CASA0018 Bird Sound Classifier

Hanpu Liu,
link to github repo with project work in:
https://github.com/HanpuLiu12138/casa0018/tree/main/Assessment,
link to Edge Impulse projects: https://studio.edgeimpulse.com/studio/200378

Introduction

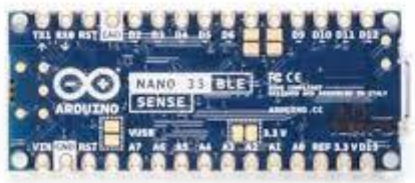
The Bird Sound Classifier project seeks to develop a system which can identify bird calls in real-time, providing users with information about which bird species are being heard. Hiral Nagda, a well-known author and physiotherapist, once said, "Birds chirping around you is a beautiful realization that life is incredibly good. Let this sound be a gentle break in your routine." When I first read this quote, I really connected with it because I love hearing birds sing and it makes me feel calm. When I hear a bird call in my normal day-to-day life, I often wonder what kind of bird it is. So I wanted to make a way to sort the sounds of birds so I could tell what they are. This project utilizes deep learning techniques, employing Edge Impulse and Arduino Nano 33 BLE Sense as tools for data collection, model development and deployment. These tools were chosen based on their compatibility and ability to support all stages of the development process, from raw data sampling through machine learning model deployment on compact development boards. The ultimate aim is to design a user-friendly, portable, accurate, and accurate bird sound identification system that helps users appreciate nature and enhance outdoor experiences. This innovative project serves as a bridge between nature and technology by teaching people more about birds around them - building up appreciation of our natural world in turn.

Research Question

Can I develop a bird sound classifier that accurately and quickly recognizes bird calls to deepen our understanding of our environment and enhance our appreciation? How can deep learning techniques be effectively utilized to address this challenge?

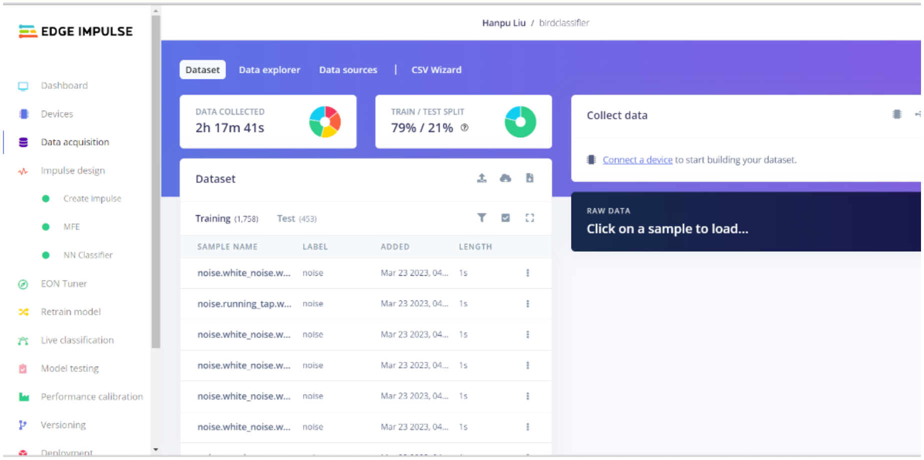
Application Overview

Data gathering, data preparation, model training, and deployment are all of the components that make up this project. The information comes from Xeno-Canto, which is a big database that contains recordings of bird calls. After the data has been preprocessed and augmented, a deep learning model is trained with the help of Edge Impulse. After that, the trained model is installed on a device called an Arduino Nano 33 BLE Sense as well as on a mobile phone. Because of its Cortex-M4 microprocessor, motion sensors, microphone, and Bluetooth Low Energy connectivity, the Arduino Nano 33 BLE Sense is an excellent choice for this project. The programme for mobile phones gives users the ability to recognise bird screams in real time, which enriches their experience of being outside. The design of this system places a strong emphasis on reaching a high level of accuracy while also ensuring that it is easy to use for the customer. This project offers an all-encompassing answer to the problem of identifying bird calls by linking the various components in a logical and consistent way.



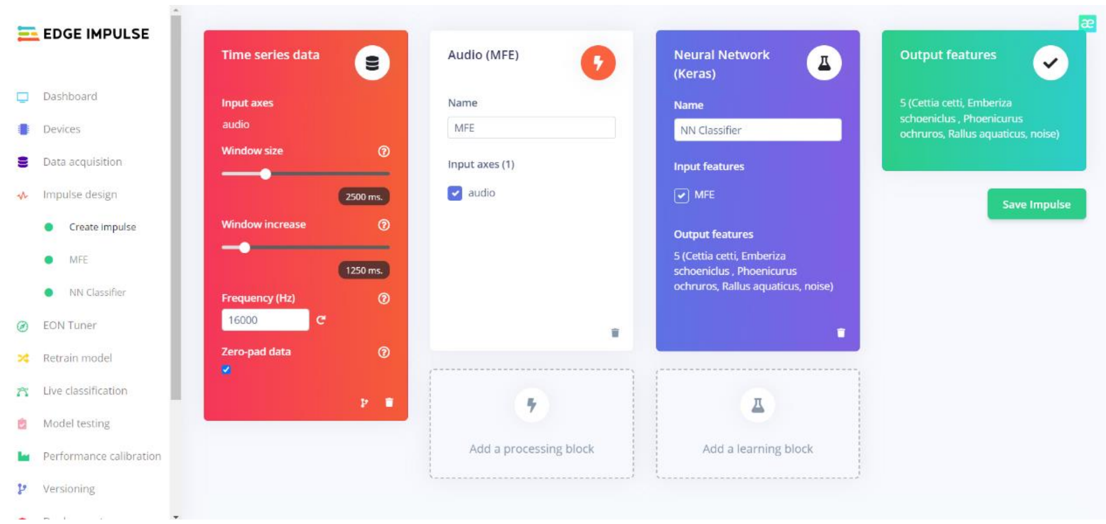
Data

Data on the bird calls of four common bird species that may be found at Queen Elizabeth Olympic Park were acquired from Xeno-Canto. These four species are the Water rail, Cetti's warbler, Black redstart, and Common reed bunting. Downloading and utilising the Lightning software to do preprocessing on around 10-15 audio recordings for each bird was done. The data were extended by producing close to 200 files, each of which was 10 seconds long and dedicated to a different kind of bird and noise. The data set was well-balanced, consisting of 20 minutes of training data and 6 minutes of test data for each label. This dataset made it possible to create a model that is varied and representative, which is necessary for successfully detecting bird cries in a variety of different contexts. The selection of these four bird species serves as a beginning point for the system, which has the potential to be modified in the future so that it may contain a greater variety of bird species. For the purpose of producing a high-quality dataset, which is an essential need for the accomplishment of any machine learning endeavour, the procedure of collecting data required the careful selection of bird calls. This study sets the groundwork for a reliable and accurate bird sound classifier by putting an emphasis on the quality and variety of the data collected.



Model

In order to properly categorise sounds, the deep learning model that was applied to this project included a combination of audio processing (MFE) and a neural network (Keras). The MFE processing block was selected because it is very efficient at managing audio data, while the Keras learning block was picked because it enables simple implementation and is highly adaptable. The model was set up and put through its paces by experimenting with a wide variety of parameters and configurations in order to identify the optimal combination for reliable bird call detection. In order to get the highest potential level of performance from the neural network, its topology, activation functions, loss functions, and optimisation methods were all fine-tuned. Convolutional neural networks (CNN), recurrent neural networks (RNN), and transformers were some of the deep learning model designs that were put through their paces in order to identify which one proved to be the most useful solution to this specific issue. The final model architecture was chosen to ensure that the system is both accurate and usable in the actual world. This was accomplished by striking a compromise between the efficiency with which the model could be computed and its level of precision.



Experiments

A number of different tests were carried out in order to evaluate the bird sound classifier and improve upon its overall performance. Iterative adjustments were made to several parameters in order to attain the best possible outcomes, including the learning rate, batch size, number of layers in the neural network, and the architecture used for the model. The performance was evaluated based on the accuracy of both training and testing, and visualisations were developed to help get a better understanding of the influence that different configurations have on the model's overall performance.

Convolutional neural networks (CNN), recurrent neural networks (RNN), and transformers were some of the model designs that were put through their paces throughout the evaluation process. The ultimate choice was decided based on a compromise between performance and the amount of computing complexity that was required, and each design had both positive and negative aspects to offer.

Automated tuning approaches such as Bayesian optimisation, which may be used to discover the optimal combination of hyperparameters in order to increase the accuracy of the model, were also investigated as part of this study. The goal of the project was to improve the performance of the model and attain a greater level of accuracy when classifying bird sounds, and one way to do this was to use these methodologies.

During the phase of experimentation, a number of different evaluation tools and scripts were utilised in order to evaluate the performance of the model. Visualisations of the outcomes of each experiment, in the form of graphs and tables, were created in order to facilitate improved comprehension of the manner in which the model performed and the influence that various configurations had. These visualisations were essential in finding the optimal mix of parameters and model architecture, which finally resulted in the model with the highest level of performance.

Results and Observations

The research was able to effectively construct a bird sound classifier with a high level of accuracy (93.59 percent testing accuracy and 97.2 percent training accuracy). Even if there were no major issues with the functioning of the system, there is still potential for improvement. In the next phases, we are going to use model fusion, hyperparameter adjustment, and multimodal information in order to improve the performance of the classifier.

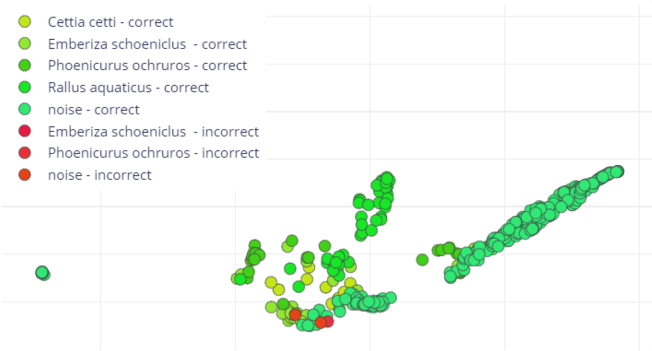
Model testing results

ACCURACY

93.59%

	CETTIA CE	EMBERIZA	PHOENICU	RALLUS AQ	NOISE	UNCERTAI
CETTIA CET	100%	0%	0%	0%	0%	0%
EMBERIZA S	0%	86.1%	2.4%	2.0%	0%	9.4%
PHOENICUR	0%	5%	86.4%	0%	1.4%	7.1%
RALLUS AQ	0.3%	0%	0%	99.7%	0%	0%
NOISE	0%	0.3%	0%	0%	97.1%	2.6%
F1 SCORE	1.00	0.90	0.92	0.99	0.98	

Feature explorer



Last training performance (validation set)

ACCURACY

97.2%

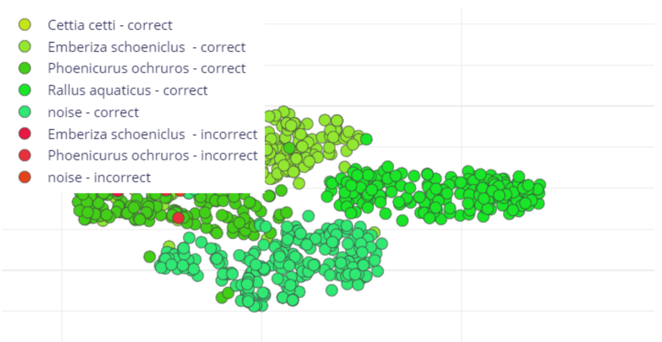
LOSS

0.07

Confusion matrix (validation set)

	CETTIA CETT	EMBERIZA S	PHOENICUR	RALLUS AQU	NOISE
CETTIA CETTI	100%	0%	0%	0%	0%
EMBERIZA SCH	0%	97.9%	1.6%	0%	0.5%
PHOENICURUS	0%	7.8%	92.2%	0%	0%
RALLUS AQUAT	0%	0%	0%	100%	0%
NOISE	0.4%	2.3%	0.4%	0%	96.9%
F1 SCORE	1.00	0.94	0.95	1.00	0.98

Data explorer (full training set)



Model fusion is the process of merging the predictions of numerous deep learning models, such as CNNs, RNNs, and transformers, and averaging or voting on their predictions. Examples of these models are CNNs and RNNs. Because different models may be better suited to different aspects of the audio data, this strategy has the potential to improve recognition accuracy.

Tuning of hyperparameters is an important component that contributes to the overall performance of deep learning models. It is possible to increase the accuracy of the model by further honing in on the selection of hyperparameters, as well as by investigating automated tuning approaches, such as Bayesian optimisation. This method calls for a great deal of testing and might ultimately result in improved performance.

Another area with room for prospective development is that of multimodal information. Not only do bird cries carry auditory information, but they also contain other information that might help in the identification process. These other pieces of information include the habitat of the bird, its size, and the colour of its feathers. It is possible to further improve the classifier's performance by merging the data from the acoustic analysis with this information.

The findings of this experiment illustrate the possibility of using deep learning methods to find solutions to issues that occur in the real world, such as the recognition of bird calls. The inventor of this bird sound classifier received excellent expertise in the application of deep learning methods to real-world applications, as well as developed independent thinking and problem-solving abilities via the process of constructing this classifier.

Bibliography

1. Nagda, H. (no date) A quote by Hiral Nagda, Goodreads. Goodreads. Available at: https://www.goodreads.com/quotes 10924854-birds-chirping-around-you-is-a-beautiful-realisation-that-life (Accessed: May 1, 2023).

2. Agnihotri, A. and Batra, N. (2021) Exploring bayesian optimization, Distill. Available at: https://distill.pub/2020/bayesian-optimization/ (Accessed: May 1, 2023).

Declaration of Authorship

I, Hanpu Liu, confirm that the work presented in this assessment is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Hanpu LIU

1/5/2023