Sound-AI: A Pedagogical Tool for Exploring AI in Audio and Bioacoustic Research

Muhammad Azeem, Hoang D. Nguyen, Rosane Minghim

School of Information Technology and Computer Science, University College Cork, Ireland.
m.azeem@cs.ucc.ie, hn@cs.ucc.ie, r.minghim@cs.ucc.ie

Abstract

Artificial intelligence offers powerful methods for audio processing and analysis. Still, complex workflows and the required programming skills often limit access for students and domain experts, such as marine bioacousticians and soundscape ecologists. We present "AI EcoSound Tutor", a code-free and interactive tool that lowers these barriers by allowing users to construct and explore a complete AI pipeline for audio data analysis. Starting from raw recordings, users can choose from various feature extraction techniques (MFCC, OpenL3), apply dimensionality reduction methods (PCA, t-SNE, UMAP), and optionally perform unsupervised clustering (K-Means, GMM, HDBSCAN). The results are displayed with an interactive 2D visualisation where the user can compare multiple plots by employing various techniques, including PCA and t-SNE. Interactive plots enable the selection of points or clusters of interest, allowing exploration of spectrograms within the desired frequency range, and playing an audio clip corresponding to the selected points. An integrated "Help" feature provides explanations of each method (i.e., what it is, how it works, and its practical use in different domains, such as bioacoustics), fostering both conceptual understanding and useful skill acquisition as learning outcomes. For precomputed features or embeddings, this tool also supports training and evaluating a variety of machine learning models, providing visual feedback on the results. By merging accessibility, interactivity, pedagogy, and domain relevance, our application demystifies AI methods for interdisciplinary education and supporting research in audio analysis.

Code —
https://github.com/malikazeemcs/AIEcoSoundTutor

Datasets —
https://zenodo.org/records/7072196/files/Orthoptera.zip
https://gist.github.com/curran/a08a1080b88344b0c8a7

Introduction

Artificial intelligence (AI) has emerged as a transformative force in a wide range of fields, enabling the automation of complex tasks, pattern recognition in large datasets, and accelerated scientific discovery (Gridach et al. 2025). AI methods have empowered researchers to process and interpret sounds at different scales and levels of detail that were previously impractical or unachievable in the domain of audio analysis (Purwins et al. 2019). This has paved the way for significant advancements in fields such as bioacoustics, marine biology, and soundscape ecology, where these AI methods are being utilised to identify species, classify acoustic events and monitor ecosystem health (Sethi et al. 2020; P, R, and R 2024). These advancements have made these fields highly promising for both scientific research and educational purposes as they provide solutions for sustainability and concrete and sensory-rich data that demonstrate profound AI capabilities (Jafarian and Kramer 2025).

Although there have been significant advancements in AI methods, several barriers remain to using these approaches effectively. Many of these powerful AI tools require programming knowledge, with some expertise and familiarity in complex data science workflows (List, Ebert, and Albrecht 2017). While various libraries, such as Pandas, Librosa, Scikit-learn, or TensorFlow, are highly useful for technically skilled users,they are not easily accessible to students or domain experts lacking coding expertise (Rathod 2025). For them, this technical hurdle can delay or impair the adoption of these advanced AI methods. Furthermore, educators who seek to teach AI concepts through hands-on examples face the unavailability of intuitive, code-free platforms that allow learners to experiment with these advanced AI techniques on real-world data (Lundberg and Lee 2017).

There is a pressing need for accessible and pedagogically oriented tools for domains where data are both complex and meaningful to the learner (Witten et al. 2011; El-Sabagh 2021; Carney et al. 2020). For instance, a wide range of anthropogenic, biological, and geological acoustic signals are captured in underwater environments. Such recordings can be used to explain and demonstrate the fundamental AI concepts, including but not limited to feature extraction, dimensionality reduction, clustering and supervised classification. However, the fragmented nature of the existing tools restricts the learning experience. These tools and applications require users to manually convert file formats, write complex codes and scripts, or configure various software packages before any useful and meaningful investigation and visualisation can be achieved.

To address the gaps and challenges in accessibility of existing tools and guidance, we have developed "AI EcoSound Tutor", a fully interactive and graphical user interface (GUI) application. It is designed with the primary objective of supporting the learning and application of AI approaches in the domain of sound analysis. Our tool enables users to upload audio files and extract features such as Mel-frequency cepstral coefficients (MFCCs) (Lakdari et al. 2024) or OpenL3 embeddings (Grollmisch et al. 2021). Users can then apply various dimensionality reduction methods, including Principal Component Analysis (PCA), t-distributed Stochastic Neighbour Embedding (t-SNE), or Uniform Manifold Approximation and Projection (UMAP), for visualisation purposes (Wang et al. 2021). Optionally, unsupervised clustering can be performed using K-Means, Gaussian Mixture Model (GMM), or Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)(Wani 2024). Finally, the results are presented in an interactive 2D visualisation, where each point is annotated with relevant information, including segment time and cluster number. Users can select various points using a free lasso tool or by clicking on a point to display the associated spectrogram, along with an option to play each sound segment. These spectrograms can be customised to focus on a specific frequency range. The application also supports applying these methods to pre-computed features, e.g. by uploading a ".csv" file containing features. For supervised datasets, it enables the application of various fundamental machine learning (ML) models, including Random Forest, Decision Tree, and voting classifier. Additionally, visual feedback is provided through accuracy and confusion matrices. Additionally, it offers learners the ability to annotate unsupervised data (whether in .wav or .csv format) based on clustering results and export these annotated datasets for further analysis.

The ability of this tool to integrate functionality with explanatory guidance is its core educational importance. Users are provided with concise and context-sensitive clarification of the selected method at each stage. The explanation for each step includes the definition of the method, its operation, such as how it works and how it might be useful in their analysis. This design guarantees that it not only presents results but also fosters a conceptual understanding of the AI approaches being used. This consequently makes it particularly suitable for use in higher education settings and interdisciplinary collaborations. This work contributes to the growing landscape of AI educational resources by offering:

- An interactive, end-to-end, code-free environment for exploring and constructing AI workflows in sound analysis.
- Direct audio analysis, seamlessly integrated with interactive visualisation.
- Contextualised and pedagogically guided interpretations that reinforce conceptual learning.
- A toolkit specifically designed to lower the entry barrier and make it easier for students, educators, and researchers to get started with acoustics analysis or related fields.

In the following section, we place our work within the context of existing educational applications and tools for conceptual AI learning and audio analysis. We will explain how our tool and proposed method expand and improve upon them to meet the needs of learners in interdisciplinary fields and non-programming experts.

Related Work

AI in Sound Analysis

Machine learning and deep learning methods have revolutionised sound analysis and related fields in recent years through various approaches. These techniques enable us to extract rich representations from raw recordings for clustering, detection, and classification tasks. To capture spectral and temporal characteristics of a sound, MFCC is widely used, being one of the classic feature extraction methods (Davis and Mermelstein 1980; Wang et al. 2024). Deep embedding approaches such as OpenL3 have emerged more recently, providing a high-level and task-independent representation that can generalise across diverse acoustic domains (Cramer et al. 2019). These methods are being applied to a variety of tasks, including species identification, call detection and classification, and soundscape characterisation in environmental monitoring and bioacoustics (Stowell et al. 2019; Kahl et al. 2021). Although these pipelines are robust, they are complex to implement, especially for beginners and researchers from non-computer science backgrounds. These tools typically require programming expertise and specialised knowledge to be implemented effectively.

Code-Free AI Platforms

Several platforms are offering code-free interfaces in an effort to overcome the limitations and reduce the entry barrier for machine learning. Google Teachable Machine (Carney et al. 2020) provides learners with the ability to quickly train classification models using a custom audio, image, or posture dataset. Similarly, Weka (Hall et al. 2009) offers a wide range of machine learning models using a graphical user interface and is standard in introductory data mining courses. Various applications, such as Orange Data Mining and KNIME, also offer a visual workflow design that allows users to combine processing nodes without the need for coding. While these tools and applications work well for general AI education, they often lack field-specific functionalities. These limitations are particularly noticeable in audio-focused pipelines, where capabilities such as spectrogram visualisation, customisation of spectrograms to focus on specific acoustic calls, and tight integration with audio playback features are lacking.

Bioacoustics and Soundscape Ecology Tools

Advanced visualisation, data annotation, and call detection capabilities for ecological and conservation research are provided by various domain-specific tools, such as Raven Pro (Bioacoustics Research Program 2014), PAMGuard (Gillespie et al. 2009), and Kaleidoscope (Wildlife Acoustics 2022). These platforms are extremely useful and practical for domain experts and technicians. The way these tools integrate into a modern AI pipeline is not accessible for learners without a programming background and expertise. Furthermore, they support spectrogram analysis and playback but rarely include interactive machine learning models or integrated explanations of AI concepts, which are crucial for pedagogical use.

Gaps in Existing Tools

While various general-purpose, code-free AI platforms and domain-specific bioacoustics analysis tools are available, there is a lack of an integrated system that provides the following capabilities:

1. A seamless and fully integrated audio analysis pipeline from raw acoustics signal representation to model implementation and interpretation within a single accessible environment.
2. A unified platform, integrating advanced visualisation with exploratory analysis, and providing users with the ability to interact with intermediate data representations meaningfully.
3. Interactive mechanisms that link data exploration with field-relevant artefacts, e.g. spectrograms and audio playback, to bridge computational outputs and perceptual understanding.
4. Embedded and pedagogically informed interpretations of AI concepts, supporting practical application and conceptual learning to users from diverse backgrounds

Methodology

The primary objective of "AI EcoSound Tutor" is to assist students and other field experts with no programming background by providing a hands-on, code-free platform allowing them to explore an end-to-end AI workflow. The methodology and flow of our tool are designed in such a way that it emphasises interactivity, transparency, and guided learning. It allows learners to understand the impact of each AI method without requiring computer programming expertise.

System Overview

Our application combines various stages of a sound analysis pipeline into a single, interactive interface. Users can upload raw audio files (.wav) or precomputed numerical features data (.csv file), and progress through a configurable workflow. This workflow includes feature extraction, dimensionality reduction, clustering, interactive visualisation, and an optional application of fundamental supervised machine learning models. The purpose of the tool is to provide learners with visual and auditory feedback, enabling them to explore and gain a concrete understanding of AI concepts for soundscape ecology. A simplified workflow diagram (Figure 1) illustrates the modular structure of "AI EcoSound Tutor".

Feature Extraction

For audio input, our tool offers two feature extraction methods, including MFCC and OpenL3 embeddings. Users can choose one of the methods and configure the segment length (from 0.1 seconds to 10 seconds, default: 1 second) to compute features. MFCC captures the spectral characteristics of an audio signal and provides compact representations suitable for speech and environmental sound analysis (Lakdari et al. 2024). For an audio input signal $x(t)$, it computes the Discrete Fourier Transform (DFT), and maps frequencies f to the Mel Scale m:

$$m = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \tag{1}$$

Finally, it applies the Discrete Cosine Transform (DCT) on log filterbank energies to output cepstral coefficients.

OpenL3 embeddings are higher-level feature representations that capture the semantic content of an audio and are derived using a pretrained convolutional neural network (Cramer et al. 2019). A spectrogram is computed from given audio input, and then for spectrogram $S \in \mathbb{R}^{T \times F}$, the network learns a mapping $\phi : S \mapsto \mathbb{R}^d$, where $\mathbb{R}$ is a real number, $\mathbb{T}$ is the number of time frames, $\mathbb{F}$ is the number of frequency bins, and d is the embedding dimension.

Users can bypass this feature extraction step when they have precomputed features (i.e. CSVs as input) and directly utilise their own data for the next steps in the learning workflow. This flexibility enables the application to accommodate both new learners and domain experts working with existing datasets.

Dimensionality Reduction

To project high-dimensional features into a 2-dimensional space for visualisation, we employ various widely used dimensionality reduction algorithms, including PCA, t-SNE, and UMAP (Wang et al. 2021). These methods aim to reduce computational complexity while preserving the essential structure of the data that helps learners understand the relationships within the dataset.

PCA is a linear method that projects high-dimensional data onto a lower-dimensional subspace while maximising variance along orthogonal axes. t-SNE is a non-linear method that preserves local structures by converting pairwise similarities into probability distributions and minimising divergence between high- and low-dimensional spaces, making it effective for revealing clusters. UMAP models data as a weighted graph, preserving the global connectivity structure, which offers better scalability and computational efficiency for large datasets.

Clustering

For exploratory analysis, particularly with an unsupervised dataset, this application provides a way to apply various clustering algorithms, including K-Means, GMM, and HDBSCAN (Wani 2024). To help users identify patterns and correlations interactively, points in each cluster are assigned a unique colour. Users may choose to visualise data without applying any clustering algorithm.

K-Means divides data points into k clusters by assigning points to the nearest centroid. It minimises the sum of squared distances to the cluster centroid:

$$J = \sum_{i=1}^{k} \sum_{x_j \in C_i} ||x_j - \mu_i||^2 \tag{2}$$

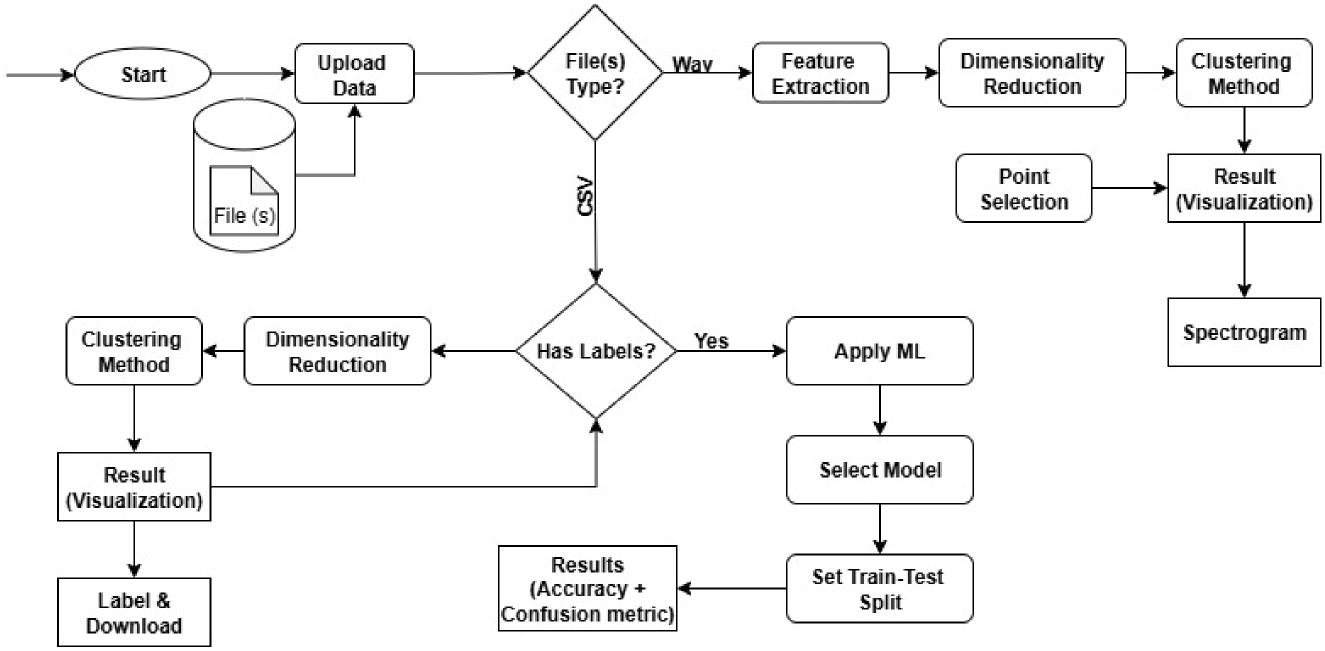

Figure 1: Proposed system workflow for analysing sound or tabular data, from feature extraction and dimensionality reduction to clustering, visualisation, and supervised learning.

where x_j is the j-th data point, C_i is the i-th cluster, $\mu_i = \frac{1}{|C_i|}\sum_{x_j \in C_i} x_j$ is the centroid of cluster C_i, and k is the total number of clusters.

GMM clustering method models data as a mixture of k Gaussian distributions, estimating the parameters $\theta = \{\pi_i, \mu_i, \Sigma_i\}$ by maximizing the log-likelihood:

$$\mathcal{L}(\theta) = \sum_{j=1}^{N} \log\left(\sum_{i=1}^{k} \pi_i \mathcal{N}(x_j \mid \mu_i, \Sigma_i)\right) \tag{3}$$

where π_i is the mixture weight, μ_i the mean, and Σ_i the covariance of the i-th Gaussian component.

HDBSCAN is a density-based clustering method that groups points into clusters based on neighbourhood density, identifying points in dense regions while labelling sparse points as noise. A point x is considered a *core point* if it has at least minPts neighbors within a distance ϵ:

$$|N_\epsilon(x)| \geq \text{minPts}, \quad N_\epsilon(x) = \{y \mid \|x - y\| \leq \epsilon\} \tag{4}$$

where $N_\epsilon(x)$ is the set of neighbours of x within radius ϵ, ϵ is the neighbourhood radius, and minPts is the minimum number of points required to form a dense region.

Data Annotation

When clustering methods are applied, the application provides an additional pedagogical feature that allows users to annotate selected points or clusters and label the resulting dataset for further analysis. This feature enables users to link exploratory analysis to downstream tasks, fostering a deeper understanding of data processing, interpretation, and real-world applications.

Interactive Visualisation

The interactive 2D visualisation plot is the core pedagogical element of "AI EcoSound Tutor". Information about individual points is displayed when hovering, and users can select points using a lasso tool, with the selected points highlighted in the plot. Learners can view spectrograms of selected points, with the option to focus on specific acoustic signals. Audio playback with an animated playhead illustrates the connection between visual features and acoustic properties. Users can generate multiple plots by modifying parameters, such as feature type or clustering method, with each plot shown in a separate window for direct comparison. These interactions enhance understanding of how AI methods process audio data.

Machine Learning Models

This application enables users to apply fundamental supervised machine learning algorithms directly, including Random Forest, Decision Tree, Linear SVC, Gradient Boosting, and Voting classifier, on labelled data (Bonaccorso 2020). Learners need to upload labelled data, adjust the train-test split using a slider, choose a model and apply the selected model. To provide visual feedback to the user, the accuracy report and confusion matrix are displayed as heatmaps, enabling the user to evaluate the model's performance. This visual feedback promotes an understanding of how data representation and model choice affect accuracy and predictive outcomes.

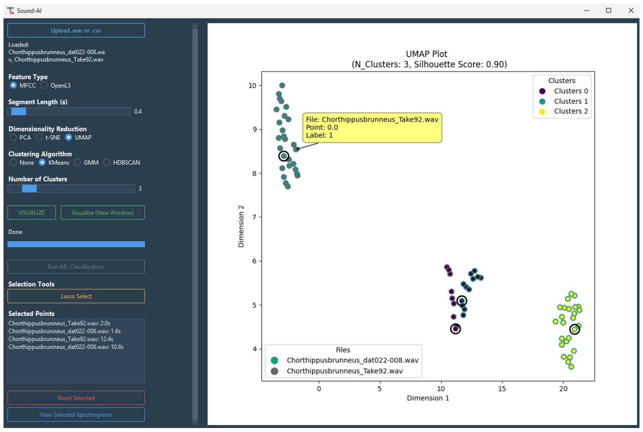

Figure 2: Application of feature extraction, dimensionality reduction, and clustering within ”AI EcoSound Tutor”. The figure illustrates how UMAP and HDBSCAN yield well-separated clusters of acoustic segments from raw audio data.

Results

To demonstrate the functionality and pedagogical potential of ”AI EcoSound Tutor”, we present two learning cases. The first uses unsupervised bioacoustic recordings to demonstrate how the workflow enables domain experts to identify vocalisations of different species. The second uses an unlabeled tabular dataset with precomputed features. Together, these studies illustrate how learners and experts can use this tool step-by-step to understand AI concepts.

Case study 1: Unsupervised Bioacoustics Recordings

For the first case study, we have used a subset of bioacoustics recordings from the Grasshopper (Orthoptera) acoustic dataset that contains recordings of Chorthippus brunneus as .wav files (Faiß 2022). This dataset was collected for species identification research and includes various Orthoptera species, with each file approximately 15 seconds in duration. To evaluate whether the system can help domain experts identify species calls within raw and unlabeled data, we have focused only on a single species in this study.

Feature Extraction: We extracted MFCC features, a widely used representation in speech and bioacoustics analysis. Segments of 0.4 seconds were used, and the first 13 MFCC coefficients were computed for each segment, following recommendations in the bioacoustics literature (Noda et al. 2019). These 13 coefficients are sufficient to capture temporal dynamics, species-relevant spectral patterns, and meaningful acoustic structure. This process transformed the raw recordings into a feature space suitable for further analysis.

Dimensionality Reduction: To explore the acoustic structure, we experimented with all three dimensionality reduction methods. UMAP revealed the clearest separation, supported by a high silhouette score. UMAP effectively captured local acoustic similarities while preserving global structure, consistent with prior research in bioacoustics.

Clustering: We experimented with various combinations of feature types, dimensionality reduction methods, and clustering algorithms available in the tool. Table 1 summarises some of these results. All experiments used a segment length of 0.4 seconds and three clusters (except for HDBSCAN, which determines the optimal number automatically). While other segment lengths and cluster numbers were tested, this configuration provided the best overall results. The combination of MFCC features with UMAP and any clustering method achieved a silhouette score of 0.9, indicating excellent clustering performance.

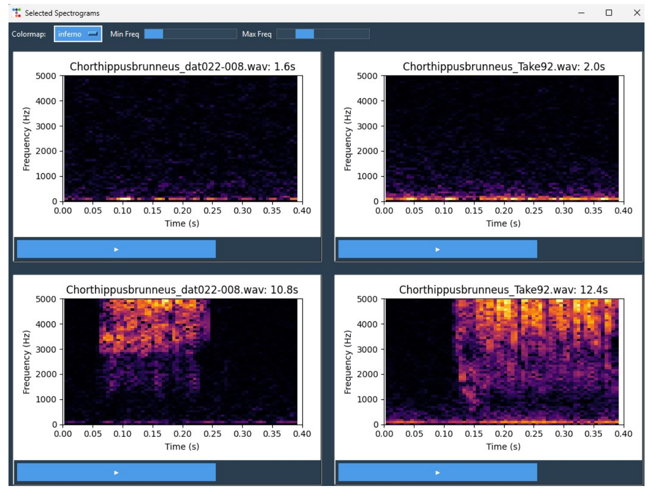

Figure 3: Spectrogram visualisation for clusters derived from MFCC–UMAP features. Cluster 0 reveals structured acoustic activity, while clusters 1 and 2 show only noise or silence, confirming cluster separation.

Results: The visualisation plot with UMAP and K-Means clustering is shown in (Figure 2). We selected some sample points from the plot and visualised their spectrograms (Figure 3) to evaluate the content of these 3 clusters. The spectrogram analysis explained a clear and consistent pattern as follows:

- Cluster 0 contained all those points where there is prominent acoustic activity corresponding to grasshopper vocalisation. The spectrogram of these points revealed distinct frequency bands, indicating the presence of an acoustic signal. We verified the existence of acoustic events by playing these segments.
- Clusters 1 and 2 contained all those segments without any specific acoustic activity or discernible call. The spectrograms of these points were flat, showing only low-frequency background noise. This indicated a silent interval or environmental noise, which was again verified by playing those segments.

Manual inspection of various additional samples across the clusters confirmed a clear pattern. Cluster 0 contained a homogeneous group of points corresponding to the acoustic

Features Type	Dimensionality Reduction	Clustering Type	Number of Clusters	Silhouette Score
OpenL3	PCA	K-Means	3	0.80
OpenL3	PCA	GMM	3	0.79
OpenL3	PCA	HDBSCAN	Auto(4)	0.57
OpenL3	t-SNE	K-Means	3	0.51
OpenL3	t-SNE	GMM	3	0.50
OpenL3	t-SNE	HDBSCAN	Auto(7)	0.53
OpenL3	UMAP	K-Means	3	0.80
OpenL3	UMAP	GMM	3	0.80
OpenL3	UMAP	HDBSCAN	Auto(3)	0.80
MFCC	PCA	K-Means	3	0.67
MFCC	PCA	GMM	3	0.64
MFCC	PCA	HDBSCAN	Auto(5)	0.71
MFCC	t-SNE	K-Means	3	0.73
MFCC	t-SNE	GMM	3	0.74
MFCC	t-SNE	HDBSCAN	Auto(5)	0.58
MFCC	**UMAP**	**K-Means**	**3**	**0.90**
MFCC	UMAP	GMM	3	0.90
MFCC	UMAP	HDBSCAN	Auto(3)	0.90

Table 1: Clustering performance (calculated using Silhouette score) across different feature types, dimensionality reduction methods, and clustering algorithms.

event of interest, demonstrating the tool's success in identifying and isolating biologically meaningful sounds. Our analysis reveals that background noise is heterogeneous; instead of forming a single cluster, it splits into two distinct groups with different acoustic signatures. The high silhouette score and HDBSCAN's consistent three-cluster solution confirm that three is the natural structure in the data. Thus, our approach not only separates acoustic events from noise but also uncovers meaningful variation within the background audio, offering potential for more detailed soundscape analysis.

For learners, this study demonstrates the utility of feature extraction, dimensionality reduction, and clustering in detecting various biological events using unsupervised acoustic data. It also illustrates the ability of "AI EcoSound Tutor" to assist users with an unsupervised analysis pipeline that can transform unlabeled recordings into meaningful biological insights. Our system not only helps in scientific analysis but also serves as a pedagogical tool for teaching core AI concepts in audio signal processing.

Case Study 2: Clustering to Supervised Classification with Iris Dataset

In this case study, we utilised the Iris dataset to demonstrate the tool's capabilities beyond audio analysis. The Iris dataset is a widely recognised benchmark in machine learning, containing 150 instances (Wu et al. 2019; ANDERSON 1935). Each instance represents an iris flower with four features — sepal length, sepal width, petal length, and petal width, belonging to one of three classes: 'setosa', 'versicolor', or 'virginica'. The class labels were encoded as numerical values 0, 1, and 2.

Dimensionality Reduction: We uploaded the Iris dataset file (.csv); therefore, the application automatically disabled feature extraction, and we proceeded directly to the dimensionality reduction phase. The label column (species) was excluded during upload to assess the effectiveness of clustering. All three dimensionality reduction methods were applied, and UMAP produced the clearest separation among the three species.

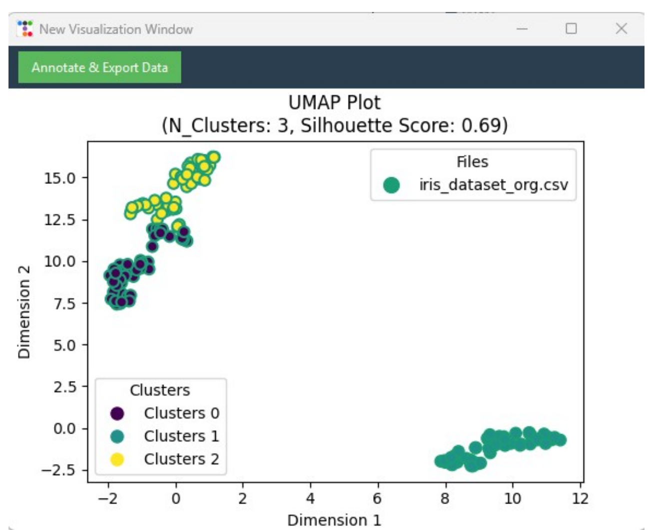

Figure 4: Projection of K-Means Clusters on Iris Dataset

Clustering: We employed various clustering techniques, but K-Means, applied to UMAP embeddings, proved to be the most effective. We set the number of clusters to 3 as there were 3 species labels in the dataset. K-Means identified three dense clusters that separated the points well and left a very small fraction of borderline points as noise (Fig-

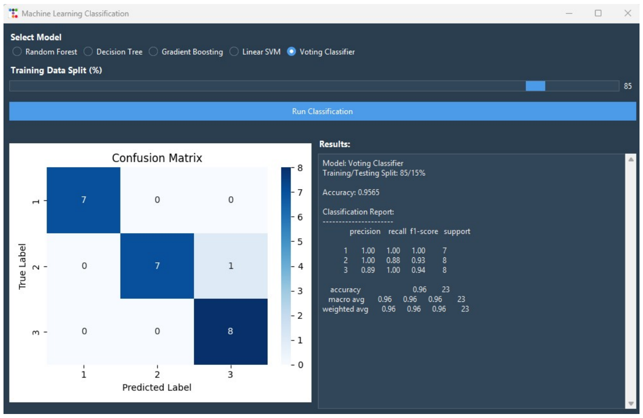

Figure 5: Confusion matrix and classification results of a Voting Classifier (ensemble of Random Forest, Decision Tree, and Linear SVC) on the Iris dataset, achieving 95.65% accuracy.

ure 4). It yielded a silhouette score of 0.69, indicating well-formed groups suitable for segregation and reflecting the natural class structure, while being robust to outliers, as clearly shown in the visualisations.

Data Annotation: To analyse the effectiveness of clustering, we assigned a label to each sample based on its cluster membership. We compared these assigned labels with ground truth species, and it revealed 145 out of 150 (96.67%) correct assignments, demonstrating a close alignment of unsupervised structure with true classes (Figure 6). This step explains a practical pathway from unsupervised to supervised data for downstream task modelling.

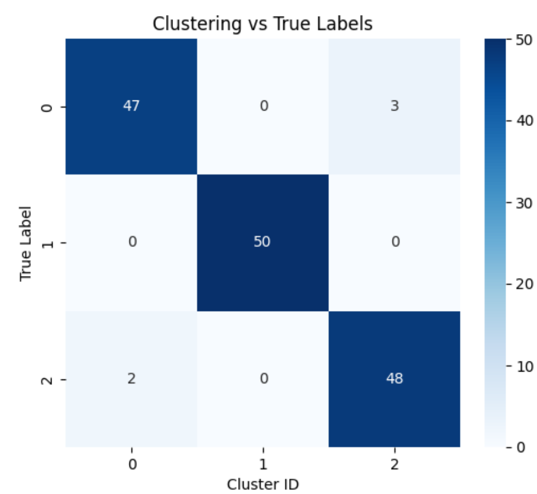

Figure 6: Comparison of Cluster Assignments with True Class Labels

Machine Learning: To train a machine learning model, we fed the Iris dataset with ground-truth labels to various models and compared the results by varying the train-test split. Voting Classifier (ensemble of Decision Tree, Random Forest, and Linear SVC) using 85% data for training and 15% for testing, achieved the best results. Figure 5 shows the achieved accuracy of 0.9565, along with a confusion matrix presented as a heatmap, which enables the learner to interpret both overall and class-wise performance.

Results: This case study highlights that learners can:

- Understand how UMAP reveals class structure from numeric features and how feature geometry separates the classes.
- Compare various linear vs. non-linear dimensionality reduction techniques and clustering methods to understand what methods work well for well-separated, non-spherical clusters.
- Validate clusters by comparing pseudo-labels to true labels, showing that clustering can help to generate annotations.
- Connect to supervised learning by relating representation quality to accuracy and analysing errors with a confusion matrix.

Conclusion and Limitations

Advancements in artificial intelligence have transformed the way audio data is analysed by providing various techniques. But many students and researchers struggle with the technical complexity required to work with these methods. The ”AI EcoSound Tutor” addresses this challenge by providing an interactive, accessible, and code-free tool for constructing and exploring AI workflows in various domains, with a focus on soundscape ecology and bioacoustics. It enables users to choose and apply different feature extraction, dimensionality reduction, clustering, and visualisation methods, fostering hands-on learning. For supervised data with precomputed feature embeddings, it enables users to directly train and evaluate fundamental ML algorithms to gain deeper insights. This application empowers students and researchers in bioacoustics and related domains to understand their data. The current version of the application is limited to a subset of commonly used methods. For feature extraction, dimensionality reduction to provide visual maps, clustering, and machine learning algorithms, we have focused on some of the most used techniques. The application supports .wav for audio and .csv format for precomputed features. While the integrated ‘Help’ feature offers valuable explanations and guidance, further enhancements in user feedback and tutorial integration will improve the learning experience. Future work will focus on expanding methods for feature extraction, clustering, and other related tasks, as well as applying deep learning algorithms to supervised data and supporting additional data types. We will focus on integrating ”AI EcoSound Tutor” into formal educational curricula to maximise its benefits and impact on AI education and research.

Acknowledgements

This publication has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland under Grant numbers 18/CRT/6222 and 12-RC-2289-P2, which are co-funded under the European Regional Development Fund.

References

ANDERSON, E. 1935. The Irises of the Gaspe peninsula. *Bulletin of American Iris Society*, 59(1): 2–5.

Bioacoustics Research Program. 2014. Raven Pro: Interactive Sound Analysis Software. http://ravensoundsoftware.com/. Accessed: 2025-08-13.

Bonaccorso, G. 2020. *Mastering Machine Learning Algorithms: Expert techniques for implementing popular machine learning algorithms, fine-tuning your models, and understanding how they work*. Birmingham, UK: Packt Publishing Ltd, 2 edition. ISBN 978-1838820299.

Carney, M.; Webster, B.; Alvarado, I.; Phillips, K.; Howell, N.; Griffith, J.; Jongejan, J.; Pitaru, A.; and Chen, A. 2020. Teachable Machine: Approachable Web-Based Tool for Exploring Machine Learning Classification. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, 1–8. New York, NY, USA: Association for Computing Machinery. ISBN 9781450368193.

Cramer, A. L.; Wu, H.-H.; Salamon, J.; and Bello, J. P. 2019. Look, Listen, and Learn More: Design Choices for Deep Audio Embeddings. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 3852–3856.

Davis, S.; and Mermelstein, P. 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4): 357–366.

El-Sabagh, H. A. 2021. Adaptive e-learning environment based on learning styles and its impact on development students' engagement. *International Journal of Educational Technology in Higher Education*, 18(1): 53.

Faiß, M. 2022. InsectSet32: Dataset for automatic acoustic identification of insects (Orthoptera and Cicadidae). https://doi.org/10.5281/zenodo.7072196. Accessed: 2025-08-14.

Gillespie, D.; Mellinger, D. K.; Gordon, J.; McLaren, D.; Redmond, P.; McHugh, R.; Trinder, P.; Deng, X.; and Thode, A. 2009. PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localization of cetaceans. *The Journal of the Acoustical Society of America*, 125(4 Supplement): 2547–2547.

Gridach, M.; Nanavati, J.; Zine El Abidine, K.; Mendes, L.; and Mack, C. 2025. Agentic AI for Scientific Discovery: A Survey of Progress, Challenges, and Future Directions. *arXiv e-prints*, arXiv:2503.08979.

Grollmisch, S.; Cano, E.; Kehling, C.; and Taenzer, M. 2021. Analyzing the Potential of Pre-Trained Embeddings for Audio Classification Tasks. In *28th European Signal Processing Conference*, 790–794.

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; and Witten, I. H. 2009. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1): 10–18.

Jafarian, N. R.; and Kramer, A.-W. 2025. AI-assisted audio-learning improves academic achievement through motivation and reading engagement. *Computers and Education: Artificial Intelligence*, 8: 100357.

Kahl, S.; Wood, C. M.; Eibl, M.; and Klinck, H. 2021. BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61: 101236.

Lakdari, M. W.; Ahmad, A. H.; Sethi, S.; Bohn, G. A.; and Clink, D. J. 2024. Mel-frequency cepstral coefficients outperform embeddings from pre-trained convolutional neural networks under noisy conditions for discrimination tasks of individual gibbons. *Ecological Informatics*, 80: 102457.

List, M.; Ebert, P.; and Albrecht, F. 2017. Ten Simple Rules for Developing Usable Software in Computational Biology. *PLOS Computational Biology*, 13(1): 1–5.

Lundberg, S. M.; and Lee, S.-I. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Noda, J. J.; Travieso-González, C. M.; Sánchez-Rodríguez, D.; and Alonso-Hernández, J. B. 2019. Acoustic Classification of Singing Insects Based on MFCC/LFCC Fusion. *Applied Sciences*, 9(19).

P, I. C.; R, D. K.; and R, M. 2024. Bird Sound Identification System using Deep Learning. *Procedia Computer Science*, 233: 597–603. 5th International Conference on Innovative Data Communication Technologies and Application.

Purwins, H.; Li, B.; Virtanen, T.; Schlüter, J.; Chang, S.-Y.; and Sainath, T. 2019. Deep Learning for Audio Signal Processing. *IEEE Journal of Selected Topics in Signal Processing*, 13(2): 206–219.

Rathod, P. R. 2025. A Research Review on Comparative Analysis of Data Mining Tools and Techniques. *International Journal on Science and Technology*, 16(3).

Sethi, S. S.; Jones, N. S.; Fulcher, B. D.; Picinali, L.; Clink, D. J.; Klinck, H.; Orme, C. D. L.; Wrege, P. H.; and Ewers, R. M. 2020. Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. *Proceedings of the National Academy of Sciences*, 117(29): 17049–17055.

Stowell, D.; Wood, M. D.; Pamuła, H.; Stylianou, Y.; and Glotin, H. 2019. Automatic acoustic detection of birds through deep learning: the first bird audio detection challenge. *Methods in Ecology and Evolution*, 10(3): 368–380.

Wang, M. Y.; Chu, Z.; Entzminger, C.; Ding, Y.; and Zhang, Q. 2024. Visualization and interpretation of mel-frequency cepstral coefficients for uav drone audio data. In *Proceedings of the 13th International Conference on Data Science, Technology and Applications*, 528–534.

Wang, Y.; Huang, H.; Rudin, C.; and Shaposhnik, Y. 2021. Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMap, and PaCMAP for Data Visualization. *Journal of Machine Learning Research*, 22(201): 1–73.

Wani, A. A. 2024. Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions. *PeerJ Computer Science*, 10(2286).

Wildlife Acoustics. 2022. Kaleidoscope Pro Software. https://www.wildlifeacoustics.com/products/kaleidoscope.

Witten, I. H.; Frank, E.; Hall, M. A.; and Pal, C. J. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Cambridge, MA: Morgan Kaufmann, 4th edition. ISBN 978-0-12-374856-0.

Wu, Y.; He, J.; Ji, Y.; Huang, G.; Yao, H.; Zhang, P.; Xu, W.; Guo, M.; and Li, Y. 2019. Enhanced Classification Models for Iris Dataset. *Procedia Computer Science*, 162(1): 946–954.