Sensors and Materials, Vol. 37, No. 3 (2025) 1073–1098
MYU Tokyo

S & M 3975

Cat and Dog Behavior Recognition Method Using Deep Learning Approach Based on Inertial Measurement Unit Sensor Data

Guanyu Chen,[1] Yoshinari Takegawa,[1*] Kohei Matsumura,[2]
Hiroki Watanabe,[1] and Keiji Hirata[1]

[1]Graduate School of Systems Information Science, Future University Hakodate,
116-2 Kamedanakano-cho, Hakodate, Hokkaido 041-8655, Japan
[2]College of Information Science and Engineering, Ritsumeikan University,
2-150, Iwakura-cho, Ibaraki, Osaka 567-8570, Japan

(Received September 2, 2024; accepted March 11, 2025)

Keywords: animal–computer interaction, IoT, wearable devices, deep learning, animal behavior recognition

With the growing interest in enhancing the well-being of pets through advanced technology, in this study, we address the challenge of accurately recognizing animal behavior using wearable sensors and machine learning techniques. Existing methods are often restricted to recognizing behavior in a single species, limiting their potential for general application across different animal types. This shortcoming hinders broader applications in multi-species environments and results in inconsistent monitoring outcomes. In this research, we present an improved approach for animal behavior recognition by developing a 1D Convolutional Neural Network Long Short-Term Memory hybrid model specifically designed to process inertial measurement unit sensor data. By targeting the most relevant movement features using accelerometer, gyroscope, and magnetometer data, our model achieves a high degree of precision in classifying common activities among cats and dogs, with recognition accuracies of 89% for cats and 94% for dogs. The results validate the applicability of our model in diverse contexts, making it a promising tool for enhancing automated behavior monitoring in animal–computer interaction. This research contributes to the development of intelligent systems that improve pet care and lay the foundation for broader applications in animal welfare and behavioral studies.

1. Introduction

In recent years, there has been a notable expansion in the pet market,[(1)] coinciding with increased pet ownership. This trend has amplified concerns about pet health and safety, emphasizing the importance of these issues. The human–animal bond has emerged as a key area of interest in recent research, particularly in understanding the impact of this bond on both animal welfare and human well-being. In this study, we aim to address specific gaps in

*Corresponding author: e-mail: yoshi@fun.ac.jp
https://doi.org/10.18494/SAM5359

ISSN 0914-4935

monitoring and analyzing pet behaviors, focusing on enhancing real-time interaction between pets and their owners. Extended periods of separation between pets and their owners have been linked to increased owner anxiety regarding the well-being of their pets, highlighting the need for continuous health monitoring.(2) Conversely, pets, particularly cats, are often adept at hiding their injuries or illnesses,(3) which poses the problem of potential dangers or health issues going undetected.

In response to these challenges, we explored the development of wearable solutions aimed at real-time health monitoring and behavior analysis for pets. The developed wearable device is equipped with a nine-axis inertial measurement unit (IMU) that captures comprehensive movement data, which is stored locally on Secure Digital (SD) cards and can be transmitted to our AI model. This setup ensures robust data collection for the monitoring of pet behaviors. Through the integration of AI for data analysis and interactive technologies, the devices are designed to be pet-friendly and informative. This technology not only helps pet owners understand their pets' current conditions, thus reducing anxiety, but also prevents health deterioration, decreases the likelihood of pets getting lost, and potentially reduces the number of stray animals.

Wearable technology presents a unique opportunity to bridge the gap in understanding pet behavior, enabling more precise health monitoring and contributing to enhanced human–animal interaction.

In the development of our activity classification model, we gathered data through home visits to cats (with owner consent), conducted adaptability tests for the wearables, and recorded synchronized sensor data and videos. This approach allowed us to capture the cats' natural interactions with their owners and other induced behaviors. After annotating the activities recorded, we compiled a dataset based on the cats' voluntary actions. The model's impressive 89% accuracy rate not only supports daily monitoring but also guarantees reliability across various datasets. It is adept at identifying issues such as insufficient exercise or improper dietary habits. Utilizing sensor data from wearables, we can also glean environmental insights, contributing to a holistic approach to pet care. Furthermore, cross-validation with publicly available canine datasets(4) confirms that this high level of accuracy is consistent across different species, suggesting its potential for broader applications in animal research.

The contributions introduced in this article are as follows.

1) Development of a 1D Convolutional Neural Network-Long Short-Term Memory (1DCNN-LSTM)-based cat activity detection system using three types of sensor data, i.e., accelerometer, gyroscope, and magnetometer data, for the detection of seven activities of cats and dogs
2) Implementation of a novel data preprocessing pipeline, including time-frequency analysis through continuous wavelet transforms (CWTs) and the introduction of synthetic data augmentation techniques to improve model generalization across species
3) Validation of the model's applicability across species through rigorous cross-validation on both self-collected feline and publicly available canine datasets, demonstrating the potential for broader applications in animal behavior research

This paper is organized as follows: In Sect. 2 (Related Works), we explore market trends and current technologies in pet health monitoring, including wearable and edge-AI applications. In Sect. 3 (Methods), we describe in detail the hardware and software for data collection and the iterative development process. In Sect. 4 (Algorithm), we evaluate the effectiveness and applicability of the model using self-collected data. In Sect. 5 (Evaluation), we summarize the system testing results and the effectiveness of analyzing pet behaviors. In Sects. 6 (Discussion) and 7 (Conclusions), we review the simulation system's utility, future enhancements, and the potential of intelligent wearables to improve pets' lives.

2. Related Works

Previous studies have utilized acceleration sensors to monitor animal posture, highlighting developments in wearable technologies for animals.[5–7] Different behaviors generate different acceleration traces. Animal activity is one of the most important indicators associated with animal health and welfare.[8] For instance, the specific equipment designed for search and rescue dogs integrates posture-sensing capabilities into collars to monitor the dog's status effectively.[9] Through Kestler and Wilson's study,[10] which came from tracking four cats, a total of 10 behaviors were collected and defined. We selected seven behaviors ("Resting," "Moving," "Sleeping," "Eating," "Jumping," "Standing," and "Using Litter Tray") for collection through an interview survey of participants and an evaluation of performance.

Currently, various pet tracking products are available on the market. Notably, Apple's Air Tag utilizes Bluetooth Low Energy (BLE) technology to form a location network with users' devices. While this technology potentially addresses the issue of lost pets, it primarily tracks the pet's location without providing additional health or status information.[11] However, this device should not be used for pet monitoring, as it is not designed for animal usage and poses many safety risks, such as the risk of the reflective shell being accidentally ingested by an animal.[12]

Commercially available pet trackers such as the FitBark[13] and Tractive [14] smart collars not only focus on outdoor tracking but also monitor pet behaviors and provide pet owners with health insights. Additionally, devices such as Whistle have been studied for their ability to issue behavioral alerts, addressing the intention-action gap.[15]

Commercial pet wearables often focus on specific monitoring aspects, leaving gaps in comprehensive behavior analysis. For example, while existing devices offer location tracking and some health monitoring, they fall short of delivering advanced behavioral anomaly detection. This limitation underscores the need for more sophisticated monitoring systems capable of providing insights into a broader range of pet behaviors.

Ito[16] developed an edge-AI-based cattle monitoring system that exemplifies the use of sophisticated wearable technologies in livestock management. The system, attached around a cow's neck, features an accelerometer and connectivity functions, enabling data collection and the analysis of individual and herd movements. This capability facilitates the early detection of heat, labor, and disease symptoms, significantly enhancing herd management.

Deep learning models such as the 1DCNN-LSTM have shown promise in enhancing detection accuracy.[17,18] Trujillo-Guerrero *et al.*[17] demonstrated the effectiveness of 1DCNN-

LSTM models in capturing spatial and temporal features for human motion classification using IMU data. Inspired by their methodologies, our work extends the application of 1DCNN-LSTM architectures to animal behavior recognition, specifically focusing on cats and dogs. This extension presents unique challenges due to the distinct movement patterns and behaviors of animals, which differ from those of humans. Many existing approaches require large datasets and significant computational resources. Our research builds on these advances by proposing a more resource-efficient workflow that maintains high performance in behavior recognition tasks.

In contrast to existing studies and products, which often fail to provide comprehensive animal health data or require complex and specialized equipment, wearable devices leveraging AI can extract a wealth of information from a single IMU. Through algorithmic analysis, these devices can deduce extensive motion characteristics and health insights.(19)

To cope with the nonstationary signals in animal motion data, we introduced more flexible time-frequency analysis methods, such as CWT and empirical wavelet transform (EWT), to enhance the accuracy and robustness of data processing.

Furthermore, adopting wearable devices in pet health management promises to enhance overall pet health and provide pet owners with detailed insights into their pets' behavioral patterns and daily activities. Such detailed monitoring can enable more informed and effective caregiving strategies.

3. Methods

In current studies, although behavior recognition research involving dogs is more prevalent than that involving cats, and there are relatively more open-source datasets available,(20) the overall volume of such research is still limited. For cats, the lack of accessible datasets and customized hardware is even more pronounced. Therefore, we had to collect cat-specific data ourselves and develop customized hardware tailored to the requirements of feline behavior recognition.

3.1 Cat-oriented wearable device design requirements

Contrary to popular belief, a significant majority of cats (approximately 70%) can wear collars for extended periods without discomfort.(21) Nonetheless, cats are notoriously selective about what they wear around their necks. A collar that is very large, itchy, or improperly fitted is likely to be removed by the cat or cause harm. To address this, our device must be sufficiently compact and lightweight to attach to a collar without causing irritation. This poses a challenge, as most commercial smart collars do not meet our size specifications. For cats accustomed to wearing collars but not devices, our gadget needs to be exceptionally unobtrusive. Additionally, the device must not harm the cat or become damaged if the cat attempts to remove it.

3.2 Wearable devices for cats

To address the challenge of feline behavior recognition, we designed a specialized wearable device equipped with a nine-axis IMU, focusing on comprehensive data collection and behavior analysis. The device was carefully designed to be compact and lightweight, ensuring that it could be comfortably worn by cats without causing irritation or discomfort.

According to results from preliminary studies, sensors positioned near the back provide a higher accuracy than those hung around the neck.[21] To accommodate this, our design includes a half-body garment instead of a simple strap [Figs. 1(a) and 1(b)], which helps secure the device and allows for positional adjustments without constricting the cat (Fig. 2).

(a) (b)

Fig. 1. (Color online) Photographs of (a) half-body garment and (b) cat wearing half-body garment.

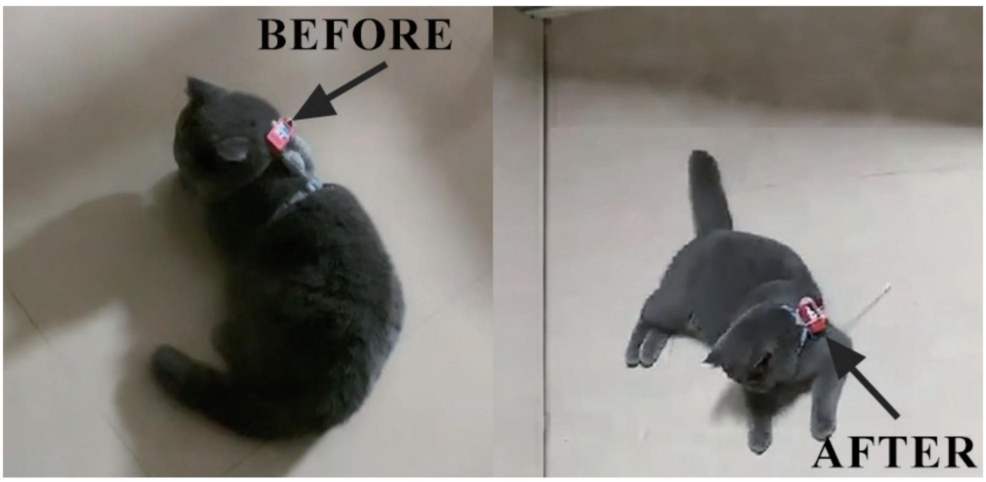


Fig. 2. (Color online) Sensor position shift from center to left side: comparison before and after experiment.

3.3 Experimental devices

To enhance versatility and cater specifically to our project's needs, the data collection hardware has been designed to focus solely on data collection, omitting unrelated hardware features to ensure ease of use by individuals unfamiliar with development. The system hardware architecture (Fig. 3) is based on the M5 capsule, augmented with a BNO055 IMU module, a BME280 temperature-humidity-barometer module, and a micro secure digital memory (microSD) card. It is powered by a gel lithium battery from Amperex Technology Limited (ATL) with a capacity of 4.15 Wh, which is detachable to facilitate short data collection sessions. This configuration enables approximately 8 h of data recording at a sampling rate of approximately 90 Hz.

To ensure that the equipment does not impede the natural behavior of the cats, we carefully managed the total weight in accordance with established guidelines. Wilson *et al.*[22] recommended that the weight of attached devices should not exceed 5% of the animal's body mass to minimize impacts on natural behavior, striving for weights below 3% whenever possible. Kenward[23] also advised that equipment should be as lightweight as possible to avoid hindering the animal's movements and behavior.

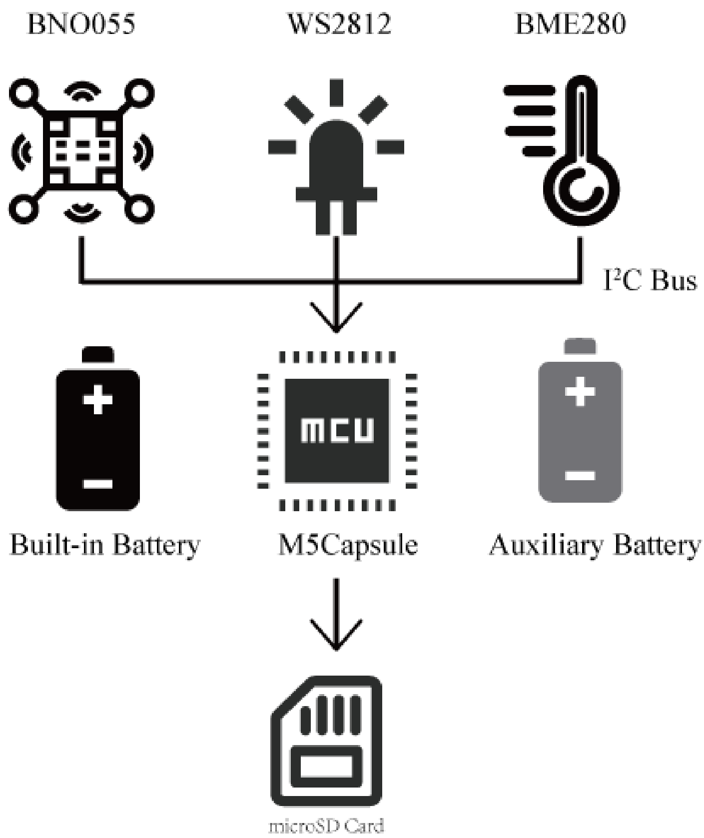


Fig. 3. Hardware architecture.

In our study, the equipment weights were as follows:

- Modified Harness: 41 g. We used a commercial cat harness and removed unnecessary components, such as heavy metal leash attachments, to reduce weight.
- Data Logger: 20 g.
- Auxiliary Battery: 30 g.
- Additional Allowance (for size variations and potential extra sensors): 10 g.

The total maximum equipment weight was 101 g. Our lightest participating cat weighed 3.5 kg, resulting in the equipment weight being approximately 2.9% of the cat's body mass, well within the recommended limits. In our standard configuration (harness and data logger only, totaling 61 g), the equipment weight remains below 3% of the body mass for cats weighing as little as 2 kg.

We programmed the M5capsule wearable device with onboard algorithms to output data such as quaternions, Euler angles, raw acceleration, and magnetic field readings. The quaternion and Euler angle data, which are automatically corrected through the BNO055's on-chip processing that also performs coordinate conversion, help adjust the sensor's orientation in response to any shifts caused by the pet's movement, ensuring accurate frame reference transformations. This adjustment is crucial to prevent the 1DCNN-LSTM model from learning incorrect features due to device displacement. Only essential data, such as acceleration and magnetic fields, are logged extensively to maintain efficiency and data integrity. The Nine Degrees of Freedom (9DOF) mode on the BNO055 also merges data from the accelerometer, gyroscope, and magnetometer to provide precise and stable orientation data, balancing the individual sensors' strengths and limitations.

Visual feedback to the user is provided via the M5capsule's built-in WS2812 RGB lighting, which helps quickly ascertain the hardware's status. In some scenarios, the LED can remain lit continuously to aid in locating the pet and monitoring the device status. Data annotation is performed using an iPhone 13 in conjunction with a Blackmagic Camera (version 1.1), capturing the experiment at 1080p 60 fps in H.265 with an added time code for synchronization.

The M5capsule, powered by an ESP32-s3fn8, supports various development frameworks, with the project utilizing the Arduino framework for simplicity. The Adafruit Sensor libraries drive the BNO055 and BME280 sensors.

During initialization, the sensors are configured with the BNO055 set to the 9DOF mode for easy calibration and low-noise data collection. After acquiring the Network Time Protocol (NTP) timestamp and initializing the microSD card, the device displays the time and enters a ready state. Upon activation by the researcher, a multi-threaded task commences, polling for data on motion states (quaternion and Euler angles) and environmental conditions (temperature and pressure). The data are then appended to a CSV file using a lock-free buffer pool and a dedicated writing thread. Throughout this process, the LED indicator remains to signal the researcher's operational status.

The recording session concludes with a second press of the record button, which prompts the LED to indicate that the recording has stopped; the file is then closed and saved, resetting the device to a ready state.

3.4 Behavior labeling definitions

To clarify the behavioral categories used for labeling the video data, we present the definitions and criteria for each class in Table 1. This table outlines the specific behaviors observed in the cats and the precise definitions used during annotation, ensuring consistency and reliability in the dataset

This detailed classification provides a standardized framework for annotating the behaviors observed in the video data, ensuring that each behavior category is clearly defined and can be consistently identified across different annotators.

4. Algorithm

4.1 Neural network model

In this section, we introduce the architecture of our dual-input neural network model, depicted in Fig. 4. This architecture is specifically designed to process and analyze IMU raw data, leveraging the strengths of both convolutional and recurrent neural networks.

Table 1
Behavior labeling definitions.

Class	Behavior	Definition and Criterion
0	Resting	The cat lies down on all fours with its body in contact with the ground. The head is centrally positioned, eyes are open, and there may be slight movements of the head and ears. The head may turn left or right, but the overall posture remains relaxed and stationary.
1	Moving	The cat is in motion, either walking or running. This includes slow walking with intermittent pauses. Continuous pauses lasting less than 2–3 s are still considered part of the moving behavior. Movement is characterized by coordinated limb activity propelling the cat forward or backward.
2	Sleeping	The cat is lying on the ground with limbs touching the ground. The body may be stretched out or curled up. The head might occasionally tilt to one side or rest on an object. Eyes are closed, indicating a state of sleep. There is minimal to no movement, and the posture is maintained for an extended period.
3	Eating	The cat lowers its head to consume food from a bowl or the ground. This behavior involves rhythmic movements associated with chewing and swallowing. To avoid confusion with other interactions, instances where food was hand-fed to the cat have been mostly excluded from the dataset. Only autonomous eating behaviors are included.
4	Jumping	The cat performs a jumping motion. The action begins with the cat initiating a leap, with the forelimbs leaving the ground first, followed by the hind limbs. The back may arch during the motion. The cat lands smoothly, absorbing impact through the limbs. This includes both jumping up onto objects and jumping down from heights.
5	Standing	The cat is stationary in an upright position with all four limbs extended and touching the ground. The body is supported, but there is no significant locomotion. This category includes periods where the cat is alert and may be observing its environment. Movements that are ambiguous or difficult to classify but involve the cat maintaining this upright posture are included here. The data that cannot be confidently classified is excluded to maintain dataset integrity.
6	Using Litter Tray	The behavior starts when the cat completely enters the litter box and ends when it fully exits. Actions include digging, posturing to urinate or defecate, and covering up waste afterwards. This category focuses on the entire duration spent within the litter box, which is significant for monitoring the cat's health and hygiene practices.

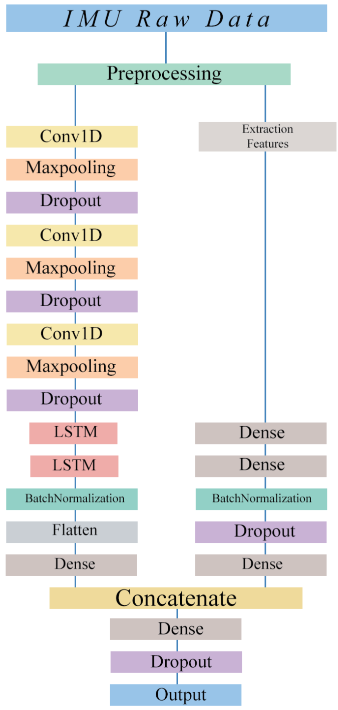


Fig. 4. (Color online) Neural network model structures.

The model consists of two primary branches, as described below.

1) Convolutional Pathway:
This pathway begins with a series of 1D convolutional layers (Conv1D) designed to automatically extract spatial features from the IMU data. Each convolutional layer is followed by a max-pooling layer, which serves to reduce the dimensionality and computational complexity while retaining the most salient features. Dropout layers are incorporated to prevent overfitting, enhancing the model's ability to generalize to unseen data. The extracted features are then passed through LSTM layers, which capture temporal dependencies and patterns within the sequential data. A batch normalization layer is applied to stabilize and accelerate the training process, followed by a flattening layer to prepare the data for dense layer processing.

2) Feature Extraction Pathway:
In parallel, the model processes extracted features through a series of dense layers. Similar to the convolutional pathway, batch normalization and dropout layers are employed to ensure robust learning and prevent overfitting.

The outputs from both pathways are concatenated, allowing the model to integrate spatial, temporal, and manually extracted features. This combined representation is further processed through additional dense and dropout layers, culminating in the final output layer. By structuring the model in this manner, we harness the strengths of CNNs and LSTMs in the convolutional pathway for automatic feature learning and sequence modeling, while integrating the feature extraction pathway. This combined approach provides a comprehensive analysis of IMU data.

4.2 Data preprocessing

The preprocessing pipeline begins by applying a Butterworth low-pass filter to the raw sensor data, effectively attenuating high-frequency noise and preserving the integrity of the relevant signals for subsequent analysis. This step is crucial for reducing the impact of sensor noise and ensuring the integrity of the subsequent analysis. The filter design is characterized by the following transfer function:

$$H(s) = \frac{1}{1+\left(\frac{s}{\omega_c}\right)^{2n}}, \tag{1}$$

where (ω_c) is the cutoff frequency and (n) is the filter order. A fifth-order filter was chosen to achieve a balance between sharp cutoff and computational efficiency.

For the 1DCNN-LSTM model input, we performed time window segmentation with each window set to a duration of 0.5 s and a stride equal to the window length. This configuration, determined through hyperparameter tuning, was found to optimize the balance between capturing rapid movements and computational efficiency.

In our experiments, we also explored the use of overlapping windows to increase robustness. However, we encountered issues with action overlap within the windows, which complicated the data processing.

Given the unpredictable nature of animal movements, we examined the distribution of each action label within the dataset to address class imbalances, thereby enhancing model robustness. To mitigate overfitting due to data redundancy, we introduced synthetic variations to augment the data using a noise injection method defined as

$$x' = x + \mathcal{N}(0,\sigma^2), \tag{2}$$

where (x') is the augmented data point, (x) is the original data point, and [$\mathcal{N}(0,\sigma^2)$] represents Gaussian noise with zero mean and standard deviation (σ). This augmentation method helps in increasing the diversity of the training data, which is particularly beneficial when dealing with limited datasets.

Time windows of 0.5 s were employed to capture rapid movements such as jumps, ensuring no overlap between consecutive windows. This windowing strategy allows for the preservation of temporal dynamics within each segment while maintaining manageable computational complexity.

4.3 Feature extraction

To further enrich the dataset, both time domain and time-frequency domain features were extracted, providing a comprehensive representation of the underlying sensor signals. The time domain features included mean, standard deviation, maximum-minimum difference, and wavelet transform. The CWT and EWT have been extensively applied to time-frequency analysis in various scientific and engineering domains. The use of the CWT with complex Morlet wavelets has been demonstrated to be effective in extracting precise time-frequency characteristics from vibration signals, facilitating the identification and diagnosis of mechanical faults in electromechanical systems.(24)

Wavelet transform was utilized for time-frequency domain features. Specifically, we used the DWT with the Daubechies 4 (db4) wavelet. The choice of db4 is motivated by its ability to provide good balance between time and frequency localization, which is essential for capturing the transient characteristics of animal movements. The wavelet coefficients are computed using

$$W_j = \sum_{t=0}^{N-1} x(t)\psi_{j,k}(t), \tag{3}$$

where (W_j) are the wavelet coefficients, ($x(t)$) is the signal, and [$\psi_{j,k}(t)$] are the wavelet functions at scale (j) and position (k). Statistical features such as mean and standard deviation were then extracted from these coefficients.

To balance the dataset, we applied random oversampling (ROS), ensuring an equal representation of all classes. This technique helps in mitigating the bias towards majority classes during model training. The final dataset was normalized using a standard scaler to ensure uniformity across features. The normalization process can be described as

$$x_{norm} = \frac{x-\mu}{\sigma}, \tag{4}$$

where (x_{norm}) is the normalized value, (μ) is the mean of the feature, and (σ) is the standard deviation. Normalization is crucial for speeding up convergence during model training and ensuring that all features contribute equally.

The final extended features are as follows:

4.3.1 Time-domain features

The time-domain features extracted from each of the nine original sensor signals include the following:

- mean: captures the average signal value over the time window (9 features);
- standard deviation: measures the signal's variability (9 features);
- maximum-minimum difference: represents the range of signal values (9 features).

This results in a total of 27 time-domain features.

4.3.2 Time-frequency domain features

For each of the nine original sensor signals, the DWT was performed up to three decomposition levels. From the wavelet coefficients at each level, two statistical features were extracted, as follows.

- Mean of wavelet coefficients
- Standard deviation of wavelet coefficients

For each original sensor signal, the number of wavelet features is 54. Combining the time-domain and time-frequency domain features yields a total of 81 features resulting in a data shape of (batch size, 81).

5. Evaluation

The project was conducted in accordance with the ethical guidelines of the World Medical Association Declaration of Helsinki for activities involving human subjects and the international standards of animal ethics from the Institutional Animal Care and Use Committee for activities involving animals, while placing value on existing working practices and priorities concerning the production workflow.

5.1 Volunteers and animals used in the experiment

The experiment involved two human participants and three cats, described in Table 2, weighing around 3 to 4 kg. Both participants were over 18 years old and experienced in handling pets ranging from 2 to 6 kg. To assess the cats' temperaments, we conducted interviews with their families—the primary caregivers—who provided comprehensive evaluations of the cats' typical behaviors and temperaments. The pets were not trained to follow commands and showed no stress symptoms when wearing the devices. Participants were provided with a comprehensive explanation of the experimental procedures beforehand, and informed consent was obtained for all activities.

All the cats were healthy, with no known movement-related conditions, and were not under any medication that could affect their behavior or activity levels. Prior to the study, each cat underwent a thorough health check to ensure their well-being, and post-experiment evaluations confirmed no adverse effects from their participation.

During the data collection sessions, the cats were engaged using a variety of stimuli, including toys, laser pointers, and treats, to encourage a range of movements. The cats were allowed to take breaks or withdraw from the study at any point if they showed signs of fatigue or

Table 2
Experimental animals.

Cat ID	Sex	Age	Weight	Breed	Temperament
C1	Male	9 months	4.1 kg	British Shorthair	Somewhat sedentary
C2	Female	14 months	3.9 kg	Domestic Shorthair	Very active
C3	Male	7 months	3.5 kg	Domestic Shorthair	Somewhat active

disinterest, ensuring a stress-free environment. We drew on established canine and feline behavioral studies as references for annotation.[4]

The environment was set up to mimic the cats' natural habitats as closely as possible.[25] The study room was equipped with food, water, fresh air circulation, litter boxes, scratching posts, and various toys, such as cardboard boxes. Continuous video monitoring allowed for the collection of natural behavior data, which were recorded alongside the IMU data.

The experimental protocol was carefully designed to minimize any potential stress for the cats while ensuring the integrity of the data collected. Before each session, the environment was prepared with all the necessary provisions to make the cats comfortable. The sessions were conducted with the cats free to move and interact with the stimuli provided, and the duration of each session was determined by the cats' willingness to participate.

5.2 Data collection

This project requires multi-state classification; hence, data collection was conducted through the recording of induced and natural-state actions. We started by ensuring that the wearable device is securely attached and activated; then, we synchronized the NTP time with the video recording's time code. The participants used pet toys and food to elicit various actions from the pets, such as jumping to reach food placed at elevated locations. The schedule was designed with reference to the experimental procedure outlined in a previous study,[26] where each experimental session lasted a maximum of 5 min, followed by a 5-min rest period.

The IMU data was synchronized with video footage using NTP to ensure accurate time-stamping. This synchronization allowed precise behavior annotation based on the video recordings. Each behavior was annotated within a minimum time window of 1 s, and any behaviors shorter than this or difficult to identify clearly from the video footage were excluded from the dataset to maintain data clarity and reliability.

In addition to induced activities, natural-state data collection involves setting up a playroom equipped with structures of varying heights and necessities such as food bowls and litter boxes. Ceiling-mounted cameras recorded the pets' natural behaviors at 1080p 25 fps to capture comprehensive behavioral data without human interaction.

5.3 Dataset composition

We collected a total of 9 h of data from each participating cat in a completely natural setting. However, one cat exhibited continuous owner-seeking behavior due to its preference for human interaction. To maintain consistency and data quality, we prematurely ended the experiment for this cat and excluded its data from our analysis.

In practical applications, wearable devices often operate at low sampling rates owing to constraints such as battery life and processing power. While many studies have achieved impressive results using high sampling rates, we aimed to explore the effectiveness of behavior recognition under low sampling conditions. By downsampling our own dataset of 5 Hz and employing extended feature extraction methods, we sought to determine if satisfactory performance could still be achieved. The guided behaviors amounted to 2 h and 6 min of recording time. After data cleaning, we segmented the dataset and selected the clearest and most standard portions of the actions for analysis. To expedite the experiment and evaluate the model's performance under conditions of limited data, we focused on a subset of behaviors. The total number of instances for each behavior category is as follows.

- Resting: 19944
- Standing: 16696
- Moving: 8944
- Eating: 8806
- Sleeping: 7396
- Jumping: 1561
- Using litter tray: 181

To balance the dataset, we employed Gaussian noise augmentation, standardizing the number of samples for each behavior category to match that of the most frequent behavior.

The annotation process involved seven labels corresponding to the behaviors listed above. Our research team performed the annotations, consulting with the cat owners to clarify any actions that were difficult to judge. This collaborative approach ensured the accuracy and reliability of the annotations.

5.4 Accuracy of the model

The data were divided into three sets: training, validation, and testing. The training and validation sets constituted 70% of the total data volume, with the remaining 30% designated as the test set, which was not exposed to augmentation in preprocessing steps. The training and validation sets comprised 48.5k samples, while the test set comprised 20.8k samples.

The 1DCNN-LSTM model was trained using categorical cross-entropy loss over 100 epochs with a batch size of 64. The model was optimized using the Adam optimizer with a learning rate of 0.0001. Model performance was evaluated on the basis of validation loss, precision, and recall metrics. The training was conducted on Google Colab's A100 computing units, with a computation time of 10 min for 100 epochs.

The training and validation loss curves shown in Fig. 5 indicate that the model is learning effectively. Both curves demonstrate a consistent decrease in loss over epochs, with the validation loss stabilizing at a lower value than the training loss, suggesting good generalization to unseen data. The accuracy curves exhibit a similar trend, with both training and validation accuracies improving steadily and converging towards a high value of around 89%. This performance suggests that the model is effectively learning to classify the data accurately. The detailed classification report in Fig. 6 shows high precision, recall, and F1 scores across most classes, with an accuracy of 89%. The confusion matrix depicted in Fig. 7 further confirms that the majority of predictions are correct, with minimal misclassifications. These results demonstrate the model's strong learning capabilities and its potential for practical application in similar tasks.

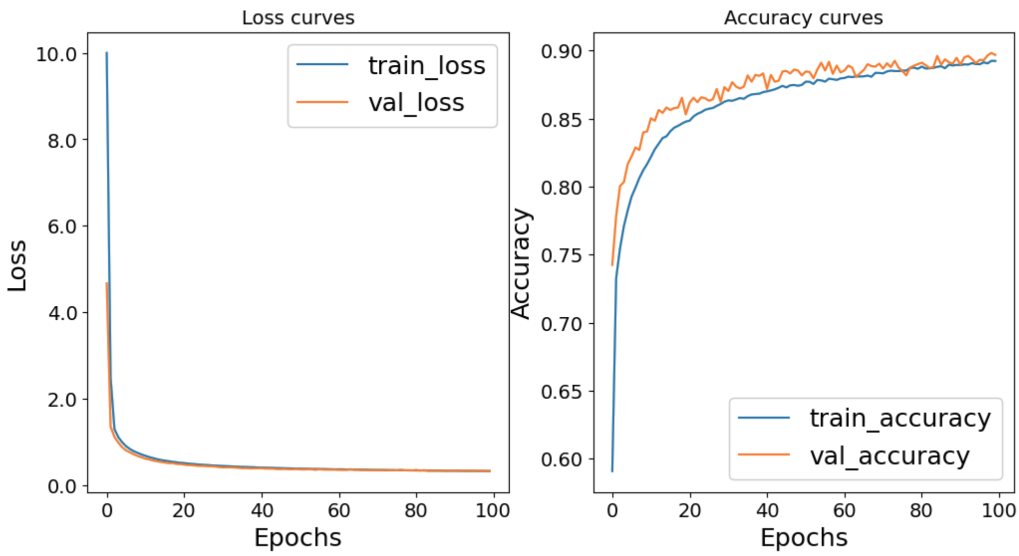


Fig. 5. (Color online) Model in cat dataset training history.

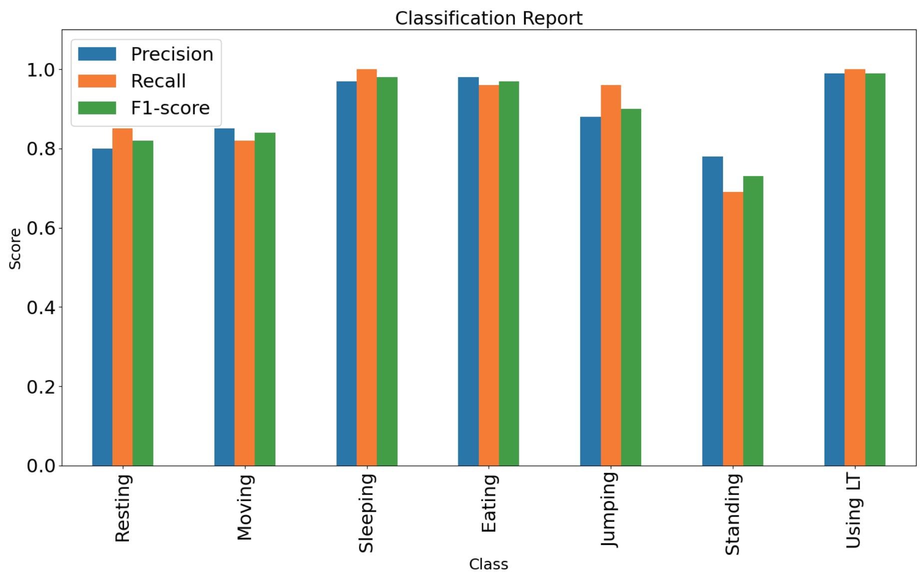


Fig. 6. (Color online) Model in cat dataset classification report.

5.5 Applicability of the model

To test the model's applicability, we employed a publicly available canine dog dataset, which involved IMU sensors from three different positions, operating at 100 Hz and fully labeled. The IMU raw datasets were concatenated as follows, producing 27 data streams: three IMUs (neck, chest, and back) containing three sensors each (accelerometer, gyroscope, and magnetometer) with three axes each (X, Y, and Z).[4] This helped verify our earlier findings about the reduced

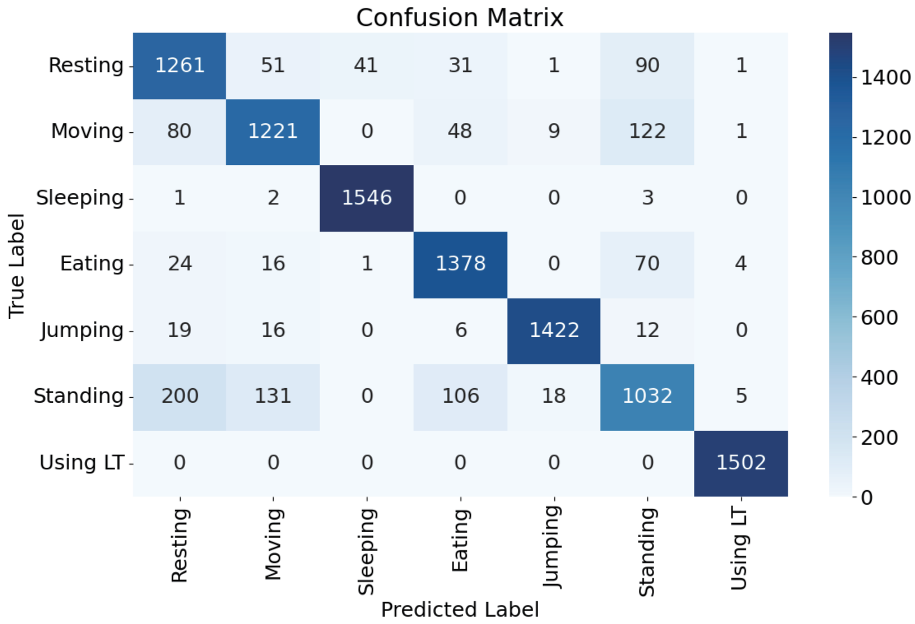


Fig. 7. (Color online) Model in cat dataset confusion matrix.

accuracy of neck-mounted sensors.[21] We used data from the back-mounted IMU as the dog's dataset to input the model. The number of data points in the dog dataset is shown in Table 3.

The training and validation loss curves shown in Fig. 8 indicate that the model is learning effectively. Both curves demonstrate a significant decrease in loss over epochs, with the validation loss stabilizing close to the training loss, suggesting good generalization to unseen data. The accuracy curves exhibit a similar trend, with both training and validation accuracies improving steadily and converging towards a high value of around 94%. The detailed classification report in Fig. 9 and Table 4 shows high precision, recall, and F1 scores across most classes and an accuracy of 94%. The confusion matrix depicted in Fig. 10 further confirms the model's applicability, indicating that the majority of predictions are correct.

To further evaluate the applicability of our model, we conducted experiments using a dataset that includes different dog breeds: Labrador Retriever (LR), Golden Retriever (GR), and a crossbreed of Labrador and Golden Retrievers (LR×GR). We trained and validated our model exclusively on data leave-one-breed-out cross-validation. The classification report shown in Fig. 11 and the confusion matrix depicted in Fig. 12 reveal that while learning is feasible, there is a notable decline in performance. Specifically, confusion arises in the classification of body shaking and sitting behaviors.

Our experimental results validated the framework's ability to accurately process and learn from animal movement data. The model achieved an accuracy of 89% on our self-collected dataset using accelerometer and gyroscope readings, and on accuracy of 94% on a publicly available dog posture recognition dataset.[4] In cross-species tests with similar data settings, it managed an accuracy of 82%, demonstrating the model's applicability. The addition of computed

Table 3
Number of data points in dog dataset.

Behavior	Data point	Instances (Windows)	After augmented instances
Standing	640700	12814	12814
Walking	466502	9330	12814
Sitting	323200	6464	12814
Lying down	235300	4706	12814
Body shake	9300	186	12814

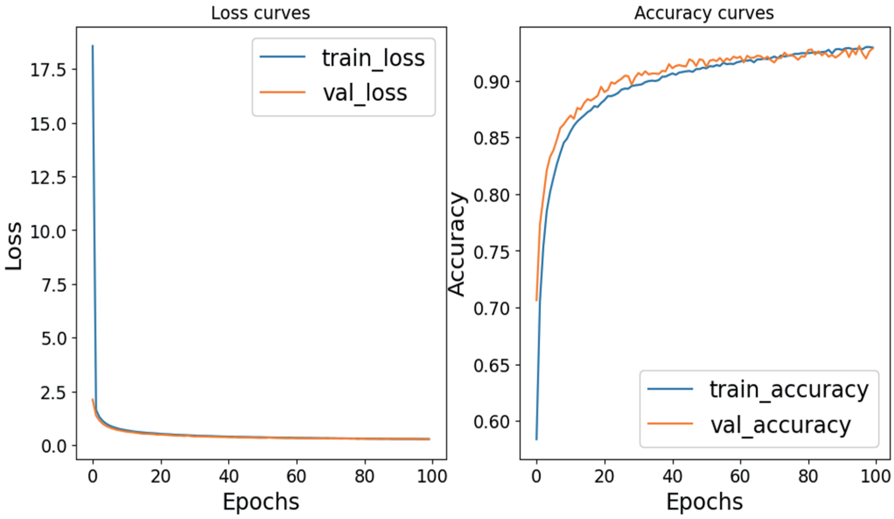


Fig. 8. (Color online) Model in dog dataset training history.

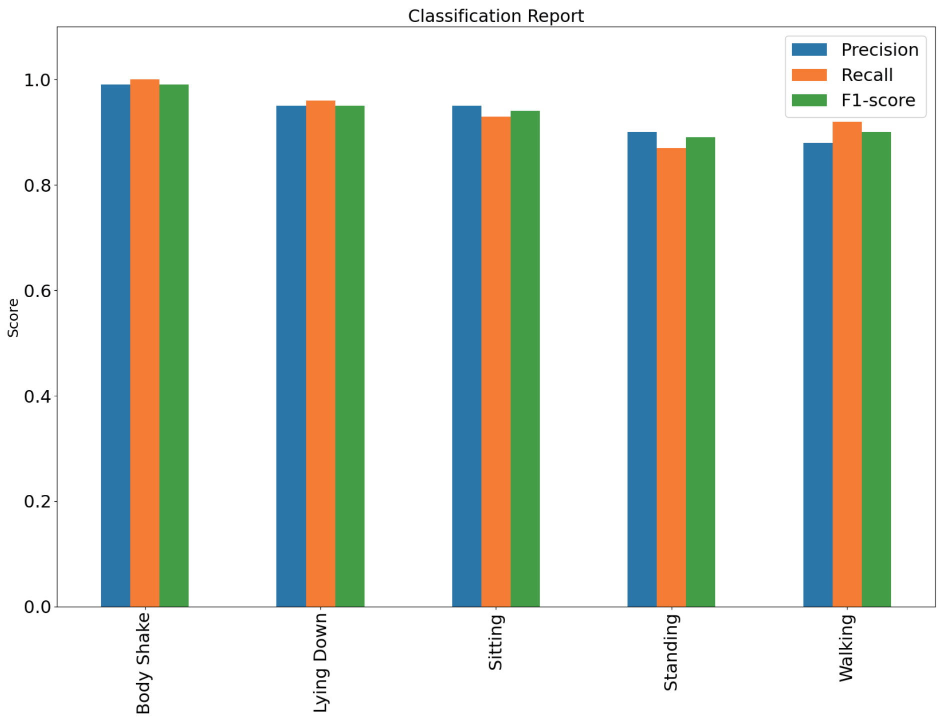


Fig. 9. (Color online) Model in dog dataset classification report.

Table 4
Model in dog dataset.

Behavior	Precision	Recall	F1 score
Body shake	0.99	1.00	0.99
Lying down	0.95	0.96	0.95
Sitting	0.95	0.93	0.94
Standing	0.90	0.87	0.89
Walking	0.88	0.92	0.90

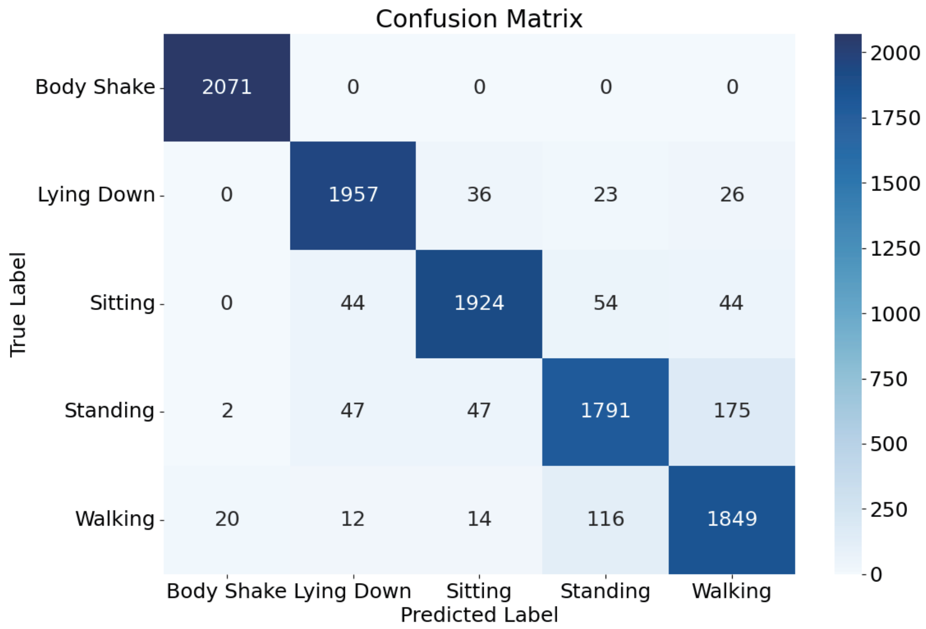


Fig. 10. (Color online) Model in dog dataset confusion matrix.

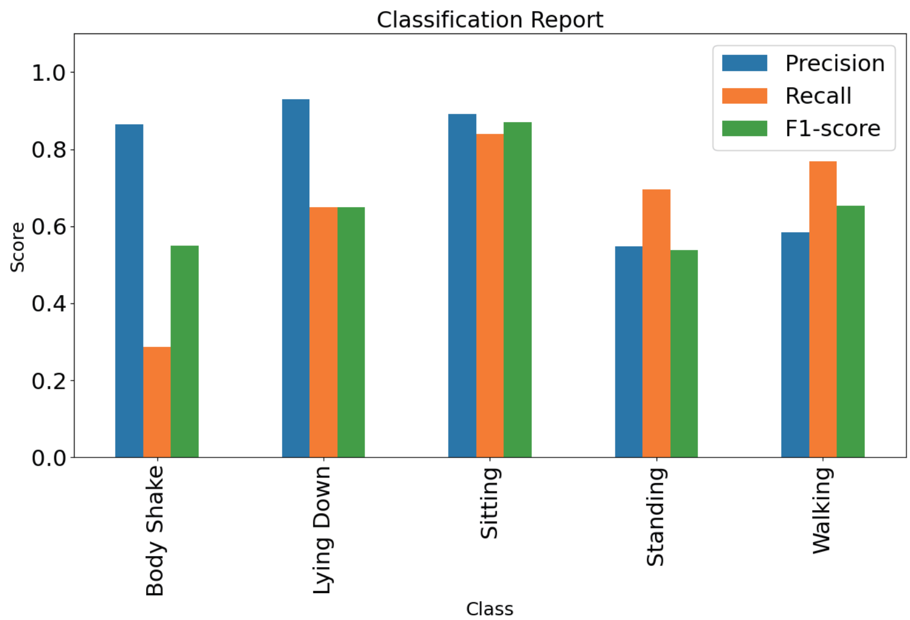


Fig. 11. (Color online) Model in dog cross-breed validation classification report.

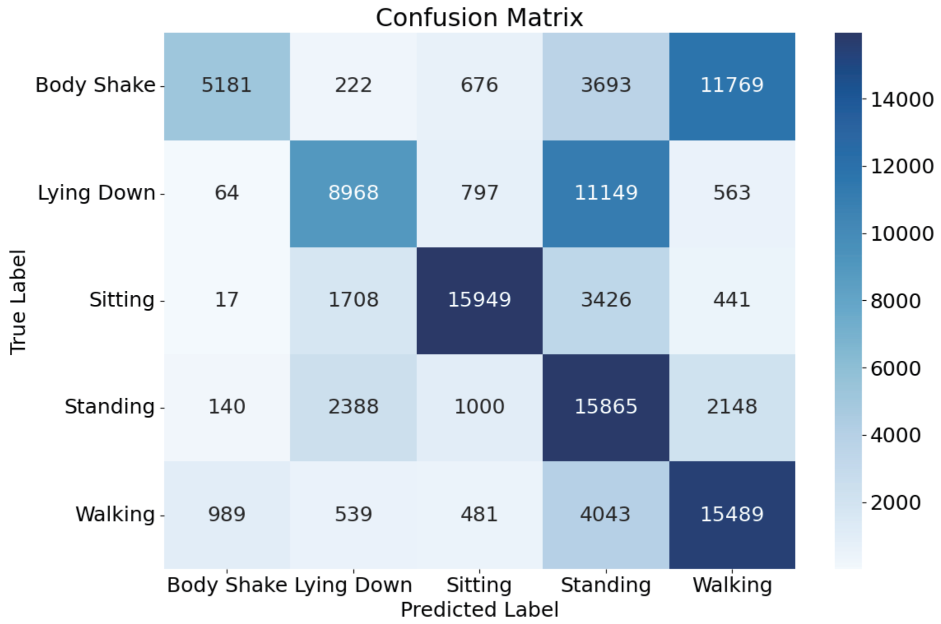


Fig. 12. (Color online) Model in dog cross-breed validation confusion matrix.

features to each time window of sensor data significantly enhanced the model's applicability and accuracy. We used the data from the dogs' back-mounted IMU dataset while the model is being trained and observed the *x*-axis at two different movements.

As illustrated in Figs. 13 and 14, wavelet features of sitting and walking actions for dogs showed clear differences in time-frequency characteristics, corroborating findings from previous studies.[27] Also, when the coefficient level is increased, we can easily find clear differences.

To identify the minimal sensor configuration required without significantly compromising model performance, we evaluated all possible combinations of the three sensor modalities: accelerometer (acc), gyroscope (gyr), and magnetometer (mag). This analysis aims to determine which sensors contribute most to activity recognition accuracy and whether any can be omitted to simplify the hardware setup. We conducted experiments using the seven possible combinations of these sensor modalities. The performance metrics for each configuration are presented in Table 5.

When utilizing individual sensor modalities, the accelerometer alone achieved an F1 score of 0.89, indicating its significant role in capturing motion features relevant to activity recognition. The magnetometer alone had an F1 score of 0.86, showing that it has utility but is slightly less effective than the accelerometer. The gyroscope alone had the lowest performance with an F1 score of 0.78, suggesting limited effectiveness when used in isolation.

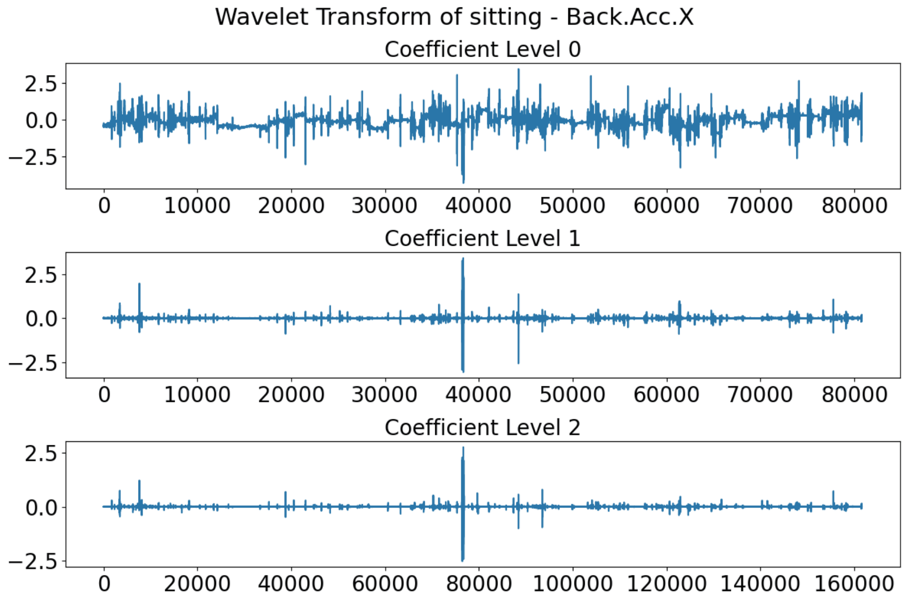


Fig. 13. (Color online) Wavelet transform of sitting.

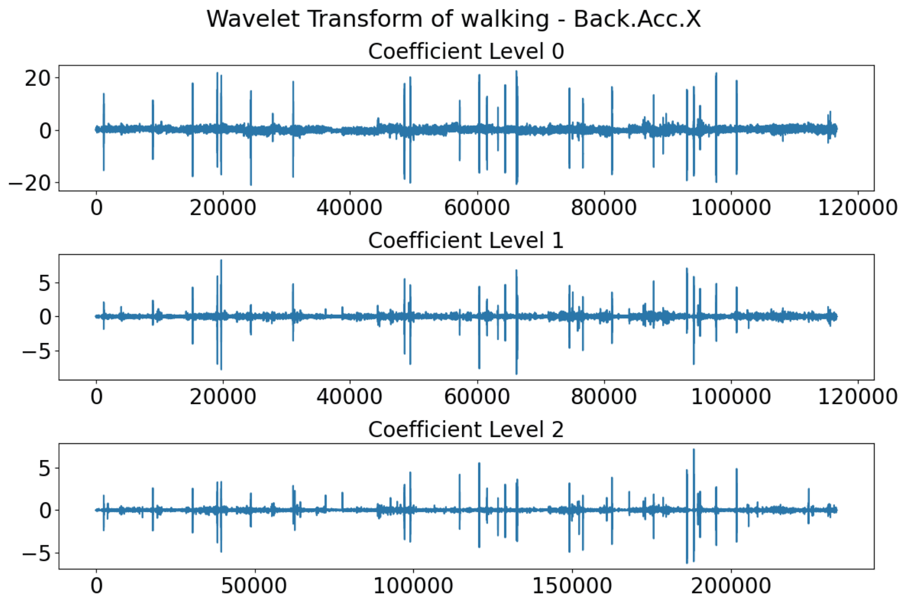


Fig. 14. (Color online) Wavelet transform of walking.

Table 5
Performance metrics for different combinations of sensor modalities.

Sensor	Precision	Recall	F1 score
acc	0.89	0.89	0.89
gyr	0.79	0.79	0.78
mag	0.86	0.86	0.86
acc + gyr	0.91	0.91	0.91
acc + mag	0.94	0.93	0.93
gyr + mag	0.88	0.88	0.88
acc + gyr + mag	0.94	0.94	0.94

6. Discussion

While prior research has focused on either species-specific wearable technologies or basic activity detection, our study bridges these gaps by providing a cross-species behavior recognition system using a novel 1DCNN-LSTM architecture. This system demonstrates robust performance across different datasets, addressing the need for more generalizable models in pet health monitoring.

6.1 Summary of validation results

Our model demonstrated commendable accuracy when tested on both our proprietary feline motion dataset and an open-source canine motion dataset. In noncritical applications, it is common to accept certain levels of error depending on the specific requirements and user context. For instance, in veterinary clinical testing, total allowable error thresholds of 20–30% are deemed acceptable for specific biochemical analyses where high precision is not critical.(28) Drawing a parallel, in our domain, an accuracy threshold of around 70–80% could be suitable for nonprofessional users, such as pet owners, who prioritize usability and convenience. Conversely, professionals such as veterinarians typically demand higher accuracy owing to the critical nature of their decision-making processes. Although these fields differ, the principle of tailoring accuracy requirements to the specific needs of different user groups remains relevant. Our model validation tests have confirmed the efficacy of our data processing and learning workflows.

The current pet market is diverse, with users no longer limited to traditional cats and dogs. Unlike existing commercial products that often focus on specific species or breeds, our data preprocessing extends features, enabling the processing and learning of raw data from a variety of animals. This capability offers researchers a plug-and-play solution.

The study results showed classification accuracies of 89% for cat activities and 94% for dog activities, validating the cross-species applicability of our model. These findings underscore the effectiveness of our preprocessing and feature engineering approaches, which enhance model generalization.

Compared with existing commercial solutions, which primarily target single species, our model's capability to generalize across multiple species aims to address a significant limitation in current pet wearables.

6.2 Animal welfare

According to the World Veterinary Association (WVA), ensuring animal welfare is a fundamental responsibility of animal caretakers. Our research aligns with the five basic principles of animal welfare, enhancing animal nutrition and health through the monitoring of activity and feeding times. The monitoring of environmental and animal body temperatures can improve living conditions, reducing illness and other issues caused by environmental factors. Additionally, assessing activity levels helps determine if animals have adequate living space and playtime. These data enhance the bond between users and their pets, fostering better connections and improving overall animal well-being. Our cost-effective hardware and software solutions promote greater societal attention to animal welfare, benefiting both healthy and sick animals. With data usage permissions, we can further refine the model to improve accuracy.

6.3 Interactivity

To enhance user experience, we developed an interactive web application that allows users to easily monitor and analyze their pets' behavior. The application enables the real-time visualization of predicted activity states, such as sleep, exercise, and rest, through intuitive animations. Users can access key daily metrics, including the duration of each activity, directly from any web-enabled device without the need to install additional software [Fig. 15(a)].

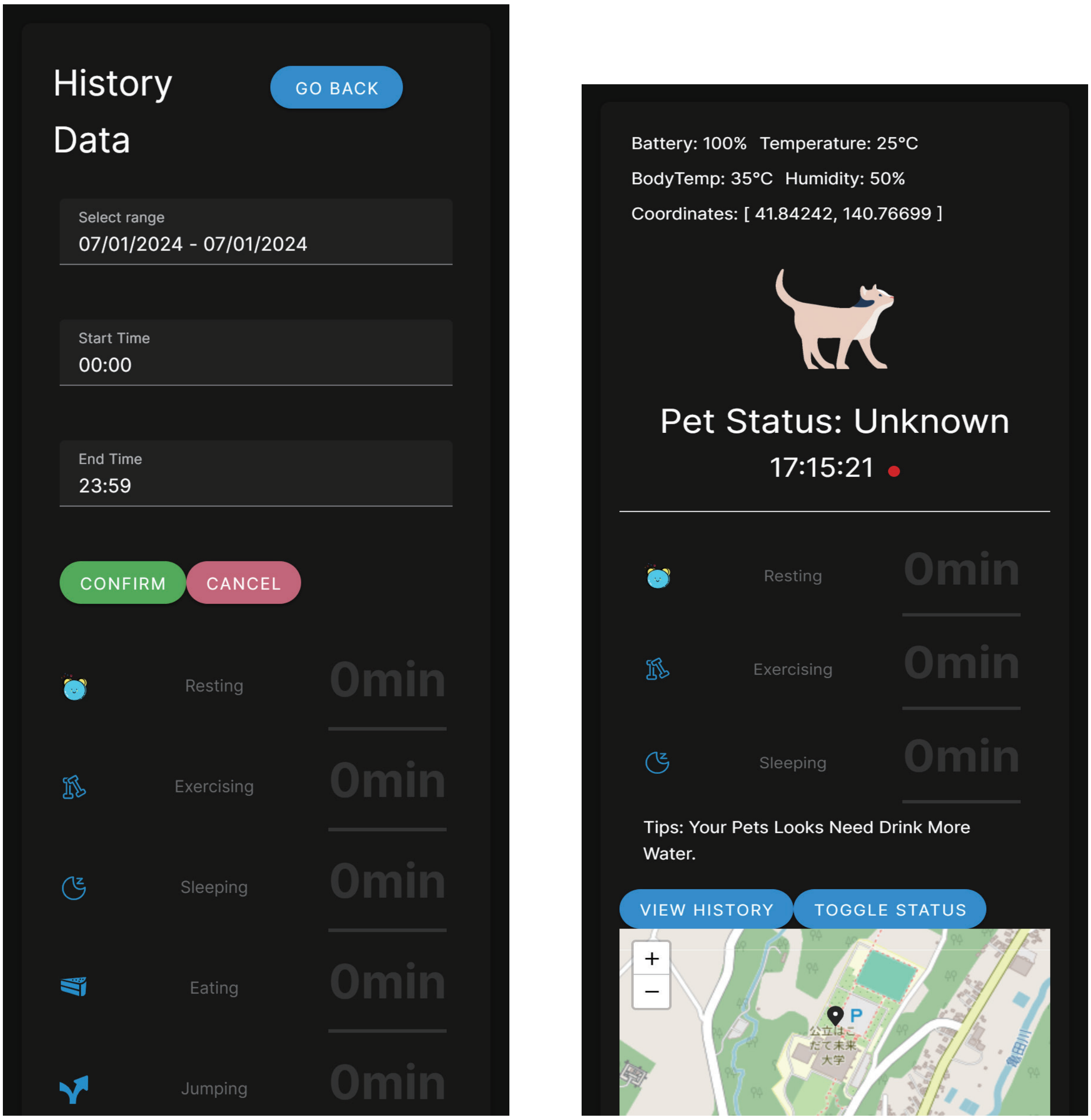
The web app provides detailed status updates on the device, including battery level, temperature, and humidity, helping users keep track of the device performance. A daily report summarizes key insights, allowing users to quickly understand their pets' behavior patterns throughout the day. The app also includes a location tracking function, displaying the pet's current position and allowing for historical data review. By selecting specific time periods, users can view the pet's various activity states and understand trends in exercise volume and rest status over time [Fig. 15(b)]. This functionality supports informed decision-making and enhances the overall pet care experience.

6.4 Limitations and future work

Despite the promising results, the current study is limited by the device size and the controlled environments in which data were collected. Future iterations should explore miniaturization and durability improvements, especially for applications involving larger or more active animals.

Our project currently faces several limitations, such as device size, which can impede animal movement. Achieving long battery life without increasing battery size is crucial, as it minimizes the need for frequent charging, enhancing user experience for both researchers and end-users. The device must also exhibit excellent weather resistance. While our experiments have focused on cats and dogs, expanding to larger animals presents challenges for device durability. Animals can generate significant accelerations of 9–13 m/s^2 (e.g., cheetahs, the fastest land mammals, can reach speeds of 100–120 km/h with accelerations of 9–13 m/s^2 in 3 s[29]). Additionally, playful behaviors such as water interaction in cats pose waterproofing challenges for devices. Future work will address these device limitations through further exploration and development.

Moreover, factors such as the animal's height, leg length, and sensor placement relative to the floor can affect the IMU sensor readings. To address these concerns, we employed the BNO055 sensor in its 9DOF mode, allowing for the collection of absolute orientation data using quaternions. The sensor's built-in magnetometer and accelerometer enable gravity compensation and orientation correction, mitigating the effects of variations in sensor height and positioning. We also worked closely with participants to adjust the elastic straps securing the sensor,



(a) (b)

Fig. 15. (Color online) (a) Real-time page and (b) history data page 2 of web application.

achieving a balance between comfort for the animal and sensor stability. During data collection, particularly during vigorous activities, we monitored the sensor's position and readjusted it as necessary to ensure consistent data acquisition. However, sensor displacement during high-intensity activities remains a potential limitation. Future work should focus on developing custom harnesses for higher sensor stability and exploring methods to further normalize data across animals of different physical characteristics.

Another limitation is that our current model assumes that all relevant behaviors are represented in the training data. In practical applications, the classifier may encounter data streams containing various behaviors not considered during the training phase. This scenario could lead to the misclassification of unknown behaviors as known categories, affecting the overall classification quality and system reliability. During our experiments, we observed that the model exhibits a certain misclassification rate when dealing with unknown behaviors, particularly those that were excluded during the data cleaning phase due to the difficulty of clear definition. This misclassification underscores the need for the model to handle unseen behaviors effectively. To address this, future work should explore open-set recognition techniques, integrate anomaly detection mechanisms, and develop incremental learning frameworks that allow the model to adapt to new behaviors without complete retraining.

Regarding the model and interaction, our current 1DCNN-LSTM neural network structure, while effective, requires optimization to reduce hardware dependence and resource consumption. Although we developed a Vue-based web app for visualizing experimental data, it remains rudimentary, lacking intuitive weekly reports or professional insights. The current web app also struggles with multi-user (multi-pet) environments, unable to simultaneously display or select desired subjects for viewing. This presents a significant limitation in user interaction. Future work aims to enhance user experience through more intuitive data visualization and support for multiple users, such as family group sharing, allowing all family members to access pet status data.

In terms of animal welfare, our data currently comes from healthy animals. We have yet to recruit animals with conditions such as diabetes, renal insufficiency, or arthritis, preventing us from determining if these conditions affect data presentation or if the model can maintain accuracy in predicting actions. Future work should incorporate diverse data, including from FitBark's public datasets, which suggest that dogs with arthritis exhibit different activity patterns. Analyzing large datasets reveals varied exercise preferences among different dogs and regions. Thus, future efforts should focus on collecting diverse data to enhance model applicability and database completeness. In this study, the behavior definitions were based on previous works and participant feedback; however, input from animal experts was lacking. We excluded behaviors that were undefined or ambiguous to ensure clarity and accuracy. In future research, we aim to collaborate with animal experts to expand the range of recognized behaviors. Additionally, we plan to leverage the model's capability for detailed feature extraction to analyze and obtain more comprehensive information about the animals' states. In conclusion, we demonstrated that a well-structured preprocessing pipeline combined with a cross-species adaptable neural network can effectively monitor and analyze pet behavior.

7. Conclusions

In this study, we revealed that observing animal movement states through IMU units is feasible. The application of AI technology for the learning and analysis of data enables the accurate classification and judgment of movement states. With this technology, it is possible to equip pets with intelligent wearable devices similar to those used by humans. These devices are not only inexpensive and lightweight but also enhance the care and attention given to pets while maximizing the value of the data generated by the wearables. Moreover, the extensibility of these wearable devices makes this research applicable not only to animals such as cats and dogs but also to a broader range of species in the future. This can lead to a wider application, yielding more valuable data and contributing to the formation of a positive ecosystem.

Acknowledgments

The authors would like to express their gratitude to Ryuta Okuda and Tatsuya Ishikawa from Fuji Safari Park for their extensive cooperation regarding the design and implementation of the proposed hardware.

References

1 W. Zhang, H. Cao, and L. Lin: Proc. 2022 7th Int. Conf. Financial Innovation and Economic Development (Atlantis Press, 2022) 1682–1689. https://doi.org/10.2991/aebmr.k.220307.275
2 T. Rehn and L. J. Keeling: Appl. Anim. Behav. Sci. **129** (2011) 129. https://doi.org/10.1016/j.applanim.2010.11.015
3 P. M. Taylor and S. A. Robertson: J. Feline Med. Surg. **6** (2004) 313. https://doi.org/10.1016/j.jfms.2003.10.003
4 M. Marcato, S. Tedesco, C. O'Mahony, B. O'Flynn, and P. Galvin: PLOS ONE **18** (2023) 1. https://doi.org/10.1371/journal.pone.0286311
5 D. D. Brown, R. Kays, M. Wikelski, R. Wilson, and A. P. Klimley: Anim. Biotelem. **1** (2013) 20. https://doi.org/10.1186/2050-3385-1-20
6 B. Fan, R. Bryant, and A. Greer: J **5** (2022) 435. https://doi.org/10.3390/j5040030
7 E. L. C. Shepard, R. P. Wilson, F. Quintana, A. G. Laich, N. Liebsch, D. A. Albareda, L. G. Halsey, A. Gleiss, D. T. Morgan, A. E. Myers, C. Newman, and D. W. Macdonald: Endanger. Species Res. **10** (2008) 47. https://doi.org/10.3354/esr00084
8 L. S. Liu, J. Q. Ni, R. Q. Zhao, M. X. Shen, C. L. He, and M. Z. Lu: Biosyst. Eng. **176** (2018) 162. https://doi.org/10.1016/j.biosystemseng.2018.10.011
9 G. Valentin, J. Alcaidinho, A. Howard, M. M. Jackson, and T. Starner: Proc. 2016 ACM Int. Symp. Wearable Computers (ACM, 2016) 100–107. https://doi.org/10.1145/2971763.2971788
10 J. Kestler and M. Wilson: Ph.D. Dissertation, Acceleration Derived Feral Cat (Felis catus) Behaviour During Ground Nesting Bird-Breeding Season on the Island of Schiermonnikoog, Van Hall Larenstein (2015).
11 L. T. Xia, L. Zhang, and F. Wu: Measurement Control Technology and Instruments. **2** (2021) 28. https://doi.org/10.16157/j.issn.0258-7998.200125
12 D. Brown: The Wall Street Journal (Jan. 30, 2023) https://www.wsj.com/articles/when-your-dog-eats-your-apple-airtag-11675091820 (accessed July 27, 2024).
13 FitBark, Inc.: https://www.fitbark.com/ (accessed May 4, 2024).
14 Tractive: https://tractive.com/en/pd/gpstracker-cat (accessed June 4, 2024).
15 Whistle, Inc.: https://www.whistle.com/ (accessed July 4, 2024).
16 H. Ito: 2020 Int. Symp. VLSI Design, Automation and Test (IEEE, 2020) 1. https://doi.org/10.1109/VLSI-DAT49148.2020.9196303
17 M. F. Trujillo-Guerrero, S. Román-Niemes, M. Jaén-Vargas, A. Cadiz, R. Fonseca, and J. J. Serrano-Olmedo: IEEE Access **11** (2023) 106650. https://doi.org/10.1109/ACCESS.2023.3318563
18 J. Xu, Z. He, and Y. Zhang: J. Phys. Conf. Ser. **1267** (2019) 1. https://doi.org/10.1088/1742-6596/1267/1/012044
19 F. J. Ordóñez and D. Roggen: Sensors **16** (2016) 1. https://doi.org/10.3390/s16010115
20 P. Kumpulainen, A. V. Cardó, S. Somppi, H. Törnqvist, H. Väätäjä, P. Majaranta, Y. Gizatdinova, C. H. Antink, V. Surakka, M. V. Kujala, O. Vainio, and A. Vehkaoja: Appl. Anim. Behav. Sci. **241** (2021) 105393. https://doi.org/10.1016/j.applanim.2021.105393
21 W. Goodall, F. Wallace, and M. Jackson: Proc. 10th Int. Conf. Animal-Computer Interaction (ACM, 2023) 1–8. https://doi.org/10.1145/3637882.3637894
22 R. P. Wilson, K. A. Rose, R. Gunner, M. D. Holton, N. J. Marks, N. C. Bennett, S. H. Bell, J. P. Twining, J. Hesketh, C. M. Duarte, N. Bezodis, M. Jezek, M. Painter, V. Silovsky, M. C. Crofoot, R. Harel, J. P. Y. Arnould, B. M. Allan, D. A. Whisson, A. Alagaili, and D. M. Scantlebury: R. Soc. B **288** (2021) 2005. https://doi.org/10.1098/rspb.2021.2005
23 R. E. Kenward: A Manual for Wildlife Radio Tagging, (Academic Press, New York, 2001) 2nd ed.
24 D. Łuczak: Electronics **13** (2024) 1. https://doi.org/10.3390/electronics13020452
25 M. Amat, T. Camps, and X. Manteca: J. Feline Med. Surg. **18** (2016) 577. https://doi.org/10.1177/1098612X15590867

26 A. Hussain, S. Ali, M.-I. Joo, and H.-C. Kim: IEEE Sens. J. **24** (2024) 1996. https://doi.org/10.1109/JSEN.2023.3324665
27 Y. Xue and L. Jin: Proc. Int. Conf. Artificial Intelligence (WorldComp, 2011) 1–5.
28 B. Flatland, M. S. Camus, and R. M. Baral: Vet. Clin. Pathol. **47** (2018) 527. https://doi.org/10.1111/vcp.12649
29 A. M. Wilson, J. C. Lowe, K. Roskilly, P. E. Hudson, K. A. Golabek, and J. W. McNutt: Nature **498** (2013) 185. https://doi.org/10.1038/nature12295