Li et al. *Zool. Res.* 2026, 47(6): 1828−1838
https://doi.org/10.24272/j.issn.2095-8137.2025.642
https://cstr.cn/32248.14.zr.2025.642

ZR | Zoological Research

Article Open Access

Deep learning reveals persistent individual signatures in bat echolocation calls of the greater leaf-nosed bat

Ao-Qiang Li[1,#], Wan-Tao Huang[2,#], Xin-Yuan Xie[2,#], Wei-Ming Wen[2], Liang Ji[2], Hui-Jun Zhang[1], Cheng-Yun Zhang[2,*], Jin-Hong Luo[1,3,*]

[1] *Key Laboratory of Pesticide & Chemical Biology of Ministry of Education*, *Hubei Key Laboratory of Genetic Regulation and Integrative Biology*, *School of Life Sciences*, *Central China Normal University*, *Wuhan*, *Hubei* 430079, *China*

[2] *School of Electronics and Communication Engineering*, *Guangzhou University*, *Guangzhou*, *Guangdong* 510006, *China*

[3] *Hubei International Science and Technology Cooperation Base for Ecology and Environment*, *Central China Normal University*, *Wuhan*, *Hubei* 430079, *China*

ABSTRACT

Intraspecific variation is a prerequisite for natural selection and can manifest in various phenotypic traits, including vocal signals. However, classifying individuals based on their vocalizations, or acoustic individual identification (AIID), remains a significant challenge. This is particularly true for species that use rapidly varying echolocation calls for orientation. Here, we demonstrate that deep learning can overcome the limitations of traditional methods and reveal persistent individual signatures within bat echolocation calls. We recorded 17 584 echolocation calls from 34 individuals of the greater leaf-nosed bat (*Hipposideros armiger*) under controlled laboratory conditions, with 19 individuals recorded repeatedly over three months. We show that a convolutional neural network (CNN) significantly outperforms a traditional method, achieving an average identification accuracy of 84% for single calls and 91% for call sequences. In contrast, the traditional discriminant functional analysis method achieved accuracies of only 39% and 47%, respectively. By systematically altering the temporal structure of echolocation calls in input sequences, we found that temporal patterning enhances individual classification accuracy, suggesting it contributes to the encoding of individual-specific information. This study reveals that echolocation calls of *H. armiger* can contain stable, individual identity information that was previously undetectable. Our findings highlight the potential of deep learning for non-invasive AIID and provide a methodological basis for future studies aiming to monitor animals in more dynamic environments.

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2026 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Keywords: Acoustic communication; Artificial intelligence; Biomarker; Bioacoustics; Wildlife

INTRODUCTION

Intraspecific variation in genetic traits can provide individuals with adaptive advantages to drive natural selection (Darwin, 1859). Individual identification serves as the cornerstone of ecological and evolutionary research, providing critical insights into population dynamics, social structures, and movement ecology (Cantor et al., 2021; Clutton-Brock & Sheldon, 2010; Sheldon et al., 2022). To date, individual identification and tracking in the wild have relied heavily on conventional techniques like mark-recapture and telemetry (Dechmann et al., 2009; Goldshtein et al., 2024; Greif et al., 2014). However, these methods suffer from significant limitations: they are labor-intensive, invasive, and susceptible to observer bias or behavioral alterations in the study subjects (Camacho et al., 2017; McClintock & White, 2009; Terry et al., 2005).

Acoustic Individual Identification (AIID) represents a promising alternative approach for individual identification, given that vocalizations encode individual-specific features through genetic, environmental, and cultural influences (Aplin, 2019; Hyacinthe et al., 2023; Smith-Vidaurre et al., 2021; Teixeira et al., 2019). AIID involves identifying individuals by their calls, eliminates the need for making physical contact with the individual, allows efficient long-term monitoring, and reduces personnel costs. Early AIID implementations using traditional machine learning, including support vector machines, hidden Markov models, or Discriminant Function Analysis (DFA) (Fagerlund, 2007; Putland et al., 2018; Tan et al., 2022), demonstrated conceptual promise. However, they suffered from several critical constraints: reliance on expert knowledge and experience, vulnerability to ambient

Received: 11 December 2025; Accepted: 25 May 2026; Online: 04 June 2026

Foundation items: This work was supported by the Fundamental Research Funds for the Central Universities (CCNU26PTFZ007; JC2026TS-001), the National Natural Science Foundation of China (32270535 and 32400395), and the China Postdoctoral Science Foundation (30205202371)

[#]Authors contributed equally to this work

*Corresponding authors, Email: cyzhang@gzhu.edu.cn; jluo@ccnu.edu.cn

noise contamination and inter-class spectral overlap, constrained generalizability under small-sample conditions, and performance degradation caused by low signal-to-noise ratios (SNR) (Stowell, 2022; Xie et al., 2021). These limitations critically impair identification fidelity, often resulting in classification accuracy that falls below 50%, which is insufficient for robust ecological inference and restricts researchers to studying individuals only in specific contexts.

Recent advances in deep learning, particularly convolutional neural networks (CNNs), offer paradigm-shifting capabilities for pattern recognition, garnering increased attention from ecologists, CNNs are widely used for automated species identification, environmental monitoring, ecological modelling, and behavioral studies (Borowiec et al., 2022; Knight et al., 2024; Oestreich et al., 2024; Pichler & Hartig, 2023; Stowell et al., 2019). Unlike traditional machine learning approaches that rely on hand-crafted feature extraction, CNNs autonomously extract discriminative features through hierarchical learning of spectrotemporal patterns, bypassing error-prone manual feature selection while maintaining robustness to acoustic variability (Rundo & Militello, 2024). Such tools can automate the analysis of various types of data, ranging from species abundance to behavioral patterns and derived from multiple data sources such as images or audio recordings. Moreover, the use of deep learning methods for individual identification has also become a subject of heightened research involving diverse taxa, such as humans, primates, and birds (Cao et al., 2024; Ferreira et al., 2020; Ranjan et al., 2018; Schofield et al., 2019). However, most existing applications focus on image-based individual identification, while the use of acoustic signals for individual identification remains underdeveloped. A key unresolved question is whether deep learning is suitable for AIID, considering the pronounced intra-individual variability in vocal signals in many taxa (Knight et al., 2024).

Bat echolocation calls are well-known for their remarkable plasticity, varying across behavioral tasks, environments, and time (Li et al., 2026; Ma et al., 2025; Moss & Surlykke, 2010; Schnitzler et al., 2003), thus representing a highly challenging case for robust AIID. The high flexibility of their calls has led to enduring scientific controversy over whether bat echolocation calls truly encode individual signatures (Jones & Siemers, 2011; Russo et al., 2018). Additionally, the physical and ecological demands of echolocation may impose evolutionary constraints on vocal inter-individual variability. Because echolocation calls must maintain acoustic features optimized for target detection and spatial navigation, the extent of inter-individual variability may be more limited than in many non-echolocating species, where variability in vocal signatures is important for individual recognition. Determining whether stable individual signatures can still be extracted under such constraints represents a key challenge for acoustic individual identification in echolocating mammals. Notably, stability does not imply invariance of acoustic signals, but rather the persistence of individual-specific information despite substantial within-individual variability across time and context. Addressing this question would help to reconcile the debate regarding whether individual signatures in echolocation calls exist, and if so, facilitate the application of AIID as a powerful tool for studying behavior and ecology in a wide range of vocally active species.

Here, we first conducted a systematic literature search of peer-reviewed studies applying deep learning to the acoustic identification of species and individuals, establishing baselines to evaluate the performance of our models. The objective of this study is to evaluate the effectiveness of deep learning for AIID under controlled laboratory conditions. We demonstrate that, comparable to species-level classification, deep learning is a powerful method for AIID under controlled laboratory conditions, achieving up to 91% accuracy even when applied to the challenging domain of bat echolocation calls.

MATERIALS AND METHODS

Literature survey

We first investigated whether deep learning has been successfully applied to acoustic identification of species and individuals across taxa, and if so, with what accuracy. We conducted a systematic literature search of peer-reviewed studies that applied deep learning to species and individual identification based on acoustic signals. The literature search was performed using the Web of Science "all databases" on February 28, 2025, with the following search terms: TS="individual identification" AND TS="deep learn" or TS="species identification" AND TS="deep learn" or TS="call classification" or TS="call identification". After screening titles, abstracts, and keywords to exclude irrelevant publications, with particular attention to removing studies based on image-based identification. Ultimately, we retained 98 studies focusing on acoustic identification across multiple taxa, including birds, bats, amphibians, marine mammals, and invertebrates (Supplementary Table S1). For each relevant study, we extracted reported classification accuracy. When data were presented in figures, we used the function "grabit.m" in MATLAB (v.R2021b, MathWorks, USA) to extract the values (Li et al., 2025). Furthermore, to compare the performance of AIID between deep learning and non-deep learning methods, we also searched the Web of Science "all databases" search and performed a search with the following terms: TS="individual identification" AND TS="bats" to find studies on bat acoustic individual identification based on traditional methods. We excluded results with small sample sizes (fewer than three individual bats), as these are unlikely to provide reliable or generalizable performance estimates, which yielded 14 studies (Supplementary Supplementary Table S2). Here, sample size (n) is defined as the number of individual bats. We extracted the classification accuracy (A) and sample size (N), and additionally calculated the chance level or baseline chance accuracy (C) and the normalized accuracy (Cohen's Kappa, κ) following equations:

$$C = 1/N \tag{1}$$

$$\kappa = (A - C)/(1 - C) \tag{2}$$

Echolocation call recording

A total of 34 adult greater leaf-nosed bats (*Hipposideros armiger*) were used for acoustic individual identification. The *H. armiger* emits a CF component followed by a short FM component and the pulse energy is predominantly distributed in the second harmonic. These bats were captured from two geographical locations separated by approximately 900 km. Fifteen individuals were caught from a cave located in Hanzhong County, Shaanxi Province, China and transferred to the housing facility at Northeast Normal University (NENU). Nineteen individuals were caught from a cave located in Xianning County, Hubei Province, China and transferred to the

housing facility at Central China Normal University (CCNU). Bats were kept in temperature and humidity regulated rooms with 12 h light-dark cycles. Bats had *ad libitum* access to water and food. Capture, housing and experimental protocols were approved by the Institutional Animal Care and Use Committee of the respective university (Permits: NENU-2018-03 and CCNU-IACUC201902191427).

We recorded vocalizations of individual bats hanging from a perch in echo-attenuated rooms, with the walls and ceiling covered by thick acoustic foam and the floor by a nylon blanket (Ding et al., 2023; Lu et al., 2020; Wang et al., 2022). Under these controlled conditions, bats predominantly emit echolocation calls. At CCNU, recordings were conducted in a flight room (5 m long×4 m wide×2.2 m high) (Supplementary Figure S1A). Individual bats were recorded from a tripod approximately 1.5 m above the floor. A custom-made ultrasonic microphone (based on SPU0410LR5H, Knowles Corporation, Itasca, IL, USA), positioned approximately 50 cm from the bat, was directed toward the bat. The microphone signals were amplified via microphone preamplifiers (OctaMic II, RME, Haimhausen, Germany) and sampled with an audio interface (Fireface 802, RME, Germany) at a rate of 192 kHz. The bats were recorded multiple times per week over three months. At NENU, the recordings were provided by our collaborators and were originally collected for a different experimental purpose. Echolocation calls were recorded at a sampling rate of 250 kHz with an array of 16 ultrasonic microphones (NEUmic, Ultra Sound Advice, UK) using a multichannel data acquisition board (PXIe 6358, National Instruments, USA) via custom LabVIEW programs (Supplementary Figure S1B). The recordings were conducted in a flight room measuring 7 m long×4.8 m wide×2.6 m high and the bat hung from a perch attached to the ceiling and positioned 30 cm below it. The microphones were placed on the floor and pointed towards the bat. The bats were recorded for five days in total. Although recording distances and setups differed between the two sites, these variations did not affect the extracted call and sequence features, with the only noticeable variation being in SNR. The recording time lasts for 5 s. A comprehensive summary of the dataset is shown in Supplementary Table S3. To combine the two recording datasets, recordings sampled at 250 kHz were down sampled to 192 kHz before further processing, a step that will not affect call and sequence features.

Call sequence segmentation

This study utilized two distinct datasets derived from labeled audio recordings. The first was composed of standard 5 s call sequences, referred to as sequence-level identification. The second dataset consisted of single calls, which were uniformly zero-padded to 1 s to maintain dimensional consistency, referred to as call-level identification. The single calls were obtained by processing call sequences through an Endpoint Detection Algorithm. Briefly, the call sequences first underwent several preprocessing steps, including the removal of the DC component and amplitude normalization, which collectively serve to enhance the stability and robustness of subsequent signal analysis and feature extraction. A Hamming window (length=200 points) was employed for audio framing, with adjacent frames overlapping by 120 points (60% overlap ratio) to maintain signal continuity. Following the framing process, energy-entropy ratio features were extracted, with high and low thresholds set to 0.1 times and 0.05 times the dynamic range, respectively. To mitigate transient noise, a 10-point median filtering technique was employed. Furthermore, dynamic thresholds were established by analyzing a leading 0.25 s silent segment, combined with constraints of a minimum valid segment (5 frames) and maximum silent interval (8 frames), to precisely extract individual vocalization clips from each sequence.

Discriminant function analysis

Discriminant Function Analysis (DFA), a statistical classification method designed to identify optimal linear combinations for distinguishing between two or more predefined groups, was implemented for individual identification. We selected DFA due to its historical prevalence in bat acoustic studies, allowing for a direct reference to prior literature. Two distinct models for bat individual identification were constructed, utilizing call sequence datasets and single call datasets, respectively. For call-level identification, the frequency spectrum surrounding the dominant second harmonic of the echolocation calls (60–80 kHz) was first transformed to the Mel scale via a Discrete Fourier Transform (DFT), which provides a nonlinear resampling of the frequency axis and a compact representation of spectral information for machine-learning analysis. Mel-frequency representations were chosen due to their widespread application in the machine learning community. Although the Mel scale is essentially a logarithmic transformation that de-emphasizes higher frequencies, restricting the frequency range to 60–80 kHz effectively preserves the spectral detail relevant to these echolocation calls. Then, feature extraction employed a Discrete Cosine Transform (DCT) to derive Mel-Frequency Cepstral Coefficients (MFCCs), Zero Crossing Rate (ZCR), and Spectral Centroid (SC). Mean and standard deviation values of these features were calculated to generate 44-dimensional feature vectors (comprising 40 dimensions from 20 MFCCs, 2 dimensions from ZCR, and 2 dimensions from SC), ensuring stability for subsequent model training and classification. For individual identification, Linear Discriminant Analysis (LDA) was employed using Python's librosa library with the following parameters: no pre-emphasis, n_mfcc=20, frame size=2 048 samples, hop length=512 samples, and Hann window spectral smoothing. This pipeline maintained dimensional consistency across training and test sets while capturing stable acoustic characteristics for classification. During analysis, the total dataset was partitioned with 10% allocated as the test dataset, while the remaining 90% was divided into training and validation datasets through 5-fold cross-validation. The same methodology was applied to the sequence-level identification.

Deep learning

The experiment used the same datasets as the DFA model for comparative analysis, with experimental conditions grouped into two main experiments: call- and sequence-level identification. We investigated whether deep learning-based individual identification models rely on spectral features and temporal cues. At the call-level, the following conditions were included: single call (original single call datasets), multiple consecutive calls (sequences of 2 to 23 calls were extracted in their natural, original order from the call sequences dataset), and multiple randomly selected calls (2 to 23 calls randomly sampled from different positions within the call sequences dataset, with no regard to their original order). For the

sequence-level, the conditions included: 5 s call sequences (original single call datasets), 5 s time-reversal sequences (reversing the order of calls within the sequences), 5 s position-random sequences (calls within the sequences randomly reordered, disrupting natural temporal structure), and call-swapping sequences (hybrid sequences constructed by replacing the spectral content of one individual's calls with that of another, while preserving the original temporal structure). By comparing classification performance across these conditions, we tested three hypotheses. First, if temporal cues contribute encoding individual identity, disrupting the natural order of calls should reduce classification accuracy. Second, if spectral features are the primary carriers of identity information, preserving an individual's spectral content while altering its temporal structure should still yield high accuracy. Conversely, if temporal patterns dominate, sequences with preserved temporal structure but foreign spectral content should show better performance. Third, the call-swapping experiment was designed to directly evaluate the relative contribution of spectral and temporal features to individual identification. For all conditions except single calls and 5 s call sequences, we employed endpoint detection using a ratio method to identify the calls in each audio file, obtaining the starting and ending frames for each call. We performed echo filtering and saved each call individually, when the interval between the previous call and the current call was less than 0.01 s and the maximum amplitude of the previous call was more than three times that of the current call.

For individual identification, each condition was first partitioned with 10% allocated as the test datasets, while the remaining 90% was divided into training and validation dataset through 5-fold cross-validation. Each audio segment underwent Short-Time Fourier Transform (STFT) with a 4 096-point Fourier window and 2 048-point frame shift to compute time-frequency spectrograms. A bandpass filter (55 to 85 kHz) was applied to retain frequency bands containing the dominant second harmonic component. Concurrently, the spectrograms were converted to log-Mel spectrograms using 128 Mel filter banks for the 60–80 kHz frequency range to achieve dimensionality reduction (Zhang et al., 2023). The ResNest50d network was selected as the backbone for feature extraction and classification (Zhang et al., 2022a). This choice was motivated by three main considerations (Azhari et al., 2025; Maclean & Triguero, 2023). First, CNN architectures with integrated attention mechanisms have been shown to substantially improve classification performance in bioacoustic contexts, where features are subtle and background noise is complex. However, it is worth noting that the split-attention mechanism in ResNeSt operates on grouped features along the channel dimension rather than on spatial positions within the input spectrogram. As a result, the attention weights do not directly indicate which time–frequency regions are most influential, limiting the direct interpretability of the model's decisions in bioacoustic terms. Second, empirical evidence indicates that ResNeSt50 outperforms other mainstream CNN architectures when processing complex animal vocalization spectrograms, demonstrating its strong ability to capture intricate time–frequency patterns. Third, the ResNeSt50d variant incorporates architectural enhancements from the ResNet-D design, enabling the model to reduce feature map dimensionality while preserving the structural integrity of acoustic signals. This helps minimize the loss of key discriminative information, which is crucial for individual-level identification in bat echolocation calls. Nevertheless, we also selected EfficientNet-B0 as the backbone network for feature extraction and classification to provide a more comprehensive evaluation. During the training process, the model parameters were iteratively updated using a Stochastic Gradient Descent (SGD) optimizer with momentum, with an initial learning rate set to 0.01, and a minimum learning rate of 0.0001. A cosine annealing learning rate scheduling strategy was implemented, where the annealing cycle T_{max} was set to 100, matching the total training epochs. To explore the acoustic basis of individual signatures, we separated each echolocation call into its constant frequency (CF) and frequency-modulated (FM) components and performed individual recognition on each component independently. The CF and FM components were separated by estimating the end of the CF component and the start of the FM component, using custom-written scripts in MATLAB (Liu et al., 2026; Wang et al., 2026). We trained our models on an NVIDIA RTX 4090 GPU over 100 epochs with a batch size of 32 and binary cross-entropy based on focal loss.

Statistical analysis

All statistical analyses were conducted using R Studio (v.2022.10.0+353, R v.4.2.2) and MATLAB (R2021b, MathWorks, USA). Principal component analysis was used to examine the variability of individual calls using the *FactoMineR* package based on acoustic parameters, including peak frequency, call duration, bandwidth, and pulse interval. We employed the Bhattacharyya coefficient to quantify the degree of overlap between individuals' call distributions in the reduced feature space. A higher coefficient value indicates greater overlap, meaning that individuals are difficult to distinguish based solely on these linearly transformed features. We also employed the Mann-Whitney *U* test (as data normality was rejected according to the Shapiro-Wilk test in the package *stats*) to compare the differences in average accuracy between groups.

RESULTS

Accuracy for acoustic identification of animal species by deep learning

We conducted a comprehensive review of 98 studies that evaluated acoustic identification of species based on deep learning. These studies encompassed a broad range of species, from well-studied taxa such as birds (63.28%) and bats (25.78%) to less frequently studied amphibians, marine mammals, and invertebrates (Figure 1A; Supplementary Table S1). Regarding the classification performance, the overall average identification accuracy for all taxa was 0.89±0.08 (mean±SD), with the highest average accuracy for marine mammals at 0.94±0.04 (mean±SD). For bats, the average species-level identification accuracy was also high at 0.90±0.07 (mean±SD), and the lowest for invertebrates at 0.86, which was derived from only a single study (Figure 1A). Additionally, we systematically reviewed 14 studies focusing on bat acoustic individual identification based on traditional methods. We found that the identification accuracy for individual bats based on traditional methods was much lower, at 0.48±0.22 (mean±SD), with a normalized accuracy of 0.46±0.23 (mean±SD; Figure 1B; Supplementary Table S2).

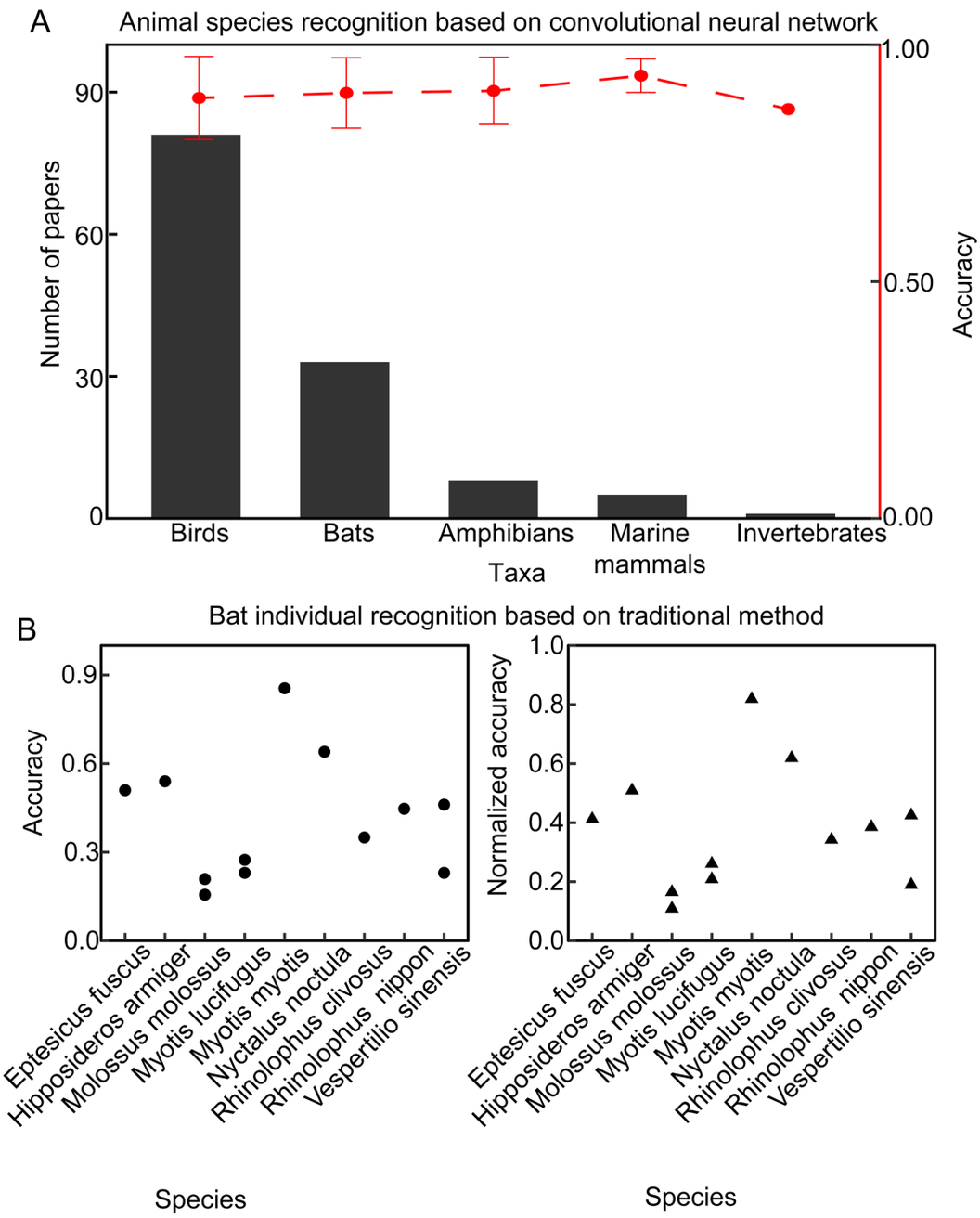


Figure 1 Acoustic species identification in representative taxa and acoustic individual identification in echolocating bats

A: Acoustic species identification across taxa using convolutional neural networks revealed an average accuracy of 0.89±0.08 (red; mean±SD). B: All previous researches on acoustic individual identification in bats have used traditional machine learning methods, with an average accuracy of 0.48±0.22 (mean±SD).

Substantial intra-individual variability in bat echolocation calls across time

Apart from the individuals from NENU, we conducted three-month acoustic recordings of *H. armiger* (Figure 2A), yielding a total of 17 584 standardized 5 s vocalization sequences. Figure 2B displays an example of the waveform of a 5 s call, as well as the waveform and spectrogram of a single call. We found that call parameters such as peak frequency, duration, bandwidth, and pulse interval varied considerably across different recording days, indicating that even for the same individual, there is a high level of variability in multiple call features (Figure 2C). Principal component analysis further revealed that most calls clustered together, with an average Bhattacharyya coefficient, which is a measure of the amount of overlap between two statistical samples or populations, of 0.80±0.14 (mean±SD) for 20 individuals recorded on the same day, suggesting limited acoustic distinctiveness among individuals (Figure 2D). Similar results were obtained for all individuals.

AIID by the traditional DFA

We first employed traditional methods to investigate AIID of bats. The results of the DFA showed that the average classification accuracy based on single calls for the 34 individuals was 0.39±0.21 (mean±SD). Furthermore, no significant difference in classification accuracy was observed between the CCNU and NENU sites (Mann-Whitney *U* test: P=0.19; Figure 3A). For 5 s call sequence recognition, the

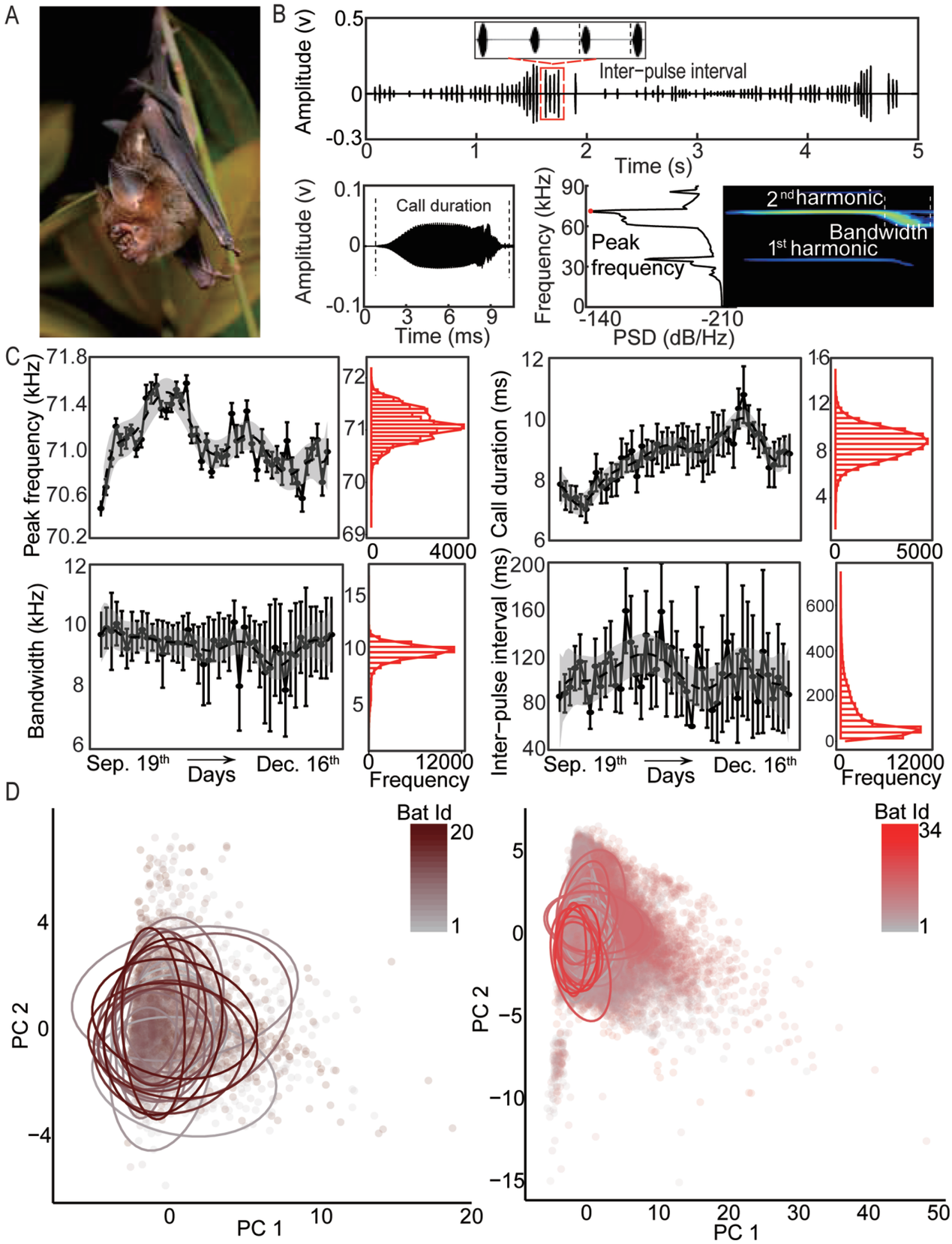


Figure 2 Intra-individual vocal variability in *Hipposideros armiger* under controlled laboratory conditions

A: A portrait of *H. armiger*. B: Examples of the echolocation call. The top panel displays the waveform of a 5 s echolocation sequence, while the bottom panel shows the waveform of a single call alongside the corresponding spectrogram and power spectrum. C: Illustrations of daily variation in four call parameters (peak frequency, duration, bandwidth, and pulse interval) measured from one individual, representative of the general pattern observed across all individuals. D: Principal Component Analysis (PCA) of 20 individuals recorded on the same day (left panel) and 34 individuals recorded across multiple days (right panel), with ellipses representing 95% confidence intervals. The degree of overlap between individuals' calls was quantified by the Bhattacharyya coefficient, where a higher value indicates greater overlap.

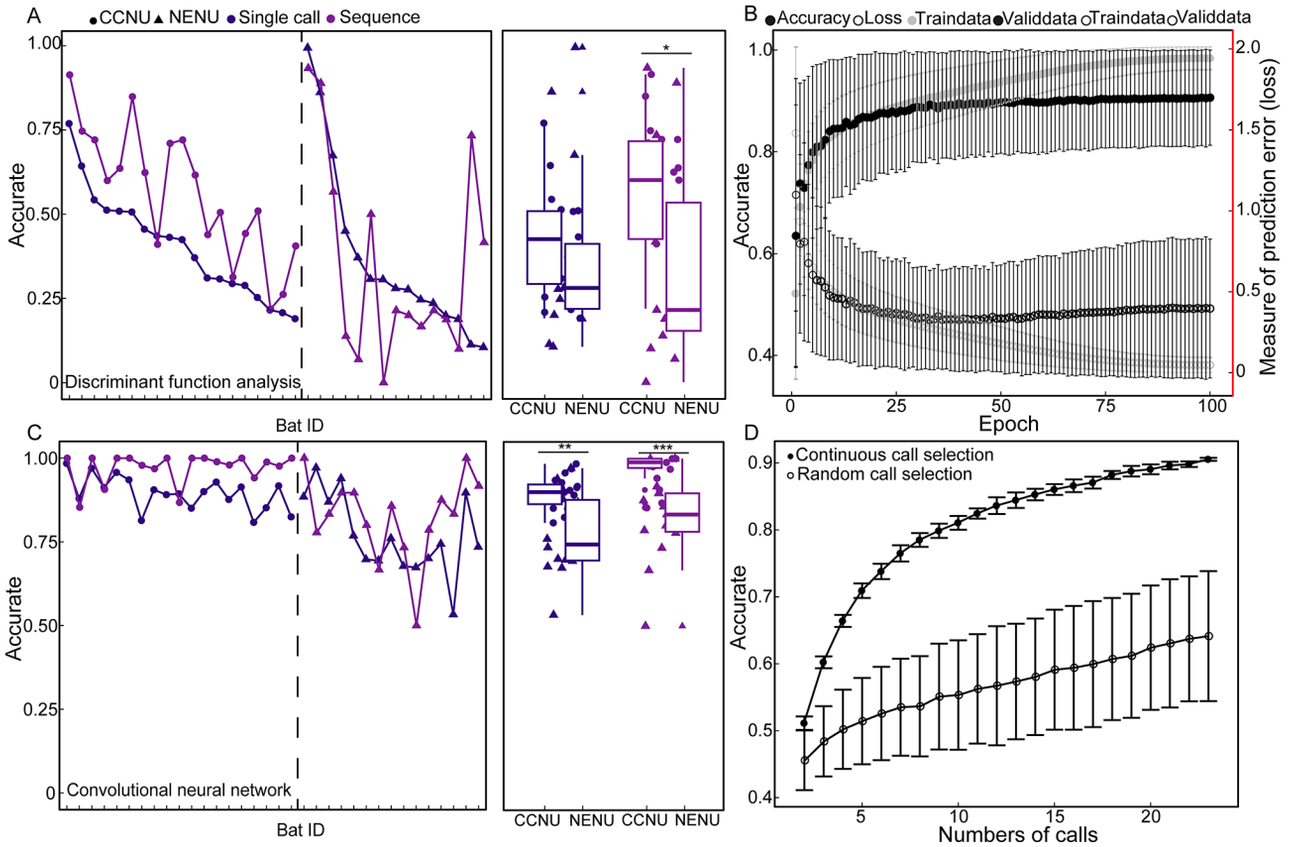


Figure 3 Comparison of acoustic individual identification performance in bats using traditional methods versus deep learning

A: A comparison of classification accuracy between CCNU (circle) and NENU (triangle) based on DFA for single-call (blue) and 5 s call sequence (purple) paradigms; B: Changes in accuracy (black axis) and loss (red axis) for both the training dataset and validation dataset as epochs progress (mean±SD); C: A comparison of classification accuracy between CCNU and NENU based on deep learning for single-call and 5 s call sequence paradigms; D: The variation in accuracy as the number of selected calls changes, with solid circles indicating consecutively selected calls and hollow circles representing randomly selected calls (mean±SD).

average classification accuracy improved to 0.47±0.26 (mean±SD), with classification accuracy at the CCNU site significantly higher than that at NENU (Mann-Whitney *U* test: *P*=0.02).

AIID by deep learning

In contrast, deep learning achieved markedly higher performance. The model showed stable convergence throughout training, with classification accuracy consistently increasing and cross-entropy loss decreasing as epochs progressed (Figure 3B). Although classification accuracy varied among individuals based on single calls, the majority exhibited relatively high classification accuracy (>0.50), achieving an average classification accuracy of 0.84±0.11 (mean±SD) and the normalized accuracy of 0.84 (Figure 3C; Figure S2A). Notably, accuracy at CCNU was significantly higher than at NENU (0.90 vs. 0.77: Mann-Whitney *U* test: *P*=0.002). Furthermore, we observed a progressive increase in classification accuracy with the number of calls incorporated into the analysis. Multiple consecutive calls consistently outperformed those constructed from multiple randomly selected calls (Figure 3D), suggesting that temporal sequence information may help to enhance classification performance.

Compared to single-call analysis, the classification accuracy improved for most individuals under the 5 s call sequence paradigm, culminating in an overall mean accuracy of 0.91±0.12 (mean±SD) and the normalized accuracy of 0.91 (Figure 3C). The classification accuracy at CCNU was also significantly higher than at NENU (Mann-Whitney *U* test: *P*<0.001), reaching 0.97 compared to 0.83. When using EfficientNet-B0 as the backbone network, the mean accuracy showed a slight decrease but remained at 0.89±0.10 (mean±SD), and the classification accuracy at CCNU remained higher than that at NENU (0.94 vs. 0.84: Mann-Whitney *U* test: *P*=0.02; Figure S2B).

Figure 4A displays the schematic representation of 5 s time-reversal sequences and 5 s position-random sequences. We observed that both time-reversed and position-random calls led to statistically significant reductions in accuracy compared to natural sequences (Mann-Whitney *U* test, all *P*<0.01; Figure 4B), while no significant difference was observed between these two modifications (Mann-Whitney *U* test; *P*=0.14), suggesting that temporal cues play a key role in encoding individual identity and that maintaining natural call order is critical for accurate recognition. In addition, Figure 4C displays the schematic representation of call-swapping sequences. We found that when synthesizing novel call profiles through recombination of spectral features and temporal sequence information, individuals identified based on spectral features (feature dominated) exhibited significantly higher accuracy than those identified from temporal sequence information (position dominated) (Figure 4D). This indicates that the importance of spectral features in recognition is greater than that of temporal sequence information. However, when each call was decomposed into its CF and FM

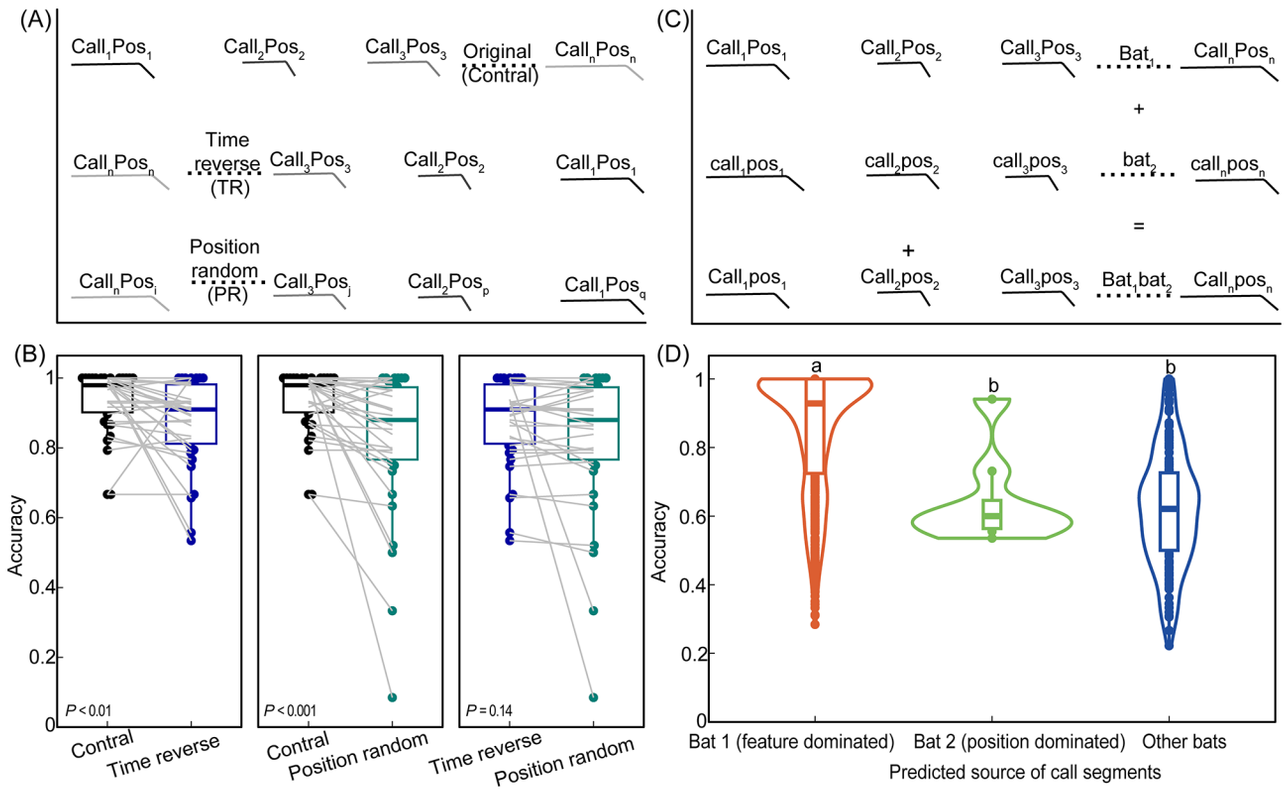


Figure 4 Exploring the importance of spectral features and temporal sequence information in individual identification

A: Schematic representation of 5 s time-reversed (reversing the order of calls within the sequences) and randomly repositioned call sequences (calls within the sequences randomly reordered, disrupting natural temporal structure), illustrating the experimental design; B: Comparison of accuracy among different groups; C: Schematic representation of call-swapping sequences (hybrid sequences constructed by replacing the spectral content of one individual's calls with that of another, while preserving the original temporal structure), demonstrating the approach used for synthesizing novel call profiles through the recombination of spectral features and temporal sequence information; D: Comparison of accuracy among different groups. The numbers represent the corresponding identification counts.

components and evaluated independently, the classification accuracy deteriorated markedly (CF-only: 0.34±0.28; FM-only: 0.30±0.39; Figure S3).

DISCUSSION

In this study, we successfully applied deep learning to AIID under controlled laboratory conditions and achieved a recognition accuracy of up to 0.91, which is comparable to the benchmark for species identification (0.89 on average), despite the inherent challenge of discerning minute individual vocal differences. This achievement is particularly striking given that individual identification demands resolving acoustic variations orders of magnitude smaller than those between-species. In contrast, traditional methods, such as discriminant function analysis (DFA), achieved accuracy approximately two times lower than that of deep learning, underscoring deep learning's capacity to decode nonlinear spectrotemporal interactions impervious to manual feature engineering. Our dataset, comprising 34 bats monitored under controlled laboratory conditions over three months and 17 584 standardized 5 s vocalization sequences, demonstrates the natural range of vocal plasticity in bat echolocation calls. This variability, driven by factors such as environmental noise (Tressler et al., 2011), behavioral states (Schoeppler et al., 2022), and ontogenetic changes (Luo et al., 2017), and social factors (Hiryu et al., 2006), has historically posed challenges for conventional methods (Guo et al., 2015; Jones & Siemers, 2011).

The primary objective of this study was to establish a proof-of-concept under controlled conditions, specifically to evaluate whether deep learning is suitable for AIID. Our results suggest that deep neural networks effectively leverage hierarchical feature abstraction to isolate stable "vocal fingerprints", that is individual signatures, even in groups with excessive intra-individual variability in acoustic signals, achieving robustness where traditional methods tend to struggle (Bai & Zhang, 2021). Although individual call features such as frequency or duration vary substantially across time, the relative configuration and joint distribution of these features appear to form a robust identity signature. Deep learning models are particularly well suited to capture such distributed representations, enabling reliable identification even when individual acoustic components are highly variable. In this context, our use of "reveals" refers to the model's ability to detect and utilize identity-relevant information, rather than to explicitly identify the precise acoustic features involved. However, the noticeably higher classification accuracy at CCNU compared to NENU suggests that site-specific factors may have influenced model performance, potentially arising from acoustic artifacts introduced by differences in recording environments and equipment, coupled with a lower signal-to-noise ratio (SNR) in the NENU recordings.

Our research highlights the crucial role of spectral features in individual identification, providing compelling evidence for the encoding of individual information in the echolocation calls of bats. Spectral features, such as frequency and bandwidth of

echolocation calls, serve as distinctive biological markers that allow researchers to differentiate between individuals within a species (Barclay & Jacobs, 2023; Jiang et al., 2010; Yovel et al., 2009). Studies have shown that variations in the frequency components of echolocation calls can reflect individual identity, with certain bats exhibiting unique call characteristics that can be consistently recognized across different contexts (Amichai et al., 2015; Suga et al., 1987). Furthermore, the application of advanced bioacoustic analysis techniques has enabled researchers to automate the identification process by analyzing these spectral features, thereby improving the efficiency and accuracy of individual recognition (Lu et al., 2020; Luo et al., 2015; Ma et al., 2025). The encoding of individual labels through spectral characteristics is not merely an analytical tool, but a critical aspect in the broader context of understanding bat ecology and behavior.

To further investigate the acoustic basis of individual signatures, we conducted a preliminary analysis by separating each echolocation call into its CF and FM components and performing recognition on each independently. Both CF-only and FM-only models yielded markedly lower accuracy (with averages below 35%), far below the 91% achieved with full-spectrum calls. This suggests that individual identity is not encoded in isolated acoustic components but rather distributed across the full spectrotemporal structure. This may also explain the poor performance of traditional approaches such as DFA, which rely on manually extracted features. However, we believe this result should not be taken as direct evidence that FM bat echolocation calls cannot encode individual signatures, for several reasons. First, we did not specifically fine-tune the model to optimize performance; therefore, these results likely represent an underestimation of the potential accuracy. Second, the FM component in CF-FM bats appears more stereotyped than in purely FM bats. For instance, both the duration and bandwidth of the FM component in *H. armiger* exhibit a much narrower dynamic range compared to those of many FM bats, such as the big brown bat (Lu et al., 2020; Surlykke & Moss, 2000). Third, there is solid evidence that the echolocation calls of FM bats contain strong individual signatures (Barclay & Jacobs, 2023; Yovel et al., 2009), although the number of individuals involved in previous analyses has typically been limited. Furthermore, it is critical to interpret these findings within the constraints of the methodology. The deep learning model identifies a robust acoustic signature that could facilitate individual identification. This demonstrates the information encoding capacity of the calls but does not constitute evidence that bats perceptually discriminate or cognitively utilize these cues. The fundamental question of whether bats use these specific, model-derived features for social recognition remains unanswered.

In addition to spectral features, when both spectral and temporal features are integrated, the classification performance improves significantly, highlighting the value of temporal sequence information in individual recognition tasks. This temporal pattern provides context and relational dynamics that are essential for distinguishing between individuals, particularly in species where vocalizations are complex and varied (Bøttcher et al., 2018; Carlson et al., 2020). Thus, the encoding of individual labels through time-based characteristics is not just an auxiliary feature; it is a fundamental element that shapes how organisms interact and communicate within their species (Amit & Yovel, 2023). By incorporating temporal patterns into identification frameworks, such as recurrent neural networks (RNNs) (Salehinejad et al., 2017), researchers can better understand social structures, mating systems, and communication strategies in various species, lending insight into behavioral ecology and evolution. However, it is worth noting that in this study, all recordings were obtained from perching individuals. In many bat species, the timing of echolocation pulses during flight is partially coupled with the wingbeat cycle or foraging phases, which can influence call emission patterns and temporal structure (Xia et al., 2025). Consequently, the importance of pulse interval as a discriminative feature identified in our analysis may differ under flight conditions.

This study establishes bats as model organisms for advancing AIID development, whereby surmounting the limitations of traditional methods through deep learning, our framework enables non-invasive individual studies at under controlled laboratory conditions. These results show great promise for extrapolating to other acoustically active taxa. Echolocation signals are primarily under stabilizing selection for reliable echo detection, which should theoretically minimize individual variation. In contrast, many other animal vocalizations are under diversifying selection to encode information such as individual identity, group membership, or reproductive status. With appropriate adaptation to species-specific acoustic characteristics, such methods may be extended to other vocally complex systems (e.g., cetaceans and songbirds), enabling novel insights into adaptive signal evolution, dynamic social network topologies, host-pathogen coevolution, and anthropogenic impact assessments (Sheldon et al., 2022).

Nevertheless, application of AIID to freely-ranging wild animals still faces considerable challenges. First, our experiments were conducted on recordings of single bats in a resting state under controlled laboratory conditions. In real ecological environments, bats often fly and move in groups. In such cases, the complexity of acoustic data will significantly increase, potentially leading to a decrease in identification accuracy. Moreover, echolocation calls produced across different behavioral contexts, such as flight, foraging, or prey capture, may exhibit context-dependent acoustic variation driven by task-specific sensorimotor demands, which could further influence the performance of CNN-based individual identification. In addition, environmental complexities such as reflections and reverberation in cluttered habitats and calls from conspecific and heterospecific individuals can introduce spectral-temporal distortions, which require advanced algorithms for separating overlapping calls (Zhang et al., 2022b). A further limitation is that the present study employed a closed-set design, in which all individuals included in the test set were also represented in the training set. In contrast, real-world applications, such as long-term acoustic monitoring of wild animals, will inevitably involve open-set scenarios, where calls from previously unobserved individuals must either be identified as unknown or incorporated into the recognition system. The closed-set framework adopted here therefore represents an initial step toward validating the feasibility of deep learning-based individual identification under controlled conditions. Extending this approach to open-set recognition will be an important direction for future research, potentially involving strategies such as novelty detection or incremental learning to enable robust performance in more realistic

ecological settings. Future research should focus on optimizing algorithms, expanding datasets, and enhancing the robustness of models to evaluate whether individual signatures remain detectable across diverse behavioral contexts and ecological environments, a prerequisite for applying AIID to freely behaving wild populations.

DATA AVAILABILITY

The datasets and Rcode during the current study are available in the Figshare repository (10.6084/m9.figshare.19613934).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare no competing interests.

AUTHORS' CONTRIBUTIONS

A.Q.L. and H.J.Z. performed the literature survey. A.Q.L. analyzed the survey data. W.T.H., X.Y.X, W.M.W., and L.J. conducted the classification tasks. A.Q.L. and J.H.L. drafted the manuscript. C.Y.Z. and J.h.L. supervised the study. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank Man-Man Lu and Xin-Yao Yang for collecting the recordings, Ai-Qing Lin and Jiang Feng for permitting us to use the recordings from NENU, and Jia-Nan Ding and Xue-Jiao Qin for their valuable discussions. We are grateful to Long-Gang Pang for performing the initial individual classification tests, which motivated this study.

REFERENCES

Amichai E, Blumrosen G, Yovel Y. 2015. Calling louder and longer: how bats use biosonar under severe acoustic interference from other bats. *Proceedings of the Royal Society B: Biological Sciences*, **282**(1821): 20152064.

Amit Y, Yovel Y. 2023. Bat vocal sequences enhance contextual information independently of syllable order. *iScience*, **26**(4): 106466.

Aplin LM. 2019. Culture and cultural evolution in birds: a review of the evidence. *Animal Behaviour*, **147**: 179–187.

Azhari MF, Syah GM, Nurdiati S, et al. 2025. Bioacoustic sound classification based on noise and vocalizations using convolutional neural networks: a comprehensive systematic review. *Turkish Journal of Mathematics and Computer Science*, **17**(2): 472–484.

Bai ZX, Zhang XL. 2021. Speaker recognition based on deep learning: an overview. *Neural Networks*, **140**: 65–99.

Barclay RMR, Jacobs DS. 2023. Interindividual communication by bats via echolocation. *Canadian Journal of Zoology*, **101**(3): 128–143.

Borowiec ML, Dikow RB, Frandsen PB, et al. 2022. Deep learning as a tool for ecology and evolution. *Methods in Ecology and Evolution*, **13**(8): 1640–1660.

Bøttcher A, Gero S, Beedholm K, et al. 2018. Variability of the inter-pulse interval in sperm whale clicks with implications for size estimation and individual identification. *The Journal of the Acoustical Society of America*, **144**(1): 365–374.

Camacho C, Canal D, Potti J. 2017. Lifelong effects of trapping experience lead to age-biased sampling: lessons from a wild bird population. *Animal Behaviour*, **130**: 133–139.

Cantor M, Maldonado-Chaparro AA, Beck KB, et al. 2021. The importance of individual-to-society feedbacks in animal ecology and evolution. *Journal of Animal Ecology*, **90**(1): 27–44.

Cao Z, Wang KH, Wen JW, et al. 2024. Fine-grained image classification on bats using VGG16-CBAM: a practical example with 7 horseshoe bats taxa (CHIROPTERA: Rhinolophidae: *Rhinolophus*) from Southern China. *Frontiers in Zoology*, **21**(1): 10.

Carlson NV, Kelly EM, Couzin I. 2020. Individual vocal recognition across taxa: a review of the literature and a look into the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **375**(1802): 20190479.

Clutton-Brock T, Sheldon BC. 2010. Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, **25**(10): 562–573.

Darwin C. 1859. On the Origin of Species. London: John Murray.

Dechmann DKN, Heucke SL, Giuggioli L, et al. 2009. Experimental evidence for group hunting via eavesdropping in echolocating bats. *Proceedings of the Royal Society B: Biological Sciences*, **276**(1668): 2721–2728.

Ding JN, Han FJ, Zhang KK, et al. 2023. Performance of Doppler shift compensation varies with environmental temperature and humidity in bats. *Animal Behaviour*, **205**: 35–46.

Fagerlund S. 2007. Bird species recognition using support vector machines. *EURASIP Journal on Advances in Signal Processing*, **2007**(1): 038637.

Ferreira AC, Silva LR, Renna F, et al. 2020. Deep learning-based methods for individual recognition in small birds. *Methods in Ecology and Evolution*, **11**(9): 1072–1085.

Goldshtein A, Chen X, Amichai E, et al. 2024. Acoustic cognitive map–based navigation in echolocating bats. *Science*, **386**(6721): 561–567.

Greif S, Borissov I, Yovel Y, et al. 2014. A functional role of the sky's polarization pattern for orientation in the greater mouse-eared bat. *Nature Communications*, **5**: 4488.

Guo X, Luo B, Liu Y, et al. 2015. Cannot see you but can hear you: vocal identity recognition in microbats. *Zoological Research*, **36**(5): 257–262.

Hiryu S, Katsura K, Nagato T, et al. 2006. Intra-individual variation in the vocalized frequency of the Taiwanese leaf-nosed bat, *Hipposideros terasensis*, influenced by conspecific colony members. *Journal of Comparative Physiology A*, **192**(8): 807–815.

Hyacinthe C, Attia J, Schutz E, et al. 2023. Acoustic signatures in Mexican cavefish populations inhabiting different caves. *PLoS One*, **18**(8): e0289574.

Jiang TL, Liu R, Metzner W, et al. 2010. Geographical and individual variation in echolocation calls of the intermediate leaf-nosed bat, *Hipposideros larvatus*. *Ethology*, **116**(8): 691–703.

Jones G, Siemers BM. 2011. The communicative potential of bat echolocation pulses. *Journal of Comparative Physiology A*, **197**(5): 447–457.

Knight E, Rhinehart T, de Zwaan DR, et al. 2024. Individual identification in acoustic recordings. *Trends in Ecology & Evolution*, **39**(10): 947–960.

Li AQ, Tian XT, Ma NN, et al. 2025. Habitat preference contributes to explaining the varied sensitivity of bats to anthropogenic noise. *Biological Conservation*, **302**: 110974.

Li AQ, Zhang HJ, Lu MM, et al. 2026. Robust Doppler shift compensation in freely flying bats under acoustic interference. *Integrative Zoology*, **21**(4): 752–762.

Liu ZX, Cao QH, Lu MM, et al. 2026. Echolocation behaviour is highly sensitive to noise interference in flying Pratt's roundleaf bats. *Animal Behaviour*, **233**: 123465.

Lu MM, Zhang GM, Luo JH. 2020. Echolocating bats exhibit differential amplitude compensation for noise interference at a sub-call level. *Journal of Experimental Biology*, **223**(19): jeb225284.

Luo JH, Goerlitz HR, Brumm H, et al. 2015. Linking the sender to the receiver: vocal adjustments by bats to maintain signal detection in noise. *Scientific Reports*, **5**: 18556.

Luo JH, Lingner A, Firzlaff U, et al. 2017. The Lombard effect emerges early in young bats: implications for the development of audio-vocal integration.

Journal of Experimental Biology, **220**(6): 1032–1037.

Ma NN, Xia HJ, Zheng HJ, et al. 2025. Prey evasiveness and masking noise jointly promote the ultrahigh call rate in echolocating bats. *BMC Biology*, **23**(1): 323.

Maclean K, Triguero I. 2023. Identifying bird species by their calls in Soundscapes. *Applied Intelligence*, **53**(19): 21485–21499.

McClintock BT, White GC. 2009. A less field-intensive robust design for estimating demographic parameters with mark–resight data. *Ecology*, **90**(2): 313–320.

Moss CF, Surlykke A. 2010. Probing the natural scene by echolocation in bats. *Frontiers in Behavioral Neuroscience*, **4**: 33.

Oestreich WK, Oliver RY, Chapman MS, et al. 2024. Listening to animal behavior to understand changing ecosystems. *Trends in Ecology & Evolution*, **39**(10): 961–973.

Pichler M, Hartig F. 2023. Machine learning and deep learning—A review for ecologists. *Methods in Ecology and Evolution*, **14**(4): 994–1016.

Putland RL, Ranjard L, Constantine R, et al. 2018. A hidden Markov model approach to indicate Bryde's whale acoustics. *Ecological Indicators*, **84**: 479–487.

Ranjan R, Sankaranarayanan S, Bansal A, et al. 2018. Deep learning for understanding faces: machines may be just as good, or better, than humans. *IEEE Signal Processing Magazine*, **35**(1): 66–83.

Rundo L, Militello C. 2024. Image biomarkers and explainable AI: handcrafted features *versus* deep learned features. *European Radiology Experimental*, **8**(1): 130.

Russo D, Ancillotto L, Jones G. 2018. Bats are still not birds in the digital era: echolocation call variation and why it matters for bat species identification. *Canadian Journal of Zoology*, **96**(2): 63–78.

Salehinejad H, Sankar S, Barfett J, et al. 2017. Recent advances in recurrent neural networks. *arXiv*, https://doi.org/10.48550/arXiv.1801.01078.

Schnitzler HU, Moss CF, Denzinger A. 2003. From spatial orientation to food acquisition in echolocating bats. *Trends in Ecology & Evolution*, **18**(8): 386–394.

Schoeppler D, Denzinger A, Schnitzler HU. 2022. The resting frequency of echolocation signals changes with body temperature in the hipposiderid bat *Hipposideros armiger*. *Journal of Experimental Biology*, **225**(3): jeb243569.

Schofield D, Nagrani A, Zisserman A, et al. 2019. Chimpanzee face recognition from videos in the wild using deep learning. *Science Advances*, **5**(9): eaaw0736.

Sheldon BC, Kruuk LEB, Alberts SC. 2022. The expanding value of long-term studies of individuals in the wild. *Nature Ecology & Evolution*, **6**(12): 1799–1801.

Smith-Vidaurre G, Perez-Marrufo V, Wright TF. 2021. Individual vocal signatures show reduced complexity following invasion. *Animal Behaviour*, **179**: 15–39.

Stowell D. 2022. Computational bioacoustics with deep learning: a review and roadmap. *PeerJ*, **10**: e13152.

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

Suga N, Niwa H, Taniguchi I, et al. 1987. The personalized auditory cortex of the mustached bat: adaptation for echolocation. *Journal of Neurophysiology*, **58**(4): 643–654.

Surlykke A, Moss CF. 2000. Echolocation behavior of big brown bats, *Eptesicus fuscus*, in the field and the laboratory. *The Journal of the Acoustical Society of America*, **108**(5): 2419–2429.

Tan X, Lin AQ, Sun KP, et al. 2022. Greater horseshoe bats recognize the sex and individual identity of conspecifics from their echolocation calls. *Animals*, **12**(24): 3490.

Teixeira D, Maron M, van Rensburg BJ. 2019. Bioacoustic monitoring of animal vocal behavior for conservation. *Conservation Science and Practice*, **1**(8): e72.

Terry AM, Peake TM, McGregor PK. 2005. The role of vocal individuality in conservation. *Frontiers in Zoology*, **2**(1): 10.

Tressler J, Schwartz C, Wellman P, et al. 2011. Regulation of bat echolocation pulse acoustics by striatal dopamine. *Journal of Experimental Biology*, **214**(19): 3238–3247.

Wang HM, Zhou YX, Li HH, et al. 2022. Sensory error drives fine motor adjustment. *Proceedings of the National Academy of Sciences of the United States of America*, **119**(27): e2201275119.

Wang QY, Ye H, Luo JH. 2026. Task learning drives adaptive vocal adjustments in echolocating bats. *Zoological Research*, **47**(4): 1124–1133.

Xia HJ, Ma NN, Li AQ, et al. 2025. Call production and wingbeat coupling is flexible and species-specific in echolocating bats. *Annals of the New York Academy of Sciences*, **1547**(1): 105–115.

Xie J, Colonna JG, Zhang JL. 2021. Bioacoustic signal denoising: a review. *Artificial Intelligence Review*, **54**(5): 3575–3597.

Yovel Y, Melcon ML, Franz MO, et al. 2009. The voice of bats: how greater mouse-eared bats recognize individuals based on their echolocation calls. *PLoS Computational Biology*, **5**(6): e1000400.

Zhang CY, Li QR, Zhan HS, et al. 2023. One-step progressive representation transfer learning for bird sound classification. *Applied Acoustics*, **212**: 109614.

Zhang H, Wu CR, Zhang ZY, et al. 2022a. ResNeSt: split-attention networks. *In*: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans: IEEE, 2735–2745.

Zhang KK, Liu T, Song SJ, et al. 2022b. Separating overlapping bat calls with a bi-directional long short-term memory network. *Integrative Zoology*, **17**(5): 741–751.