Advanced Intelligent Systems

ADVANCED INTELLIGENT SYSTEMS Open Access

www.advintellsyst.com

RESEARCH ARTICLE OPEN ACCESS

Deep Learning Methods for Assessing Time-Variant Nonlinear Signatures in Clutter Echoes

Ibrahim Eshera[1] | Gaurav Duggal[1] | Rolf Müller[2]

[1]Department of Electrical & Computer Engineering, Virginia Tech, Blacksburg, Virginia, USA | [2]Department of Mechanical Engineering, Virginia Tech, Blacksburg, Virginia, USA

Correspondence: Rolf Müller (rolf.mueller@vt.edu)

Received: 24 October 2025 | **Revised:** 22 December 2025 | **Accepted:** 8 January 2026

Keywords: biosonar | classification | clutter | cyber-physical systems | deep learning | doppler | robotics

ABSTRACT

The biosonar systems of bats in families of horseshoe and Old-World leaf-nosed bats include peripheral dynamics where the outer ears undergo fast rotations and deformations during echo reception. These motions impart time-variant linear and nonlinear effects on the received echoes. In the present study, we have investigated whether such time-variant effects create discriminable and reliable signatures in clutter, i.e., echoes that are created by a superposition of reflections from multiple, unresolved scatterers. We have used a laboratory setup with artificial foliage that was agitated by fans to create large data sets of clutter echoes. These echoes were triggered by pulses with different time–frequency signatures (constant-frequency, frequency-modulated, and a compound of the two) and received by flexible biomimetic pinna that was actuated via strings to create two different motion shapes at five different motion speeds. Different deep-learning architectures (ResNets, transformers, and a 2D convolutional neural network) were tested for their ability to classify the different motions based on single clutter echoes. The achieved performances (up to 97% overall correct classifications) demonstrated that the time-variant signatures in the clutter echoes could form a reliable substrate for the encoding of sensory information that may provide functional advantages for navigating complex natural environments.

1 | Introduction

Developing engineered systems—such as robots anddrones—that can reliably navigate and interact with complex natural environments remains an unresolved challenge [1, 2]. Much of the gravity of this challenge arises from the difficulty of obtaining sensory information that is capable of supporting such operations [3, 4]. This raises the question of whether, and how, current paradigms for environmental sensing could be extended—or perhaps fundamentally reimagined-to meet these demands.

A promising biological model for addressing this challenge is provided by the sophisticated biosonar systems of bats, particularly those species that inhabit densely vegetated environments where they routinely navigate through confined spaces among foliage [5–7] relying on echolocation as their primary sensing modality [8, 9]. Echoes originating from dense clouds of scatterers, such as leaves in foliage, are known as clutter [10]. The defining feature of acoustic clutter is that the echo waveforms arise from superpositions of returns from numerous unresolved scatterers whose individual positions, orientations, and shapes remain unknown [11]. Consequently, clutter echoes must be regarded as inherently random and unpredictable making them difficult to interpret [11]. Nevertheless, recent work has demonstrated that clutter echoes can contain task-relevant information that biosonar systems might exploit: It has been shown that clutter echoes enable passageway detection in foliage without requiring resolution of individual scatterers [12, 13]. Similarly, location identification has been demonstrated in forest environments—both at large scales (tens of km [14]) as well as at fine spatial resolutions (sub-10 m accuracy [15]).

The present work is motivated by findings that certain families of bats (Rhinolophidae [16]) and Old World round-leaf bats (Hipposideridae [17]), noted for their exceptional ability to navigate dense vegetation [18–20], also exhibit distinctive peripheral dynamics in their biosonar systems: The animals possess baffle-like emitter and receiver structures, namely, the noseleaf [21] and the external ears (pinnae [22]) that can both undergo rapid deformations during pulse emission and reception respectively. These deformations are precisely timed to coincide with pulse emission or echo reception respectively [22–24] and fast enough to result in substantial shape changes within the duration of pulse/echo (approximately 10 to 100 ms [22]). Furthermore, the deformations are actuated by specialized musculatures [25] suggesting the hypothesis that these deformations are a substrate for active sensing that has been shaped by evolutionary pressure to improve sensory information encoding in the densely vegetated habitats that are the hallmark of the sensory ecology of these species.

To investigate this hypothesis, previous work by the authors [26] has evaluated the impact of static differences in pinna shape on the reception of clutter echoes. Changes to the pinna shape result in different acoustic sensitivity patterns over direction and frequency (i.e., beampatterns [22, 27]). Since beampatterns are linear system characterizations [28], their impacts on the clutter echoes can be investigated using a collection of different static shapes. The primary objective of these prior efforts [26] has been to determine whether different pinna shapes exert consistent, i.e., discriminable and reliable, effects on clutter echoes, despite the inherently random nature of these signals. The rationale of this approach has been that any functional relevance of different pinna shapes must rest on their ability to produce acoustic effects that are both of substantial magnitude as well as reliable. In the previous work, this was confirmed by the ability of a deep-neural-network classifier to determine which of ten tested pinna shapes had been used to record a given clutter echo.

However, in a pinna that is continuously deforming while receiving the incoming echoes, the linear system characteristics are no longer stationary, but time-variant. In addition, the acoustic effects of pinna motions in rhinolophid and hipposiderid bats go beyond altering the linear characteristics of the pinna: Since the pinna motions in these species are fast, they impart substantial Doppler shifts (up to ± 2kHz) on the received echoes [23]. As nonlinear phenomena, Doppler shifts may offer a basis for sensory information encoding that extends beyond the linear framework provided by beampattern characterizations. To wit, it has been demonstrated that Doppler-shift signatures created by a biomimetic deforming pinna could be used to estimate target direction with sub-degree accuracy [29].

Furthermore, ample evidence highlights the importance of Doppler shifts in the biosonar systems of rhinolophid and hipposiderid bats. The time–frequency structure of the biosonar pulses emitted by these species includes narrow-band components that are particularly well suited for detecting and characterizing Doppler shifts [30, 31]. To achieve the necessary frequency resolution, these bats possess an acoustic fovea—a frequency region with neural overrepresentation due to a large number of narrowly tuned frequency channels [31, 32]. This specialization is complemented by an auditory-vocal control loop that compensates for Doppler shifts caused by the bat's own motion, thereby maintaining the narrow-band components of returning echoes within the acoustic fovea [30].

So far, sensory information conveyed by Doppler shifts has only been investigated in the context of small, deterministic targets, e.g., for direction-finding [29], acoustic flow [33], or characterization of insect prey [30]. The potential role of Doppler shifts in extracting sensory information from clutter echoes has received little attention as of now.

In the present work, the suitability of Doppler-shift patterns created by pinna motions for encoding sensory information into the waveforms of clutter echoes has been investigated. The approach taken for this purpose closely followed that used previously to evaluate the effects of pinna shape on the system's linear acoustic characteristics. In this study, deep-learning classifiers were trained to identify different pinna motion shapes and speeds from the distinctive signatures they imposed on the clutter echoes. Success in doing so can be seen as experimental confirmation that these nonlinear signal transformations have consistent effects on the clutter echoes.

Whereas the linear effects of different static pinna shapes can be characterized completely in the frequency domain, the time-variant and Doppler-shift signatures produced by deforming pinnae exhibit inherently joint time–frequency properties. This suggests that the time–frequency structure of the emitted pulses may influence how sensory information is encoded in the resulting echoes. To examine this hypothesis, the present study employed pulses with distinct time–frequency profiles to classify the corresponding pinna motion patterns.

As in the previous study, the results can be interpreted as follows: If different motion shapes and speeds cannot be reliably identified by a state-of-the-art deep neural network, then consistent signatures capable of supporting the encoding of useful sensory information are either absent or were undetectable by state-of-the-art deep learning classifiers. Conversely, successful classification would indicate the presence of a reliable substrate for nonlinear encoding of information into clutter echoes—although its functional utility would still require further demonstration.

2 | Experimental Section

2.1 | Acoustic Setup

The acoustic elements of the biomimetic sonar system used in this research were a single emitter and a single receiver, each surrounded by a biomimetic baffle that was designed to mimic the respective interfaces in horseshoe bats (family *Rhinolophidae*). The emitter structure was designed to mimic the noseleaf, i.e., a "megaphone-like" emission baffle, of the greater horseshoe bat (*Rhinolophus ferrumequinum* [16]). The noseleaf was scaled up in size by a factor of approximately 1.7 relative to the biological model. As a result of this scaling, the noseleaf had a total height (tip of lancet to base of anterior leaf) of approximately 50 mm and the "nostrils," i.e., the outlets for the ultrasound, were approximately 3.5 mm in diameter and spaced 5.5 mm apart.

Two electrostatic ultrasonic transducers (Series 600 open-face ultrasonic transducer, diameter 38 mm, SensComp, Livonia, MI, USA) with a -6 dB passband from approximately 45 to 75 kHz and a maximum response at approximately 55 kHz were used

to generate the ultrasonic emissions. One transducer each was connected to the nostrils of the biomimetic noseleaf via a conical waveguide approximately 10 cm in length. Both pulse generation (digital-to-analog conversion) and echo recording (analog-to-digital conversion) were performed using a single data acquisition system (PXIe-6356, National Instruments, Austin, Texas, USA) operating at 1.25 MHz with 16-bit resolution.

The receiver structure was a reception baffle designed to mimic the pinna, i.e., the outer ear, of a horseshoe bat. The geometry of the pinna model was adapted from previous simplifications of the pinna geometry in greater horseshoe bats (*Rhinolophus ferrumequinum* [27]). The physical realization of the pinna model was manufactured from silicone elastomer (Dragon Skin 30, Smooth-On Inc., Macungie, PA, USA), which provided the durability and flexibility that are required for repeated dynamic actuation. The aperture of the pinna, which represents the sound-receiving region relevant for diffraction toward the ear canal, measured approximately 50 mm in height, representing a 1.7× scaling relative to the biological pinna of greater horseshoe bats (approximately 29 mm aperture height). The ultrasonic echoes were recorded with MEMS capacitive microphones (Monomic, Dodotronic, Rome, Italy) placed at the end of an artificial "ear canal" (length 10 mm) that was attached to the pinna model.

To replicate the dynamic nature of pinna motions observed inhorseshoe bats [22], the pinna model was fitted with an actuation system that produced deformations to coincide with echo reception. The actuation system was set up to produce two distinct motion types at five different speeds each (Figure 1). Deformations were generated by virtue of motor-driven tendon actuation with two Kevlar strings (diameter, 2 mm) attached to the pinna tip: one oriented to produce a straight downwards motion ("downwards") when pulled, and the other to create a diagonal, downward-sideways motion trajectory ("sideways"). The maximum deformation amplitude for both motion types was approximately 2.5 cm.

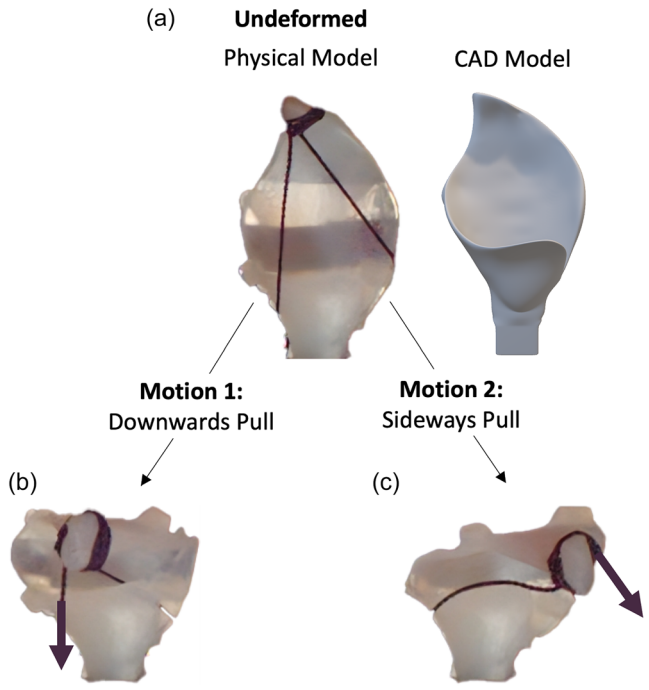

FIGURE 1 | Illustration of the two motion types applied to the biomimetic pinna model: (a) undeformed configuration showing physical model (left) and CAD geometry (right); (b) downwards deformation; and (c) sideways deformation. The arrows superposed on the photos indicate the actuation force applied to deform the pinna.

Five discrete speed levels (labeled "1" through "5" from slow to fast) were established for each motion type by varying the motor rotation amplitude, which directly controlled the magnitude of string displacement. The speeds were calibrated to produce maximum Doppler shifts in increments of approximately 400 Hz, ranging from 400 Hz (speed "1") to 2 kHz (speed "5") for a carrier frequency of 50 kHz. The maximum of this range was selected to match the maximum Doppler shifts (up to ± 2kHz) previously observed in these species [23]. A static baseline condition (no motor actuation) was included as a reference, yielding a total of eleven motion profiles for classification: two motion types, each at five speeds, plus the static reference.

All pinna deformations were driven by brushed DC motors equipped with 64 counts-per-revolution encoders (37D series, Pololu Corp., Las Vegas, NV, USA, 24V, 10,000 RPM no-load speed, 0.55 kg·cm stall torque) for feedback. Motor control was accomplished by virtue of a dedicated controller (RoboClaw, BasicMicro Inc., Lake Forest, CA, USA) with a tuned PID loop ($P = 52.62$, $I = 0.0$, $D = 25.0$). The motor controller was interfaced with the host computer via a microcontroller (Teensy 4.0, PJRC, Sherwood, OR, USA). Actuation and ultrasonic transmission were synchronized via a trigger signal (transistor–transistor logic, TTL) from the digital to analog converter (PXIe-6356 DAQ, National Instruments, Austin, Texas, USA). A single transmit pulse was transmitted per deformation cycle.

To ensure that the echo recordings coincided with the maximal Doppler shifts, preliminary calibration experiments were conducted. For these experiments, a speaker was placed approximately 1.4 m downrange to transmit continuous-frequency (CF) signals (carrier frequency 50 kHz, duration 300 ms) directly at the receiver (Figure 2). In the recorded signals, the highest Doppler shift frequency was defined as the highest frequency in the spectrogram that was associated with a power-spectral amplitude of not less than 20 dB below maximum. For each motion type and speed, the time of occurrence of the largest Doppler shift defined in this way was noted. Based on this data, pulse transmission was timed to account for both the deformation kinematics and the round-trip acoustic propagation delay such that the maximum Doppler shift occurred during the central portion of the 6 ms echo recording window.

Three types of ultrasonic transmit pulses were employed, each with a duration of 6 ms and tapered in amplitude with a Hamming-window envelope: The constant-frequency (CF) pulses contained only a single carrier frequency at 50 kHz. The carrier signal of the frequency-modulated (FM) pulses was a linear downward sweep from 50 to 35 kHz. The combined CF-FM pulse consisted of a 4 ms CF tone at 50 kHz immediately followed by a 2 ms FM sweep from 50 kHz to 35 kHz. The CF-FM structure mimics the general time–frequency shape of the echolocation pulses seen in horseshoe bats [34].

The choice of a center frequency of 50 kHz for the CF component was informed by both morphological scaling and comparative data from horseshoe bats. The greater horseshoe bat (*Rhinolophus ferrumequinum*) typically emits a CF component near 80 kHz

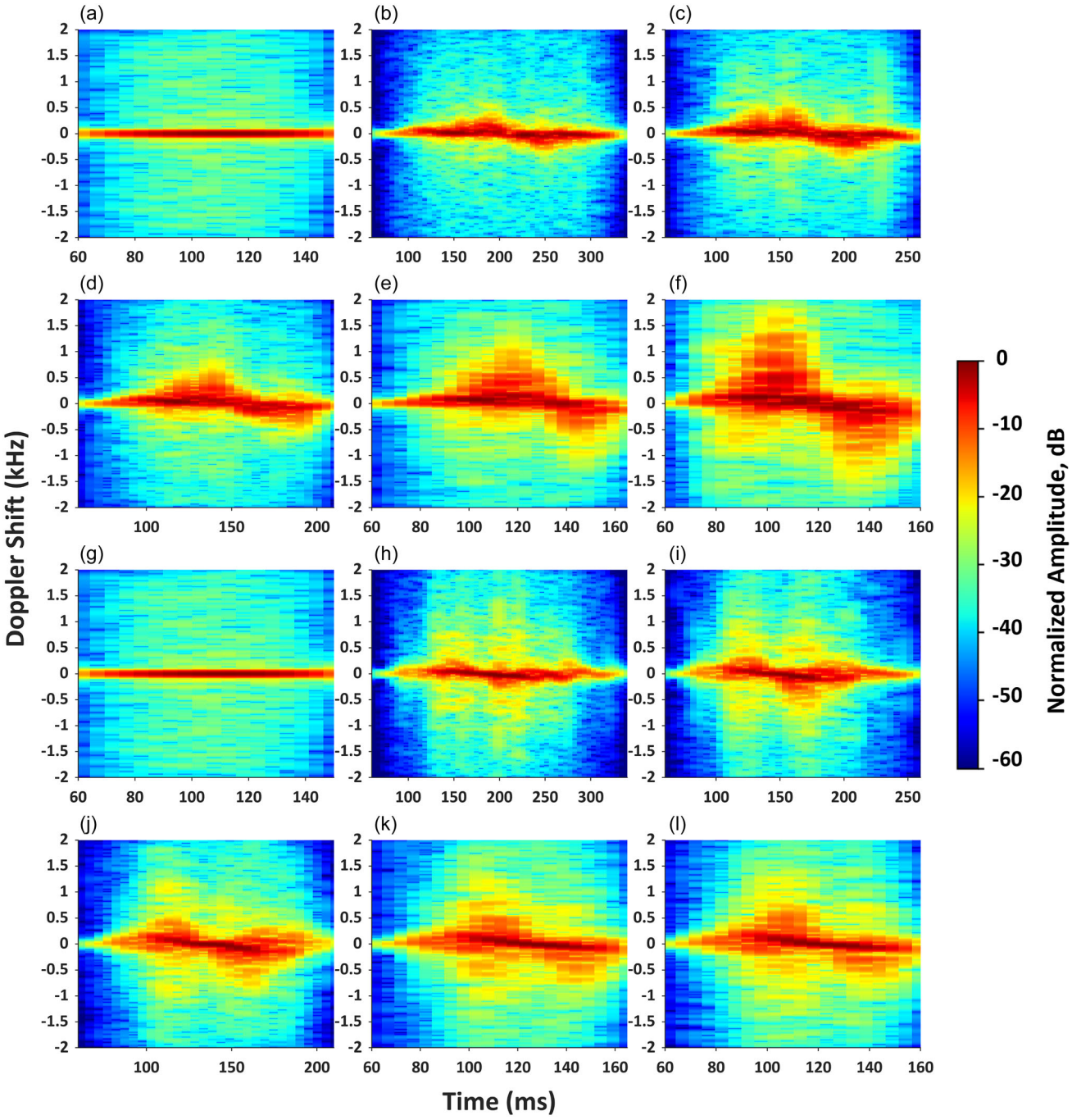

FIGURE 2 | Doppler-shift patterns associated with the two motion shapes and six speeds tested: The top set of spectrograms (a–f) corresponds to the "downwards" deformation and the bottom set (g–l) corresponds to the "sideways" deformation. Within each spectrogram set, Doppler patterns for speeds from static (a,g) to the largest tested speed ("speed 5", f,l) are shown. Doppler patterns were created using a CF pulse (50 kHz, duration 300 ms) directly transmitted from a loudspeaker to the biomimetic pinna. The time window varies between speeds: For slower motions (b–d, h–j) more time is needed to fully develop both positive and negative Doppler shifts, while for faster motions (e,f,k,l) shorter windows are sufficient, so the time axis was adjusted accordingly.

[35]. Since the biomimetic pinna was scaled up by a factor of approximately 1.7, direct allometric scaling [36, 37] would yield a proportionally lower CF frequency of approximately 47 kHz (80 kHz divided by 1.7). The selected center frequency of 50 kHz that was picked for convenience falls within the deviations from strict allometric scaling that have been observed across different rhinolophid bats: Several species (e.g., *R. luctus*, *R. clivosus*, *R. mehelyi*) exhibit CF frequencies that differ by up to 50% from their size-based predictions [37]. The 15 kHz FM bandwidth similarly corresponds to typical biological values when scaled according to the 1.7× size factor.

2.2 | Clutter Environment

The sonar setup was placed in an enclosure with approximate dimensions of 2 × 2 × 2 m (length × width × height) that was surrounded by artificial foliage on three of the side walls (Figure 3).

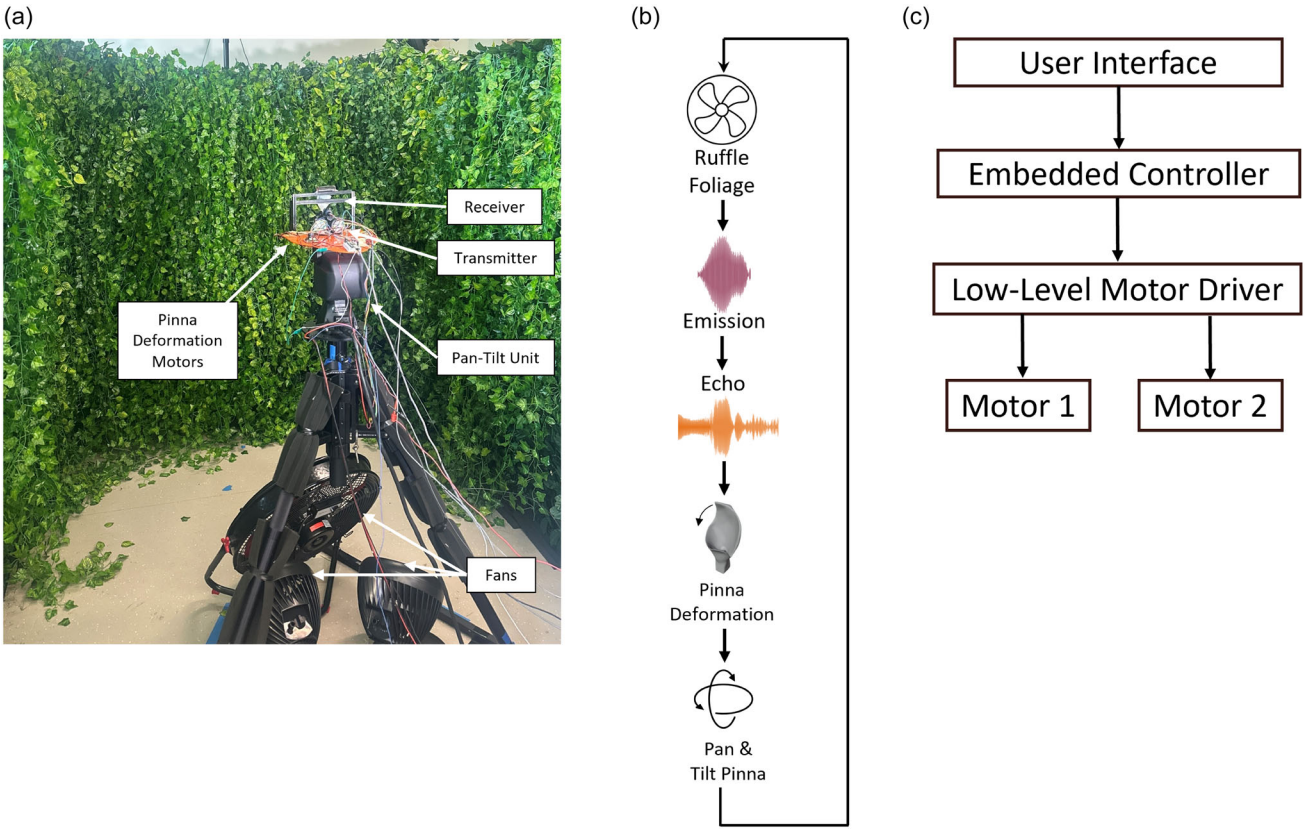

FIGURE 3 | Data collection setup: (a) view of the physical setup; (b) control scheme for the experiments; and (c) system control hierarchy from the user interface to the low-level motor drivers.

The artificial foliage consisted of plastic leaves that were approximately elliptical in shape, with an average length of 5 cm, a width of 3 cm, and a thickness of 0.2 mm. Since the specific acoustic impedances of plastic materials are similar to those of biological soft tissues (one to a few MRayl [38, 39]) and hence four orders of magnitude larger than air (413 Rayl [40]) for any of these materials, the diffraction behavior of the plastic leaves can be expected to be identical to that of biological leaves to a very good approximation.

The total thickness of the foliage layer was approximately 0.30 m. The density of the foliage was estimated by manually counting the number of leaves within representative 10×10 cm segments of the foliage layer, yielding an estimated density of approximately 6 leaves per 100 cm^2, corresponding to 6,000 leaves per cubic meter. Based on the measured beampattern characteristics of the pinna models, the −6 dB beamwidth illuminated a circular footprint with radius approximately 1 m on the foliage surface. With the foliage depth of 0.30 m, an estimated 16,000 leaves could contribute to each echo at this gain level.

The floor and ceiling were not covered with artificial foliage due to the low sonar gains observed in these directions. In this setup, the noise floor in the ultrasonic recording was around −66 dB relative to the maximum amplitude of the echoes. The minimum and maximum Fraunhofer distances [41] of the biomimetic sonar emitter were estimated as 0.5 and 0.75 m, respectively, based on a receiver aperture width of 5 cm and the operating frequency range from 35 to 50 kHz. The sonarhead was placed at a distance of 1.5 m from the artificial foliage, ensuring that all reflections were received under far-field conditions. All ultrasound recordings were made under laboratory conditions (temperature ~20°C, relative humidity ~30%).

The complete biomimetic sonar assembly (emitter, receiver, and pinna actuation) was mounted on a pan-tilt unit (PTU-46-17.5, FLIR, Wilsonville, Oregon, USA) that was used to vary the orientation of the sonarhead systematically. Data was collected across a grid of orientations by stepping the azimuth from −10° to +10° in increments of 1°, and the elevation from −3° to +3° in increments of 1°, yielding a total of 147 unique angular positions (21 azimuth × 7 elevation). This dense spatial sampling ensured that the majority of reflections originated from the foliage layer rather than from the enclosure boundaries (Figure 4).

Data acquisition was controlled (Figure 3) so that between each pair of subsequent echo recordings, three large high-velocity drum fans (fan diameter 50.8 cm, Hyper Tough, SFDE-500B3-1, Libertyville, Illinois, USA), with an air throughput of 3.3 $m^3\,s^{-1}$ each, were activated to perturb the positions and orientations of the leaves in the foliage. This randomization procedure ensured that sequential echo recordings sampled different foliage configurations, creating uncorrelated clutter echoes. Previous work [26] has demonstrated that this fan activation reduces the average off-diagonal correlation between sequential echoes from 0.76 (static foliage) to 0.22 (agitated foliage), approaching the correlation structure observed in field recordings from natural vegetation [12, 14].

2.3 | Data Collection

For the clutter echo experiments, a total of 11,000 individual echo samples were collected, equally distributed across all 11 motion

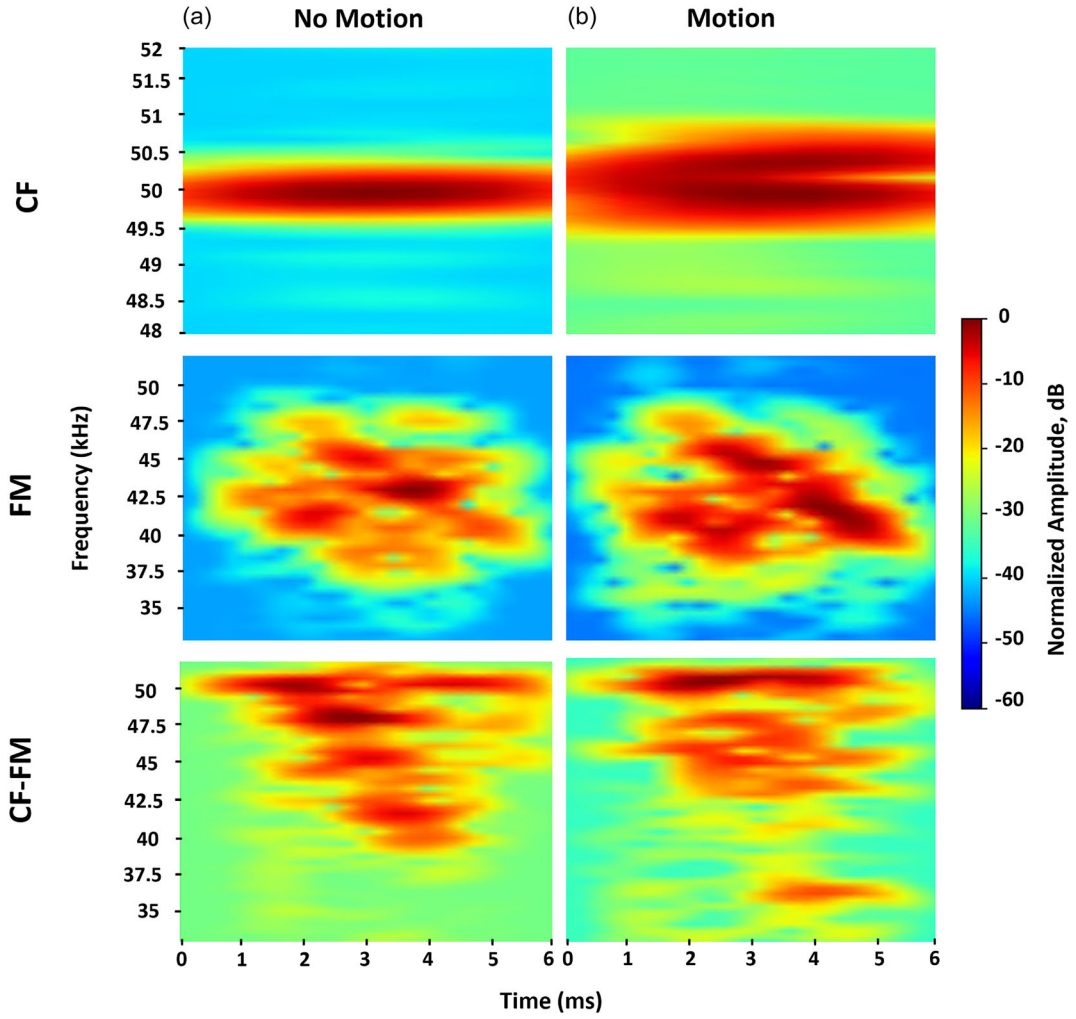

FIGURE 4 | Example clutter-echo recordings obtained for the three studied signal types: constant frequency (CF 50 kHz carrier frequency, 6 ms duration, top row), frequency-modulated (50–35 kHz sweep, 6 ms duration, middle row), and combined CF-FM (50 kHz CF carrier, 4 ms duration, followed by a 50–35 kHz sweep with a 2 ms duration, bottom row). Spectrograms are shown for (a) the static-pinna condition and (b) a motion condition (downward, speed 5). Each panel shows the spectrogram of the received clutter-echo segment (approximately 6 ms duration). Clutter-echo segments like these examples were used as input to the deep learning classifier for the pinna motion.

profiles (1,000 samples per profile). Data collection proceeded systematically through the angular position grid. The recording sequence was as follows: (1) The sonarhead was positioned at successive azimuth-elevation angles from the 147-position grid, following a systematic raster scan from −10° to +10° azimuth and −3° to +3° elevation, (2) for nonstatic motion profiles, the pinna motion was initiated and allowed to stabilize for one complete deformation cycle before pulse transmission, (3) echoes were recorded sequentially for all three signal types (CF, FM, CF-FM) at that angular position with the pinna motion synchronized to each pulse, (4) the pinna motion was stopped, (5) the fans were activated for 5 s to alter the foliage configuration, (6) the fans were deactivated for 2 s to allow the foliage motions and setup vibrations to die down, (7) the sonarhead was repositioned to the next angular location in the grid, and (8) Steps 2–7 were repeated. This procedure cycled through the angular grid repeatedly until 1,000 echo samples were collected for each motion profile. Metadata were logged for each echo, including: motion profile (static, downwards speeds 1–5, or sideways speeds 1–5), azimuth angle, elevation angle, signal type (CF, FM, or CF-FM), acquisition timestamp, and environmental conditions (temperature, humidity).

2.4 | Signal Processing and Representation

Echo segments for classification were extracted from the recorded signals with temporal alignment to the time of occurrence of the maximum Doppler shift. Each echo segment was clipped to a duration of 6 ms, corresponding to the transmitted pulse duration, beginning at the expected echo arrival time based on the round-trip acoustic propagation delay (1.5 m distance, ≈9 ms round-trip time). This ensured that all analyzed echo segments captured the portion of the received signal with the maximum motion-induced Doppler modulation.

Bandpass filters (6^{th}-order IIR Butterworth design) were applied to the echo signals to filter out frequencies not covered by the spectrum of the employed pulses and the expected maximum Doppler shifts induced in the experiments: 48–52 kHz for CF signals, and 32–52 kHz for both FM and CF-FM signals. As input for the motion classification experiments, the bandpass-filtered echoes were converted into spectrogram representations using short-time Fourier transforms (STFT) with a Hanning window (length varied from 250 to 6,250 samples, corresponding to frequency resolutions from 200 Hz to 5 kHz) and 90% overlap between

adjacent windowed segments. Spectrograms for all employed time–frequency resolution settings were used separately in the data analysis to investigate the impact of time–frequency resolution on classification accuracy.

Each spectrogram was computed over the full 6 ms signal segment, corresponding to 7,500 samples at the employed 1.25 MHz sampling rate. At the highest frequency resolution (200 Hz), each window covered approximately 5 ms of data, stepping forward in 0.5 ms increments, effectively creating an approximation of a pure frequency-domain representation with minimal temporal resolution. At the lowest frequency resolution (5 kHz), each window covered approximately 0.2 ms, stepping forward in 20 µs increments, approximating a pure time-domain representation. Intermediate resolutions (500 Hz, 1 kHz, 2 kHz) were also tested.

To investigate whether Doppler-induced spectral broadening contributed to motion classification, a second set of spectrogram representations was prepared for the FM and CF-FM echoes only. In these representations, the echoes were subjected to bandpass filtering (6^{th}-order IIR Butterworth design, passband 35–50 kHz) in order to restrict the spectral representations of the echoes to the frequency band of the original pulses.

All spectrograms were standardized to a fixed size of 120×35 pixels using linear interpolation. Since different time–frequency resolutions produce spectrograms of varying native dimensions (e.g., 114×11 for 200 Hz, 1×31 for 5 kHz), this standardization ensured all neural network models received identical input dimensions. Following interpolation, min–max normalization was applied to scale the power-spectral density values of each spectrogram into the range [0, 1].

2.5 | Data Analytics

Deep-learning network architectures with varying complexity and numbers of parameters were evaluated for their ability to identify the motion type and speed from the spectrogram representations of the clutter echoes (Figure 5). The tested architectures included residual networks (ResNet-18, ResNet-34, ResNet-50, and ResNet-152 [42]), a simple 2D convolutional neural network (2D-CNN) [43], and several transformer variants (Basic Transformer, Vision Transformer, Lightweight Transformer, Hybrid CNN-Transformer, Tiny Transformer).

All tested ResNet architectures (Figure 5a) incorporated the standard convolutional neural network design with groups of convolution blocks (Figure 5b) and identity blocks (Figure 5c). Each convolution block employed either a basic block (used for ResNet-18 and ResNet-34) or a bottleneck block (used for ResNet-50 and ResNet-152). The basic block contained two convolutional layers with kernel sizes of 3×3 and 1×1, followed by batch normalization and ReLU activation [44]. The bottleneck block included three convolution layers: a 1×1 convolution for channel reduction, a 3×3 convolution for spatial feature extraction, and another 1×1 convolution for channel expansion. The identity blocks mirrored the convolution block structure but retained the input dimensions via skip connections, ensuring that the residual learning path propagated through the network. To provide a reference for the performance of the deep-neural networks, a simpler 2D-CNN (Figure 5d) was tested as well. This network consisted of a 3×3 convolution stage, followed by 2×2 max pooling, and a series of dense layers leading to a softmax classifier with eleven outputs (one per motion profile).

Several transformer-based architectures (Figure 5e) were evaluated to assess the suitability of alternative processing strategies including attention mechanisms for motion classification based on clutter echoes. The Vision Transformer [45] followed a patch-based architecture, dividing input spectrograms (120×35 pixels) into 8×8 patches, yielding 60 patches with 64-dimensional representations each (Figure 5g). The patches were linearly embedded into 512 dimensions and augmented with learnable positional encodings to preserve spatial structure. Four transformer blocks [46] processed the patch sequence, each containing multihead attention layers (implemented as dense transformations with residual connections) followed by feed-forward networks with 512-dimensional hidden layers, concluding with global average pooling and a classification head. The lightweight transformer (Figure 5f) employed a frequency-focused design: Spectrograms were transposed from the standard time–frequency representation (120 time steps × 35 frequency bins) to a frequency-time representation (35×120) and processed through three 1D convolutional layers (64, 128, 256 channels respectively, kernel size 3 for all layers) to extract frequency-domain features along the 120-sample temporal axis, treating each of the 35 frequency bins as a sequence element. The resulting 35-length frequency sequence was augmented with learnable positional encodings, where each of the 35 frequency bins received a position-specific embedding to preserve frequency-domain structure, and processed by three transformer blocks with 256-dimensional embeddings, residual connections, and feed-forward networks. The Hybrid CNN-Transformer (Figure 5h) combined spatial feature extraction with attention mechanisms: three 2D convolutional layers (32, 64, 128 channels respectively with 3×3 kernels and 2×2 max-pooling between layers) extracted hierarchical features, which were flattened to convert spatial positions into a sequence format (each spatial location in the downsampled feature map became a sequence element) augmented with learnable positional encodings that encoded the spatial position of each feature within the downsampled feature map, and processed through three transformer blocks (Figure 5i) with 128-dimensional embeddings, attention layers, and feed-forward networks. Two simpler dense network architectures provided baseline comparisons. The basic transformer flattened spectrograms into 4,200-dimensional vectors and processed them through dense layers with 512 and 256 neurons, followed by global average pooling and a classification head. The Tiny Transformer employed a similar but more compact design with dense layers of 128 and 64 neurons to test the minimal viable architecture complexity. All transformer-inspired architectures used ReLU activations, dropout (rate = 0.2), and softmax classification.

All networks were implemented in TensorFlow ([47], version 2.20.0) via the Keras interface library ([48], version 3.10.0) and the Python programming language (version 3.12.11). All training and inference were conducted on a graphics card (H200 Tensor Core GPU, NVIDIA, Santa Clara, California, USA) with the CUDA application programming interface ([49], version 12.7) on the Virginia Tech Advanced Research Computing high-performance computing cluster [50].

The classification results were obtained using a k-fold cross-validation paradigm [51] with $k = 5$, whereby the data were partitioned into five equal parts (i.e., "folds"). The samples in each fold were split into five subsets, four of these subsets (i.e., 80% of

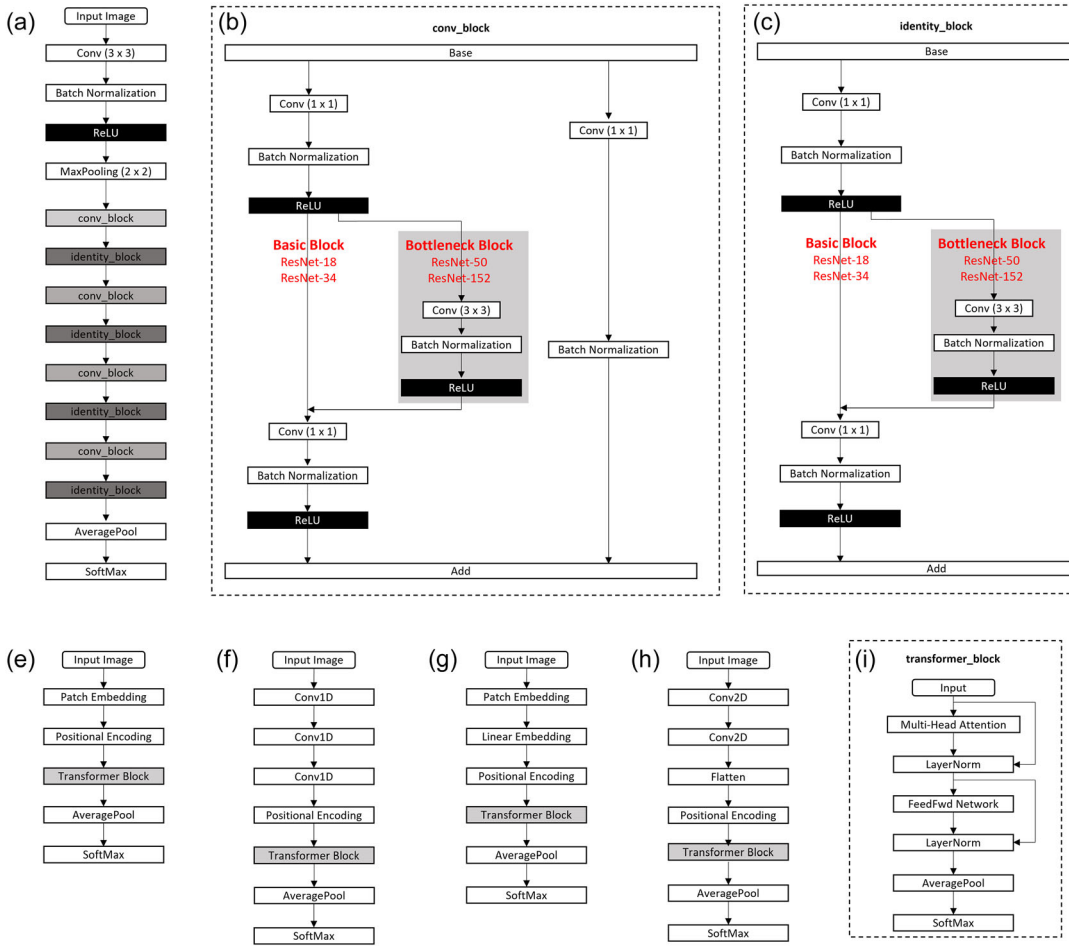

FIGURE 5 | Deep-learning network architectures used to identify the pinna motion given the spectrogram of a single clutter echo: (a) overall ResNet architecture; (b) architecture of an individual convolution block showing the basic block used for ResNet-18 and ResNet-34 and the bottleneck block used for ResNet-50 and ResNet-152; (c) the identity block architecture with three convolutional layers and the original input propagated in parallel; (d) the 2D convolutional neural network architecture used as a reference; (e) general transformer architecture showing the general encoder structure with input embedding, positional encoding, stacked transformer blocks, and classification head; (f) Lightweight Transformer with frequency-focused 1D convolutions processing frequency bins as sequence elements; (g) vision transformer (ViT) architecture used showing patch embedding, positional encoding, and transformer blocks; (h) hybrid CNN-Transformer architecture used combining 2-D convolutional feature extraction with transformer processing; and (i) transformer block architecture with multihead self-attention, layer normalization, and feed-forward network with residual connections.

the data) were used for training and the remaining subset (i.e., 20% of the data) for testing. In each fold, a different subset was designated for testing. This ensured that all samples contributed to both training and testing. The average classification accuracy was computed across the five folds to provide the reported performance metric.

Model training used the Adam optimizer [52] with an initial learning rate of 0.001 and categorical cross-entropy loss [53]. Networks were trained for up to 500 epochs with a batch size of 64. Dropout (rate = 0.2) was applied in the classifier heads and transformer blocks to mitigate overfitting. Early stopping was employed based on validation accuracy, with patience set to 100 epochs; upon triggering, model weights were restored to the epoch with the highest validation accuracy rather than retaining the final epoch's weights. Adaptive learning-rate scheduling was implemented whereby the learning rate was reduced by a factor of 0.5 when validation loss failed to improve for 25 consecutive epochs, with a minimum learning rate floor of 1×10^{-6}. Model checkpointing was enabled to save the best-performing model from each fold based on validation accuracy, and the saved model was then evaluated once on the corresponding held-out test set.

To investigate the separability of motion-induced signatures at different stages of processing, three different uniform manifold approximation and projection (UMAP) analyses were performed [54] to produce 2D representations of the samples (Figure 6): First, unsupervised UMAP was applied directly to the raw input spectrograms (CF-FM signal type at 400 Hz frequency resolution, corresponding to the optimal configuration for ResNet-18 with 97.19% overall classification accuracy) to assess whether motion-related patterns were readily accessible in the raw inputs. This baseline analysis revealed the natural clustering structure of the clutter echoes before any learned transformations. Second, supervised UMAP [54] was applied to the same input spectrograms, incorporating motion profile labels to examine the maximum achievable separation when the dimensionality reduction

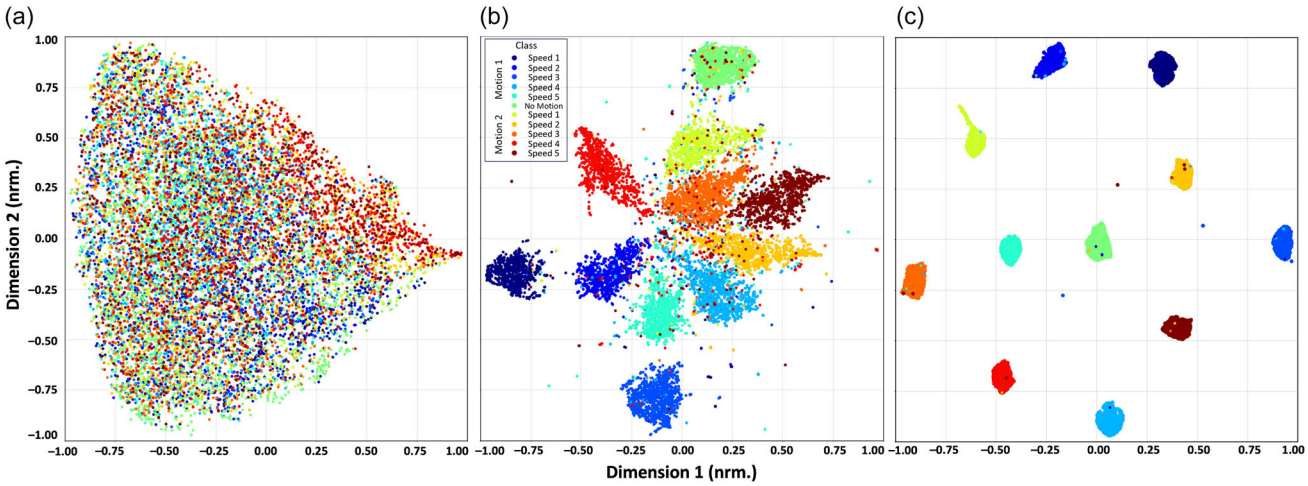

FIGURE 6 | Separation of the clutter-echo dataset for the CF-FM signal type of the different pinna motions based on uniform manifold approximation and projection (UMAP) analysis: (a) spectrogram inputs without supervision; (b) spectrogram inputs with supervision; and (c) feature representations at the final layer of the deep neural network, showing distinct cluster separation for the classes in two dimensions without supervision.

was informed by the ground truth. Finally, unsupervised UMAP was applied to the activations from the final hidden layer of the trained ResNet-18 network to visualize how the deep learning model transformed the input space.

The UMAP algorithm was configured as follows: The neighborhood size parameter (15) determined the number of neighboring points considered when constructing the high-dimensional topological representation, with this moderate value balancing preservation of local structure (fine-scale relationships between similar samples) against global structure (broad organization of clusters). The minimum distance parameter (0.1) controlled how tightly UMAP was permitted to pack points together in the two-dimensional embedding, with this relatively small value encouraging tight clustering of similar samples while allowing some separation for visualization clarity. The number of components (2) specified the dimensionality of the output embedding space, set to two dimensions to enable visualization. Euclidean distance was used as the metric for computing distances in the high-dimensional input space.

3 | Results

The motions of the deforming biomimetic pinna were found to impart distinct Doppler signatures on the narrowband signals that were received via direct transmission from a speaker (Figure 2): For the downward deformations, fairly clear sinusoidal modulation patterns were observed, with positive and negative frequency shifts approximately equal in magnitude, and the maximum positive Doppler shifts scaling linearly with actuation speed as configured in the calibration: approximately 400 Hz (Speed 1), 800 Hz (Speed 2), 1.2 kHz (Speed 3), 1.6 kHz (Speed 4), and 2 kHz (Speed 5). The Doppler patterns that were observed for the sideways deformations had a similar general shape, but appeared less symmetric. Likewise, the maximum Doppler shifts (above the −20 dB threshold) followed the same linear scaling from 400 Hz to 2 kHz. However, high power-spectral amplitudes tended to be concentrated in narrower frequency bands than was the case for the downward motions.

In the spectrograms of the clutter echoes (Figure 4), the observable impact of the pinna motions varied with the type of the pulse that was employed to trigger the respective echoes: The spectral broadenings due to the Doppler shifts (up to the maximum of +2 kHz) were obvious in echoes triggered by the CF pulses (Figure 4b, top), but were much less discernible in the FM and CF-FM echoes (Figure 4b, center and bottom) due to the random distribution of the signal energy across the time–frequency plane. However, the spectral broadenings were still detectable at edges of the signal band upon careful inspection.

The tested classifier architectures differed in their ability to identify the different pinna motion conditions (Figure 7). The four tested ResNet architectures consistently delivered some of the highest classification performances. For example, when trained on CF-FM signals and the most favorable time–frequency resolution of the spectrograms, the ResNet-18 architecture achieved the highest classification accuracy (97.19%), followed closely by ResNet-34 (97.09%), ResNet-50 (97.08%), and ResNet-152 (96.05%). The simple 2D-CNN architecture achieved 90.82%. Transformer-based architectures yielded widely varying performances: the Vision Transformer achieved 96.05%, the Hybrid CNN-Transformer reached 93%, and the Lightweight Transformer attained 92.21%, while the Tiny Transformer and the Basic Transformer architectures failed to perform as effectively (below 40%).

In general, the performance level of the tested classifier architectures tended to increase with the number of model parameters (Figure 8): On the high end, the tested ResNet architectures had the highest number of model parameters and also achieved the highest levels of performance. On the opposite end, the transformer architectures with the least number of parameters also displayed the lowest levels of performance. However, there were exceptions from this trend: Within the different ResNet architectures, there was no discernible trend for an improved performance with increased model size. Several lightweight architectures were able to match architectures with higher parameter count in terms of the achieved classification performances: The Hybrid CNN-Transformer and Lightweight Transformers architectures, for example, achieved 92% to 93% accuracy for CF-FM signals with

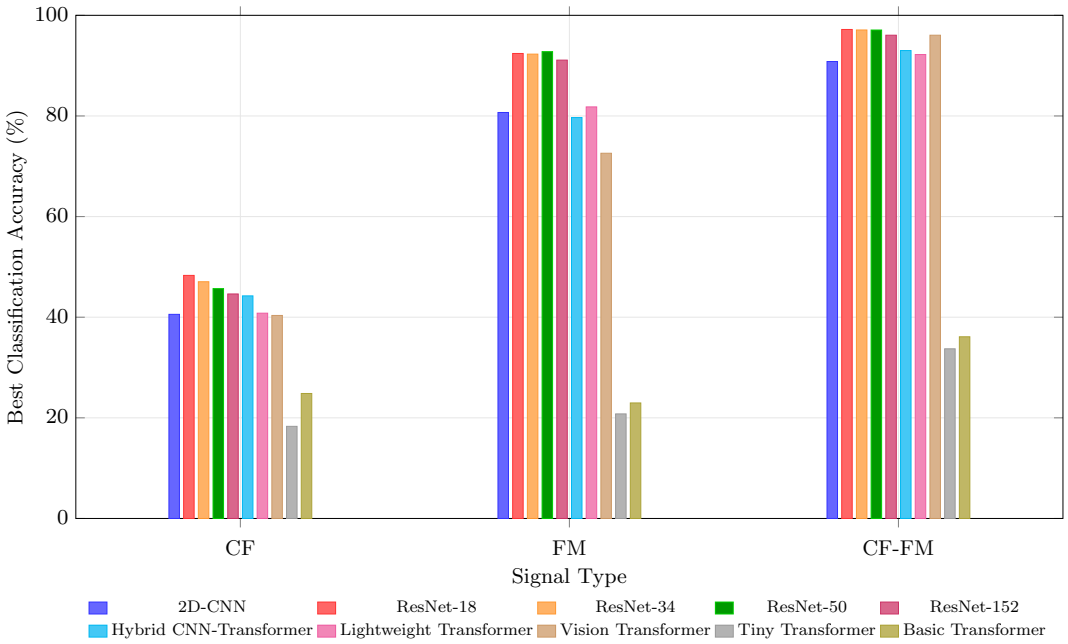

FIGURE 7 | Best average classification accuracy achieved by each of the tested classifier network architectures (color, see legend) for each of the three employed signal types (CF, FM, CF-FM). The best performance was determined across all tested time–frequency resolutions of the echo spectrograms (Figure 10).

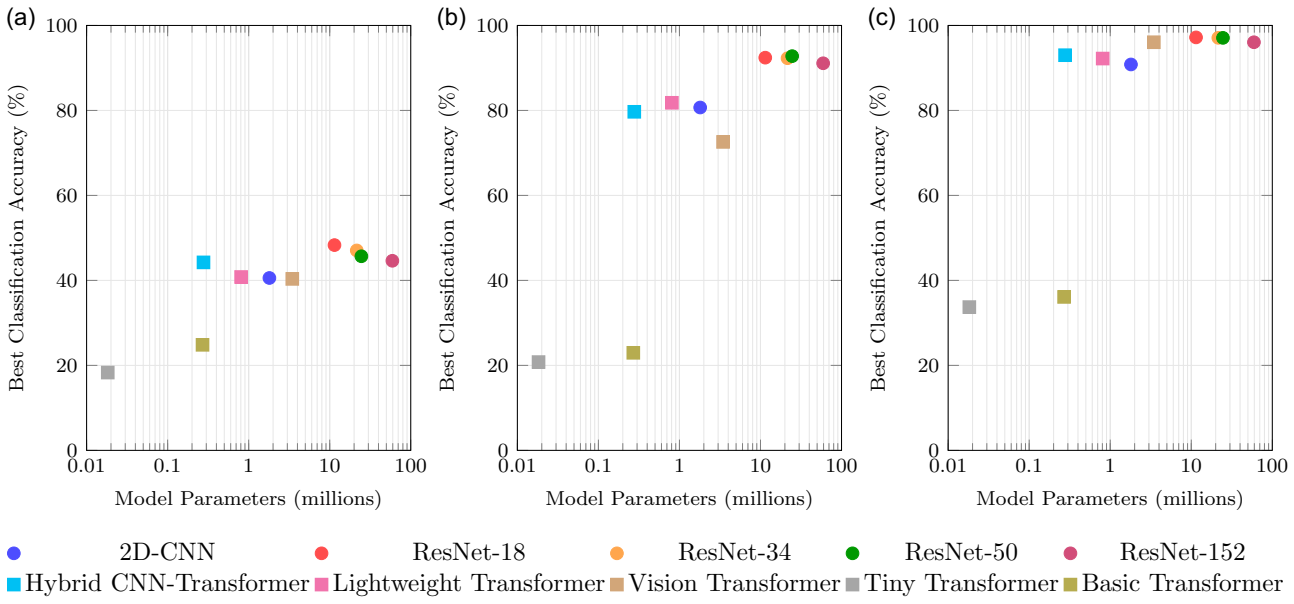

FIGURE 8 | Classification accuracy versus model parameter count for the different tested classifier architectures (colors, legend) and signal types: (a) CF signals, (b) FM signals, and (c) CF-FM signals.

less than one million parameters. This represents a 14–41× parameter reduction when compared with ResNet-18 (97.19%). Similarly, the Vision Transformer matched the performance of ResNet-152 (96% accuracy) using 94% fewer parameters.

Classification accuracy varied substantially across the three signal types tested. With the best-performing network architecture (ResNet-18) and the most favorable frequency resolution (400 Hz), CF signals achieved 48.31% overall accuracy, FM signals reached 92.78%, and CF-FM signals attained 97.19% (Figure 9).

While the CF echoes supported the lowest motion identification accuracies, the performance level achieved (48.31% overall) was still well above chance level (9.1% for the 11 motion classes). The confusion matrix revealed systematic patterns in this performance (Figure 9a): The static shape achieved the highest individual accuracy at 93.7%, followed by some of the conditions with the highest speeds (downwards at speed 5, 67.3%; sideways at speed 4, 55.5%). Mid-range speeds showed substantially lower accuracies (30–45%), with the highest confusions between adjacent speed levels of the same motion type. Confusions within

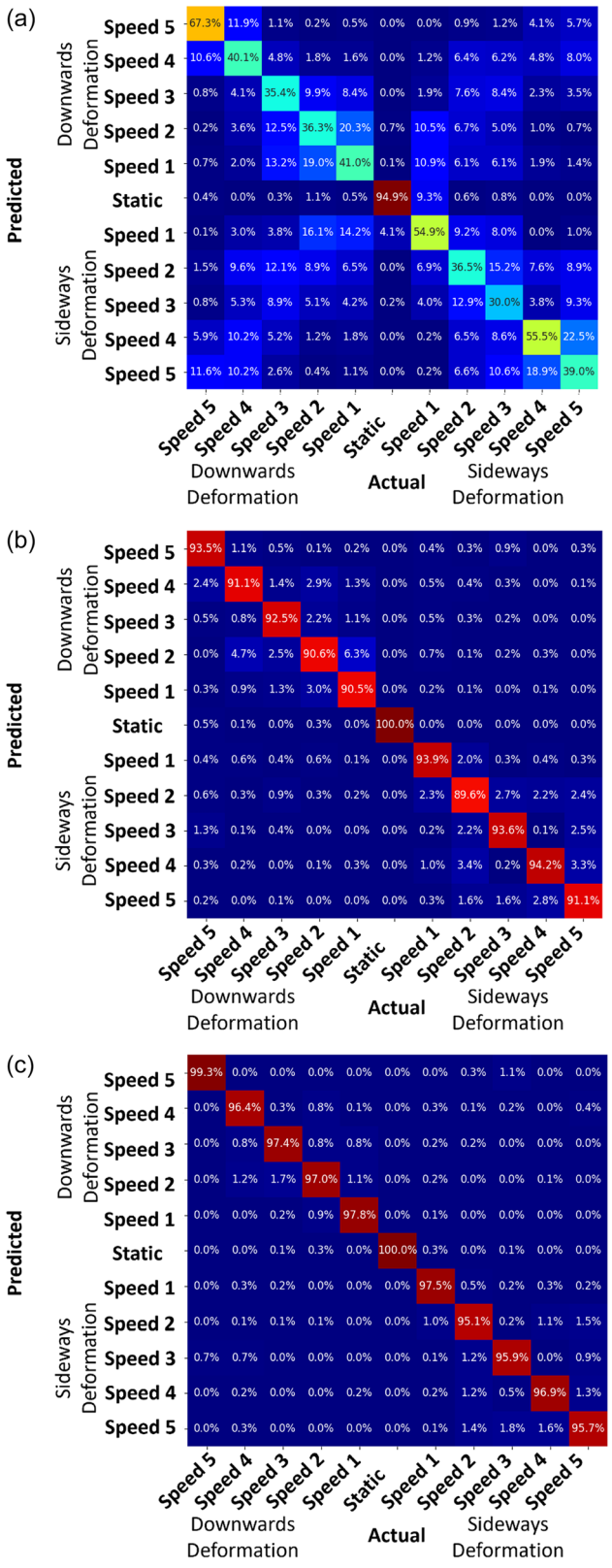

FIGURE 9 | Performance of the pinna motion classifier for different signal types. Average confusion matrices for the best performing network architecture (ResNet-18), operating on the 6 ms clutter-echo segments (Figure 4) using k-folds validation, with $k = 5$: (a) CF (50 kHz carrier, 6 ms duration), with an average accuracy of 48.31%; (b) FM (50–35 kHz sweep, 6 ms duration), with an average accuracy of 92.78%; and (c) CF-FM (50 kHz tone, 4 ms duration followed by a 50–35 kHz sweep, 2 ms duration), with an average accuracy of 97.19%.

motion types (i.e., downward versus sideways) tended to be greater than between motion types, but this difference was not statistically significant (average confusion 7.20% within motion types vs. 5.26% between motion types, two-sample t-test: $t(88) = 1.80$, $p > 0.05$). Confusion between neighboring speeds (i.e., one level apart) was significantly higher than between speeds that were two or more levels apart (average confusion 11.94% for neighboring speeds vs. 4.04% for distant speeds, two-sample t-test: $t(38) = 5.03$), $p < 0.001$).

For FM signals, true positive rates ranged from 89.6% to 100% across motion types and speeds, with the static shape achieving perfect classification (Figure 9b). Misclassifications occurred primarily between adjacent speeds within the same motion type, similar to the results for the CF signals, but at overall substantially reduced rates of confusion. Within-motion-type confusion averaged 1.63%, significantly higher than cross-motion-type confusion (0.28%; two-sample t-test: $t(88) = 6.75$, $p < 0.001$).

The classifier performance for CF-FM signals was only slightly better than for FM signals: The overall accuracy was 97.19%, with true positive rates ranging from 95.1% to 100% across all motion types and speeds (Figure 9c). The static condition again achieved perfect classification. Errors were confined almost exclusively to neighboring speed levels within the same motion type, with the largest confusions (1.8%) occurring between speed levels 3 and 5 for the sideways deformation. Within-motion-type confusion averaged 0.60%, significantly higher than cross-motion-type confusion (0.12%, two-sample t-test: $t(88) = 5.40$, $p < 0.001$), demonstrating robust motion-type discrimination with errors confined to fine-grained speed differentiation within each motion category.

Bandpass filtering the spectrogram representation of the FM and CF-FM clutter echoes to restrict their passband to that of the original pulses resulted in only negligible changes to the classification performance: When using the ResNet-18 classifier with a spectrogram frequency resolution of 400 Hz, the estimated overall classification accuracy after bandpass filtering changed from 92.36% to 92.15% for the FM signals and from 97.19% to 97.20% for the CF-FM signals.

Classification accuracy exhibited a strong dependence on the time–frequency resolution of the spectrogram representations (Figure 10) that followed a similar pattern across all three signal types: Classification accuracies tended to peak in the region of high frequency resolutions and to decrease monotonically from there as the frequency resolution decreased. Exceptions from this trend were only found in the lowest-performing architectures. For FM and CF-FM signals, the peak performances occurred at frequency resolutions that were reduced from the maximum values tested, whereas for the pure CF signals, performances tended to decrease monotonically from the highest frequency resolutions tested.

UMAP analysis displayed a three-stage progression in motion signature separability from raw spectrograms to learned network representations (Figure 6): Unsupervised UMAP applied to raw CF-FM spectrograms showed heavy overlap between the different motions with no apparent structure (Figure 6a). Supervised UMAP based on motion-labels resulted in improved separation between different motion. However, the clusters remained closed together and a substantial number of samples were scattered outside of their respective clusters (Figure 6b).

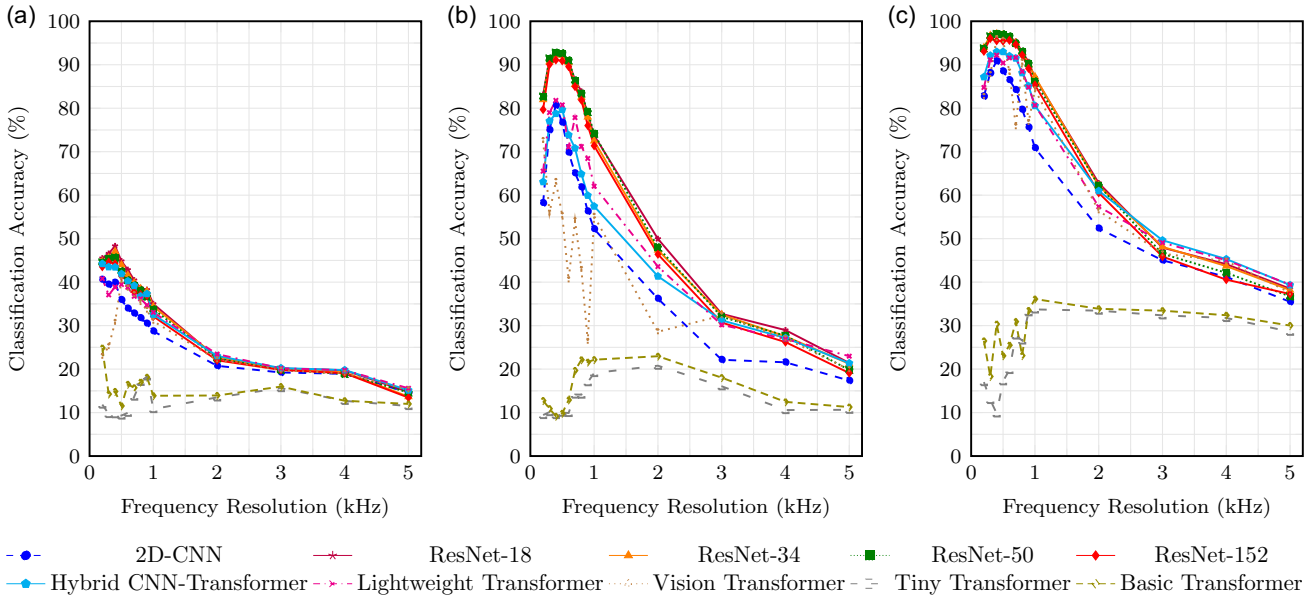

FIGURE 10 | Effect of the time–frequency resolution of the clutter-echo spectrograms on classification accuracy for the three signal types: (a) CF (50 kHz carrier frequency, 6 ms duration), (b) FM (50–35 kHz sweep, 6 ms duration), and (c) CF-FM (50 kHz carrier frequency with 4 ms duration followed by a 50–35 kHz sweep with 2 ms duration).

Unsupervised UMAP applied to final hidden layer activations from the trained ResNet-18 network demonstrated separation of the different motions into compact regions with large inter-cluster distances (Figure 6c).

4 | Conclusion

The successful classification of pinna motion types and speeds from clutter echo spectrograms demonstrates that dynamic pinna motions create consistent, i.e., discriminable and reliable, signatures despite the inherently stochastic nature of the clutter echoes. These fundamental properties should allow these motion-induced effects to serve as a suitable substrate for the encoding of sensory information. Furthermore, the fine-grained discrimination of both motion types and speeds could offer an opportunity for information encoding in subtle differences as well as give fine control over the process of information encoding. Hence, the current research extends previous findings on the potential role of pinna mobility in bats. Whereas prior work has been limited to rigid rotations that reorient the beampattern [55, 56] or static differences in beampattern [26], the current work covers nonstationary linear as well as nonlinear (Doppler) effects associated with the fast, continuous deformations of the pinnae.

The classification results revealed marked differences in performance across the three signal types, with implications for understanding how motion information is encoded in clutter echoes: The comparatively poor performance of pure CF signals (48% overall accuracy) appears to be counter to prior findings and expectations in the context of deterministic targets, where the task has been to detect small Doppler shifts [29, 33]. In these cases, a narrowband CF signal provides a well defined marker along the frequency axis that is conducive to high accuracy. However, the stochastic distributions of the reflectors (leaves) in a foliage over a wide angular range will degrade the efficiency of any narrow frequency marker. Regardless of the input signal type, the echo generated by such a scatterer arrangement will always be blurred out by the random distribution of the Doppler shifts created by the different scatterers. This may have been the reason why CF signals supported reliable detection of motion (evident in the 94% correct classification of the static reference), but allowed only for a poor resolution of motion types and speeds, particularly when dealing with adjacent speeds.

The superior performance of the FM and CF-FM echoes (93% and 97% overall classification accuracy respectively) indicates that these signals conveyed more information about the different motions to the classifier networks. This is presumably due to these signals filling a much wider area in the time–frequency plane. The finding that restricting the echoes to the original passband of the pulses did not degrade performance for the FM and CF-FM echoes further corroborates that the classifiers were not looking at an overall spectral broadening due to the Doppler shifts. Instead, motion discrimination appears to rely on spectro-temporal structure distributed across the time–frequency representation. Similarly, the results of experimenting with different time–frequency resolutions of the echo spectrograms provided indication that some time resolution did contribute to the improved classification performance associated with the FM and CF-FM signals: For both of these signal types, the maximum performance was not achieved for the highest frequency resolution, but for a frequency resolution (300–500 Hz) that was somewhat lower than maximum (200 Hz). This finding could point to either nonlinear effects arising from Doppler shifts or linear time-variant effects evident in the time–frequency plane being important for the superior classification performance associated with the FM and CF-FM signals.

The comparisons of the results obtained with the different classifier architectures revealed that accurate classification of the motion signatures could be accomplished with comparatively small networks (277–807 K parameters). This finding supports

the notion that the motion-induced signatures are readily accessible in the spectrotemporal representations and should make utilizing them on resource-constrained platforms possible. However, the failure of unsupervised UMAP to separate motion classes in raw spectrograms and the poor performance (down to below 40% overall accuracy) of some of the smallest networks (Tiny and Basic Transformers with less than 270K parameters) suggests that a certain level of effort and sophistication may be required to extract sensory information from the motion-induced signatures. The much smaller spread between the worst and best performing classifier architectures for CF signals ($5.02\times$) on one side and FM and CF-FM signals on the other ($10.14\times$ and $9.91\times$, respectively) could be taken as an indication that the performance in the case of the CF signals is more likely limited by the information content of the signals than the capacity of the classifier networks.

While the present work has demonstrated that motion signatures exist in clutter echoes and are readily discriminable by deep-learning classifiers, a critical gap remains: Discriminability does not establish functional utility. Future work should evaluate pinna motions in actual sonar tasks by comparing static versus dynamic receiver configurations on ecologically relevant challenges such as target detection, localization, or classification in clutter. Such experiments would establish whether motion-induced signatures translate to measurable performance improvements beyond classification benchmarks, thereby bridging the gap between demonstrating that motion information exists in clutter echoes and establishing its functional value for biosonar-based sensing in complex acoustic environments.

Acknowledgments

This research has been supported by the Naval Engineering Education Consortium (NEEC, contract number N001741910001) and the Office of Naval Research (ONR, MURI N00014-17-1-2736). The authors would like to thank Sanmeel Lagad, Joseph Sutlive, Max Engel, Colin Fox, Henry Forsyth, Sam Klemic, and Matthew Long for contributions to the research.

Funding

This study was supported by Naval Engineering Education Consortium (N001741910001) and Office of Naval Research (N00014-17-1-2736).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. G. Solaimalai, K. J. Prakash, S. S.K., A. Bhagyalakshmi, P. Siddharthan, and K. R. Senthil Kumar, in *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)* (2024): 1–5.

2. M. R. Rezaee, N. A. W. A. Hamid, M. Hussin, and Z. A. Zukarnain, "Comprehensive Review of Drones Collision Avoidance Schemes: Challenges and Open Issues," *IEEE Transactions on Intelligent Transportation Systems* 25, no. 7 (2024): 6397.

3. A. T. Azar, A. Koubaa, N. Ali Mohamed, et al., "Drone Deep Reinforcement Learning," *Electronics* 10, no. 9 (2021): 999.

4. E. Cetin, C. Barrado, G. Muñoz, M. Macias, and E. Pastor, in *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)* (IEEE, 2019): 1–7.

5. U. M. Norberg and J. M. V. Rayner, "Ecological Morphology and Flight in Bats (Mammalia; Chiroptera): Wing Adaptations, Flight Performance, Foraging Strategy and Echolocation," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 316 (1987): 335.

6. B. Falk, J. Kasnadi, and C. F. Moss, "Tight Coordination of Aerial Flight Maneuvers and Sonar Call Production in Insectivorous Bats," *Journal of Experimental Biology* 218 (2015): 3678.

7. B. Tian and H. U. Schnitzler, "Echolocation Signals of the Greater Horseshoe Bat (*Rhinolophus ferrumequinum*) in Transfer Flight and During Landing," *Journal of the Acoustical Society of America* 101, no. 4 (1997): 2347.

8. H.-U. Schnitzler and E. K. V. Kalko, "Echolocation by Insect-Eating Bats: We Define Four Distinct Functional Groups of Bats and Find Differences in Signal Structure that Correlate with the Typical Echolocation Tasks Faced by Each Group," *BioScience* 51, no. 7 (2001): 557.

9. G. Neuweiler and E. Covey, *The Biology of Bats* (Oxford University Press, 2000).

10. R. J. Urick, *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, 1983), reverberation/Clutter modeling.

11. R. Müller and R. Kuc, "Foliage Echoes: A Probe into the Ecological Acoustics of Bat Echolocation," *Journal of the Acoustical Society of America* 108, no. 2 (2000): 836.

12. R. Wang and R. Müller, "Bioinspired Solution to Finding Passageways in Foliage with Sonar," *Bioinspiration and Biomimetics* 16, no. 6 (2021): 066022.

13. R. Wang, Y. Liu, and R. Müller, "Detection of Passageways in Natural Foliage using Biomimetic Sonar," *Bioinspiration and Biomimetics* 17, no. 5 (2022): 056009.

14. L. Zhang and R. Müller, "Large-Scale Recognition of Natural Landmarks with Deep Learning based on Biomimetic Sonar Echoes," *Bioinspiration and Biomimetics* 17, no. 2 (2022): 026011.

15. L. Zhang, A. Farabow, P. Singhal, and R. Müller, "Small-Scale Location Identification in Natural Environments with Deep Learning based on Biomimetic Sonar Echoes," *Bioinspiration and Biomimetics* 18, no. 2 (2023): 026009.

16. G. Csorba, P. Ujhelyi, and N. Thomas, *Horseshoe Bats of the World: (Chiroptera: Rhinolophidae)* (Alana Books, 2003).

17. S. Solari and R. J. Baker, "Mammal Species of the World: A Taxonomic and Geographic Reference by D. E. Wilson; D. M. Reeder," *Journal of Mammalogy* 88, no. 3 (2007): 824.

18. J. Habersetzer, G. Schuller, and G. Neuweiler, "Foraging Behavior and Doppler Shift Compensation in Echolocating Hipposiderid Bats, Hipposideros Bicolor and Hipposideros Speoris," *Journal of Comparative Physiology A* 155 (1984): 559.

19. G. Neuweiler, W. Metzner, U. Heilmann, R. Rübsamen, M. Eckrich, and H. Costa, "Foraging Behaviour and Echolocation in the Rufous Horseshoe Bat (*Rhinolophus rouxi*) of Sri Lanka," *Behavioral Ecology and Sociobiology* 20 (1987): 53.

20. T. A. Vaughan, "Foraging Behaviour of the Giant Leaf-Nosed Bat (*Hipposideros commersoni*)," *African Journal of Ecology* 15, no. 4 (1977): 237.

21. L. Feng, L. Gao, H. Lu, and R. Müller, "Noseleaf Dynamics during Pulse Emission in Horseshoe Bats," *PLoS ONE* 7, no. 5 (2012): e34685.

22. L. Gao, S. Balakrishnan, W. He, Z. Yan, and R. Müller, "Ear Deformations Give Bats a Physical Mechanism for Fast Adaptation of

Ultrasonic Beam Patterns," *Physical Review Letters* 107, no. 21 (2011): 214301.

23. X. Yin and R. Müller, "Fast-Moving Bat Ears Create Informative Doppler Shifts," *Proceedings of the National Academy of Sciences of the United States of America* 116 (2019): 25–12270.

24. L. Zhang, L. Yang, R. Zhang, and R. Müller, "An Experimental Link between Fast Noseleaf Deformations and Biosonar Pulse Dynamics in Hipposiderid Bats," *Journal of the Acoustical Society of America* 148, no 2 (2020): 954.

25. H. Schneider and F. P. Möhres, "Die Ohrbewegungen der Hufeisenfledermäuse (*Chiroptera, Rhinolophidae*) und der Mechanismus des Bildhörens," *Zeitschrift für Vergleichende Physiologie* 44 (1960): 1.

26. I. Eshera, S. Lagad, and R. Müller, "Impact of Biomimetic Pinna Shape Variation on Clutter Echoes: A Machine Learning Approach," *Advanced Intelligent Systems* n/a (n/a): e202500442.

27. P. B. Caspers, "*Analysis of Bat Biosonar Beampatterns: Biodiversity and Dynamics*," (PhD diss., Virginia Polytechnic Institute and State University (Virginia Tech), 2017), http://hdl.handle.net/10919/74425.

28. R. Müller, A. K. Gupta, H. Zhu, et al., "Dynamic Substrate for the Physical Encoding of Sensory Information in Bat Biosonar," *Physical Review Letters* 118 , no. 15 (2017): 158102.

29. X. Yin and R. Müller, "Integration of Deep Learning and Soft Robotics for a Biomimetic Approach to Nonlinear Sensing," *Nature Machine Intelligence* 3, no. 6 (2021): 507.

30. R. Kober and H.-U. Schnitzler, "Information in Sonar Echoes of Fluttering Insects Available for Echolocating Bats, *Journal of the Acoustical Society of America* 87, no. 2 (1990): 882.

31. G. Neuweiler, "Foraging Ecology and Audition in Echolocating Bats," *Trends in Ecology and Evolution* 4, no. 6 (1989): 160.

32. S. Radtke-Schuller and G. Schuller, "Auditory Cortex of the Rufous Horseshoe Bat: 1. Physiological Response Properties to Acoustic Stimuli and Vocalizations and the Topographical Distribution of Neurons," *European Journal of Neuroscience* 7, no. 4 (1995): 570.

33. R. Müller and H.-U. Schnitzler, "Acoustic Flow Perception in CF-bats: Properties of the Available Cues," *Journal of the Acoustical Society of America* 105, no. 5 (1999): 2958.

34. H.-U. Schnitzler, C. F. Moss, and A. Denzinger, "From Spatial Orientation to Food Acquisition in Echolocating Bats," *Trends in Ecology and Evolution* 18, no. 8 (2003): 386.

35. M. M. Andrews, P. T. Andrews, D. F. Wills, and S. M. Bevis, "Ultrasound Social Calls of Greater Horseshoe Bats (*Rhinolophus ferrumequinum*) in a Hibernaculum," *Acta Chiropt* 8, no. 1 (2006): 197.

36. Z. Huihua, Z. Shuyi, Z. Mingxue, and Z. Jiang, "Correlations between Call Frequency and Ear Length in Bats Belonging to the Families Rhinolophidae and Hipposideridae," *Journal of Zoology* 259, no. 2 (2003): 189.

37. H. Wu, T.-L. Jiang, R. Müller, and J. Feng, "The Allometry of Echolocation Call Frequencies in Horseshoe Bats: Nasal Capsule and Pinna Size are the Better Predictors than Forearm Length," *Journal of Zoology* 297, no. 3 (2015): 211.

38. H. Tohmyoh, T. Imaizumi, and M. Saka, *Key Engineering Materials* (2007), 353-358 2349, Asian Pacific Conference for Fracture and Strength (APCFS'06); Conference date: 22-11-2006 Through 25-11-2006.

39. F. Duck, *Physical Properties of Tissues: A Comprehensive Reference Book* (Academic Press, 1990).

40. L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. (Wiley, 2000).

41. E. Hecht, *Optics*, 4th intern ed. (Addison Wesley, 2002).

42. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016): 770–778.

43. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems* 25 (2012): 1097.

44. X. Glorot, A. Bordes, and Y. Bengio, In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, volume 15 of Proceedings of Machine Learning Research*, ed. G. Gordon, D. Dunson and M. Dudík (PMLR, 2011): 315–323, https://proceedings.mlr.press/v15/glorot11a.html.

45. A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., *International Conference on Learning Representations* (2021).

46. A. Vaswani, N. Shazeer, N. Parmar, et al., *Advances in Neural Information Processing Systems* (2017): 30.

47. M. Abadi, A. Agarwal, P. Barham, et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," *CoRR* (2016): abs/1603.04467.

48. F. Chollet, *Deep Learning with Python*, 2nd ed., (Manning Publications, 2021).

49. NVIDIA, P. Vingelmann, and F.H. Fitzek, Cuda, release: 10.2.89, 2020, https://developer.nvidia.com/cuda-toolkit.

50. *Virginia Tech Advanced Research Computing, High performance computing resources, Virginia Tech, Blacksburg, VA* (2024), https://arc.vt.edu.

51. R. Kohavi, et al., in *Ijcai*, Vol. 14 (1995): 1137–1145.

52. D. P. Kingma and J. Ba, "*Adam: A Method for Stochastic Optimization*," (2017), https://arxiv.org/abs/1412.6980.

53. K. P. Murphy, *Machine Learning : A Probabilistic Perspective* (MIT Press, 2013).

54. L. McInnes, J. Healy, N. Saul, and L. Großberger, "UMAP: Uniform Manifold Approximation and Projection," *Journal of Open Source Software* 3 (2018): 29–861.

55. V. A. Walker, H. Peremans, and J. C. T. Hallam, "One Tone, Two Ears, Three Dimensions: A Robotic Investigation of Pinnae Movements used by Rhinolophid and Hipposiderid Bats," *Journal of the Acoustical Society of America* 104, no. 1 (1998): 569.

56. X. Yin, P. Qiu, L. Yang, and R. Müller, "Horseshoe Bats and Old World Leaf-Nosed Bats Have Two Discrete Types of Pinna Motions," *Journal of the Acoustical Society of America* 141, no. 5 (2017): 3011.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.