was the hardware of the time. Computers didn't have enough memory to store the massive amounts of information needed, nor the computational power to process it.

Over the next decades the field of AI went through hype cycles: Enthusiasm at first, but eventually leading to disappointment, followed by criticism, funding cuts and ultimately renewed interest years later. Despite these setbacks, AI research thrived in the background.

With the ever increasing power of our modern computers, in the past two decades AI technologies have become better than humans in some areas, for example games such as chess and go, and facial recognition (The Physics arXiv Blog 2014). Today AI is ubiquitous in our lives. With self-driving cars, virtual assistants like Siri and Alexa, and most recently ChatGPT, it seems that companies and researchers are going to utilize AI in every niche possible.

One of these niches this thesis is going to explore is the recognition of birds, from their songs and sounds, by an AI technology called BirdNET, and the use cases of such a tool. Questions this research will explore are whether an AI system could completely replace a human observer, and how and if it could enhance the abilities of a human observer.

Birding, the hobby of studying, surveying and, or watching birds, is very popular. In 2022, a survey (Devokaitis, 2024) found that approximately 96 million people in the United States have participated in some form in birding.

While most birders probably enjoy the hobby for its relaxing qualities, and just like being in nature, some keep a list of all the birds they have observed. For such a birder, a tool like BirdNET could be invaluable. Especially a beginner, who is unsure of their observations, can find validation and guidance from an objective perspective. Since an audio recording is needed, a permanent record of the birder's experience is created.

For a more ambitious use case, BirdNET could be a helpful tool in conservation or monitoring of environments and ecosystems. To assess an ecosystem's health and to make educated decisions regarding its conservation, it is essential to be aware of the ecosystem's animal diversity and population levels of its indicator species. An indicator species, usually an animal, is a species which gives insight to the overall status and condition of the ecosystem and of its species.

Birds work excellently as an indicator species (Forest and Rangeland Ecosystem Science Center, 2022), because they are found almost everywhere in the world. They can live and thrive in almost any environment, in deserts, the arctic, rainforests, in cities with us humans, and many more. Birds also serve in many important ecological roles in their habitats, with a vast amount of different foraging strategies (Mathialagan et al., 2023). Some eat insects, some small vertebrae, and even other birds, and some even scavenge. Many birds serve as pollinators by consuming nectar, others eat seeds and fruits, serving as seed dispersers. They are also conspicuous compared to other groups of animals, being colourful, relatively abundant, and easily detected by their vocalization.

Traditionally, an ecosystem's diversity and abundance of birds is observed by point count. A point count is a method used by birders, where you have a human observer in a predetermined location counting all birds seen or heard within a certain area, during a time interval, usually a few minutes.

Passive acoustic monitoring (PAM) using autonomous recording units (ARUs) offers a potential alternative to point counts, allowing for continuous data

collection over long periods of time, without the need of a human observer. Recent studies have shown that using ARUs have comparable results to traditional point counts, both methods having their own unique advantages and disadvantages.

Deploying a large amount of ARUs allows for a significant increase in spatial and temporal coverage compared to a human observer. The upfront cost of such a deployment could be costly, but in the long run the cost effectiveness of ARUs is comparable (Shonfield & Bayne, 2017) to classical point counts. The obvious advantage of point counts is to visually detect birds. Some birds, such as woodpeckers, raptors, hummingbirds, and many waterfowl can be difficult to survey on vocalization only and may require a visual confirmation (Darras, K. et al. 2018). The skill of the observer also plays a role, short or rare vocalizations can be hard for an inexperienced observer to recognize, while the permanent record of an ARU can help with identification later.

The main advantage of ARUs compared to point counts is the incredible amount of data that can be collected. With increased amount of data however, comes a different problem, how to effectively analyse it all? Manually verifying these vast sound archives is impractical and not time efficient, thus researchers are generally forced to choose a lower number of recordings or focus on just a few species. New technologies, like BirdNET, could be a solution to a problem like this.

BirdNET could eliminate another problem with point counts, their subjectivity. A study (Strickfaden, K. et al. 2019) divided a group of observers into two groups: Experts and amateurs, and played audio clips of birds to each individual in both groups. Of the expert group, an individual averaged a false positive rate of 9,5% and the amateur group 49,1%. With a sufficiently reliable tool such as BirdNET, these kinds of errors could be minimised, if not removed entirely.

## 2 BirdNET

BirdNET is a machine learning algorithm, utilizing a deep neural network, for the purpose of detection and classification of birds from audio files. It was created as a joint project of the K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology, and Chemnitz University of Technology.

Inspiration for creating BirdNET came from the annual LifeCLEF Bird Detection Challenge (BirdCLEF). BirdCLEF launched in 2014, and it has become one of the largest bird sound identification challenges. In the 2016 BirdCLEF a convolutional neural network (CNN) was introduced, which outperformed all competing systems. Since 2017 all teams participating in the challenge have created a CNN, based on the success of the previous solution. Figure 1 shows the workflow of a complete analysis.

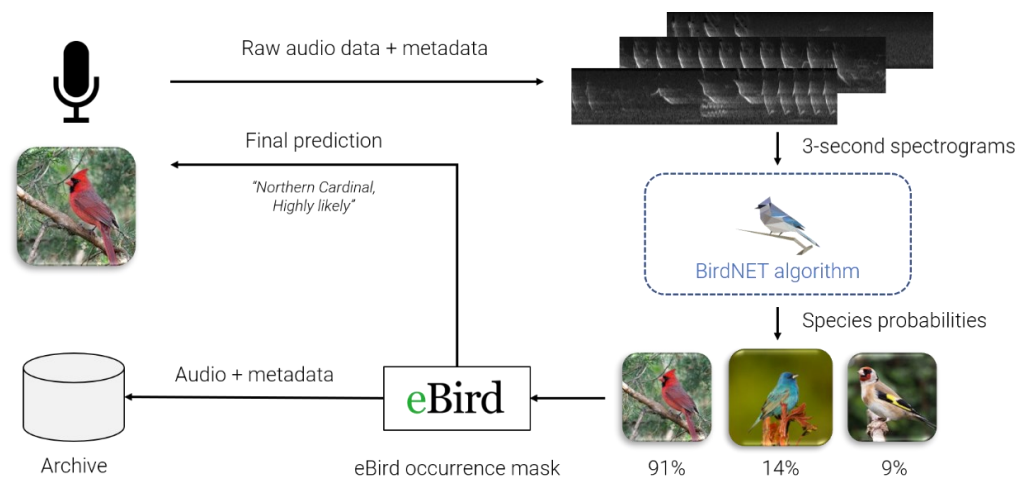


Figure 1. Workflow of BirdNET.

BirdNET works by receiving raw audio data, which is transformed into spectrograms and cut into 3-second chunks, and metadata, such as location and time of year of the recording. Since BirdNET is dealing with spectrograms, the audio recognition task turns into an image recognition task, which has more available tools for analysis (Cornell Lab of Ornithology, 2020, 8:55). Each chunk is then processed by the algorithm and species probabilities are given for each.

Then meta data is used to filter out species unlikely to occur at the given location at the given time of year.

## 2.1 Using BirdNET

The BirdNET algorithm is available on multiple platforms, such as smart phones, web browsers and desktop PCs, allowing anyone without any computer science or programming background to utilize the tool. It can also be integrated into more sophisticated systems with Arduino microcontrollers, Raspberry Pi, or cloud services.

### 2.1.1 Android

The Android or iOS application provides the user with an easy to use interface to start recording and analysing audio. Figure 2 shows the steps to record and analyse for the Android application.

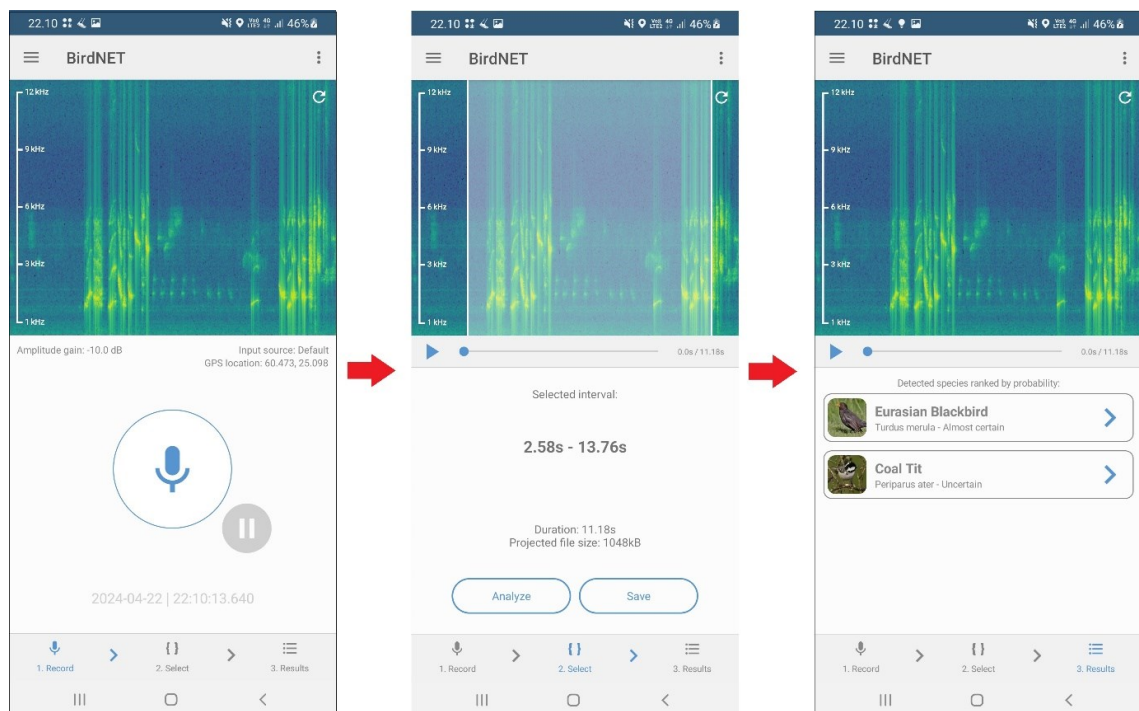


Figure 2. BirdNET UI on Android.

Press the microphone button to start recording and press stop when done. Select an interval on the spectrogram and press Analyze. After a few seconds the application will show the results.

This data and results are archived by BirdNET developers for further scientific research.

### 2.1.2 Website API

If the user wants to analyse pre-recorded audio files, the website version is a simple tool to get fast results. Figure 3 shows a screenshot how to upload a file.

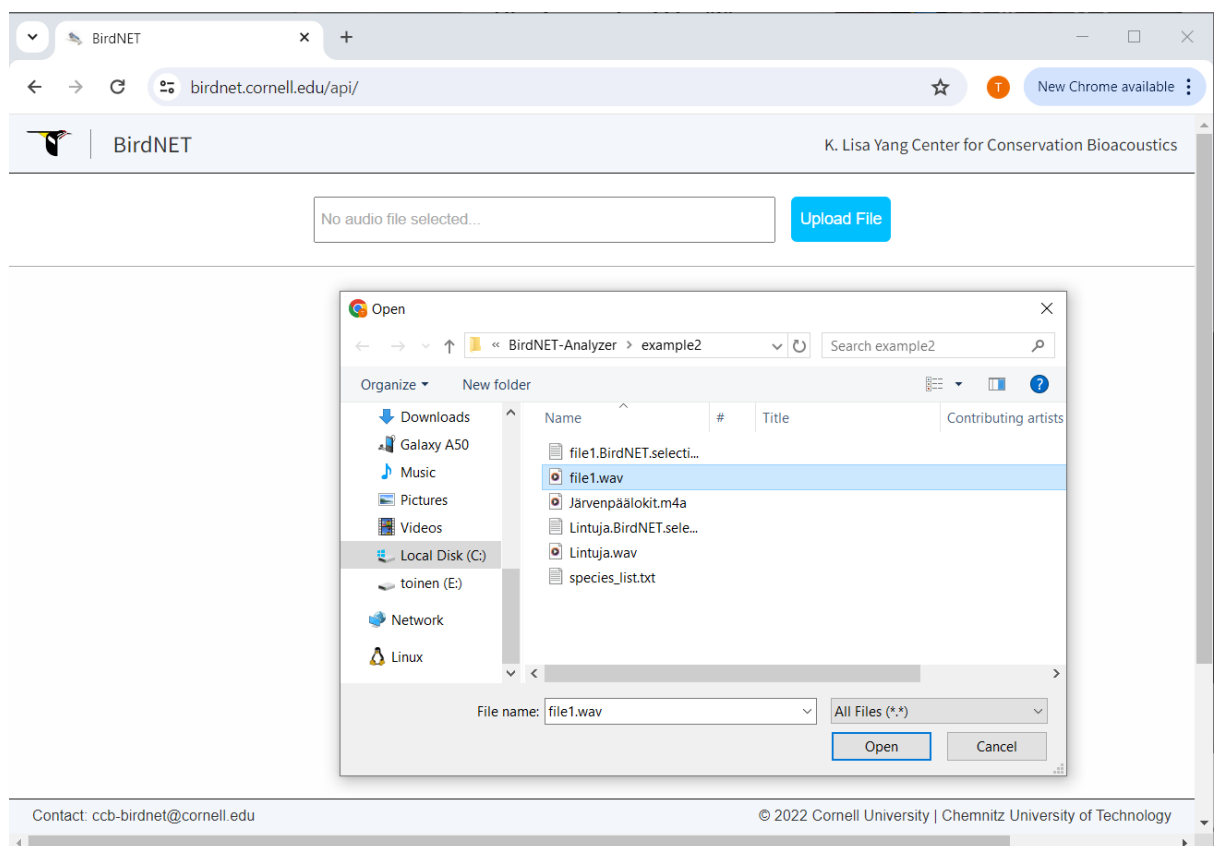


Figure 3. Website API UI.

Choose a file from your computer to upload for analysis. In figure 4, the analysis results are shown for the file.

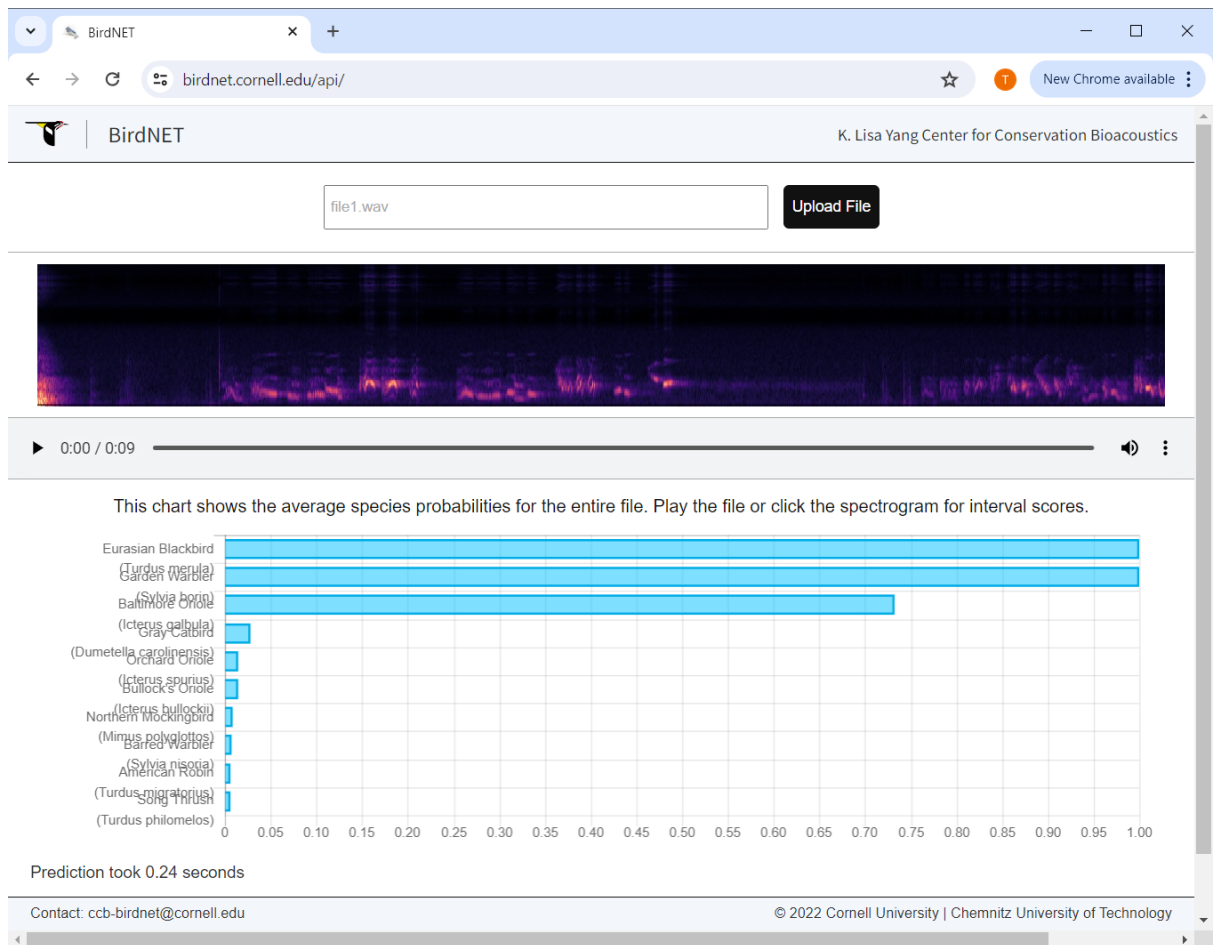


Figure 4. Analysis results.

After the analysis is complete, a chart will be shown with average probabilities of species appearing on the file. Playing the sound file or clicking on the spectrogram will show predictions for that specific time.

### 2.1.3 Windows Graphical User Interface

To analyse multiple files at the same time and receive more information of the analysis, the Windows GUI or the command line versions should be used. Installation guides can be found at the BirdNET github page (Kahl, S. 2021). Figure 5 shows a screenshot of the GUI.

BirdNET-Analyzer

K. Lisa Yang Center for Conservation Bioacoustics

**Select input and output paths:**

C:\BirdNET-Analyzer\example5

C:\BirdNET-Analyzer\example5

**Select species list file or specify latitude, longitude and week:**

C:\BirdNET-Analyzer\tests\species\_list.txt

Latitude: Longitude: Week:

-1 -1 -1

**Set analysis parameters:**

Overlap (s): Sensitivity: Min. confidence: Threads:

0,0 1,0 0,1 4

Output format: Locale:

CSV EN

**Status:**

```
Starting analysis...
Species list contains 225 species
Processing file C:/BirdNET-Analyzer/example5/Lintuja.wav
Finished C:/BirdNET-Analyzer/example5/Lintuja.wav in 3.87 seconds
Processing file C:/BirdNET-Analyzer/example5/file1.wav
Finished C:/BirdNET-Analyzer/example5/file1.wav in 0.41 seconds
Analysis done!
```

Contact: ccb-birdnet@cornell.edu © 2022 Cornell University | Chemnitz University of Technology

Figure 5. Windows GUI.

Select the folder where the files to be analysed are in the input path, and folder for the results to be printed in the output path. To reduce the number of false-positives and to get overall better results, it is highly recommended to provide a species list. This list contains all the birds in the geographic area at that time of the year. If there is no species list, the GUI will generate a virtual list with the given latitude, longitude, and week parameters.

## 2.1.4 Command Line

To use the command line interface to analyse files, you must run the analyzer.py script. Listings 1, 2 and 3 show the basic command line commands to start analysing. For arguments, the least you need to provide is the folder where the audio file(s) are, and an output folder for the results. For example:

```
python3 analyze.py --i /path/to/inputFolder/ --o /path/to/outputFolder
```

Listing 1. Analysis command without extra parameters.

It is also recommended to include a species list:

```
python3 analyze.py --i /path/to/inputFolder/ --o /path/to/outputFolder --slist /path/to/speciesList
```

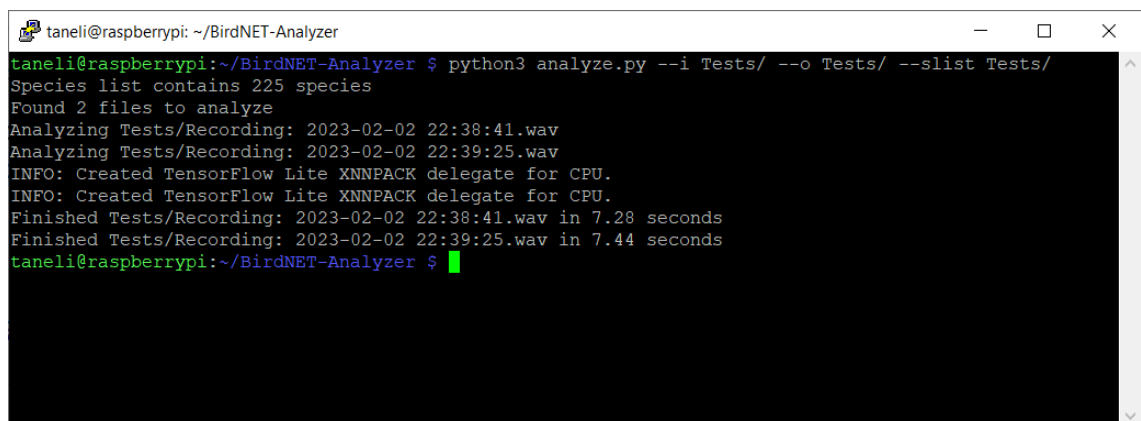
Listing 2. Analysis command with species list parameter.

Or include location and time of year data, the algorithm will generate a virtual species list:

```
python3 analyze.py --i /path/to/inputFolder/ --o /path/to/outputFolder --lat 60.48 --lon 25.06 --week 26
```

Listing 3. Analysis command with location and time of year parameters.

In figure 6, output to the command line is shown after analysis.



```
taneli@raspberrypi: ~/BirdNET-Analyzer
taneli@raspberrypi:~/BirdNET-Analyzer $ python3 analyze.py --i Tests/ --o Tests/ --slist Tests/
Species list contains 225 species
Found 2 files to analyze
Analyzing Tests/Recording: 2023-02-02 22:38:41.wav
Analyzing Tests/Recording: 2023-02-02 22:39:25.wav
INFO: Created TensorFlow Lite XNNPACK delegate for CPU.
INFO: Created TensorFlow Lite XNNPACK delegate for CPU.
Finished Tests/Recording: 2023-02-02 22:38:41.wav in 7.28 seconds
Finished Tests/Recording: 2023-02-02 22:39:25.wav in 7.44 seconds
taneli@raspberrypi:~/BirdNET-Analyzer $
```

Figure 6. Example of using BirdNET with the command line interface.

After the analysis, the results will be created in the output folder.

Table 1. Example data of an output file.

Start (s)	End (s)	Scientific name	Common name	Confidence
0.0	3.0	Cyanistes caeruleus	Eurasian Blue Tit	0.7775
6.0	9.0	Cyanistes caeruleus	Eurasian Blue Tit	0.5966
6.0	9.0	Poecile palustris	Marsh Tit	0.1203
9.0	12.0	Cyanistes caeruleus	Eurasian Blue Tit	0.4016
9.0	12.0	Erithacus rubecula	European Robin	0.1417

Table 1 shows the results of a twelve second audio file. The predictions are divided into three second chunks with each prediction given a confidence rating, the percentage chance the given species is heard in the chunk.

## 2.2 Hardware and Software used

The majority of work done for this thesis was with a Raspberry Pi 4 model B, shown in figure 7.



Figure 7. Raspberry Pi 4 model B.

All recordings were done with a plug and play USB microphone, shown in figure 8.



Figure 8. USB microphone.

A simple python script, using PyAudio, was used to record audio, and the wave package to write the audio into wav files. In figure 9 the recording and analysis process using the Raspberry Pi is shown.

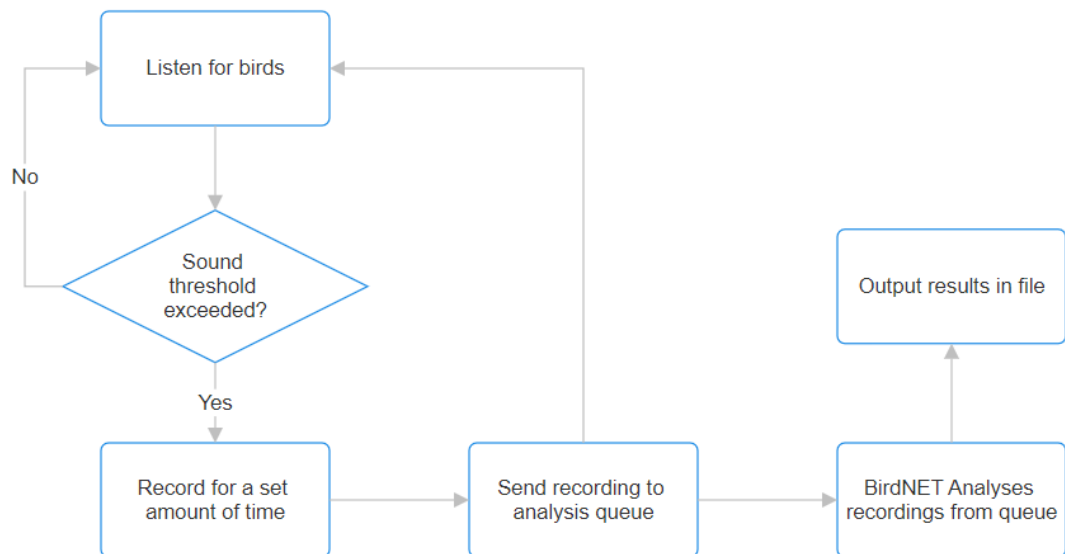


Figure 9. Recording and analysis process.

The Raspberry Pi was set to listen until a sufficiently loud noise was heard, then recording for a set amount of time, usually thirty seconds. This recording was then given to BirdNET to analyse. The Raspberry Pi was powerful enough to record and analyse simultaneously, always being ready to analyse when a new recording was put into the queue.

### 3 Experiments

To investigate the efficacy of BirdNET as a tool to enhance the abilities of a human birder, two main categories of tests were used: Pre-recorded bird songs from publicly available databases, which were played from speakers and then captured, (e.g. lintujen-äänet.net, xeno-canto.org) and recordings created in actual natural environments in several places around Finland.

To choose appropriate recordings for the pre-recorded tests, two criteria were considered:

- Audibility of the chosen species.

- Minimal background noise, from other birds and the environment.

Two common Finnish birds, Great Tit and Eurasian Blackbird were chosen. Their recordings from lintujen-äännet.net have the species clearly audible in the foreground. Considering these are recordings in a real environment, there is some background noise from other birds and the environment, but compared to the focal species they are minimal.

Of the following artificial environment experiments and their results, the reader should note that the tables presented were created as an average of several measurements, because the algorithm can be very sensitive to initial conditions. For example, table 2 shows the analysis of a file, and table 3 shows the same file analysed, but timeshifted 10 milliseconds forward. Additionally, the confidence rating must be above 10% to be included in the table.

Table 2. Unshifted results.

Selection	Start (s)	End (s)	Name	Confidence
1	0	3.0	Eurasian Blackbird	0.6421
2	6.0	9.0	Eurasian Blackbird	0.8795
3	9.0	12.0	Redwing	0.1655
4	9.0	12.0	Eurasian Blackbird	0.1301
5	15.0	18.0	Eurasian Blackbird	0.7680
6	18.0	21.0	Eurasian Coot	0.1036
7	24.0	27.0	Eurasian Blackbird	0.8307

Next results are from the exact same file but shifted 10 milliseconds forward.

Table 3. Shifted results.

Selection	Start (s)	End (s)	Name	Confidence
1	0	3.0	Eurasian Blackbird	0.5670
2	6.0	9.0	Eurasian Blackbird	0.8966
3	9.0	12.0	Redwing	0.2412
4	9.0	12.0	Eurasian Blackbird	0.1933
5	15.0	18.0	Eurasian Blackbird	0.6459
6	21.0	24.0	Eurasian Blackbird	0.1192
7	24.0	27.0	Eurasian Blackbird	0.7806

The differences aren't drastic, but the addition of multiple environmental factors around the testing area, such as the hum of an electrical appliance, the sound of a car driving on the street, or a neighbour could potentially skew the results in unwanted directions.

### 3.1 Artificial environments

In the first experiments, the recordings were played back in different volume levels, emulating distance between the birds and the observer.

Table 4 shows the results of a baseline capture for the Great Tit, where the experimenter determined the volume level to be normal. Not too low, but high enough for all the relevant elements to be audible.

Table 4. Experiment #1.

Selection	Begin (s)	End (s)	Name	Confidence
1	0	3.0	Great Tit	0.8887
2	3.0	6.0	Great Tit	0.2157
3	3.0	6.0	Eurasian Coot	0.1496
4	6.0	9.0	Great Tit	0.9947
5	9.0	12.0	Great Tit	0.9486
6	12.0	15.0	Great Tit	0.9744
7	15.0	18.0	Great Tit	0.9849
8	18.0	21.0	Great Tit	0.8490
9	21.0	24.0	Great Tit	0.9953
10	24.0	27.0	Great Tit	0.8185
11	24.0	27.0	Common Chaffinch	0.1279

The focal species was detected nine times with an average confidence of ~85,2%. Two additional species were also detected, Eurasian Coot and Common Chaffinch with ~15,0% and ~12,8% confidence respectively.

Lowering the playback volume to a minimal, while still being audible is shown in the table 5.

Table 5. Experiment #2.

Selection	Begin (s)	End (s)	Name	Confidence
1	0	3.0	Great Tit	0.2294
2	0	3.0	White-backed Woodpecker	0.1895

Selection	Begin (s)	End (s)	Name	Confidence
3	3.0	6.0	Eurasian Coot	0.1091
4	6.0	9.0	Great Tit	0.8596
5	9.0	12.0	Great Tit	0.8373
6	12.0	15.0	Great Tit	0.7083
7	15.0	18.0	Great Tit	0.9339
8	18.0	21.0	Great Tit	0.4153
9	21.0	24.0	Great Tit	0.9743

The focal species was detected only seven times, with an average confidence falling to ~70,8%. Two additional species were detected with low confidence, similar to the baseline.

The exact same volume experiments were done for the Eurasian Blackbird. Results for the baseline is in table 6.

Table 6. Experiment #3.

Selection	Begin (s)	End (s)	Name	Confidence
1	3.0	6.0	Song Thrush	0.1001
2	6.0	9.0	Eurasian Blackbird	0.9013
3	9.0	12.0	Bluethroat	0.1467
4	12.0	15.0	Eurasian Blackbird	0.1972
5	15.0	18.0	Eurasian Blackbird	0.8688
6	18.0	21.0	Eurasian Blackbird	0.7142
7	21.0	24.0	Eurasian Blackbird	0.7393

Selection	Begin (s)	End (s)	Name	Confidence
8	27.0	30.0	Eurasian Blackbird	0.9530

The focal species (Eurasian Blackbird) was detected six times, with an average confidence of ~62,5%. Two additional species with very low confidence values were also detected.

Lowering the volume to a minimal, but still audible levels result in table 7.

Table 7. Experiment #4.

Selection	Begin (s)	End (s)	Name	Confidence
1	3.0	6.0	Eurasian Curlew	0.5204
2	6.0	9.0	Eurasian Blackbird	0.3961
3	12.0	15.0	Eurasian Coot	0.1247
4	15.0	18.0	Eurasian Blackbird	0.1664
5	18.0	21.0	Eurasian Blackbird	0.4983
6	21.0	24.0	Eurasian Curlew	0.1306

The detections of the focal species are greatly reduced, halving it to three times, with an average confidence of only ~35,4%. Three additional species were detected, with one of them (Eurasian Curlew) having the highest confidence with ~52%.

The previous experiments concentrated on near-ideal situation, with practically no environmental background noise, except for minimal vocalizations from other birds. The focus on the following experiments is on adding naturally occurring

noises, wind, and rain, to see how it affects the algorithm. Wind is simulated by using a fan directed at the capturing microphone, and rain by simultaneously playing a recording of rainfall.

For the wind portion of this experiment, the sound levels were kept identical to the first experiment, with two “levels” of wind (low and high settings on a typical standing fan). Using the baseline volume settings, the sound of the fan is very noticeable on the recording but doesn’t drown out the bird’s song.

Results for baseline volume and fan on low are in table 8.

Table 8. Experiment #5.

Selection	Begin (s)	End (s)	Name	Confidence
1	0	3.0	Great Tit	0.9712
2	6.0	9.0	Great Tit	0.9272
3	9.0	12.0	Great Tit	0.8174
4	12.0	15.0	Great Tit	0.7614
5	15.0	18.0	Great Tit	0.9715
6	18.0	21.0	Great Tit	0.7846
7	21.0	24.0	Great Tit	0.9648

The focal species was detected seven times with an average confidence of ~88,5% and with no additional species. At first glance it seems as though the noise of the fan improved our accuracy, but a more probable explanation is that the confidence for selections 2,3 and 11 from the original table were lowered below our threshold of 10%, thus not even entering our table.

Results for baseline volume and fan on high are in table 9.

Table 9. Experiment #6.

Selection	Begin (s)	End (s)	Name	Confidence
1	0	3.0	Great Tit	0.7380
2	6.0	9.0	Great Tit	0.7845
3	9.0	12.0	Great Tit	0.6413
4	12.0	15.0	Great Tit	0.5575
5	15.0	18.0	Great Tit	0.8816
6	18.0	21.0	Great Tit	0.4491
7	21.0	24.0	Great Tit	0.9396

Again, seven detections, with an average confidence of ~71,3%. Unsurprisingly the higher fan setting increases the noise in the recording and lowered the accuracy on all intervals.

Volume low, fan high in table 10:

Table 10. Experiment #7.

Selection	Begin (s)	End (s)	Name	Confidence
1	0	3.0	Great Tit	0.6960
2	6.0	9.0	Great Tit	0.3256
3	9.0	12.0	Great Tit	0.1162
4	12.0	15.0	Great Tit	0.3865
5	18.0	21.0	Great Tit	0.1877
6	21.0	24.0	Great Tit	0.3067

Six detections, with an average confidence of ~33,6%. Detection accuracy continues to deteriorate.

For the rain portion, the sound levels of the bird song were kept the same as in previous experiments. The volume for the rain was kept high enough to be noticeable, while keeping the bird songs clearly audible. Results in table 11.

Table 11. Experiment #8.

Selection	Begin (s)	End (s)	Name	Confidence
1	0	3.0	Great Tit	0.8898
2	3.0	6.0	Great Tit	0.8544
3	6.0	9.0	Great Tit	0.9593
4	9.0	12.0	Great Tit	0.3415
5	12.0	15.0	Great Tit	0.9673
6	18.0	21.0	Great Tit	0.9578
7	21.0	24.0	Great Tit	0.7406
8	24.0	27.0	Great Tit	0.9248

Eight detections with an average confidence of ~83,2% with no additional species. Just like in the first wind experiment, it seems like the accuracy is increased compared to the baseline, but a more probable explanation is the same: the confidence for selections 2,3 and 11 from the original table were lowered below our threshold of 10%.

For experiment #9 the volume of the focal species' song was lowered to the same as the baseline, experiment #2. Now the resulting table is empty, zero detections with confidence above 10%. The Great Tit's song is still audible, even if very faint, but apparently not loud enough for the algorithm to detect.

In a real environment there will often be more than one different bird species vocalizing at the same time. The next experiments will use recordings of the same two species, Great Tit and Eurasian Blackbird, played at the same time. The recordings used are slightly different, to achieve more overlapping vocalizations.

The baseline result tables for the individual songs are in table 12.

Table 12. Experiment #10.

Selection	Begin (s)	End (s)	Name	Confidence
1	0	3.0	Great Tit	0.4651
2	3.0	6.0	Great Tit	0.8262
3	6.0	9.0	Great Tit	0.9838
4	9.0	12.0	Great Tit	0.8112
5	12.0	15.0	Great Tit	0.9937
6	15.0	18.0	Great Tit	0.2950
7	18.0	21.0	Great Tit	0.9893
8	21.0	24.0	Great Tit	0.8908
9	24.0	27.0	Great Tit	0.9807

Nine detections, average accuracy ~80,4%.

In table 13, the results for the Eurasian Blackbird are shown.

Table 13. Experiment #11.

Selection	Begin (s)	End (s)	Name	Confidence
1	0	3.0	Eurasian Blackbird	0.9252
2	6.0	9.0	Eurasian Blackbird	0.8164
3	9.0	12.0	Eurasian Blackbird	0.9121
4	15.0	18.0	Eurasian Blackbird	0.7359
5	18.0	21.0	Eurasian Bullfinch	0.3222

Selection	Begin (s)	End (s)	Name	Confidence
6	18.0	21.0	Coal Tit	0.3098
7	21.0	24.0	Eurasian Curlew	0.4675
8	24.0	27.0	Eurasian Blackbird	0.9495

Five detections, average accuracy ~86,8%.

Playing the individual audio files simultaneously we get these results (Appendix 1)

Great Tit: Eight detections, accuracy of ~49,0%.

Eurasian Blackbird: Five detections, accuracy of ~54,4%.

The detection rate for both species drops significantly.

### 3.2 Real environments

The first experiment in a natural environment was recorded from a balcony in an apartment on the fifth floor, with some environmental sounds, like slight wind and the occasional noise of a car.

The results (Appendix 2) show that the Great Tit is heard throughout the recording with varying confidence levels. Selections 1, 7, 10, 14 and 15 seem to have significantly lower detection rate, possibly because their detection intervals coincide with some environmental noise, an accelerating car or vocalization from other birds. Still the Great Tit is easily recognizable to the human ear in those intervals.

The next experiment's recording (Appendix 3) had multiple different birds vocalizing simultaneously, with constant slight background noise of traffic.

The constant overlapping of bird sounds and background noise from traffic, seem to make detection harder, with confidences varying from ~11% to ~82%, with an average of ~36%.

Table 14 shows the result of a recording during moderate rain, recording place being the same.

Table 14. Experiment #15.

Selection	Begin (s)	End (s)	Name	Confidence
1	18.0	21.0	Eurasian Jackdaw	0.2273
2	24.0	27.0	Common Gull	0.1727
3	45.0	48.0	Common Gull	0.5089

From the minute long recording, several different species are heard, but the algorithm has a difficult time recognizing them. Only three detections, while at least a dozen birds could easily be heard.

Table 15 shows the result of recording a flock of Eurasian Jackdaws, with multiple birds of the same species vocalizing at the same time.

Table 15. Experiment #16.

Selection	Begin (s)	End (s)	Name	Confidence
1	0	3.0	Eurasian Jackdaw	0.1512
2	9.0	12.0	Canada Goose	0.1146
3	12.0	15.0	Eurasian Jackdaw	0.9867
4	15.0	18.0	Eurasian Jackdaw	0.9917

Selection	Begin (s)	End (s)	Name	Confidence
5	18.0	21.0	Eurasian Jackdaw	0.9652
6	21.0	24.0	Eurasian Jackdaw	0.8950
7	24.0	27.0	Eurasian Jackdaw	0.9272

The algorithm seems to have no problem recognizing overlapping birds of the same species.

The next experiment was done at a park. With results in table 16.

Table 16. Experiment #17.

Selection	Begin (s)	End (s)	Name	Confidence
1	0	3.0	Black-headed Gull	0.2486
2	6.0	9.0	Black-headed Gull	0.2608
3	9.0	12.0	Black-headed Gull	0.1020

In the background there is people talking, slight wind and constant vocalization of black-headed gulls. Even though the songs of the gulls are ubiquitous, the detection rate is very low, with three detections averaging ~20,4% confidence.

## 4 Conclusions

From the experiments done for this thesis, one could make several conclusions about the efficacy of BirdNET. Experiments #1 and #2 show, that the algorithm is confident in its detections of the focal species, but adding background noise can drastically lower the detection rate and confidence. Experiment #8, which is

the same as experiment #2 but with added wind background noise, shows lowered detection rate and more than halving the detection confidence.

Experiment #13 overlaps two different recordings, whose results are experiments #11 and #12. The individual recordings have high detection rates, but combining the recordings nearly halves the detection rate for each species.

It is not surprising, that adding noise or overlapping recordings will make it harder for BirdNET to make correct detections, but the focal species in these experiments is never completely drowned out. A sufficiently competent human could easily detect these birds, even when the algorithm resulted in a confidence as low as 12%.

Experiments in real environments seem to show similar results. Experiment #14 was recorded from a fifth floor balcony, with mostly one species of bird vocalizing and occasional sounds from an accelerating car. Like experiments #1 and #2, the detection confidence of the focal species, when its clearly heard, is high. Low confidence rates coincide with an accelerating car or another, undetected bird. Experiment #16 was recorded on a rainy day, several different species can be heard, but detection rates and confidence are low.

The recording with overlapping bird songs in a real environment in experiment #15 show similar results to overlapping of two pre-recorded songs in experiment #13. In experiment #15 many different species of bird can be heard vocalizing constantly, with detection confidence varying wildly, low detection rates coinciding with simultaneous bird songs. The overlapping of many birds of the same species in experiment #17 however, where a flock of Eurasian Jackdaws was recorded, shows high detection confidence.

Experiment #18 was done at a park. In the background there is the constant vocalization of black-headed gulls and no other species, with moderate wind noise and people talking. Again, sounds of the gulls are easily recognized by a person, but the algorithm has trouble with detection rate and confidence.

At the moment of writing this, the experiments done for this thesis suggest that an AI like BirdNET has not surpassed the skills of a veteran birder, or probably even a moderately skilled one. But that does not mean it couldn't be a valuable tool for them. For example, using ARUs with days, perhaps even days long recordings, it is not practical for a human to listen to them and make detections. An AI, even if not yet powerful enough to be trusted completely on its own, could still be used to mark sections where it believes a bird has been heard, to be verified later by a human.

The progress of developing different kinds of AIs has been incredibly fast over the last decade or so and many AIs have already surpassed the skills of humans. For example, in 1997 Russian champion Gary Kasparov lost to IBM's Deep Blue at chess, and in 2015 Microsoft researchers created an AI that beat a human in an image recognition task (Tanz, O. 2017). Since BirdNET's algorithm works on analysing spectrograms, which is an image recognition task, there is no reason for an AI to not surpass even the most skilled human eventually, given enough research.

The question is, will audio recognition be given enough research. Those working on AI solutions, ranked audio recognition one of the least used AI technologies, because there are fewer uses for it at the moment. Companies are less likely to invest in something customers are not expressing an interest in (Roumeliotis, R. 2021).

A study compiled the results of 21 studies in avian ecological research, with two different methods of surveying birds: Human observers, mostly using point counts, and humans listening to recordings of birds. Different studies used different metrics, such as species richness, composition, abundance, presence or absence, call counts and detection probabilities, with species richness being the most used. Majority of the studies found the two methods having equal performance, with some favouring humans and some the use of recordings.

A way to quantify the efficacy of an AI like BirdNET, would be to do more studies like these with a few additions. Give some of the people who are listening to recordings the ability to also analyse the recordings with an AI. If the people with access to an AI also provide better results, it is evidence for the efficacy of the AI.

Further research could be done on the type of equipment used. This thesis was done using a Raspberry Pi, which is not an ideal device for doing recordings in the field. The microphone used has decent audio quality, but especially wind affected many recordings badly, and the results of the experiments show this.

## References

Cornell Lab of Ornithology (2020) *What's that bird song? ID birds by sound with BirdNET*. 9 July. Available at: <https://www.youtube.com/watch?v=MQHunTLt1TI> (Accessed June 2022)

Darras, K. et al. (2017) 'Comparing the sampling performance of sound recorders versus point counts in bird surveys: A meta-analysis' *Journal of Applied Ecology* 55 (6) pp. 2575-2586 [Online] Available at: <https://doi.org/10.1111/1365-2664.13229> (Accessed 22 March 2023)

Devokaitis, Marc (2024) *A Third Of American Adults Are Birdwatchers, According To Nationwide Survey* Available at: <https://www.allaboutbirds.org/news/a-third-of-american-adults-are-birdwatchers-according-to-nationwide-survey/> (Accessed 15 February 2024)

Forest and Rangeland Ecosystem Science Center (2022) *Birds as Indicators of Ecosystem Health* Available at: <https://www.usgs.gov/centers/forest-and-rangeland-ecosystem-science-center/science/birds-indicators-ecosystem-health> (Accessed 22 March 2023)

Kahl, S. 2021 *BirdNET-Analyzer* Available at: <https://github.com/kahst/BirdNET-Analyzer> (Accessed 10 August 2022)

Mathialagan Mariyappan, et al. (2023) 'Ecological Role and Ecosystem Services of Birds: A Review.' *International Journal of Environment and Climate Change* 13 (6) pp 1 [Online] Available at: [www.doi.org/10.9734/ijecc/2023/v13i61800](http://www.doi.org/10.9734/ijecc/2023/v13i61800) (Accessed: 10 March 2024)

Roumeliotis, R. (2021) *AI Is Booming with Image Recognition, but Audio Recognition Lags Behind* Available at: <https://www.dataversity.net/ai-is-booming-with-image-recognition-but-audio-recognition-lags-behind/> (Accessed 15 March 2024)

Shonfield, J., & E. M. Bayne (2017) 'Autonomous recording units in avian ecological research: current use and future applications.' *Avian Conservation and Ecology* 12 (1) pp 2 [Online] Available at: <https://doi.org/10.5751/ACE-00974-120114> (Accessed 10 March 2024)

Shonfield, J., & E. M. Bayne (2017) 'Autonomous recording units in avian ecological research: current use and future applications.' *Avian Conservation and Ecology* 12 (1) Table 1 [Online] Available at: <https://doi.org/10.5751/ACE-00974-120114> (Accessed 10 March 2024)

Strickfaden, K. et al. (2017) 'Dependent double-observer method reduces false-positive errors in auditory avian survey data' *Ecological Applications* 30 (2) Table 1 [Online] Available at: <https://doi.org/10.1002/eap.2026> (Accessed 12 March 2024)

Tanz, O. (2017) *Can Artificial Intelligence Identify Pictures Better than Humans?* Available at: <https://www.entrepreneur.com/science-technology/can-artificial-intelligence-identify-pictures-better-than/283990> (Accessed 15 March 2024)

The Physics arXiv Blog (2014) The Face Recognition Algorithm That Finally Outperforms Humans Available at: <https://medium.com/the-physics-arxiv-blog/the-face-recognition-algorithm-that-finally-outperforms-humans-2c567adbf7fc> (Accessed 16 March 2024)

**Results of experiment #12.**

Table 1. Experiment #12.

Selection	Begin (s)	End (s)	Name	Confidence
1	0	3.0	Eurasian Blackbird	0.7526
2	0	3.0	Great Tit	0.6278
3	3.0	6.0	Great Tit	0.2714
4	6.0	9.0	Great Tit	0.6341
5	6.0	9.0	Eurasian Blackbird	0.3414
6	9.0	12.0	Eurasian Blackbird	0.5695
7	9.0	12.0	Great Tit	0.1734
8	12.0	15.0	Great Tit	0.9925
9	12.0	15.0	Common Chaffinch	0.1812
10	15.0	18.0	Eurasian Blackbird	0.3253
11	18.0	21.0	Great Tit	0.2956
12	18.0	21.0	European Golden-Plover	0.2486
13	18.0	21.0	Eurasian Bullfinch	0.1965
14	21.0	24.0	Eurasian Curlew	0.8978
15	21.0	24.0	Great Tit	0.7089
16	24.0	27.0	Eurasian Blackbird	0.7329
17	24.0	27.0	Great Tit	0.2193

**Results of experiment #13.**

Table 2. Experiment #13.

Selection	Begin (s)	End (s)	Name	Confidence
1	6.0	9.0	Great Tit	0.2184
2	9.0	12.0	Great Tit	0.8028
3	12.0	15.0	Great Tit	0.8478
4	15.0	18.0	Great Tit	0.7996
5	15.0	18.0	Eurasian Eagle-Owl	0.3969
6	18.0	21.0	Great Tit	0.5652
7	21.0	24.0	Great Tit	0.1065
8	24.0	27.0	Great Tit	0.7550
9	27.0	30.0	Great Tit	0.7776
10	30.0	33.0	Great Tit	0.2059
11	33.0	36.0	Great Tit	0.4830
12	36.0	39.0	Great Tit	0.8838
13	39.0	42.0	Great Tit	0.3434
14	42.0	45.0	Great Tit	0.1730
15	45.0	48.0	Great Tit	0.1458
16	48.0	51.0	Great Tit	0.6373
17	51.0	54.0	Great Tit	0.7670
18	54.0	57.0	Great Tit	0.5977

**Results of experiment #14.**

Table 3. Experiment #14.

Selection	Begin (s)	End (s)	Name	Confidence
1	0	3.0	Fieldfare	0.1182
2	0	3.0	Northern Goshawk	0.1121
3	3.0	6.0	Fieldfare	0.1915
4	6.0	9.0	Fieldfare	0.4722
5	9.0	12.0	Fieldfare	0.8191
6	12.0	15.0	Fieldfare	0.4603
7	12.0	15.0	Black-headed Gull	0.1946
8	12.0	15.0	Common Gull	0.1843
9	15.0	18.0	Common Gull	0.2527
10	15.0	18.0	Fieldfare	0.1461
11	18.0	21.0	Common Gull	0.6403
12	18.0	21.0	Fieldfare	0.4498
13	18.0	21.0	European Honey-buzzard	0.1931
14	18.0	21.0	Northern Goshawk	0.1531
15	24.0	27.0	Fieldfare	0.6566
16	27.0	30.0	Fieldfare	0.7948