A LIGHTWEIGHT MINIROCKET-BASED PIPELINE FOR BAT ECHOLOCATION CALL DETECTION AND CLASSIFICATION WITH ANOMALY FILTERING

S. Alipek[1], C. Kexel[2], M. Lambat[1], M. Strack[3], M. Dietz[3], J. Moll[1], P. Kraemer[1]

[1] *Department of Mechanical Engineering, University of Siegen, Germany*
[2] *Verein für Datenwissenschaften (VEFDAWI), Frankfurt am Main, Germany*
[3] *Institut für Tierökologie und Naturbildung GmbH (ITN), Laubach-Gonterskirchen, Germany*

Abstract

Automatic detection and classification of bat echolocation calls is essential for large-scale biodiversity monitoring, yet remains challenging under real-world acoustic conditions and constrained hardware environments. This work presents a lightweight and deployable pipeline for bat call analysis that combines simple signal processing with efficient time-series machine learning. An adaptive amplitude-based segmentation using the Otsu algorithm is applied to extract dominant acoustic events from continuous recordings. These candidate events, primarily representing bat echolocation chirps, are transformed into feature representations using MiniROCKET. The resulting feature vectors are classified by a Multi-Layered-Perceptron trained on a reference dataset of bat taxa common in Germany. To improve robustness, an auxiliary anomaly detection stage is introduced. An Isolation Forest model is trained on the same feature space to identify atypical events. During inference, detected segments are first evaluated for similarity to known bat calls before being passed to the classifier. This cascaded strategy reduces false detections in acoustically challenging environments. The proposed approach provides a practical balance between performance, robustness, and deployability for scalable acoustic wildlife monitoring.

Keywords: *bioacoustics, bat call analysis, signal processing, machine learning*

1. Introduction

Bats are small flying mammals inhabiting many different ecosystems around the world and their ability to sense environmental changes makes them essential bioindicators [1], [2]. The natural activity of bats, including foraging and general locomotion, implies an important role in pest control and pollination, and is shown to provide significant monetary benefits for the agricultural industry [3]. The anthropocene, to which bats are susceptible to, is quickly changing and many species are highly endangered [3], [4].

For bats in specific, the acoustic monitoring is the most intuitive approach, since most bat species consistently rely on echolocation to navigate and forage [5]. The domain of acoustic bat monitoring has matured in several areas during the last few decades especially in the context of passive acoustic monitoring (PAM) and the theory behind methods to analyze bat echolocation calls [5]–[7]. Convolutional Neural Networks (CNNs) in particular, became very popular in recent years through their innate compatibility with spectrograms, which represent acoustic information in image-like patterns [8], [9]. PAM has become one of the most widely used methods in bat monitoring because it enables continuous, non-invasive observation of vocalizations. However, it has fundamental ecological limitations such as large intra- and interspecific call variation, habitat complexity and regional differences, that impose boundaries beyond the capabilities of classification algorithms [5]. Technical and data related limitations on the other hand, can be significantly improved with edge computing applications [10].

One major limitation is the generation of vast amounts of data, which is difficult to store and analyse without a reliable infrastructure including power supply, transmission technology, bandwidth and storage size [10]–[13]. Running processes or models directly on the edge system enables pre-screening or classification in realtime which can significantly reduce the transmitted data size. Early adoptions that highlight

Corresponding author: *sercan.alipek@uni-siegen.de.*

edge applications in the context of bat monitoring presented a functional end-to-end Internet-of-Things (IoT) pipeline with integrated edge boards to perform real-time classifications with CNNs [14]–[16]. While some setups used GPU-accelerated hardware such as NVIDIA Jetson Nano or Google Coral [15], [16], other solutions relied on CPU-focused devices like Intel Edison on Arduino [14] or certain Raspberry Pi [15]–[17]. What all approaches had in common was that the respective CNN models were pretrained offline on more performant local hardware and were only used for inference on the edge system. Furthermore, they necessarily transformed the audio data into spectrograms to train their CNNs. While spectrograms are a powerful representation of acoustic information, they come with additional computational costs, due to their two-dimensional format compared to one-dimensional raw waveforms. We therefore provide an alternative classification approach for future edge applications, that (i) does not require a transform of serial audio data into two-dimensional spectrograms and (ii) can efficiently extract and classify raw waveform segments on CPU. The integration of the MiniROCKET algorithm [18] replaces the CNN as a feature extractor. Instead of convolutional layers that are trained on the provided data, MiniROCKET uses a predefined set of simple random convolutions. The final feature vector from MiniROCKET is then passed to a conventional classifier such as Multi-Layered-Perceptron (MLP). To the best of our knowledge, this work is the first that applies MiniROCKET on bat ultrasound data.

To further reduce computation costs, Otsu's method [19] is used to adaptively extract individual pulses/chirps which are sparse compared to the recording time. In this work, the root mean square (RMS) envelope of the signal amplitude was computed to remove tiny fluctuations near the threshold level that would lead to false extractions. A similiar approach with different domain-related adjustments was taken for songbird vocalizations to separate vocals from silence [20]. Other works computed the spectrogram and applied Otsu's method on various animal species including birds and bats [21] or human subjects [22] similiar to the algorithms original purpose.

Most acoustic classification models are trained to distinguish between classes of interest that each follow a common structure, and generally lack an anomaly class. The main reason is the difficulty to anticipate what types of unwanted noise could occur in the given scenario. In the rare cases where the noise soundscape is more controllable, the acquisition of sufficient samples makes the task unfeasible. A popular approach is to instead teach a separate model what a positive or normal sample is. By learning the feature distribution of normal samples, the model evaluates structurally different samples as an anomaly. Samples that are not rejected are more likely to represent a positive class and can be passed to the regular classifier.

In this work, Isolation Forest [23] is used as a simple and efficient auxiliary model for anomaly rejection. It is trained on the MiniROCKET feature vectors of a reference bat dataset. Previous research already deployed Isolation Forest as an anomaly detector in office environments where it was trained with Mel-coefficients [24] or with a feature vector of custom acoustic parameters from normal manufacturing machines [25]. However, this is the first study to use Isolation Forest for bat acoustics and general bioacoustics.

2. Theory

2.1. Otsu's Method

Otsu's method originates in computer vision, where it represents a common ansatz for binary segmentation, separating foreground from background [19]. Rather than relying on a manually tuned and fixed cutoff, it derives a threshold directly from the (bi-modal) data distribution, which makes it inherently adaptive to varying environmental/operational conditions. This property becomes valuable for continuous machine listening, where the noise floor drifts over time and bat chirps occur only sparsely against long stretches of near-silence.

In this work the method operates on the one-dimensional RMS envelope of the waveform instead of the two-dimensional spectrogram. The split location optimizes the discrimination between the two categories of background noise and actual acoustic events, similar to unsupervised 2-means clustering. At linear cost in the number of recorded samples, splitting reduces to a single pass over the envelope histogram, keeping the complexity lightweight for on-device use.

2.2. MiniROCKET

Once segmentation has reduced the recording to isolated snippets, subsequent classification no longer needs to model long-range temporal dependencies. Concentrated waveform shapes represent a regime in which ROCKET (RandOm Convolutional KErnel Transform) performs strongly. Its large bank of fixed convolutional kernels, drawn at random in the original ROCKET, yields expressive features without learning. MiniROCKET [18] tightens this design by fixing the kernels to a single short length with two-valued weights, which lowers both computation and variability. Here, different dilations allow to probe several time scales. The resulting feature maps get condensed by one distinct pooling operation: the proportion of positive values (PPV), which effectively measures how prevalently a characteristic pattern occurs within a segment. Stacking these values across kernels yields a fixed-length feature vector.

2.3. Isolation Forest

This algorithm builds an ensemble of decision trees that recursively partition the high-dimensional feature space of the MiniROCKET vectors by randomly selecting one feature and applying the split by a value. Each split increases the depth of a tree by 1. Chirps that require fewer partitions to become isolated are structurally different than the majority, lead to higher

anomaly scores and are identified as outliers. The majority of chirps usually requires more splits due to a greater similarity and is therefore considered representative of the dataset.
Isolation Forests need neither geometric distance nor density estimation, which lets them run in linear time on small sub-samples with a modest memory footprint.

3. Dataset

The reference dataset used to train the MLP classifier and the Isolation Forest was collected in Hessen, Germany in the time between 2010 and 2024 using Batcorder audio recording devices. It contains typical echolocation calls from three of the most common German bat species, i.e. Myotis Myotis, Pipistrellus Pipistrellus and Nyctalus Noctula. They are prominent representatives of their genus (Pipistrellus, Myotis) or group (Nyctaloid). The class distribution of the dataset is shown in Fig. 1. All recordings were created with a 500 kHz sampling rate, which fully covers the frequency bandwidth of bat calls. While the number of recordings for Nyctaloid is twice the number of the other two classes, they contain fewer individual calls, called chirps or pulses. Nyctaloid contains ∼15000 pulses similar to Pipistrellus with ∼14500 and Myotis contains ∼22500 pulses.

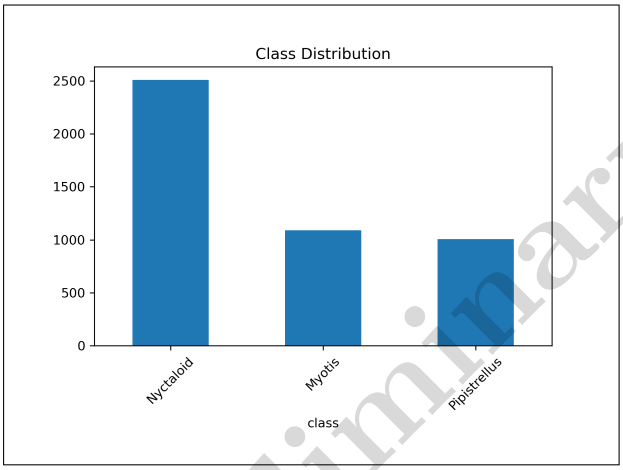

Figure 1: Class distribution of reference dataset containing three bat species. Shows the number of recordings.

4. Pipeline Design

4.1. Automated Chirp Detection

Before audio recordings can be fed to the anomaly rejection and bat classification part they must be preprocessed to maintain an efficient pipeline. With bat calls being usually sparse compared to the overall recording length, an automated detection of individual call events is described in the following to attain efficiency for edge-applications (see Fig. 2).
First, an RMS envelope of the raw waveform is computed with a 1.5ms window size and 0.1ms hop size to remove local fluctuations of the signal amplitude (top plot, Fig. 2). To enable the individual extraction of echolocation pulses, Otsu's method is applied to the RMS envelope of each recording using 512 histogram bins. The separate application of Otsu's method per recording creates a helpful adaptive thresholding process (top plot, dashed horizontal line, Fig. 2) because each recording can hold very specific signal-to-noise ratios (SNR).
An additional criteria before the final chirp extraction is the implementation of the minimum expected inter-pulse-interval (IPI) of the reference dataset (top plot, dashed vertical lines, Fig. 2). To reduce the number of false pulse detections and separate detections of pulses and nearby pulse echoes, a detection cooldown is added to the chirp selection process that allows the next pulse detection only after the minimum IPI across all expected bat species has passed (i.e. 80ms). Most clearly-audible bat echolocation calls have a much higher amplitude than the background noise floor. The adaptive threshold would then be naturally high, and could lead to missed weaker pulses. A reduction factor lesser 1 can be multiplied with the threshold to also fetch low-amplitude calls in recordings with high SNR.
The indices of detected segments within the RMS envelope are extracted, refactored by hop size and window size, to then extract the raw waveform of that chirp (mid plot, highlighted segments, Fig. 2). The spectrogram in the bottom plot of Fig. 2 is added to visualize the individual bat echolocation chirps in frequency domain.

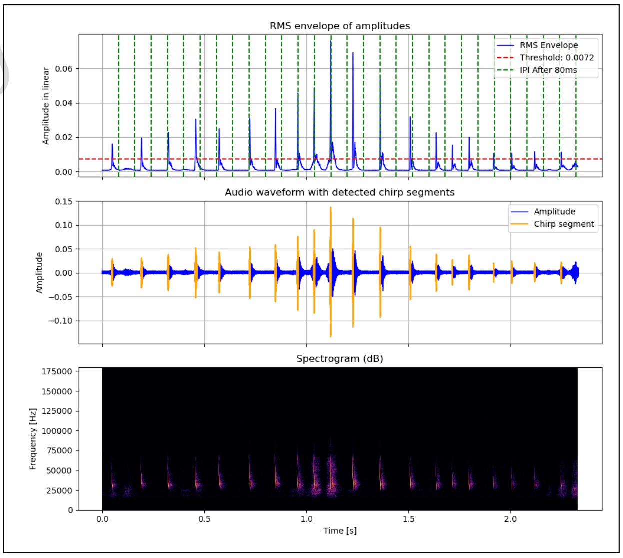

Figure 2: Visualization of automated chirp extraction process.

4.2. Complete Pipeline

The proposed pipeline design is shown in Fig. 3. The automated detection and extraction of individual chirps is illustrated in detail in Section 4.1. Those found chirps are then zero-padded to a chirp length of 4096 samples before they are separately passed through MiniROCKET.

Both anomaly rejection and chirp classification were trained with the MiniROCKET feature vectors from 80% of the reference data. The anomaly rejection model is an Isolation Forest trained to learn a correct class appearance. The model used 300 estimators and a contamination factor of 0.0075. In case that new data (unseen test data) is not considered an anomaly, because it is structurally related to the training data, the MiniROCKET output of that chirp will be fed to the classifier, which is an MLP. Otherwise, the audio recording will be discarded because it is unlikely to hold a familiar bat echolocation call.

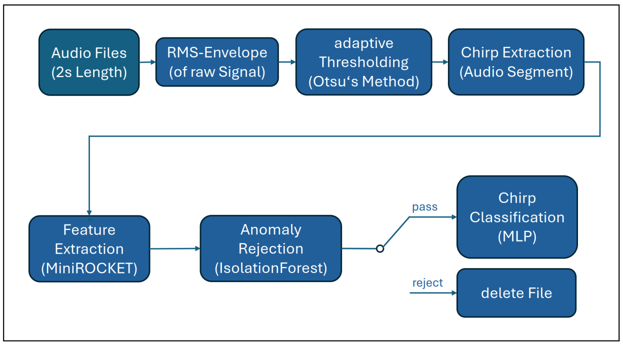

Figure 3: Diagram of the complete pipeline.

5. Results and Discussion

5.1. Isolation Forest Performance

Unwanted acoustic events can be called anomalies and embody a large and often unpredictable soundscape, especially in open field applications [9]. An Isolation Forest trained on MiniROCKET feature vectors from detected bat chirps was used as an anomaly rejection model. It was first verified on an unseen test partition (20% of the reference dataset) examining the same feature space for false negatives. No recordings from the classes Myotis and Pipistrellus were rejected. Only a small fraction (<1%) from Nyctaloid was rejected. To further evaluate the risk of bias and overfitting, the same trained Isolation Forest was tested against other, but private, bat datasets (more genera, different location, different recording devices, different sampling rates) and one anomaly noise dataset (stationary, non-stationary, different device and sampling rate).

It was observed that bat calls of the same genera were still seldomly rejected, although it was more then for the reference dataset. Bat call recordings from unknown genera were rejected more often, based on their call differences compared to learned genera. Recordings from the noise dataset contained non-silent sequences similar to cricket stridulation or machine-caused vibration. They were rejected more than 80% of the time, which implies a certain robustness of the Isolation Forest ansatz. In an actual field application the usage of this Isolation Forest is intended to increase the specificity through outlier filtering, with the risk of a reduced sensitivity. This trade-off should be considered when defining model parameters such as the contamination factor or number of estimators, also in relation to the complexity of the input data. A more detailed investigation of other datasets for benchmarking and behavioral analysis is beyond the scope of this work.

5.2. Chirp Classification Performance

For the general performance analysis of the chirp classifier, the pure reference dataset was used without a preceding anomaly rejection. In a first iteration, the focus was on individual chirp classifications. The confusion matrix in Fig. 4 shows a solid classification performance considering the diagonal from the top left to the bottom right. Striking is the bottom left and top right corner that indicate a partial confusion of Myotis calls with Pipistrellus calls and vice versa. Based on the habitat-related adaptations of both genera, there is a realistic chance to mistake one chirp for the other genus. In highly cluttered environments, both genus calls can converge into similar steep, short, and broadband pulses, especially when a Pipistrellus call represents a larger than usual bandwidth while a Myotis call is of a smaller bandwidth [26].

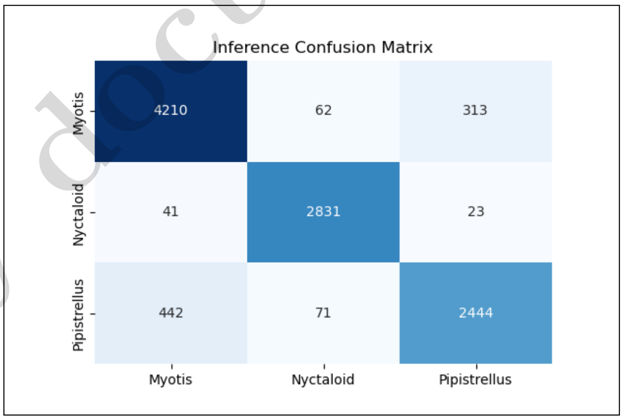

Figure 4: Confusion matrix of individual chirp classification. (Test partition)

A second iteration was based on a majority voting step (Fig. 5). Hereby, individual misclassifications of chirps within a continuous recording can be ignored towards a correct classification of the overall call sequence. It can be particularly helpful when the automatic chirp detection is exposed to low-quality audio data, where a certain amount of chirps is below the noise floor or several individual anomaly events are above the adaptive amplitude threshold. In such a scenario, if the majority of detected events is of bats, the voting process enables a correct classification of that recording.

For a better understanding, the class-based F_1 scores for both evaluation steps are listed in Table 1. It shows a slight improvement of the score for Pipistrellus, which indicates that several recordings of Pipistrellus, that held a larger proportion of individual false Myotis detections, are correctly classified as Pipistrellus on the recording level. I.e., false Myotis pulses

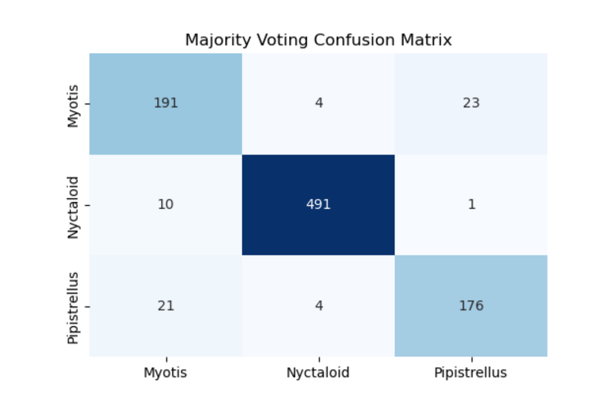

Figure 5: Confusion matrix of classification with majority voting over individual recordings. (Test partition)

in recordings of Pipistrellus no longer occur in the bottom left corner of the confusion matrix (Fig. 5) while no false majority votings as Myotis are added, effectively increasing the F_1 score for Pipistrellus. Likewise, the class Nyctaloid appears to benefit from the majority voting. The score increase resulted from falsely classified chirps of Nyctaloid recordings that are now predominantly included in the Nyctaloid class. Only for the class Myotis was a performance decrease registered. By comparing the proportions of both matrices regarding the Myotis class (left vertical line, top horizontal line) a certain trend becomes visible, which displays the opposite as for the class Pipistrellus.

Table 1: Class-based F_1 scores of individual chirps and majority voting.

Class	F_1 score (chirp)	F_1 score (majority)
Pipistrellus	0.852	0.878
Nyctaloid	0.966	0.981
Myotis	0.908	0.868

It is important to mention that this work is a qualitative proof-of-concept and is not intended as a large scale benchmarking study, which is why methods such as cross-validation are neglected, but are recommended for statistical significance tests in future studies. Only then can more nuanced conclusions be drawn about the significance of F_1 scores and their connection to chirp inference or majority voting.

5.3. Final Thoughts

Since all results of both anomaly rejection and chirp classification heavily rely on the quality of labels, the reference dataset was meticulously selected in terms of high SNR and absence of noise anomalies. This quality control enabled a robust chirp detection and prevented the extraction of anomalies that otherwise could have been stored as bat chirps. It further gives the opportunity to show the potential of both MiniROCKET-based anomaly rejection and chirp classification when all input events truly contain a bat chirp.
The simplified performance study using three individual species of three distinct bat genera showed the great potential of evaluating waveform segments, instead of chunky spectrograms, in combination with efficient adaptive thresholding via Otsu's method and feature extraction using MiniROCKET. F_1 scores between 0.85 and 0.98 per species demonstrated the effectiveness of such a domain-optimized approach. With an auxiliary anomaly rejection via Isolation Forest, the proposed end-to-end pipeline is particularly suitable for real-time and edge applications in nature.

6. Acknowledgments

This project was funded by the SAT-AND-SOUND project from the Federal Ministry for Economic Affairs and Energy (grant ID 16GM105902). The authors thank their colleagues Moritz Mälzer, Sebastian Beck and Sruthy Venu from the University of Siegen and also their project partners IMST GmbH (Kamp-Lintfort, Germany) for the technical discussions.

References

[1] D. Russo, V. B. Salinas-Ramos, L. Cistrone, S. Smeraldo, L. Bosso, and L. Ancillotto, "Do we need to use bats as bioindicators?", *Biology*, vol. 10, no. 8, p. 693, 2021.

[2] G. Jones, D. S. Jacobs, T. H. Kunz, M. R. Willig, and P. A. Racey, "Carpe noctem: The importance of bats as bioindicators", *Endangered species research*, vol. 8, pp. 93–115, 2009.

[3] C. C. Voigt and T. Kingston, "Bats in the anthropocene", in *Bats in the Anthropocene: conservation of bats in a changing world*, Springer, 2015, pp. 1–9.

[4] W. F. Frick, T. Kingston, and J. Flanders, "A review of the major threats and challenges to global bat conservation", *Annals of the new York Academy of Sciences*, vol. 1469, no. 1, pp. 5–25, 2020.

[5] L. Pino-Aedo and A. Muñoz-Sáez, "A systematic review of bat echolocation analysis methods and their implications for monitoring and conservation", *Discover Conservation*, vol. 3, no. 1, p. 17, 2026.

[6] V. Zamora-Gutierrez, M. C. MacSwiney G, S. Martínez Balvanera, and E. Robredo Esquivelzeta, "The evolution of acoustic methods for the study of bats", in *50 Years of bat research: foundations and new frontiers*, Springer, 2021, pp. 43–59.

[7] E. R. Britzke, J. E. Duchamp, K. L. Murray, R. K. Swihart, and L. W. Robbins, "Acoustic identification of bats in the eastern united states: A comparison of parametric and nonparametric methods", *The Journal of Wildlife Management*, vol. 75, no. 3, pp. 660–667, 2011.

[8] Y. Paumen, M. Mälzer, S. Alipek, J. Moll, B. Lüdtke, and H. Schauer-Weisshahn, "Development and test of a bat calls detection and classification method based on convolutional neural networks", *Bioacoustics*, vol. 31, no. 5, pp. 505–516, 2022.

[9] S. Alipek, M. Maelzer, Y. Paumen, H. Schauer-Weisshahn, and J. Moll, "An efficient neural network design incorporating autoencoders for the classification of bat echolocation sounds", *Animals*, vol. 13, no. 16, p. 2560, 2023.

[10] H. Hua, Y. Li, T. Wang, N. Dong, W. Li, and J. Cao, "Edge computing with artificial intelligence: A machine learning perspective", *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–35, 2023.

[11] S. S. Sethi, F. Fossøy, B. Cretois, and C. M. Rosten, "Management relevant applications of acoustic monitoring for norwegian nature–the sound of norway", 2021.

[12] S. S. Sethi, R. M. Ewers, N. S. Jones, A. Signorelli, L. Picinali, and C. D. L. Orme, "Safe acoustics: An open-source, real-time eco-acoustic monitoring network in the tropical rainforests of borneo", *Methods in Ecology and Evolution*, vol. 11, no. 10, pp. 1182–1185, 2020.

[13] C. L. Walters, A. Collen, T. Lucas, K. Mroz, C. A. Sayer, and K. E. Jones, "Challenges of using bioacoustics to globally monitor bats", in *Bat evolution, ecology, and conservation*, Springer, 2013, pp. 479–499.

[14] S. Gallacher, D. Wilson, A. Fairbrass, D. Turmukhambetov, M. Firman, S. Kreitmayer, O. Mac Aodha, G. Brostow, and K. Jones, "Shazam for bats: Internet of things for continuous real-time biodiversity monitoring", *IET Smart Cities*, vol. 3, no. 3, pp. 171–183, 2021.

[15] I. Zualkernan, J. Judas, T. Mahbub, A. Bhagwagar, and P. Chand, "An aiot system for bat species classification", in *2020 IEEE international conference on Internet of Things and Intelligence System (IoTaIS)*, IEEE, 2021, pp. 155–160.

[16] T. Mahbub, A. Bhagwagar, P. Chand, I. Zualkernan, J. Judas, and D. Dghaym, "Bat2web: A framework for real-time classification of bat species echolocation signals using audio sensor data", *Sensors*, vol. 24, no. 9, p. 2899, 2024.

[17] A. Vuilliomenet, S. Martínez Balvanera, O. Mac Aodha, K. E. Jones, and D. Wilson, "Acoupi: An open-source python framework for deploying bioacoustic ai models on edge devices", *Methods in Ecology and Evolution*, vol. 17, no. 1, pp. 67–76, 2026.

[18] A. Dempster, D. F. Schmidt, and G. I. Webb, "Minirocket: A very fast (almost) deterministic transform for time series classification", in *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2021, pp. 248–257.

[19] N. Otsu, "A threshold selection method from gray-level histograms", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[20] D. G. Mets and M. S. Brainard, "An automated approach to the quantitation of vocalizations and vocal learning in the songbird", *PLoS computational biology*, vol. 14, no. 8, e1006437, 2018.

[21] S. Orlandi, P. H. Dejonckere, J. Schoentgen, J. Lebacq, N. Rruqja, and C. Manfredi, "Effective pre-processing of long term noisy audio recordings: An aid to clinical monitoring", *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 799–810, 2013.

[22] M. J. Guerrero, C. L. Bedoya, J. D. López, J. M. Daza, and C. Isaza, "Acoustic animal identification using unsupervised learning", *Methods in Ecology and Evolution*, vol. 14, no. 6, pp. 1500–1514, 2023.

[23] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest", in *2008 eighth ieee international conference on data mining*, IEEE, 2008, pp. 413–422.

[24] M. Antonini, M. Vecchio, F. Antonelli, P. Ducange, and C. Perera, "Smart audio sensors in the internet of things edge for anomaly detection", *IEEE Access*, vol. 6, pp. 67 594–67 610, 2018.

[25] M. N. Ahangar, Z. Farhat, A. Sivanathan, N. Ketheesram, and S. Kaur, "Explainable ai-driven quality and condition monitoring in smart manufacturing", *Sensors*, vol. 26, no. 3, p. 911, 2026.

[26] D. Russo, L. Ancillotto, and G. Jones, "Bats are still not birds in the digital era: Echolocation call variation and why it matters for bat species identification", *Canadian Journal of Zoology*, vol. 96, no. 2, pp. 63–78, 2018.