scientific reports

OPEN

Canine gait analysis using inertial sensors and deep learning for orthopedic and neurological disorders

Netta Palez[1]✉, Léonie Straß[2], Sebastian Meller[2], Holger Volk[2], Anna Zamansky[3] & Itzik Klein[1]

Canine gait analysis using wearable inertial sensors is gaining attention in veterinary clinical settings, as it provides valuable insights into a range of mobility impairments. Neurological and orthopedic conditions cannot always be easily distinguished even by experienced clinicians. The current study explored and developed a deep learning approach using inertial sensor readings to assess whether neurological and orthopedic gait could facilitate gait analysis. Our investigation focused on optimizing both performance and generalizability in distinguishing between these gait abnormalities. Variations in sensor configurations, assessment protocols, and enhancements to deep learning model architectures were further suggested. Using a dataset of 29 dogs, our proposed approach achieved 0.96 accuracy in the multiclass classification task (healthy/orthopedic/neurological) and 0.85 accuracy in the binary classification task (healthy/non-healthy) when generalizing to unseen dogs. Our results demonstrate the potential of inertial-based deep learning models to serve as a practical and objective diagnostic and clinical aid to differentiate gait assessment in orthopedic and neurological conditions.

Keywords Inertial sensing, Canine gait analysis, Deep learning

Canine gait analysis is crucial for understanding numerous health conditions. For example, neurological gait disorders can manifest in diverse ways, including ataxia and paresis[1]. Ataxia refers to impaired coordination and may result from dysfunction in the peripheral nerves or spinal cord (proprioceptive ataxia), the vestibular system (vestibular ataxia), or the cerebellum (cerebellar ataxia)[1]. Paresis, characterized by reduced voluntary movement and muscle strength, contrasts with plegia, which denotes a complete loss of voluntary movement. Lameness is more commonly associated with pain due to orthopedic conditions[2], but mild cases of paraparesis[3] and diseases affecting the nerve root(s) can cause the same clinical presentation[3]. Orthopedic lameness is often marked by weight shifting, shortened steps, and a reduced stance phase[4], accompanied by an elongated, compensating stride in the contralateral limb[1].

Traditionally, canine gait analysis is performed by visual inspection and subjective assessment by a trained clinician. While overt lameness is often recognizable through such observation, more subtle gait abnormalities may go undetected without objective tools, making accurate diagnosis particularly challenging in early or mild cases[5]. Besides the inability of the human eye to capture the complexity of each component of movement[6], evaluation still depends on personal experience and education[7]. Moreover, studies have shown that even with objective scoring systems, inter, and intra-observer reliability remains low[8,9]. This highlights the need for more objective methods of gait assessment. In recent years, such approaches have gained growing interest in veterinary gait analysis. Notable examples include video-based kinematic analysis[10] and kinetic analysis using force plates or pressure mats[11]. Although these methods provide greater objectivity and precision, their use is limited by the high cost of equipment and the need for specialized expertise.

The use of wearable inertial measurement units (IMUs) provides an attractive alternative. IMUs are usually equipped with accelerometers and gyroscopes, which can be used to study complex dynamics not captured by kinematic or kinetic analysis[12,13]. Moreover, their compact size, low cost, and fast 3D motion acquisition, make wearable IMU sensors suitable for free-roaming canine gait studies. A growing body of research has explored the

[1]The Hatter Department of Marine Technologies, University of Haifa, Haifa, Israel. [2]Department of Small Animals Medicine and Surgery, University of Veterinary Medicine Hannover, Hannover, Germany. [3]Department of Information Systems, University of Haifa, Haifa, Israel. ✉email: npalez@campus.haifa.ac.il

use of IMUs in canine gait and locomotion analysis, including the characterization of gait patterns[14–17], detection of lameness[18], analysis of locomotion[19], and monitoring of body orientation[20]. Several studies have focused on canine gait abnormalities caused by specific diseases, including dystrophin-deficiency[21], ligament ruptures[22], lack of locomotor-cardiac coupling[23] and ataxia[24].

In most of these studies, a single IMU sensor was mounted on a dog's main body, such as the neck, back, or sternum, which only allows assessment of general body movement. Recent work has begun exploring multi-sensor placements, longitudinal kinematic tracking, and the use of machine learning for automated gait assessment[25–29]. Recently, Zhang et al.[30] investigated a four-limb IMU platform, validating its performance against a commercial pressure-sensor-based walkway system. While this multi-sensor setup offers more precise gait measurements, it may compromise the dog's comfort and potentially alter natural walking behaviour. Recent advancements in hardware and computational efficiency have demonstrated the effectiveness of deep learning (DL) methods in real-time applications, including image processing, signal processing, and natural language processing, by leveraging their ability to handle nonlinear problems[31–34]. Consequently, DL methods have started to be incorporated into inertial navigation algorithms, opening the door for using DL models with inertial sensors in general[35,36]. While deep learning has shown promising results in human gait analysis[37,38], its application to canine gait assessment remains relatively underexplored. The integration of DL with inertial sensor data has the potential to enhance early detection of orthopedic and neurological conditions in dogs, yet existing studies have not fully leveraged this capability.

In this study, we propose a practical inertial deep-learning approach for canine clinical gait analysis for orthopedic and neurological disorders. We employ wearable inertial sensors placed at multiple locations on the dog's body to investigate the differentiation between orthopedic and neurological clinical conditions, which often present with similar gait abnormalities. To this end, we explore the following research questions:

1. What is the optimal sensor configuration—both in terms of number and placement—for distinguishing between orthopedic and neurological conditions?
2. What is the most effective assessment protocol—specifically, which activities should the dog perform, such as walking or running—for optimally distinguishing between orthopedic and neurological conditions?
3. How can deep-learning techniques be applied to inertial data using the sensor configuration and assessment protocols above to accurately differentiate between orthopedic and neurological conditions?
4. How well do the developed deep learning models generalize to unseen dogs?
5. How can the developed deep learning approach performance be further enhanced?

To answer the above research questions, we employ a unique dataset consisting of 29 dogs including 17 healthy dogs, 6 dogs with orthopedic disorders and 6 dogs with neurological disorders. Addressing class imbalance in small canine gait datasets remains an open challenge, and recent studies have demonstrated that structured balancing and augmentation strategies may improve performance in underrepresented gait populations[39]. We develop a simple yet effective end-to-end neural network capable of multiclass classification of different clinical conditions.

The rest of the paper is organized as follows: "Proposed approach" introduces our proposed model, covering both its architecture and the training process. "The dataset" provides an overview of the recorded dataset, a key component of this research. "Analysis and results" presents our experimental results. Finally, "Conclusions" gives the conclusions of this research.

Proposed approach
Our approach leverages a simple, yet efficient, neural network to learn patterns from accelerometer and gyroscope readings and determine if the dog is healthy or has orthopedic or neurological abnormalities. In this section, we detail the architecture of our proposed network, including its layer composition, the unique aspects of our model, and the parameters used during training. We then describe the research methodology.

Network architecture
Our proposed architecture is illustrated in Fig. 1. It consists of two convolutional neural network (CNN) layers followed by three fully connected (FC) layers. The CNN layers incorporate max pooling, while the rectified linear unit (ReLU) activation function is applied to all layers except the final one, which uses log-softmax. To mitigate overfitting observed during development, dropout is employed as a regularization technique. Next we provide a detailed mathematical description of these components.

A FC layer is defined by[40]:

$$\boldsymbol{z}_i^{(k)} = \sum_{j=1}^{n_{k-1}} \boldsymbol{w}_{ij}^{(k)} \boldsymbol{a}_j^{(k-1)} + \boldsymbol{b}_i^{(k)} \tag{1}$$

where $\boldsymbol{w}_{ij}^{(k)}$ is the weight of the *i* neuron in the *k* layer, associated with the output of the *j* neuron in the $(k-1)$ layer ($\boldsymbol{a}_j^{(k-1)}$), $\boldsymbol{b}_i^{(k)}$ represents the bias in layer *k* of the *i* neuron, and n_{k-1} represents the number of neurons in the $(k-1)$ layer. The ReLU activation function is[41]:

$$\mathrm{ReLU}(\boldsymbol{z}_i^{(k)}) = max(0,\ \boldsymbol{z}_i^{(k)}) \tag{2}$$

By assuming a $m_1 \times m_2$ filter (or kernel) the convolutional layer can be written as follows:

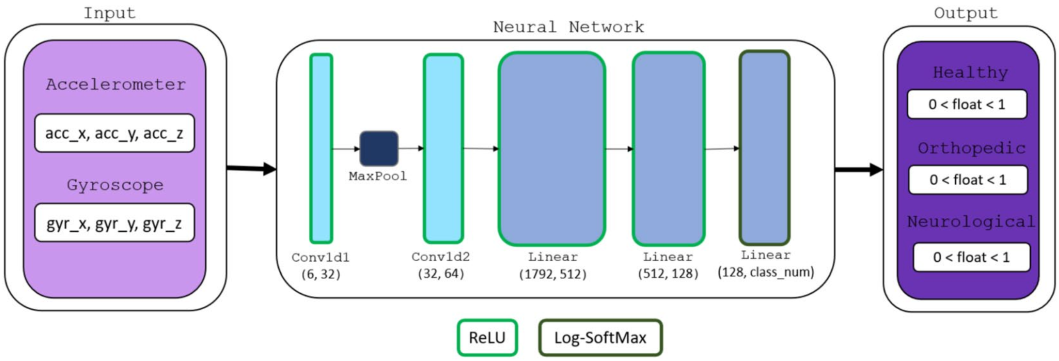


Fig. 1. Proposed neural network for classification (activation functions indicated by box outline).

$$\mathbf{C}_{ij}^{(k)} = \sum_{\alpha=0}^{m_1}\sum_{\beta=0}^{m_2} \boldsymbol{w}_{\alpha\beta}^{(l)} \boldsymbol{a}_{(i+\alpha)(j+\beta)}^{(k-1)} + b^{(l)} \tag{3}$$

where $\boldsymbol{w}_{\alpha\beta}^{(l)}$ is the weight in the (α, β) position of the l convolutional layer, $b^{(l)}$ represents the bias of the l convolutional layer and $\boldsymbol{a}_{(ij)}^{(k-1)}$ is the output of the preceding layer.

Given an input time series sample representing raw inertial measurements:

$$\boldsymbol{x} = [f_x, f_y, f_z, \omega_x, \omega_y, \omega_z]^T \in \mathbb{R}^{6\times n} \tag{4}$$

where f_x, f_y, f_z are the accelerometer readings along the x, y, and z axes, respectively, $\omega_x, \omega_y, \omega_z$ are the gyroscope readings along the x, y, and z axes, respectively, and n is the window size.

The first stage of the model aims for feature extraction, applies two one-dimensional CNN layers:

$$\boldsymbol{h}_1 = \mathrm{ReLU}\left(\mathbf{W}_0\boldsymbol{x} + \boldsymbol{b}_0\right) \tag{5}$$

where $\mathbf{W}_0 \in \mathbb{R}^{6\times 32}$ is the initial weight matrix and $\boldsymbol{b}_0$ is the initial bias vector. In the same manner, the second convolutional layer follows a similar transformation:

$$\boldsymbol{h}_2 = \mathrm{ReLU}\left(\mathbf{W}_1\boldsymbol{h}_1 + \boldsymbol{b}_1\right), \tag{6}$$

where $\mathbf{W}_1 \in \mathbb{R}^{32\times 64}$ represents the kernel weights of the second convolutional layer.

After feature extraction, the resulting feature map is flattened into a vector $\boldsymbol{z} \in \mathbb{R}^d$, where $d = 1792$ is the total number of extracted features. This vector is passed through three FC layers:

$$\boldsymbol{h}_3 = \mathrm{ReLU}\left(\mathbf{W}_2\boldsymbol{z} + \boldsymbol{b}_2\right), \tag{7}$$

$$\boldsymbol{h}_4 = \mathrm{ReLU}\left(\mathbf{W}_3\boldsymbol{h}_3 + \boldsymbol{b}_3\right), \tag{8}$$

$$\hat{\boldsymbol{y}} = \mathrm{LogSoftmax}\left(\mathbf{W}_4\boldsymbol{h}_4 + \boldsymbol{b}_4\right). \tag{9}$$

Here, $\mathbf{W}_2 \in \mathbb{R}^{1792\times 512}$, $\mathbf{W}_3 \in \mathbb{R}^{512\times 128}$, and $\mathbf{W}_4 \in \mathbb{R}^{128\times N}$ are the weight matrices of the FC layers, with the respective hidden layer sizes, and $N = 3$ is the number of output classes. The $\mathrm{LogSoftmax}$ function ensures numerical stability when computing the loss.

Training process

Using Python 3.13.5, our training process was across 60 epochs (was chosen after longer trials that converted until 50 epochs). The negative log-likelihood (NLL) loss function was used in the training process. It measures how closely our model predictions align with the ground truth (GT) labels. The NLL is defined by[40]:

$$L_{\mathrm{NLL}} = \frac{1}{n}\sum_{i=1}^{n}\big(y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)\big) \tag{10}$$

where y_i is the GT value of the ith sample, $\hat{y}_i$ is the predicted value of the ith sample, and n is the number of samples in the corresponding window. We employ the adaptive moment estimation (Adam) for the optimization process. It is a powerful optimizer that combines the strengths of two other well-known techniques—momentum and RMSprop - to deliver a powerful method for adjusting the learning rates of parameters during training[42]. In our model we use Adam with small weight decay of 0.0001 and the learning rate initialization was set to 0.0001 as well. We performed the training process 60 times and used the best model for the testing phase.

Methodology

To explore canine clinical gait analysis for orthopedic and neurological disorders, this study utilized a descriptive and exploratory methodology, combining different setups to provide a multi-faceted perspective. To this end, we distinguish between the following characteristics:

- Sensor locations: four different sensor locations are examined: head, tail, neck(collar) and, all (all the three sensors). In that manner, we can examine how gait features vary between the inertial readings from different locations.
- Assessment dynamics: we examine two different types of dog dynamics during the recording: walking and trotting. The motivation is to examine if some features are more dominant in different dynamics allowing for better classification accuracy.
- Classification task: we consider three practical classification tasks for the problem at hand, namely:

  1. Multi: classification to multiple classes healthy/orthopedic/neurological.
  2. Binary: classification to binary classes healthy/non-healthy (orthopedic + neurological).
  3. Diagnosis: classification to diagnostic classes orthopedic/neurological.

- Generalization: for our baseline models we use a train/test approach based on a random split of the dataset. However, to investigate the generalizability of the developed models to new unseen dogs, we use a stricter leave-one-out (LOO) train/test split, as discussed in[27]. This implies that in each iteration, the model is trained on all dogs except one, which is then used for testing. This process is repeated for each dog, and the aggregated performance metrics are recorded.

The chosen methodology is aimed to provide both breadth and depth in understanding the complex relationships examined.

The dataset

Data collection and preprocessing

Experimental data were collected at the Department for Small Animal Medicine and Surgery of the University of Veterinary Medicine Hanover, Germany. The study was conducted in a controlled environment, comprising an 8.75-m-long testing area. Within this space, dogs simultaneously traversed a pressure-sensing walkway, which was extended at both ends with a sensor-free running surface of the same texture and a non-slip rubber mat, ensuring a smooth transition onto and off the measurement area (Fig. 3). Dogs were led on a leash while ensuring that the handler consistently remained on the same side of the walkway, resulting in each dog being guided from both the left and right side over the course of the trials. Prior to data acquisition, dogs were permitted to ambulate freely within the testing space to facilitate environmental habituation and minimize potential stress-induced variability.

Table 1 is presenting demographic characteristics within each group among our subjects. All subjects underwent a rigorous clinical assessment encompassing general, orthopedic, and neurological examinations by a specialist or specialist in training of the respective EBVS colleges. Furthermore, a subset of neurologically affected subjects underwent magnetic resonance imaging (MRI) to corroborate clinical findings and refine diagnostic accuracy.

Based on the clinical assessment, the three classes for our models are: healthy dogs (no conditions diagnosed), orthopedic dogs (diagnosed with an orthopedic issue in one of their limbs), and neurological dogs (diagnosed with a neurological condition).

Figure 2 present the number of recorded dogs in each class with a total of 29 dogs in the dataset.

With some exceptions, each dog performed two different activities: pace (walking) and trot (running). In the neurological class, only two dogs were able to trot, leading to a significant class imbalance in the trot scenarios. The recordings of all dogs were taken on the same trajectory. It consists of three back-and-forth walks (6 passes) on a 6-m-long walkway, (as shown in Fig. 3). This results in a total trajectory length of 36 meters. However, the duration varies between dogs, due to breed differences and diseases, with some completing it faster than others. The average duration for each class, as well as the total length of the relevant data, is presented in Table 2. As previously described, each recording includes data from three sensors placed on the dog's back and collar. Since the table presents values for a single sensor, the total recorded data is three times the values shown. In total, we collected 193 min of data used for in the model, when 154 min are used for training and 39 min are for testing.

Group	Male	Male-neutered	Female	Female-neutered	Breed distribution	Median age (y)	Median height (cm)	Median Weight (Kg)
Neurologically affected	3	1	1	1	1 mixed-breed, 1 Border Terrier, 1 Tibetan Terrier, 1 Hovawart, 1 French Bulldog, 1 Dachshund	5.45	36.5	12.2
Orthopedically affected	2	0	4	0	2 mixed-breed dogs, 2 Fox Terriers, 1 Kavkazskaya Ovcharka, 1 Tibetan Terrier	2.15	40.0	10.8
Clinically healthy controls	2	5	10	0	10 Beagles, 2 mixed-breed dogs, 2 Fox Terriers, 1 Greater Swiss Mountain Dog, 1 Shar Pei, 1 Parson Russell Terrier	2.3	42.0	12.5
Total	7	6	15	1	-	3.3	40.0	12.2

Table 1. Demographic characteristics of the participant dogs.

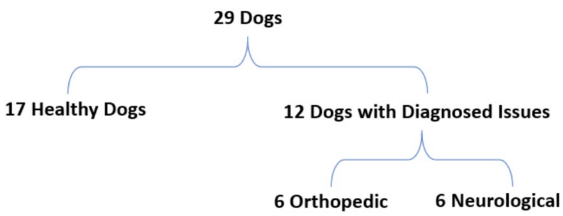


Fig. 2. Number of dogs in each class.

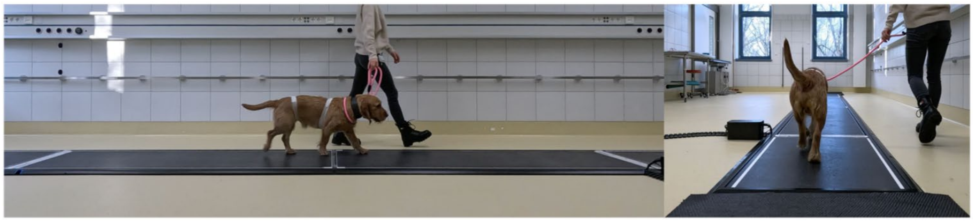


Fig. 3. Experimental setup in the gait clinic.

	Walk		Trot	
	Avg (s)	Total (s)	Avg (s)	Total (s)
Healthy	27	1356	18	859
Orthopedic	33	598	18	274
Neurological	40	679	15	91
Total [s]	–	2633	–	1224

Table 2. Average and total trajectory duration (in seconds) for each class and gait type.

Gyroscope		Accelerometer	
Bias ($^\circ$/h)	Noise ($^\circ$/s/$\sqrt{\text{Hz}}$)	Bias (mg)	Noise (μg/$\sqrt{\text{Hz}}$)
10	0.007	0.03	120

Table 3. Movella DOT IMU sensor specifications.

Sensor configurations

We used the Xsens Movella DOT IMU, a compact, wearable sensor designed for high-resolution motion tracking. The DOT operates at a sampling rate of 120 Hz, which enables accurate capture of dynamic movements[43]. Its accompanying software that supports time-synchronized recording across multiple IMUs, ensuring temporal consistency between sensor streams. Table 3 summarizes the key performance specifications of the gyroscope and accelerometer embedded in the device, including their respective bias and noise levels.

Figure 4 presents an example of the inertial raw data for healthy, orthopedic, and neurological dogs.

The sensor placements in this study build upon previous canine research where comparable configurations proved effective.[44] demonstrated that reflective markers positioned on the head and pelvis sensitively capture asymmetrical vertical displacements associated with fore- and hindlimb lameness.[18] further confirmed that IMU sensors mounted at these sites reliably detect induced lameness through alterations in vertical symmetry, with deviations in the lowest positions indicating weight-bearing and in the highest positions indicating swing-phase lameness. A study,[24], employing a smartphone with integrated inertial sensors on the cranial back successfully differentiated healthy from ataxic dogs, underscoring this region's suitability for detecting central nervous gait abnormalities. Because direct head placement can cause discomfort, ventral neck attachment via a collar offers a practical alternative that accurately reflects head-associated motion, while leg-mounted sensors were avoided as they tend to alter the natural gait. Additionally,[25] reported body-size–dependent differences in cervical and pelvic motion patterns, supporting the relevance of neck-level placement. Figure 5a illustrates the locations for placing the sensors chosen for minimization of discomfort for the tested dogs. We denote the three different locations of the sensors on the dog as follows: (1) head - top back, (2) tail - lower back, (3) neck - collar. Thus, for each dog, six different combinations of sensor location and walking type were tested. The correspondence between the canine anatomical planes and the sensor reference frame is defined relative to the IMU local earth-fixed

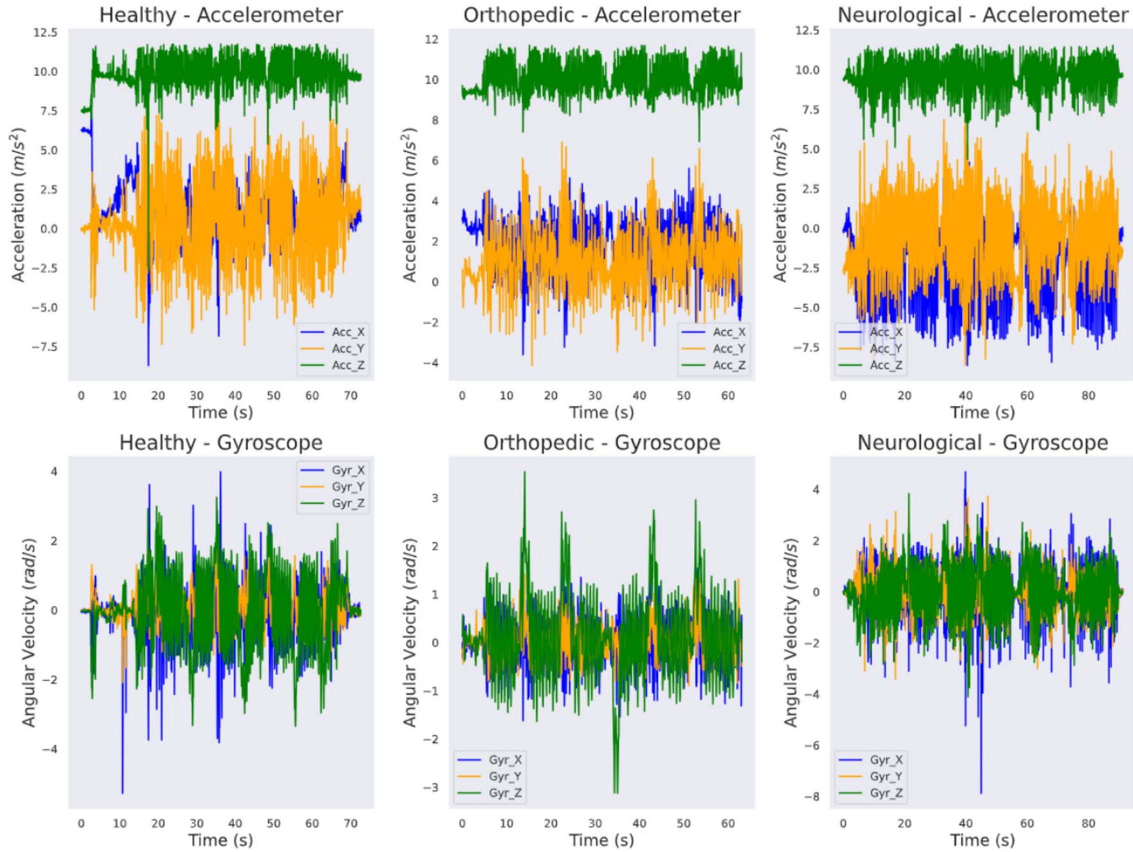


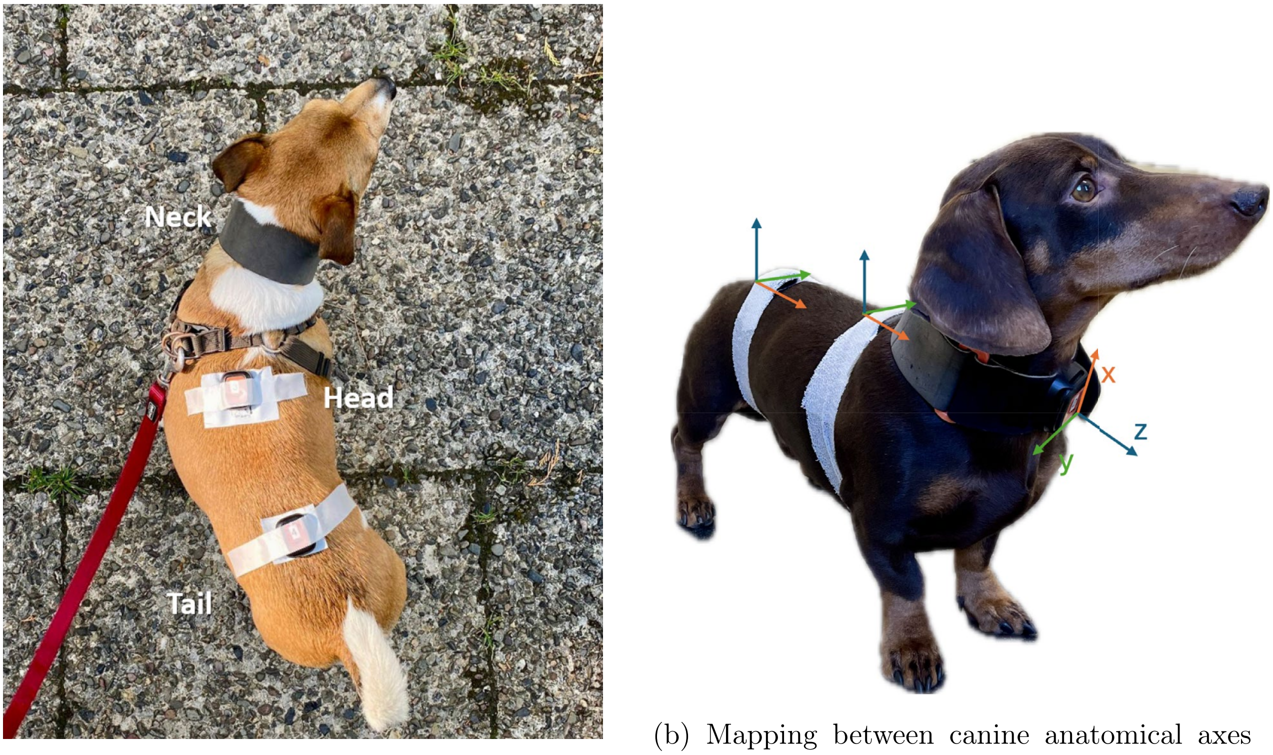
Fig. 4. Raw IMU data of the different axes for the 3 classes.

East-North-Up (ENU) coordinate system. In the canine body, the sagittal plane corresponds to motion along the craniocaudal axis, the transverse plane to dorsoventral movements, and the frontal plane to mediolateral displacements. Accordingly, when the sensors were mounted in alignment with the animal's longitudinal axis, the sensor's x-axis corresponded approximately to the craniocaudal (rostrocaudal) direction, the y-axis to the mediolateral direction, and the z-axis to the dorsoventral direction. In the default ENU frame of the Xsens system, the x-axis (East) therefore represents forward–backward motion of the dog, the y-axis (North) side-to-side motion, and the z-axis (Up) vertical displacement. Those reference frames are illustrated in Fig. 5b. It should be noted that the orientation of the sensor axes remains fixed relative to the sensor and the dog's body segment; only their orientation with respect to the global ENU frame changes during motion, affecting the direction and sign of the measured values.

The raw data was preprocessed, filtering noise (e.g., prior to the dog's walking) by analyzing the norm of the accelerometer measurements and determining a fixed window and stride size, which allows the data signal to be analyzed as a multivariate time series. We experimented with different window sizes and strides to find the optimal configuration. After a prior analysis we chose a window size of 120 and a stride of 5. Hyperparameters, including window length and stride, were selected empirically by evaluating multiple configurations (window sizes between 10–300 samples and stride values between 2–100) and choosing the combination that provided the best performance on held-out validation subjects while minimizing overfitting.

Analysis and results

We began our analysis by evaluating the performance of our approach using a known group of dogs that participated in our data collection process (baseline). Next, we examined the generalizability of our approach by extending the evaluations to unseen dogs. Lastly, we applied various techniques to improve the performance of our generalized model. In this section, we detail these steps, present the corresponding results, and summarize the most effective approaches based on the obtained outcomes. The evaluation metrics used in this section include accuracy, confidence interval accuracy, F1 score, precision, and recall. These measures collectively capture the model's performance in correctly identifying true positives (TP) and true negatives (TN). Specifically, precision quantifies the proportion of true positive predictions among all predicted positives, indicating how many of the model's positive classifications are correct, whereas recall represents the proportion of true positives among all actual positives, reflecting the model's ability to detect all relevant cases. Together, these metrics complement accuracy and F1 score by providing a more comprehensive assessment of the model's sensitivity and specificity in distinguishing between classes.



(a) Placement of the IMUs on the dog.

(b) Mapping between canine anatomical axes and IMU ENU sensor axes.

Fig. 5. Overview of IMU locations and anatomical axis mapping.

Baseline models

The first scenario focused on a closely monitored group of dogs without the need to generalize to additional dogs outside this group. As such, it involved a train/test approach based on a random split of the dataset with 80% of the data using for training and 20% for testing. In this setup, we utilized the previously described network architecture (Fig. 1) and applied it across the three classification scenarios: Multi, Binary, and Diagnosis, as described in "Methodology". The network architecture remained identical in all cases, with the only difference being the number of output units in the final layer—set to either two or three, corresponding to the classification task. The metrics we used to evaluate performance were accuracy and F1 score.

Binary classification

This scenario aimed to answer the following question: could we distinguish between healthy individuals and those exhibiting signs of illness? Figure 6 presents the classification performance for the binary classification task across the different walking protocols and sensor locations. The results showed consistently high metrics with the best performance observed using the neck sensor location, regardless of protocol type (walk/trot), reaching 0.94–0.96 across all scores. The narrow and stable confidence intervals further confirm the robustness of this performance (Fig. 7).

Diagnosis classification

In this scenario, we focused exclusively on dogs diagnosed with an illness and aimed to differentiate between orthopedic and neurological issues. Figure 8 presents the model's performance on the orthopedic vs. neurological classification task across various sensor locations and protocols. All configurations achieved strong performance across all metrics, with slightly better results observed in the trot scenario. It is important to note, however, that this case is highly unbalanced, as most of the neurological dogs were unable to trot.

Multiclass classification

In this scenario, we aimed to answer the following question: given a group of dogs without prior diagnostic information, could we accurately determine whether each dog was healthy, or if not, whether the underlying issue was orthopedic or neurological? Table 4 presents the performance of the baseline model using sensor data as input in the multiclass classification task. The results showed consistently high performances, reaching up to around 0.96 across all metrics (trotting protocol with the neck sensor). Notably, the precision and recall values were nearly identical across configurations, indicating that the model maintains a well-balanced ability to correctly identify positive cases without inflating false positives. Even the lowest-performing configuration (walking protocol with the Tail sensor) still achieved 0.917 accuracy, demonstrating the model's robustness across settings.

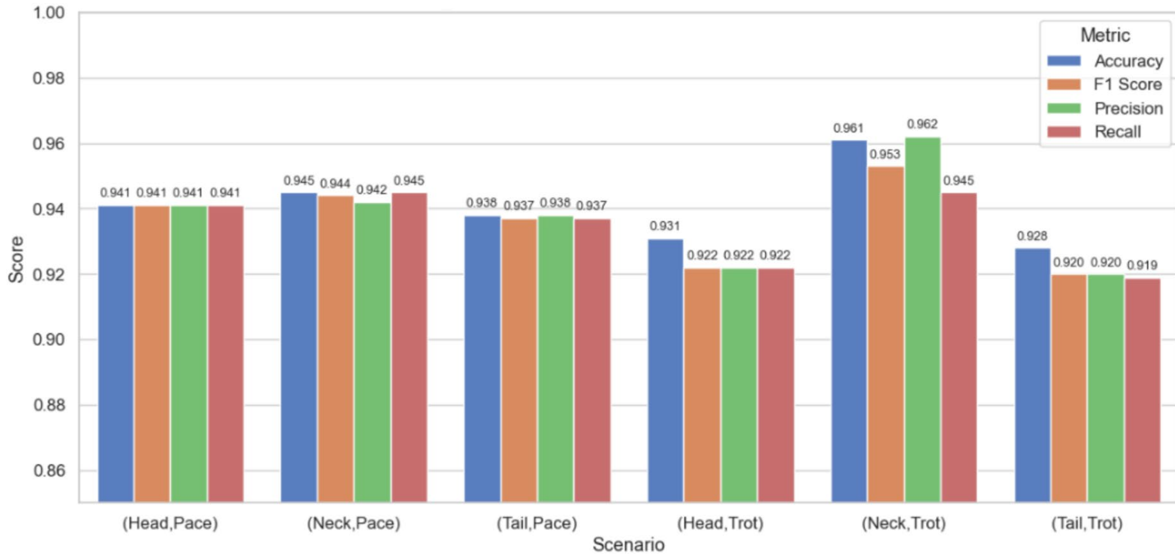


Fig. 6. Binary classification accuracy, F1, precision, and recall results across sensor locations and protocols for random split.

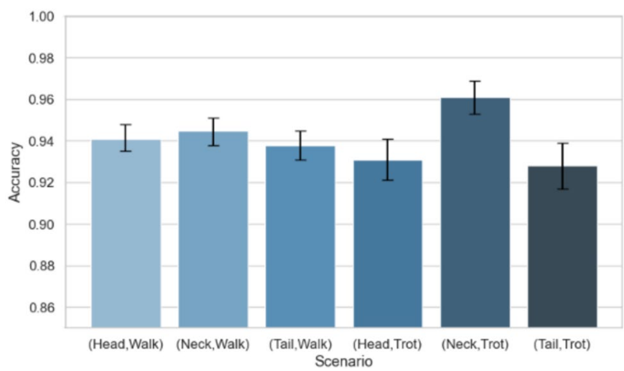


Fig. 7. Binary classification average accuracy with 95% CI for random split.

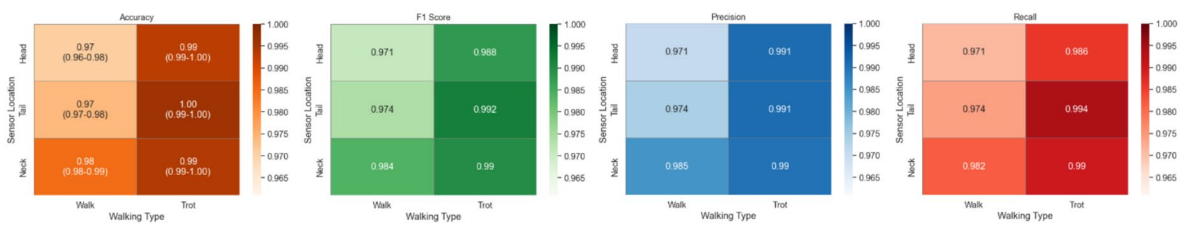


Fig. 8. Diagnostic classification performance.

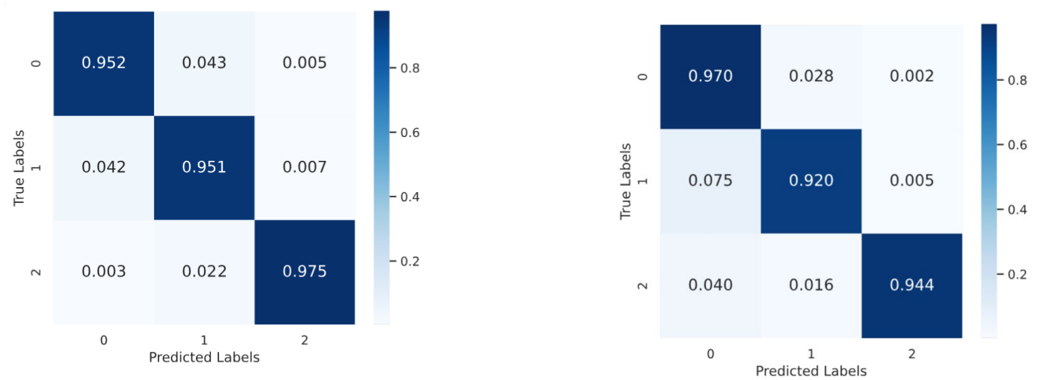
Figure 9 provides further insight into multiclass classification performance by showing the confusion matrices for the best-performing sensor placement (Neck). In both walking and trotting, the model demonstrated strong classification ability with minimal confusion between classes. Misclassifications were rare and occurred primarily between the orthopedic and neurological classes, which were naturally more similar.

Generalization

Here, we addressed the generalization issue - specifically, whether the results could be extended to any unseen dog not included in the training dataset. Generalization performance was estimated using a leave-one-out (LOO) cross validation scheme: in each iteration, the model was trained using data from all dogs except one, and the held-out dog served exclusively for testing. This procedure prevented information leakage at the animal level,

Sensor placement	Protocol	Accuracy	Accuracy CI low	Accuracy CI high	F1	Precision	Recall
Head	Walk	0.922	0.914	0.930	0.922	0.923	0.922
	Trot	0.944	0.935	0.953	0.931	0.938	0.925
Neck	Walk	0.956	0.949	0.963	0.959	0.959	0.959
	Trot	0.956	0.947	0.964	0.946	0.948	0.945
Tail	Walk	0.917	0.907	0.926	0.917	0.916	0.917
	Trot	0.923	0.912	0.935	0.919	0.915	0.923

Table 4. Multiclass classification performance results across sensor configurations: random split.



(a) Confusion matrix for (Neck, walking). (b) Confusion matrix for (Neck, trotting).

Fig. 9. Confusion matrices for neck sensor placement: (**a**) walking, (**b**) trotting.

since no windows from the test dog were present in the training set. Final results reflect the mean of the metrics computed across all LOO iterations. As an initial step in this approach, we performed the same evaluations as those conducted for the closely monitored group, as presented in "Baseline models".

Binary classification
Figure 10 presents the classification performance for distinguishing between healthy dogs and those with diagnosed issues under the different scenarios, with Fig. 11 adding a clear view of the accuracy confidence intervals. Overall, the best performance was observed with the neck sensor, particularly during walking, achieving the highest F1 score (0.82), indicating that the collar-mounted sensor was a strong choice for this task. A clear performance gap was observed between walking and trotting conditions across all sensor locations, with consistently lower recall and F1 scores during trotting. This difference is likely influenced by the imbalance in the trotting data, as several sick dogs were unable to perform the trotting protocol, reducing the model's ability to capture all affected cases. Across all scenarios, precision values were consistently higher than recall, indicating that the model made few false-positive predictions, while recall variability reflected the difficulty of identifying all true cases under more constrained motion conditions.

Diagnosis classification
Figure 12 presents the classification performance for distinguishing orthopedic from neurological cases across the three sensor locations and two walking types. The best performance was achieved using the tail sensor during the trotting protocol, with an accuracy of 0.89 (and a relatively small confidence interval) and an F1 score of 0.84. In contrast, the lowest performance was also observed with the tail sensor during the walking protocol, yielding an accuracy of 0.64 (with a substantially larger confidence interval) and an F1 score of 0.65, indicating a strong dependence of this sensor location on the walking condition. While trotting consistently resulted in higher accuracy across all sensor placements, its effect on F1 score varied by sensor location. Specifically, trotting improved the balance between precision and recall for the tail sensor, but led to reduced F1 scores for the head and neck sensors, primarily due to decreased recall. This pattern suggests that although trotting motion enhances overall class separability, it may suppress diagnostically relevant cues for certain sensor locations. Overall, these findings highlight the importance of jointly considering sensor placement and locomotion protocol when differentiating orthopedic from neurological conditions.

Multiclass classification
Table 5 shows the multiclass classification results for generalization across six scenarios that combined walking types and sensor placements. The best overall performance was achieved using the neck sensor during the trotting protocol, with 0.80 accuracy. In contrast, the lowest performance was observed with the tail sensor during the

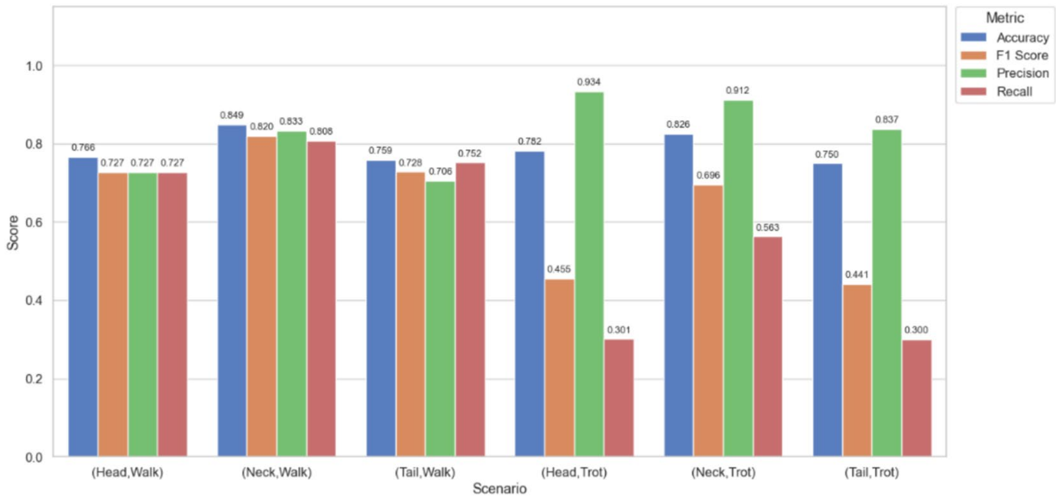


Fig. 10. Binary classification accuracy, F1, precision, and recall results for our model generalization.

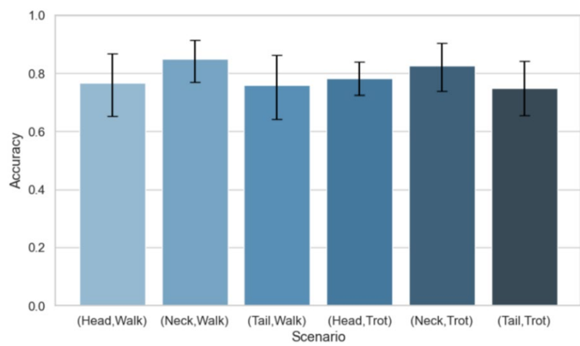


Fig. 11. Binary classification average accuracy with 95% CI.

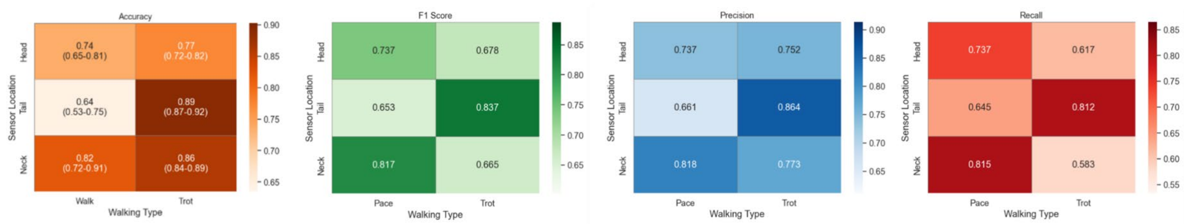


Fig. 12. Diagnostic classification performance, including accuracy (with 95% CI), F1, precision and recall.

walking protocol, yielding 0.612 accuracy. Across all configurations, precision values were moderate rather than uniformly high, ranging between approximately 0.50 and 0.74, indicating that the model did not overly favor conservative predictions in the multiclass setting. Recall values were comparable to precision, suggesting a more balanced trade-off between false positives and false negatives compared to the binary classification scenarios. This balance reflects the increased difficulty of the multiclass generalization task, in which distinguishing between healthy, orthopedic, and neurological conditions reduced both specificity and sensitivity when applied to unseen dogs.

Summary

While the model achieves high accuracy when evaluated on data from dogs included in training, performance decreases when applied to previously unseen dogs (e.g., from 0.96 to 0.85 accuracy in binary classification). This drop highlights the natural variability in gait patterns across individual dogs, including differences in size, morphology, and movement strategy. Although the architecture was designed to be lightweight and to minimize overfitting, limited dataset size and diversity may constrain generalization. These results emphasize

Sensor placement	Protocol	Accuracy	Accuracy CI low	Accuracy CI high	F1	Precision	Recall
Head	Walk	0.641	0.532	0.744	0.635	0.630	0.641
	Trot	0.726	0.688	0.768	0.546	0.497	0.615
Neck	Walk	0.751	0.654	0.837	0.738	0.735	0.746
	Trot	0.800	0.728	0.862	0.585	0.535	0.650
Tail	Walk	0.612	0.497	0.723	0.604	0.600	0.612
	Trot	0.729	0.665	0.793	0.572	0.538	0.626

Table 5. Multiclass classification performance results for the generalization scenario.

Classification task	Setting	Best sensor placement	Best walking protocol	Accuracy	F1 score
Binary	Baseline	Neck	Trot	0.961	0.953
Binary	Generalization	Neck	Walk	0.849	0.820
Diagnosis	Baseline	Tail	Trot	0.996	1.000
Diagnosis	Generalization	Tail	Trot	0.893	0.837
Multi-class	Baseline	Neck	Trot	0.960	0.951
Multi-class	Generalization	Neck	Trot	0.800	0.585

Table 6. Summary of the main classification results across tasks and settings, highlighting the best-performing sensor-protocol combination by accuracy.

the importance of collecting larger, more diverse datasets and suggest that performance on new subjects may improve with additional training data.

Table 6 summarizes the most significant results of this research. It is important to note that several dogs, primarily those in the affected groups, were unable to trot. Therefore, while identifying the configurations that yield the best results is valuable, it is equally crucial to balance performance with practicality-ensuring that every dog can benefit from a simple, accessible, and efficient model. In the closely monitored group approach, results across all cases are consistently strong and comparable. Therefore, the most reliable and convenient choice is the multi-class classification using a single collar-mounted sensor (Neck), which achieves excellent performance across all the different metrics with scores of around 0.95.

Conclusions

Wearable IMU systems have increasingly been used to quantify canine gait, primarily focusing on stride analysis, kinematic features, or simple lameness detection. While these studies have demonstrated the feasibility of using sensors for clinical monitoring, most rely on handcrafted features or traditional machine-learning models, and few attempt to classify complex gait abnormalities, such as distinguishing orthopedic from neurological conditions. The broader field is moving toward the use of multiple sensor sites, longer-term kinematic evaluation, and data-driven methods for automated gait interpretation. Building on these efforts, our study applies deep learning directly to raw multichannel accelerometer and gyroscope signals, enabling automated classification without extensive feature engineering. We systematically evaluated multiple sensor placements and walking protocols. Our results show that the model can reliably distinguish between healthy, orthopedic, and neurological gait patterns, highlighting both the practical potential and the generalizability of deep learning approaches in real-world veterinary applications, and provides a framework for reliable and deployable AI-assisted gait diagnostics, addressing limitations noted in prior studies and positioning our approach as a reference for future research in the field.

The results of this study demonstrate the strong potential of our model to classify non-healthy dogs from healthy ones, as well as to distinguish whether the cause of the problem is neurological or orthopedic. We can extract sufficient insights into the dog's gait patterns by using a single sensor placed comfortably on the upper back or neck of the dog to collect inertial readings (accelerometer and gyroscope) while walking the dog on a straight path. In preliminary experiments, we compared the proposed deep-learning model with traditional machine-learning approaches including support vector machines, random forests, logistic regression and others. Across both binary and multiclass settings, their classification performance was lower than that of our 1D CNN-based model. Notably, despite being a deep learning method, our model remains lightweight, with relatively few layers, suggesting that modest architectural complexity is sufficient to capture the relevant biomechanical patterns. Thus, although more interpretable models may be appealing, deep learning offered measurable performance benefits in this context.

Specifically, for the Neck-mounted sensor at a Walk, our model achieved superior performance in both the diagnosis classification (accuracy of 0.82 vs. 0.766) and multiclass classification (accuracy of 0.757 vs. 0.713). The performance gap was most pronounced in the binary Healthy vs. Sick classification using the Tail/Walk scenario, where the deep learning model achieved an accuracy of 0.759, compared to just 0.67 for the traditional baseline. While traditional methods remained competitive during trotting (where signal amplitude is higher), our model proved essential for detecting the subtler, non-linear gait abnormalities present at a walk, which manual feature engineering failed to fully capture. In the Baseline case, where our model is applied to a known and fixed group of dogs, it achieves high performance across all tasks, with up to 0.96 accuracy in the Multi-class classification. This setting is simplifies clinical decision-making by reducing the need for extensive diagnostic procedures. In the more challenging Generalization case, where the model is evaluated on new, unseen dogs, it still shows strong results: achieving 0.85 accuracy in the Binary classification (healthy vs. non-healthy) and 0.82 accuracy in the Diagnosis classification (orthopedic vs. neurological), when dogs are walking and wearing the Neck sensor. This demonstrates the model's potential as a practical diagnostic support tool.

While our study demonstrates the feasibility of using deep learning directly on raw IMU signals for automated gait classification, several limitations should be noted. The dataset comprises a limited number of dogs and varying breeds, which may restrict the generalizability of the model to broader canine populations or diverse gait patterns. In addition, class imbalance may bias the model toward majority classes, potentially reducing performance for underrepresented categories. Deep learning models, while effective at capturing temporal patterns and multichannel interactions, are typically more prone to overfitting compared to traditional machine-learning approaches. Therefore, results should be interpreted in the context of the dataset size. Recordings were collected in controlled conditions, and further studies in more varied or free-living environments are needed to confirm robustness. Future work will focus on expanding the dataset and evaluating practical deployment workflows to strengthen translational potential. To facilitate clinical deployment and build trust with practitioners, future studies should also explore explainable AI tools (e.g., saliency maps or SHAP-based feature attribution) to enhance transparency and provide interpretable insights into the model's decision-making process. Overall, this non-invasive approach, using a single wearable sensor and requiring only natural walking, can help veterinarians detect and distinguish between gait-related pathologies early, offering more targeted and timely treatment.

Several additional evaluations can be conducted to enhance the model's performance further. For example, the trotting data, which has shown to enhance classification accuracy in some cases, but is usually not feasible with neurological impaired dogs. Nevertheless, it offers a great opportunity to evaluate and monitor orthopedically impaired dogs whenever possible. While our current approach demonstrates the efficacy of a lightweight CNN, we acknowledge that in scenarios with small datasets, techniques such as transfer learning, domain adaptation, and structured generative augmentation (GAN-based) are increasingly utilized to stabilize training and further improve generalization in both animal and human gait studies. Thus, while our model shows promising results, it also highlights the potential for further exploration into different sensor placements, gait types, and data augmentation techniques, which could lead to even better performance and broader clinical applicability.

In conclusion, this work provides a novel, non-invasive approach for the early detection and classification of gait abnormalities in dogs, with the potential to significantly enhance diagnostic accuracy in veterinary practice. By leveraging wearable inertial sensors and deep learning models, the proposed method offers an efficient, cost-effective, and accessible tool for identifying orthopedic and neurological conditions. This work stands to benefit clinical veterinarians, researchers in animal biomechanics, and dog owners alike - contributing to earlier intervention, more informed treatment decisions, and ultimately, improved welfare and quality of life for canine patients.

Data availability
The datasets generated and/or analysed during the current study are not publicly available due to several constraints, but areavailable from the corresponding author on reasonable request.

Received: 22 July 2025; Accepted: 16 February 2026
Published online: 18 March 2026

References
1. Paluš, V. Neurological examination in small animals. *Macedonian Vet. Rev.* **37**(1), 95–105. 10.14432_j.macvetrev (2014).
2. Garosi, L. & Lowrie, M. The neurological examination. In *BSAVA Manual of Canine and Feline Neurology*. 1–24. (BSAVA Library, 2013).
3. Park, C., Sawyere, D. M., Pancotto, T. E., Lanz, O. I. & Werre, S. R. Characterization of spatiotemporal and kinetic gait variables in dogs with hindlimb ataxia and bilateral hindlimb lameness. *BMC Vet. Res.* **20**(1), 405 (2024).
4. Witte, P. & Scott, H. Investigation of lameness in dogs: 2. Hindlimb. *In Practice* **33**(2), 58–66 (2011).
5. Kerwin, S. C. & Taylor, A. R. Assessment of orthopedic versus neurologic causes of gait change in dogs and cats. *Vet. Clin. N. Am. Small Anim. Pract.* **51**(2), 253–261 (2021).
6. Weyer, A. et al. Reliability and validity of the scale for the assessment and rating of ataxia: A study in 64 ataxia patients. *Mov. Disord.* **22**(11), 1633–1637 (2007).
7. Waxman, A. S. et al. Relationship between objective and subjective assessment of limb function in normal dogs with an experimentally induced lameness. *Vet. Surg.* **37**(3), 241–246 (2008).
8. Quinn, M. M. et al. Evaluation of agreement between numerical rating scales, visual analogue scoring scales, and force plate gait analysis in dogs. *Vet. Surg.* **36**(4), 360–367 (2007).
9. Lee, C.-S., Bentley, R. T., Weng, H.-Y. & Breur, G. J. A preliminary evaluation of the reliability of a modified functional scoring system for assessing neurologic function in ambulatory thoracolumbar myelopathy dogs. *BMC Vet. Res.* **11**, 1–7 (2015).
10. Fu, Y.-C., Torres, B. T. & Budsberg, S. C. Evaluation of a three-dimensional kinematic model for canine gait analysis. *Am. J. Vet. Res.* **71**(10), 1118–1122 (2010).
11. Gillette, R. L. & Angle, T. C. Recent developments in canine locomotor analysis: A review. *Vet. J.* **178**(2), 165–176 (2008).
12. Hayati, H., Mahdavi, F. & Eager, D. Analysis of agile canine gait characteristics using accelerometry. *Sensors* **19**(20), 4379 (2019).
13. Piche, E. et al. Validity and repeatability of a new inertial measurement unit system for gait analysis on kinematic parameters: Comparison with an optoelectronic system. *Measurement* **198**, 111442 (2022).
14. Jenkins, G. J., Hakim, C. H., Yang, N. N., Yao, G. & Duan, D. Automatic characterization of stride parameters in canines with a single wearable inertial sensor. *PloS one* **13**(6), e0198893 (2018).
15. Ladha, C., O'Sullivan, J., Belshaw, Z. & Asher, L. Gaitkeeper: A system for measuring canine gait. *Sensors* **17**(2), 309 (2017).

16. Altermatt, M., Kalt, D., Blättler, P. & Schkommodau, E. Extraction of canine gait characteristics using a mobile gait analysis system based on inertial measurement units. *Vet. Anim. Sci.* **21**, 100301 (2023).
17. Clark, K., Caraguel, C., Leahey, L. & Béraud, R. Evaluation of a novel accelerometer for kinetic gait analysis in dogs. *Can. J. Vet. Res.* **78**(3), 226–232 (2014).
18. Rhodin, M., Bergh, A., Gustås, P. & Álvarez, C. G. Inertial sensor-based system for lameness detection in trotting dogs with induced lameness. *Vet. J.* **222**, 54–59 (2017).
19. Gerencsér, L., Vásárhelyi, G., Nagy, M., Vicsek, T. & Miklósi, A. Identification of behaviour in freely moving dogs (Canis familiaris) using inertial sensors. *PloS one* **8**(10), e77814 (2013).
20. Britt, W. R., Miller, J., Waggoner, P., Bevly, D. M. & Hamilton, J. A. An embedded system for real-time navigation and remote command of a trained canine. *Pers. Ubiq. Comput.* **15**, 61–74 (2011).
21. Barthélémy, I. et al. Longitudinal ambulatory measurements of gait abnormality in dystrophin-deficient dogs. *BMC Musculoskelet. Disord.* **12**, 1–12 (2011).
22. Pillard, P., Gibert, S. & Viguier, E. 3D accelerometric assessment of the gait of dogs with cranial cruciate ligament rupture. *Comput. Methods Biomech. Biomed. Eng.* **15**(sup1), 129–131 (2012).
23. Simmons, A. D., Carrier, D. R., Farmer, C. G. & Gregersen, C. S. Lack of locomotor-cardiac coupling in trotting dogs. *Am. J. Physiol.-Regul. Integr. Comp. Physiol.* **273**(4), R1352–R1360 (1997).
24. Engelsman, D. et al. Measurement of canine ataxic gait patterns using body-worn smartphone sensor data. *Front. Vet. Sci.* **9**, 912253 (2022).
25. Reinstein, R. D. S. et al. Analysis of dog movement using a single accelerometer in different body positions-A new approach. *Front. Vet. Sci.* **12**, 1551341 (2025).
26. Blättler, P., Altermatt, M., Röhrich, M. & Grütter, N. Serial kinematic analysis using inertial measurement units in growing dogs at risk of hip dysplasia. *Vet. Anim. Sci.* **25**, 100385 (2024).
27. Lee, K. & Song, C. Walking and trotting gait of domestic dogs: A scoping review of markerbased, inertial measurement units, and markerless methods. *Phys. Ther. Rehabil. Sci.* **14**(2), 242–255 (2025).
28. Mao, A., Huang, E., Wang, X. & Liu, K. Deep learning-based animal activity recognition with wearable sensors: Overview, challenges, and future directions. *Comput. Electron. Agric.* **211**, 108043 (2023).
29. Figueirinhas, P. et al. Development of an artificial neural network for the detection of supporting hindlimb lameness: A pilot study in working dogs. *Animals* **12**(14), 1755 (2022).
30. Zhang, X., Jenkins, G. J., Hakim, C. H., Duan, D. & Yao, G. Four-limb wireless IMU sensor system for automatic gait detection in canines. *Sci. Rep.* **12**(1), 4788 (2022).
31. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**(7553), 436–444 (2015).
32. Goodfellow, I., Bengio, Y., Courville, A. & Bengio, Y. Deep Learning. Vol. 1. (MIT Press, 2016).
33. Shinde, P. P. & Shah, S. A review of machine learning and deep learning applications. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*. 1–6 (IEEE, 2018).
34. Mahrishi, M., Hiran, K. K., Meena, G. & Sharma, P. Machine Learning and Deep Learning in Real-Time Applications. (IGI Global, 2020).
35. Cohen, N. & Klein, I. Inertial navigation meets deep learning: A survey of current trends and future directions. *Results Eng.* 103565 (2024).
36. Golroudbari, A. A. & Sabour, M. H. Generalizable end-to-end deep learning frameworks for real-time attitude estimation using 6DoF inertial measurement units. *Measurement* **217**, 113105 (2023).
37. Shah, V. R. & Dixon, P. C. Gait speed and task specificity in predicting lower-limb kinematics: A deep learning approach using inertial sensors. *Mayo Clin. Proc. Digit. Health* **3**(1), 100183 (2025).
38. Santos, T. M. et al. A low-cost wireless system of inertial sensors to postural analysis during human movement. *Measurement* **148**, 106933 (2019).
39. Trabassi, D. et al. Optimizing rare disease gait classification through data balancing and generative AI: Insights from hereditary cerebellar ataxia. *Sensors* **24**(11), 3613 (2024).
40. Prince, S. J. *Understanding Deep Learning*. (The MIT Press, 2023).
41. Agarap, A. F. Deep learning using rectified linear units (RELU). arXiv preprint arXiv:1803.08375 (2018).
42. Bock, S. & Weiß, M. A proof of local convergence for the Adam optimizer. In *2019 International Joint Conference on Neural Networks (IJCNN)*. 1–8 (IEEE, 2019).
43. Xsens. Xsens DOT. https://www.xsens.com/xsens-dot.
44. Álvarez, C. G., Gustås, P., Bergh, A. & Rhodin, M. Vertical head and pelvic movement symmetry at the trot in dogs with induced supporting limb lameness. *Vet. J.* **229**, 13–18 (2017).

Acknowledgements
N. P. was supported by the Maurice Hatter Foundation.

Author contributions
N.P., I.K., and A.Z. contributed to the conceptualization and methodology development of the study and were involved in writing the manuscript. L.S., S.M., and H.V. conducted the experiments and contributed veterinary expertise. N.P. implemented the code and performed the data analysis. All authors reviewed and approved the final version of the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to N.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.