Coding Project Report

Student: Alexandra Antonica
SNR: 2089359

Introduction

Insects play a vital role in maintaining the balance of Earth's ecosystems. They are essential for pollination, decomposition, soil aeration, and nutrient and energy cycling. Insects surpass all other forms of animal life in numbers, diversity, and biomass, underscoring their critical importance to ecosystem functionality (Black & Vaughan, 2009). Maintaining healthy ecosystems depends on accurately identifying and tracking insect populations.

This project aims to develop a deep learning model for identifying insect species from raw audio recordings, contributing to biodiversity monitoring through a robust and automated acoustic analysis framework. The proposed method employs a Convolutional Neural Network (CNN) architecture fine-tuned for insect sound classification. Raw audio recordings are preprocessed into mel-spectrograms, which serve as the input for the CNN. This approach transforms the audio classification task into an image recognition problem, allowing the model to extract complex patterns and features from the spectrograms effectively.

Method

The dataset used for developing this project is InsectSet32 (Faiß, 2022) . This dataset consists of 335 files representing recordings of 32 sound-producing insect species, summing up to 57 minutes of insect sounds. The recordings are split into two datasets. 147 recordings belong to nine species of the order Orthoptera. The remaining recordings (188) are of 23 species in the family Cicadidae. The dataset is accompanied by annotation files that contain information for each recording (file name, species name and identifier).

For this project, the species Platypleura "sp12 cf. hirtipennis" was removed from the dataset, as the audio files were corrupted while downloaded from the source website. Consequently, the dataset contains recordings from 31 species. Figure 1 shows a histogram of the class IDs.

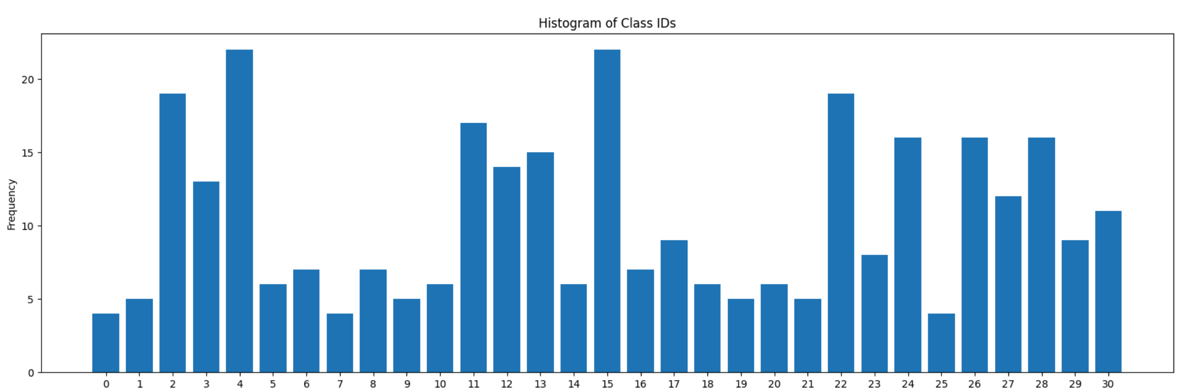

Figure 1. Histogram of class IDs

The chosen model is based on the EfficientNetB3 architecture, pre-trained on a large image dataset (ImageNet), which is known to be a CNN suitable for audio classification tasks. This architecture was used effectively for a similar task by Kortas (2020), who employed it as an approach to classifying bird songs. Given their successful results, EfficientNetB3 has been adapted in this project for classifying insect sounds.

The audio files were preprocessed as follows. Each recording was segmented into audio files of 5 seconds with stride of 1. The stride represents the overlap between the segments. The motivation of choosing these numbers stems from the study of Kortas (2020). Next, the segments were transformed into mel-spectrograms.

The mel-spectrograms serve as data for training, testing and validating the model. The data was split 60% for training, 20% for testing and 20% for validation. This is equivalent to 272 spectrograms for training, 78 for testing, and 78 for validation. This split, a common approach in machine learning, provides a balance between effective training and sufficient data for unbiased validation and testing.

The dataset was augmented during the training process to improve the model's robustness and prevent overfitting. Data augmentation was performed on the mel-spectrograms using image transformations, including: width and height shifting (horizontally or vertically by 20% of its size), shearing, and zooming (randomly by up to 10%).

As shown in Figure 1, the dataset exhibits class imbalance. For this reason, class weights were employed during training to improve the model's capability of identifying underrepresented species.

Results

The model was evaluated using the following metrics: accuracy, precision, recall, F1-score, and ROC-AUC score. These metrics ensure reliability across overall correctness, class-specific behavior, balance between precision and recall, and the model's ability to differentiate between classes effectively. Table 1 shows the results of evaluation on the test set. Figure 2 illustrates the Precision-Recall curve, while Figure 3 represents the ROC curve.

Table 1. Evaluation Results

Accuracy	Precision	Recall	F1-score (weighted)	ROC-AUC
0.73	0.78	0.73	0.72	0.98

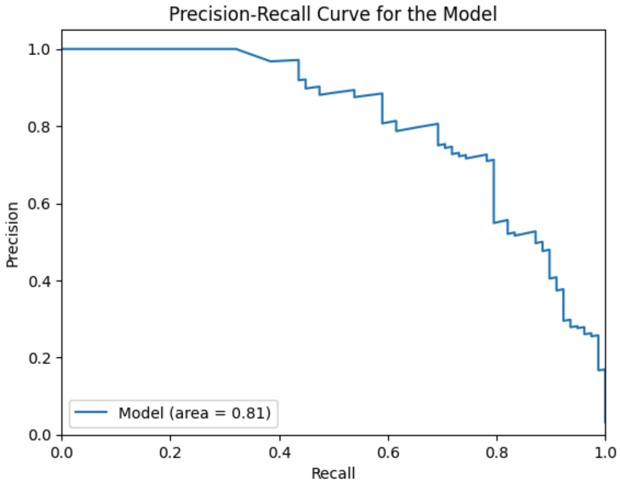

Figure 2. Precision-Recall curve

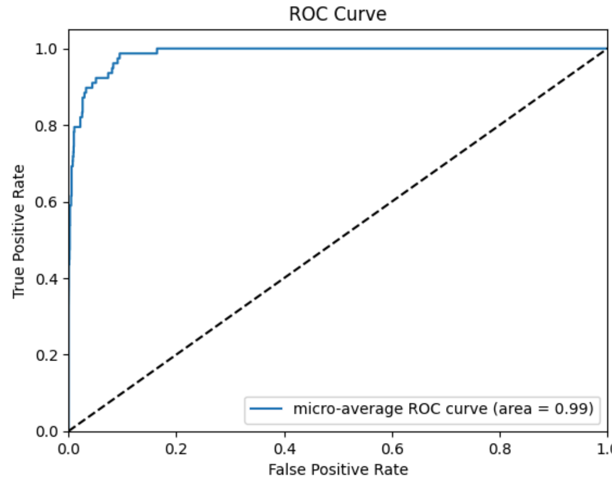

Figure 3. ROC curve

Discussion

The evaluation results suggest that the model is well-performing overall. Despite the relatively small-sized dataset, achieving an accuracy of 73% is a promising outcome. The high values of precision and recall indicate that the model is able to avoid false positives effectively, while correctly identifying a significant amount of true positives. An F1 score of 72% highlights that the model has a reasonable balance between precision and recall; this is also visible in Figure 2. Furthermore, the high micro-average ROC-AUC score demonstrates the model's strong ability to differentiate between classes effectively.

Reflecting on the results, there are some areas of improvement that are worth mentioning. As stated before, the used dataset is small and clearly imbalanced. Training on more instances could lead to better performance; data augmentation on the audio files can be a good approach for addressing this issue. Moreover, the limited computational resources restricted extensive hyperparameter tuning, which may have affected the model's overall performance. Proper optimization of hyperparameters is needed to achieve the best possible results.

The successful development and widespread adoption of the proposed model could have significant positive impacts across various fields. In the biodiversity conservation domain, the model could allow researchers and conservationists to monitor biodiversity across wider geographical areas and over longer time periods. At the same time, the model can be deployed for continuous monitoring, enabling early detection of invasive species arrival and facilitating rapid responses to lessen their impact on ecosystems. A need for this kind of system was pointed out in previous literature (Kasinathan et al, 2021).

Another potential application lies in pest management. A system capable of automated insect species detection would be highly beneficial in agriculture, helping to prevent widespread crop damage and enabling timely interventions (Kasinathan et al, 2021).

In conclusion, this study demonstrates the potential of deep learning models for automating insect species identification using acoustic data. While there are areas for improvement, the current model serves as a strong foundation for further development.

References

Faiß, M. (2022). *InsectSet32: Dataset for automatic acoustic identification of insects (Orthoptera and Cicadidae) (0.1)* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7072196

Kortas, M. (2020, January 6). *SOUND-BASED BIRD CLASSIFICATION - towards data science*. Medium. https://towardsdatascience.com/sound-based-bird-classification-965d0ecacb2b

Scott Hoffman Black, Mace Vaughan, Chapter 88 - Endangered Insects, Encyclopedia of Insects (Second Edition), Academic Press, 2009, Pages 320-324, ISBN 9780123741448, https://doi.org/10.1016/B978-0-12-374144-8.00097-7.

Thenmozhi Kasinathan, Dakshayani Singaraju, Srinivasulu Reddy Uyyala, *Insect classification and detection in field crops using modern machine learning techniques*, Information Processing in Agriculture, Volume 8, Issue 3, 2021, Pages 446-457, ISSN 2214-3173, https://doi.org/10.1016/j.inpa.2020.09.006.