frontiers | Frontiers in Computer Science

TYPE Original Research
PUBLISHED 13 August 2025
DOI 10.3389/fcomp.2025.1597143

OPEN ACCESS

EDITED BY
Xiao Liu,
Deakin University, Australia

REVIEWED BY
Hassène Gritli,
Carthage University, Tunisia
Roberto Francesco Pitzalis,
Italian Institute of Technology (IIT), Italy

*CORRESPONDENCE
Omar Coser
✉ omar.coser@unicampus.it

RECEIVED 20 March 2025
ACCEPTED 23 July 2025
PUBLISHED 13 August 2025

CITATION
Coser O, Tamantini C, Tortora M, Furia L, Sicilia R, Zollo L and Soda P (2025) Deep learning for human locomotion analysis in lower-limb exoskeletons: a comparative study. *Front. Comput. Sci.* 7:1597143. doi: 10.3389/fcomp.2025.1597143

Deep learning for human locomotion analysis in lower-limb exoskeletons: a comparative study

Omar Coser[1,2]*, Christian Tamantini[2,3], Matteo Tortora[1,4], Leonardo Furia[1], Rosa Sicilia[1], Loredana Zollo[2] and Paolo Soda[1,5]

[1]Unit of Artificial Intelligence and Computer Systems, Università Campus Bio-Medico di Roma, Rome, Italy, [2]Unit of Advanced Robotics and Human-Centered Technologies, Università Campus Bio-Medico di Roma, Rome, Italy, [3]Institute of Cognitive Sciences and Technologies, National Research Council of Italy, Rome, Italy, [4]Dipartimento di Ingegneria Navale, Elettrica, Elettronica e delle Telecomunicazioni, Università degli Studi di Genova, Genova, Italy, [5]Department of Diagnostics and Intervention, Radiation Physics, Biomedical Engineering, Umeå University, Umeå, Sweden

Introduction: Wearable robotics for lower-limb assistance is increasingly investigated to enhance mobility in individuals with physical impairments and to augment performance in able-bodied users. A major challenge in this domain is the development of accurate and adaptive control systems that ensure seamless human-robot interaction across diverse terrains. While neural networks have recently shown promise in time-series analysis, no prior work has tackled the combined task of classifying ground conditions into five terrain classes and estimating high-level locomotion parameters such as ramp slope and stair height.

Methods: This study presents an experimental comparison of eight deep neural network architectures for terrain classification and locomotion parameter estimation. The models are trained on the publicly available CAMARGO 2021 dataset using inertial (IMU) and electromyographic (EMG) signals. Particular attention is given to evaluating the performance of IMU-only inputs versus combined IMU+EMG data, with an emphasis on cost-efficiency and sensor minimization. The tested architectures include LSTM, CNN, and hybrid CNN-LSTM models, among others. Model explainability is assessed via SHAP analysis to guide sensor selection.

Results: IMU-only configurations matched or outperformed those using both IMU and EMG, supporting a more efficient setup. The LSTM model, using only three IMU sensors, achieved high terrain classification accuracy (0.94 ± 0.04) and reliably estimated ramp slopes ($1.95 \pm 0.58^{\circ}$). The CNN-LSTM architecture demonstrated superior performance in stair height estimation, achieving a accuracy of 15.65 ± 7.40 mm. SHAP analysis confirmed that sensor reduction did not compromise model accuracy.

Discussion: The results highlight the feasibility of using lightweight, IMU-only setups for real-time terrain classification and locomotion parameter estimation. The proposed system achieves an inference time of ~2 ms, making it suitable for real-time wearable robotics applications. This study paves the way for more accessible and deployable solutions in assistive and augmentative lower-limb robotic systems. Code and models are publicly available at: [https://github.com/cosbidev/Human-Locomotion-Identification].

KEYWORDS

deep learning, explainable AI, human locomotion, multimodal learning, neural networks

1 Introduction

The field of lower limb robotics for rehabilitation and assistance leverages advanced robotic technologies to support individuals with lower limb impairments or disabilities (Díaz et al., 2011). Robotic applications in clinical practice include wearable exoskeletons that augment natural movement (Lee et al., 2020), robotic prostheses that restore ambulation in users with artificial limbs (O'Keeffe and Rout, 2019), and rehabilitation robots that deliver targeted exercises (Valgeirsdóttir et al., 2022). These systems are increasingly used in stroke recovery, spinal cord injury rehabilitation, and age-related mobility support, where precise locomotion control and adaptability to varied terrain are critical for safety and therapeutic outcomes. By enhancing rehabilitation, fostering independence, and enabling personalized therapy, these innovations show promise in improving patient outcomes. However, optimal performance in these systems requires robust data-driven parameter extraction and context-aware interventions. Control strategies must incorporate gait parameters to dynamically adapt to changing terrain conditions (Gao et al., 2023).

A key component of such control strategies is human locomotion analysis, particularly terrain recognition and the quantification of slope or stair height (Coser et al., 2024). Real-time identification of terrain transitions—such as stair ascent or ramp descent—enables adaptive control policies that reduce fall risk and improve user confidence in daily, unsupervised environments. Wearable sensors are instrumental in capturing locomotion data, offering real-time insight into human movement. Among the most widely used are Inertial Measurement Units (IMUs), which capture acceleration and angular velocity (Ribeiro and Santos, 2017; Bartlett and Goldfarb, 2017), and electromyography (EMG) sensors, which monitor muscle activation (Gupta and Agarwal, 2017; Sorkhabadi et al., 2019).

Traditional machine learning techniques, including k-Nearest Neighbor (kNN), Support Vector Machines (SVM), Gaussian Mixture Models (GMM), and Random Forests (RF), have been extensively applied to classify locomotion data from IMU and EMG sensors. However, these methods typically require handcrafted features, which are time-consuming, domain-dependent, and prone to human bias (Attal et al., 2015). In contrast, deep learning models can automatically extract features directly from raw sensor data, improving model generalization and performance, especially for complex tasks like terrain classification and parameter estimation in wearable robotics. Deep learning models have consistently outperformed classical machine learning techniques in both classification accuracy and adaptability, making them the de facto standard for state-of-the-art performance in human locomotion tasks, as it is shown in Zhang et al. (2020). This methodological shift reflects a broader trend, and our study contributes to consolidating this transition through a structured evaluation.

Table 1 summarizes existing work in human locomotion analysis, including sensor modalities, model types, datasets, and validation approaches. It covers traditional models (LDA, kNN, DT, RF, SVM) (Negi et al., 2020; Zhang et al., 2020; Zheng et al., 2022) and deep learning architectures such as CNN, LSTM, CNN-LSTM hybrids, ResNet, ANN, DANN, MCD, and attention-based CNN-BiLSTM (Narayan et al., 2021; Amer and Ji, 2021; Wang et al., 2022; Zhao et al., 2022; Jing et al., 2022; Liang et al., 2023; Kang et al., 2022; Le et al., 2022). These methods address tasks such as gait analysis (Zhao et al., 2019), locomotion intent prediction (Le et al., 2022), terrain classification (Anantrasirichai et al., 2014), and joint moment estimation (Molinaro et al., 2024). While most report accuracies above 90%, lower performance (e.g., 74%) often appears in transfer learning scenarios. Dataset sources vary, with public datasets like CAMARGO (Camargo et al., 2021), ENABL3S (Hu et al., 2018), and DSADS (Barshan and Yüksek, 2014), and participant numbers ranging from 5 to 500. Validation strategies—such as k-fold, hold-out, leave-one-subject-out, or leave-one-terrain-out—play a key role in assessing model generalizability. Moreover, two main strategies are commonly adopted to split data for training and evaluation: stratified random sampling and participant-based splitting. While stratified sampling ensures a balanced class distribution, it may inflate performance estimates by allowing information leakage across subjects. In contrast, participant-based splitting evaluates models on entirely unseen individuals, providing a more realistic and reproducible assessment of generalizability (Mekni et al., 2025a,b). Although this approach may yield slightly lower accuracy, it is better suited for assessing robustness in real-world deployment.

Despite growing interest in terrain classification and parameter estimation using wearable sensor data, the current literature faces key limitations. First, validation procedures are often inconsistent and insufficiently described, making it difficult to assess model robustness. Second, many studies rely on proprietary datasets, which hinders reproducibility and benchmark comparisons. These issues impede progress toward real-world-ready systems.

This study is motivated by the need to enable wearable robotic systems to accurately interpret terrain characteristics and adapt in real-time. Accurate classification of ground conditions and estimation of parameters such as ramp inclination and stair height are essential for effective human-robot interaction. Our review of current literature revealed a lack of systematic performance comparisons across state-of-the-art deep learning models for this task. To address this, we evaluate a broad set of deep learning models, including some adapted from other time-series domains, and apply Explainable AI techniques to identify key sensor inputs for optimal performance. The objective of this comparison was to overcome the key limitations identified in the extant literature. Prior studies often evaluate only one or two models in isolation, use non-public datasets that hinder reproducibility, or apply inconsistent validation strategies such as unspecified hold-out methods. In contrast, our work systematically compares eight deep learning architectures under uniform conditions using the publicly available CAMARGO 2021 dataset. To facilitate a fair comparison, the following methodologies are adopted: standardized preprocessing, consistent cross-validation, and shared training settings. Additionally, we address the limited attention to regression tasks and sensor optimization by incorporating parameter estimation and Explainable AI-based sensor reduction.

In this paper, we propose a neural-network-based system for terrain classification (level ground, stair ascent/descent, ramp ascent/descent) and parameter regression (inclination, step height)

TABLE 1 Summary of methodologies for human locomotion analysis for terrain and slope recognition.

References	Modality	Algorithm	Brief description	Acc	Dataset	# of classes	# of subject	Validation approach
Negi et al. (2020)	EMG and acceleration, * early fusion	kNN SVM, RF, LDA, DT	Gait analysis using surface electromyography and acceleration sensors classified five terrains. Signals from tibialis anterior and gastrocnemius muscles were processed with machine learning models. The goal was to optimize classification accuracy with minimal computation time and muscle signals.	0.97, 0.98, 0.79, 0.80, 0.99	CWP	5 classes (LG, RA, RD, SA, SD)	15 subjects	k-fold cross-validation
Zhang et al. (2020)	EMG and IMU, * early fusion	LDA, SVM, ANN, CNN, DANN, MCD	Predicting human locomotion intent aids in controlling wearable robots and assisting movement on various terrains. This study introduces an unsupervised cross-subject adaptation method to predict locomotion intent without labeled data.	0.92, 0.90, 0.93, 0.96, 0.95, 0.95	ENABL3S (Hu et al., 2018) and DSADS (Barshan and Yüksek, 2014)	5 classes (LG, SA, SD, RA, RD)	10 subjects	leave-one-subjects out
Narayan et al. (2021)	IMU	CNN	This study evaluates hierarchical classification for real-time locomotion mode recognition in wearable robotic prostheses and exoskeletons. A CNN-based classifier trained on inertial sensor data achieves stable and accurate mode classification, including smoother transitions. The method enables real-time operation, improving control adaptation in wearable robots.	0.94	CWP	10 classes (sit, stand, walk str, walk cur-left, walk cur-right, down s, up s, left t, right t, unknown)	8 subjects	k-fold cross-validation
Amer and Ji (2021)	IMU	CNN	This study introduces a CNN-based algorithm for classifying human locomotion activities using inertial measurement unit (IMU) data. Spectral analysis transforms inertial signals into time-frequency representations, which are then classified as images.	0.99	CWP	6 classes (Norm walking, Walking upstairs, Walking downstairs, Sitting, Standing, Laying)	30 subjects	k-fold cross-validation
Wang et al. (2022)	IMU	CNN	A residual network-based method is proposed for locomotion mode recognition in lower limb exoskeletons. Using inertial sensor data, the network autonomously learns mixed features, eliminating the need for manual feature extraction.	0.97	CWP	5 classes (LG, SA, SD, RA, RD)	5 subjects	leave-one-subject out
Zhao et al. (2022)	EMG and IMU, * early fusion	CNN	A multi-channel separated encoder-based convolutional neural network is proposed for locomotion intention recognition. Inertial sensor data are processed through spectral transformation and classified using CNNs to enhance recognition accuracy.	0.94	ENABL3S (Hu et al., 2018)	5 classes (LG, RA, RD, SA, SD)	10 subjects	k-fold cross-validation
Jing et al. (2022)	EMG	CNN	A neural network trained on HDsEMG data achieved higher accuracy and robustness against electrode shifts compared to bipolar electrodes. The approach enhances gait mapping reliability, supporting real-time control of assistive technologies.	0.97	HDsEMG (Jiang et al., 2021)	6 classes (Standing, LG, SA, SD, RA, RD)	7 subjects	k-fold cross-validation
Zheng et al. (2022)	IMU	CNN-SVM	A hybrid CNN-SVM model is proposed for locomotion mode recognition using multi-channel inertial measurement unit (IMU) signals. The approach integrates a feature mapping layer with error correction from a finite state machine (FSM) to improve accuracy and generalization.	0.98	CWP	5 classes (LW, SA, SD, RA, RD)	10 subjects	leave-one-subject out
Le et al. (2022)	IMU	CNN	A deep convolutional neural network is developed for locomotion intent prediction in powered prosthetic legs, with both subject-dependent and subject-independent validations. Transfer learning is applied to improve subject-independent performance using a small portion of data from a new subject, significantly reducing error rates.	0.74	ENABL3S (Hu et al., 2018)	5 classes (LG, SA, SD, RA, RD)	9 subjects	hold-out

(Continued)

10.3389/fcomp.2025.1597143

TABLE 1 (Continued)

References	Modality	Algorithm	Brief description	Acc	Dataset	# of classes	# of subject	Validation approach
Kang et al. (2022)	IMU	CNN	A deep convolutional neural network-based locomotion mode classifier is developed for hip exoskeletons using an open-source gait biomechanics dataset. The model operates independently of user-specific data, ensures smooth mode transitions, and relies only on minimal wearable sensors.	0.93	CAMARGO (Camargo et al., 2021)	5-classes LG, SA, SD, RA, RD	21 subejcts	leave-one-terrain-out
Son and Kang (2023)	EMG	CNN-LSTM, LSTM-CNN	Data from electromyograms (EMGs) and robot sensors were used to compare the performance of two hybrid model LSTM-CNN and CNN-LSTM	0.94, 0.95	CWP	5-classes LG, SA, SD, RA, RD	500 subjecs	hold-out
Liang et al. (2023)	IMU	Attention-based CNN-BiLSTM	A deep-learning approach is developed for estimating lower-limb joint moments during locomotive activities using inertial measurement units (IMUs). The model accurately predicts hip, knee, and ankle joint moments with a single IMU, with the shank identified as the optimal placement.	0.85	CAMARGO (Camargo et al., 2021)	4 classes: LG, Ramp, Stair, Treadmill	19 subjects	hold-out

LG, level ground; RA/D, Ramp Ascent/Descent; SA/D, Stair Ascent/Descent; CWP, Collected within the paper.
*Indicate multimodal analysis.

using multimodal sensor data from the CAMARGO dataset. Our contributions are as follows:

1. We perform a comparative analysis of eight deep learning models used for time-series classification and regression in multimodal locomotion data. Leave-one-subject-out cross-validation ensures robustness and generalizability.
2. We investigate unimodal vs. multimodal configurations, employing Explainable AI and ablation studies to determine minimal sensor setups that maintain high performance.
3. We utilize the publicly available CAMARGO 2021 dataset, which provides a high-quality, ethically sound basis for locomotion recognition research.

The remainder of this paper is structured as follows. Section 2 describes the dataset and preprocessing methods. Section 3 outlines the deep learning architectures and experimental setup. Section 4 presents the findings, and Section 6 concludes the paper with directions for future research.

2 Materials

In this work, we utilize the CAMARGO dataset (Camargo et al., 2021), a publicly available multimodal repository including sensor data from 21 subjects. Each participant was equipped with four IMUs (IMU type: Xsens MTw Awinda—a wireless 9-DOF (Degrees of Freedom) IMU, manufactured by Xsens Technologies B.V.). positioned on the trunk, thigh, shank, and foot, along with 11 EMG sensors (Type: Surface EMG by Delsys and model Trigno Wireless System) that monitored the activities of the gastrocnemius medialis, tibialis anterior, soleus, vastus medialis, vastus lateralis, rectus femoris, biceps femoris, semitendinosus, gracilis, gluteus medius, and right external oblique muscles. The participants completed multiple trials across five locomotion modes: level ground walking, ramp ascent and descent, and stair ascent and descent. Stair trials were performed at four different heights (102mm, 127mm, 152mm, 178mm), while ramp trials included six inclination angles (5.2°, 7.8°, 9.2°, 11°, 12.4°, and 18°). In particular, the sensors are positioned along the body as shown in Figure 1, We can see that there are eleven Electromyography (EMG) sensors on the right side, targetting major lower limb muscles and four Inertial Measurement Units (IMUs) attached to the torso, thigh, shank and foot, capturing acceleration and angular velocity, For a total of 35 signals recorded over an estimated 40 minutes, capturing high-resolution IMU and EMG data across various locomotion conditions. For a more detailed description of a dataset and the exact sensor placements, please refer to the original paper (Camargo et al., 2021).

The authors of the dataset preprocessed the collected dataset: whereas the IMU data, comprising acceleration and angular velocity from the three-axis accelerometer and gyroscope sampled at 200 Hz, were processed using a lowpass filter with a 100 Hz cutoff frequency (Butterworth order 6), the raw EMG data, sampled at 1000 Hz, was digitally conditioned using a bandpass filter with a cutoff frequency of 20 Hz to 400 Hz (Butterworth order 20).

In our work, we use the data as described so far, utilizing a rolling window approach to segment the data into non-overlapping temporal windows of 500ms to generate ∼1,450 samples for each subject with a class probability of 20 ± 2.5% (the number of samples

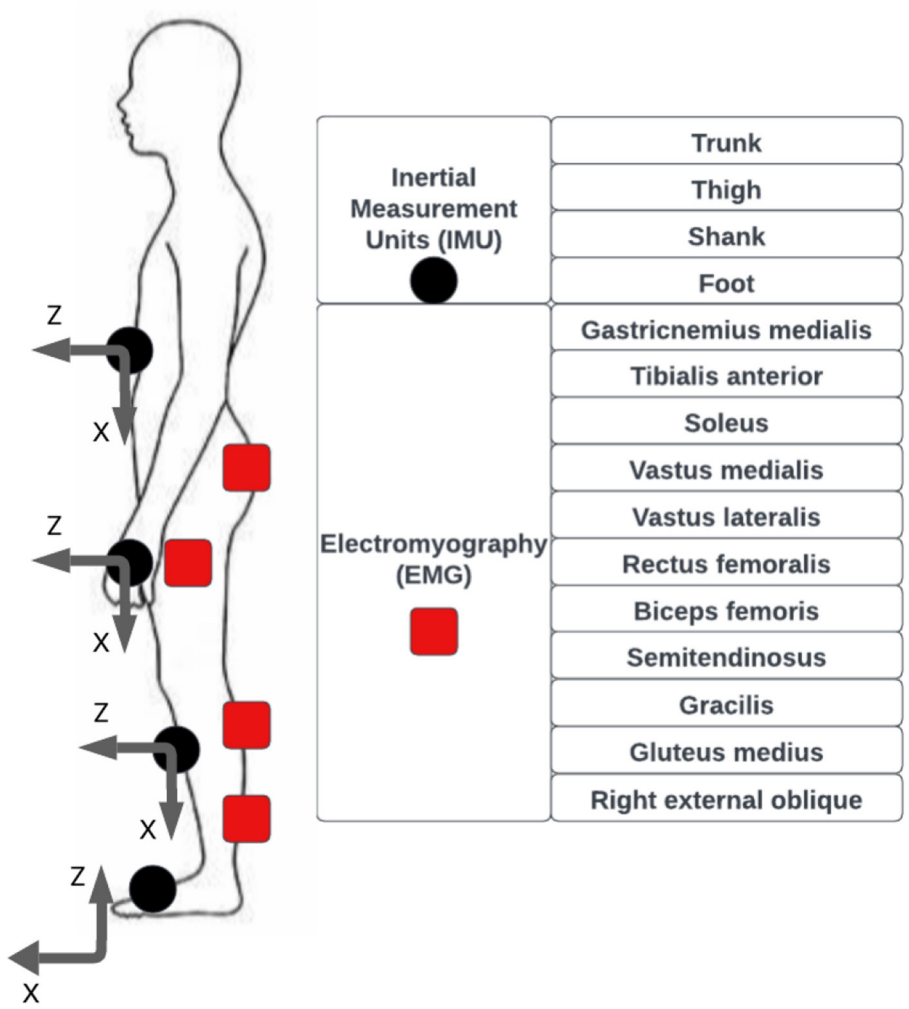


FIGURE 1
Position of the sensors along the body (image adapted from Camargo et al., 2021, licensed under CC-BY).

and the class probability depend slightly on the subject) to feed the different NN models.

3 Methods

This section outlines the methodology for developing a system architecture to classify terrain types and estimate terrain parameters using multimodal sensory data. The system comprises two separate stages: a classification network for detecting the five terrain types and two regression models to estimate the slope of a ramp or the height of a stair. We compared state-of-the-art Deep Learning (DL)-based models for time series analysis to identify the optimal configuration. We selected the best-performing sensor modality and model architecture based on the results. Additionally, we used Explainable Artificial Intelligence (XAI) techniques to determine the most influential features, providing insights into sensor importance and the minimal sensor setup required for accurate classification and regression as shown in Figure 2.

3.1 Supervised models for human locomotion parameters identification

The selection of the eight deep learning models included in this study was guided by the goal of covering a diverse and representative spectrum of architectures commonly used for time-series analysis. Each model class was chosen based on its theoretical strengths and empirical relevance to human locomotion tasks. Convolutional models (CNN and XceptionTime) are widely employed in the gait and activity recognition literature due to their efficiency in extracting local and hierarchical features from IMU signals (Müller et al., 2024). Recurrent architectures (LSTM and GRU) were selected for their strong ability to model temporal dependencies, a key characteristic of gait dynamics (Su and Gutierrez-Farewik, 2020; Lee et al., 2025). To capture both spatial and temporal features, we included hybrid models (CNN-LSTM and LSTM-CNN), which have shown promising results in various sensor-based motion recognition tasks (Sadeghzadehyazdi et al., 2021). Additionally, we integrated two recent and advanced architectures: the Time Series Transformer (TST), which applies self-attention mechanisms

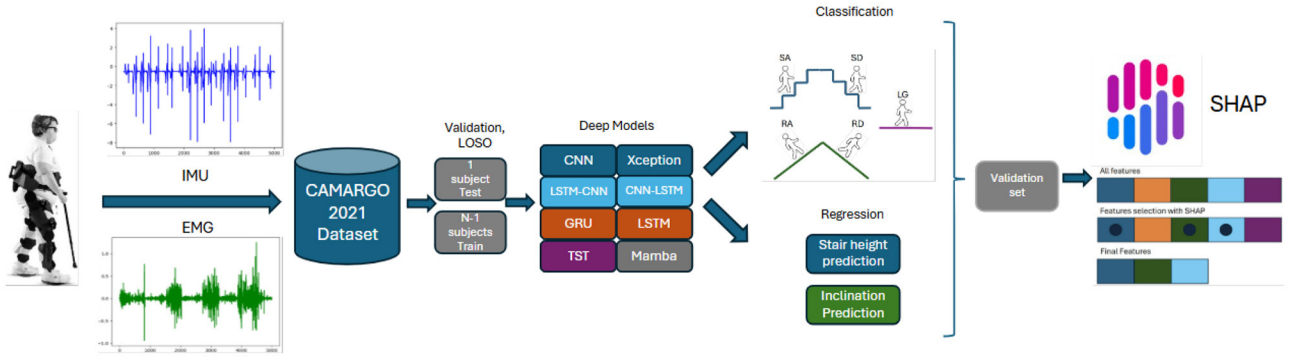


FIGURE 2
Visual methodology.

for global dependency modeling, and Mamba, a state-of-the-art State Space Model designed for efficient long-range sequence learning (Ahamed and Cheng, 2024). This comprehensive model selection ensures that our evaluation encompasses a wide array of inductive biases and learning strategies, enabling a robust comparison across modeling paradigms in the context of wearable sensor-based terrain recognition and parameter estimation. In particular, those models can be categorized in five major class, the Convolutional Neural Network (CNN, XceptionTime), hybrid models (CNN-LSTM and LSTM-CNN), Recurrent neural network (LSTM and Gated Recurrent Unit), Transformer (TST), and a variant of RNN, the State Space Model (MAMBA). They are:

- Convolutional neural networks (CNN, XceptionTime): CNNs (Kang et al., 2022) and XceptionTime Rahimian et al. (2019) are renowned for their ability to efficiently extract local and hierarchical features from time series data. While CNN provides a solid baseline for feature extraction, XceptionTime enhances this capability with depthwise separable convolutions and parallel temporal paths, improving computational efficiency and multi-scale pattern recognition.
- Hybrid models (CNN-LSTM, LSTM-CNN): combining convolutional and recurrent layers allows hybrid models to leverage both spatial feature extraction and temporal sequence modeling. CNN-LSTM (Son and Kang, 2023) first captures local patterns using CNNs before modeling temporal dependencies with LSTMs, while LSTM-CNN (Son and Kang, 2023) reverses this flow to prioritize sequential information before feature extraction. These models are particularly effective in tasks where both spatial and temporal structures are crucial.
- Recurrent neural netoworks (LSTM, GRU): LSTM (custom) and GRU (Chung et al., 2014) are well-established architectures for handling sequential data. LSTMs excel at capturing long-term dependencies through their gating mechanisms, while GRUs offer a simplified structure with comparable performance and reduced computational overhead. Their inclusion ensures a solid benchmark for traditional sequence modeling approaches.
- Transformer-based model (TST): the Time Series Transformer (TST) (Zerveas et al., 2021) introduces self-attention mechanisms to time series analysis, enabling the model to capture long-range dependencies without relying on recurrent structures. This approach enhances parallelization and scalability, addressing the limitations of RNN-based models in handling large and complex datasets.
- State space model (MAMBA): MAMBA (Gu and Dao, 2023) represents a modern variant of RNNs, leveraging structured State Space Models to efficiently capture long-range dependencies and continuous-time dynamics. Its inclusion allows for the exploration of state-of-the-art techniques in time series modeling, particularly for datasets with irregular intervals or complex temporal relationships.

These architectures encompass a wide range of modeling techniques, from convolutional approaches focused on local feature extraction to recurrent and state-space models designed for capturing temporal dependencies, and finally, transformer-based models that excel at modeling global relationships within sequences. Depending on the modality, the models take as input n features, leveraging multimodal strategies to enhance predictive performance. Among these strategies, early fusion is frequently employed, where raw sensor data is merged into a shared embedded space via concatenation, addition, pooling, or gated units (Tortora et al., 2023). Our choice to use early fusion follows from the data's homogeneous nature and consistent sampling frequencies across modalities, eliminating the need for complex alignment or resampling. These uniform data characteristics allow the model to capture meaningful cross-modal interactions from the outset, enhancing predictive capability while minimizing noise or redundancy. Additionally, early fusion offers a simpler and more efficient architecture by removing the requirement for separate processing pathways for each modality, thereby reducing computational overhead. While multimodal sensor fusion can pose challenges—such as managing redundancies and handling missing modalities—the inherent synchronization and similarity in our data mitigate these issues, making early fusion an optimal choice for effective feature integration. All models were trained using a learning rate of 0.001 and a batch size of 32, values that are commonly adopted in time-series deep learning tasks and have

been shown to provide a good balance between convergence speed and training stability (Ismail Fawaz et al., 2019). In order to guarantee a valid and reliable comparison of performance, it was necessary to ensure consistency of these hyperparameters across all architectures.

3.2 Explainable AI

Understanding and interpreting machine learning models is essential for ensuring transparency, trust, and fairness in AI-driven decision-making. There are various interpretability methods, categorized based on different aspects. Model-specific methods leverage the internal structure of models, whereas model-agnostic approaches, like SHapply Addictive exPlanations (SHAP), treat models as black boxes (Molnar, 2020; Lundberg and Lee, 2017; Ribeiro et al., 2016). Global explanations analyze overall model behavior, while local explanations focus on individual predictions (Doshi-Velez and Kim, 2017; Ribeiro et al., 2016). *Post-hoc* methods provide insights without altering the model, whereas intrinsically interpretable models (e.g., decision trees) are inherently transparent (Guidotti et al., 2018; Lipton, 2018). Feature attribution techniques, such as SHAP and LIME, quantify the impact of input features (Lundberg and Lee, 2017; Ribeiro et al., 2016), while example-based explanations use representative instances to justify predictions (Caruana et al., 1999; Molnar, 2020). Finally, visualization techniques, including Grad-CAM and saliency maps, enhance interpretability in deep learning by highlighting influential input regions (Simonyan et al., 2013; Selvaraju et al., 2017). These diverse methods play a crucial role in making AI systems more understandable and accountable. SHAP was chosen because it is specifically well-suited for explaining deep learning models used in time series tasks. It provides theoretically grounded feature attributions based on Shapley values, ensuring local accuracy, consistency, and reliability. In LSTMs and CNN-LSTMs, each time step can be treated like a separate "player," capturing how each segment influences the overall prediction. Its model-agnostic framework simplifies integration with any Python-based deep learning library without requiring internal access to the models. This attribute is critical when dealing with complex multi-layer architectures, as in the case of time series forecasting and classification. Moreover, SHAP's clear visualization tools help pinpoint which periods or features are most influential, a vital need in time series analysis. Compared to other interpretability tools, SHAP enforces consistency, meaning that more influential features receive higher attributions. Finally, the SHAP Python implementation is straightforward, making it a practical and efficient choice for researchers and practitioners alike, as illustrated in the pseudocode presented in Algorithm 1, which outlines the essential steps required to integrate SHAP into a subject-wise cross-validation framework.

Algorithm 1 outlines the step-by-step procedure adopted to compute SHAP values in a subject-wise validation setting. In each fold, the model is trained on a training set of 18 subjects. A `GradientExplainer` is then instantiated using the trained model and the corresponding training data. SHAP values are subsequently computed for the remaining three subjects in the validation set, enabling feature attribution on unseen data. This process is repeated across all cross-validation folds to ensure subject-level coverage and robust interpretability.

```
for each split of Subjects into TrainSet (18) and ValidationSet (3) do
    model ← TrainModel(TrainSet)
    explainer ← shap.GradientExplainer(model, TrainSet)
    shap_values ← explainer.shap_values(ValidationSet)
end for
```

Algorithm 1. Computing SHAP values for subject-wise validation.

Figure 3 visually summarizes the SHAP-based XAI procedure adopted in our study. The upper section illustrates the data flow in a single fold of subject-wise cross-validation, including model training on 18 subjects, SHAP explainer initialization, and evaluation on the 3 held-out subjects. Distinct subject icons emphasize the separation between training and validation data. The visual representation reinforces the methodological integrity of the process and complements the formal steps detailed in Algorithm 1.

3.3 Experimental setup

We validated all experiments using the Leave-One-Subject-Out (LOSO) cross-validation method, conducting a number of runs equal to the number of subjects in our study 21. In each run, the training set included data from $N - 1$ subjects, while the test set consisted of data from the remaining subject. Although each model was allowed to train for up to 100 epochs, we applied an early stopping criterion that halted training after just 10 epochs, as the validation loss showed no improvement during that period. Essentially, by the time we reached 10 epochs, the network had plateaued in terms of performance, triggering the early stopping condition and preventing any further, unproductive epochs. All experiments and model training were performed on Google Colab Pro, utilizing 52 GB of RAM (CPU) and a 15 GB GPU, except for Mamba that was trained on an NVIDIA A100 GPU with 80Gb of Ram resources. For the classification task, we assessed performance using accuracy, precision, recall, F1 score, given the balanced nature of the dataset. However, for regression tasks, we evaluated performance using the Mean Absolute Error (MAE). To further assess model performance under different conditions, we performed the Wilcoxon signed rank test, a nonparametric statistical hypothesis test (Gehan, 1965), and applied the Bonferroni correction to address the issue of multiple comparisons (Napierala, 2012).

4 Results

4.1 Performance indicators

In this section, we present the main results of the experiments conducted. The primary metric used for evaluating classification performance is accuracy, used because the a priori class distribution is balanced. However, additional metrics, including recall and the

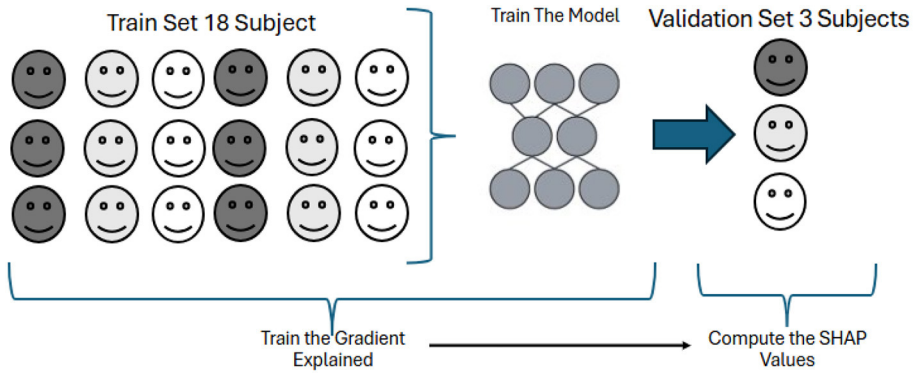


FIGURE 3
Visual overview of the Explainable AI (XAI) pipeline using SHAP values within a subject-wise cross-validation framework. The model is trained on 18 subjects and evaluated on three held-out subjects. The SHAP explainer is initialized using only the training data to avoid data leakage, and SHAP values are computed on the validation set to interpret model predictions and assess feature relevance.

F1 score, are reported in the Supplementary Material. Furthermore, for the regression analysis, we employed the Mean Absolute Error (MAE) as the primary metric.

Additionally, to determine the contribution of each sensor to the model's predictions, we computed the SHapley Additive exPlanations (SHAP) values (Lines et al., 2012), enabling an in-depth assessment of sensor importance and model interpretability. This analysis provided insights into the relative influence of each sensor on the overall system performance (Ramírez-Mena et al., 2023).

4.2 Best-performing architecture

Our first analysis compares the performances of the eight NNs selected, both in the case of classification and regression tasks which are reported in Table 2 across different sensor combinations. The combined results indicate that there is a statistically significant difference between the performance achieved with Inertial Measurement Unit sensors and that with electromyography (EMG) sensors ($p = 3.35 * 10^{-5}, p = 2.20 * 10^{-3}, p = 2.98 * 10^{-5}, p = 8.32 * 10^{-4}$). Across all performance metrics—accuracy, precision, recall, and F1 score—the IMU sensor configuration consistently outperforms the EMG sensor configuration. For example, architectures such as CNN and LSTM show higher values when using IMU data, which likely reflects the robustness of the kinematic information (such as acceleration and angular velocity) captured by these sensors. In contrast, EMG signals, which measure muscle activation, tend to be more susceptible to variability due to factors like electrode placement, skin impedance, and signal cross-talk, leading to a noisier performance profile. Moreover, the standard deviations observed with EMG data are generally higher, reinforcing the notion that its measurements are less reliable under the conditions tested. When comparing the IMU-only setting to the combined IMU+EMG setting, the differences in performance are negligible across all metrics. This lack of statistically significant improvement suggests that the inclusion of EMG data does not contribute meaningful additional information beyond what is already captured by the IMU sensors. The inertial data appears to encapsulate the essential dynamic features needed for accurate classification, thereby rendering the additional complexity of EMG integration unnecessary. It is also possible that the nature of the movements being classified is such that the gross motion captured by the IMU is sufficiently discriminative, reducing the potential benefits of fusing EMG data. The redundancy in the information provided by the EMG signals may not offer any extra value in this specific scenario, especially when considering the potential for increased noise and processing overhead. In summary, these results support the conclusion that the IMU sensors are adequate for the intended application, providing robust performance without the need for additional EMG data. The LSTM architecture consistently outperformed the other models across all evaluated performance metrics. In particular, for both IMU and IMU+EMG configurations, the LSTM achieved higher accuracy, precision, recall, and F1 scores compared to the second best performing architecture (blue in table). The computed p-value ($p = 0.0015$, $p = 0.0018$, $p = 0.0014$, $p = 0.0016$) confirms that this improvement is statistically significant and not merely a result of random variation. This robust performance can be attributed to the LSTM's inherent ability to capture long-term dependencies in sequential data. Its memory cells effectively filter noise and model temporal dynamics, which is especially important in handling the complex patterns found in sensor signals.

Moreover, the LSTM's architecture allows it to generalize better by retaining relevant contextual information over time, a feature that is crucial for accurately classifying time-series data. This capability gives it an edge over architectures such as CNNs, which are generally better at spatial feature extraction but may fall short in capturing temporal correlations. The lower variability in performance metrics for the LSTM further supports the idea that its superiority is intrinsic to its design rather than due to chance. Consequently, the statistically significant improvement, as evidenced by the p-value, highlights the robustness and effectiveness of the LSTM for this classification task.

The average training time of the neural network architecture is about 21.25 second per epochs, with a range going from 11.42 s ± 0.15 s to 55.19 s ± 0.57 s. On the other hand, all

TABLE 2 Classification results: each cell reports the average score followed by standard deviation.

Architecture	Accuracy			Precision			Recall			F1 Score		
	IMU	EMG	IMU + EMG	IMU	EMG	IMU + EMG	IMU	EMG	IMU + EMG	IMU	EMG	IMU + EMG
CNN	0.92 ± 0.05	0.73 ± 0.10	0.92 ± 0.04	0.92 ± 0.03	0.79 ± 0.05	0.92 ± 0.06	0.89 ± 0.08	0.73 ± 0.04	0.91 ± 0.06	0.89 ± 0.05	0.74 ± 0.07	0.91 ± 0.06
LSTM	**0.94 ± 0.04**	0.60 ± 0.06	**0.94 ± 0.03**	**0.95 ± 0.04**	0.39 ± 0.07	**0.95 ± 0.05**	**0.94 ± 0.06**	0.60 ± 0.07	**0.94 ± 0.08**	**0.94 ± 0.05**	0.46 ± 0.07	**0.94 ± 0.06**
CNN-LSTM	0.90 ± 0.05	0.75 ± 0.08	0.93 ± 0.04	0.93 ± 0.04	**0.80 ± 0.04**	0.93 ± 0.07	0.92 ± 0.05	**0.75 ± 0.03**	0.90 ± 0.07	0.92 ± 0.08	**0.76 ± 0.03**	0.91 ± 0.07
LSTM-CNN	0.91 ± 0.05	0.75 ± 0.10	0.91 ± 0.04	0.93 ± 0.05	0.78 ± 0.04	0.92 ± 0.08	0.91 ± 0.03	0.75 ± 0.09	0.90 ± 0.06	0.91 ± 0.04	0.76 ± 0.06	0.90 ± 0.05
GRU	0.78 ± 0.08	0.61 ± 0.03	0.79 ± 0.07	0.78 ± 0.07	0.61 ± 0.02	0.78 ± 0.06	0.77 ± 0.08	0.60 ± 0.03	0.77 ± 0.06	0.79 ± 0.07	0.60 ± 0.03	0.79 ± 0.10
TST	0.92 ± 0.04	0.74 ± 0.03	0.91 ± 0.05	0.89 ± 0.08	0.60 ± 0.03	0.89 ± 0.06	0.92 ± 0.04	0.74 ± 0.02	0.91 ± 0.04	0.92 ± 0.05	0.74 ± 0.03	0.91 ± 0.05
XceptionTime	0.88 ± 0.09	0.73 ± 0.03	0.89 ± 0.09	0.88 ± 0.09	0.73 ± 0.03	0.88 ± 0.01	0.90 ± 0.05	0.74 ± 0.03	0.89 ± 0.05	0.88 ± 0.09	0.73 ± 0.03	0.88 ± 0.09
MAMBA	0.92 ± 0.04	**0.76 ± 0.05**	0.91 ± 0.06	0.92 ± 0.05	0.75 ± 0.05	0.91 ± 0.05	0.89 ± 0.05	0.73 ± 0.05	0.91 ± 0.05	0.89 ± 0.05	0.75 ± 0.05	0.92 ± 0.05
Mean	0.89 ± 0.05	0.71 ± 0.06	0.90 ± 0.05	0.90 ± 0.05	0.70 ± 0.05	0.90 ± 0.05	0.89 ± 0.06	0.71 ± 0.06	0.89 ± 0.07	0.89 ± 0.06	0.78 ± 0.05	0.89 ± 0.07

In black bold and blue bold the best and second-best results, respectively.

TABLE 3 Regression analysis of slope inclination: MAE performance.

Architecture	MAE [°] (slope)
CNN	2.21 ± 0.58
LSTM	**1.93 ± 0.53**
CNN-LSTM	2.10 ± 0.68
LSTM-CNN	2.14 ± 0.73
GRU	2.32 ± 0.61
TST	2.21 ± 0.60
XceptionTime	2.08 ± 0.54
Mamba	2.03 ± 0.46
Mean	2.12 ± 0.59

In black bold and blue bold the best and second-best results, respectively.

models demonstrated an inference time of 1ms over a 500ms lookback window.

In summary, our analysis demonstrates that IMU sensors consistently outperform EMG sensors, likely due to their robust capture of dynamic kinematic information. Furthermore, the negligible differences between the IMU-only and IMU+EMG configurations indicate that adding EMG data does not provide a significant advantage. Among the eight architectures evaluated, the LSTM emerged as the top performer. For this reason, the regression analysis was conducted solely with IMU data, as the architectures showed no benefit when incorporating EMG signals, aligning with our aim to minimize the sensor setup.

Moreover, Table 3 demonstrates that the LSTM architecture achieves an MAE of for slope prediction, which is not only the lowest among all architectures but also statistically significantly different from the second-best performer (in blue in table). This significance confirmed by computed p-values ($p = 0.015$) well below the standard threshold—indicates that the improvement is not due to chance. The LSTM's ability to capture long-term temporal dependencies and effectively filter noise in sequential data likely underpins this robust performance. Consequently, its lower error is a direct result of its architectural strengths rather than random variation.

On average, the training process takes 24.00 ± 0.33 s per epoch for slope prediction ranging from 13.57 ± 0.17 to 64.34 ± 0.23. The inference time remains consistent across all models for both classification and regression tasks, averaging 1 ms.

Lastly, the Table 4 indicates that the CNN-LSTM architecture achieves the lowest MAE for stair height prediction, at 15.65 ± 6.33mm, outperforming the other architectures. In particular, its performance is statistically significantly better than that of the second-best architecture (LSTM, which has an MAE of 15.97 ± 6.29mm), as confirmed by the computed p-values ($p = 0.32$). This robust improvement is likely due to the CNN-LSTM's ability to combine the spatial feature extraction capabilities of CNNs with the temporal modeling strengths of LSTMs. By integrating these two approaches, the network is better equipped to capture the complex patterns and variations in stair height measurements, leading to a more accurate regression outcome that is not attributable to random chance.

On average, the training process takes 20.14 ± 0.33 s per epoch for slope prediction ranging from 10.09 ± 0.36 to 58.45 ± 0.36. The inference time remains consistent across all models for both classification and regression tasks, averaging 1 ms.

4.3 Most informative sensor

In this section, we delve deeper into the analysis of the results, focusing on identifying the most informative sensors using the SHAP methodology. Building on our previous findings, we conduct this analysis exclusively with IMU data and the LSTM and CNN-LSTM model to enhance interpretability and insight.

Figure 4 shows how much each sensor influences the ground type prediction where f, s, t, and k represent the foot, shank, thigh, and trunk, respectively. Additionally, the superscript denotes the axis (x, y, or z), while the subscript indicates the sensor type: G for the gyroscope and A for the accelerometer. The SHAP analysis indicates that foot sensors have the greatest influence on ground type prediction, followed by those on the thigh, shank, and trunk. This ranking aligns closely with the biomechanics of human locomotion and how different body segments interact with the ground.

It is evident that the foot-mounted IMU provides the most informative kinematic data for terrain classification, as the foot undergoes the most significant motion variations during locomotion. Being the primary point of contact with the ground, it is logical to deduce that the foot experiences changes in acceleration and angular velocity that directly reflect terrain properties (Shi et al., 2024). Step transitions, foot placement, and ankle dynamics induce distinctive kinematic signatures, which are more pronounced than those captured by sensors positioned elsewhere on the body. Consequently, a Foot-mounted IMU is capable of detecting terrain-induced variations more effectively, making them the optimal choice for extracting kinematic features relevant to surface characterization.

IMU sensors positioned on the shank and thigh are capable of capturing key kinematic adaptations to terrain changes, though their contribution is less pronounced than foot-mounted sensors. Variations in surface incline or height have been shown to influence knee flexion, stride length, and hip motion, leading to distinct acceleration and angular velocity patterns. These segments have been demonstrated to play a role in adjusting limb trajectories and modulating propulsion, particularly during slope ascent, descent, or stair negotiation.

In contrast, trunk-mounted sensors provide the least informative kinematic data for terrain classification. While the trunk stabilizes movement and reacts to lower-limb dynamics, its motion is less directly affected by ground properties. As a result, kinematic variations measured at the trunk are less pronounced, leading to lower relevance for identifying terrain characteristics compared to sensors positioned on the foot and lower limbs.

The observed sensor hierarchy is consistent with how humans adjust their movement to different terrains. Foot sensors capture direct interactions with the ground, thus providing the most informative data, while the shank and thigh assist in adjusting movement. The trunk plays a more supportive role in stabilization

TABLE 4 Regression analysis of stair height: MAE performance.

Architecture	MAE [mm] (height)
CNN	18.00 ± 8.51
LSTM	15.97 ± 6.29
CNN-LSTM	**15.65 ± 6.33**
LSTM-CNN	16.05 ± 7.40
GRU	21.10 ± 9.11
TST	17.56 ± 6.65
XceptionTime	17.30 ± 7.57
Mamba	16.02 ± 6.89
Mean	17.20 ± 7.34

In black bold and blue bold the best and second-best results, respectively.

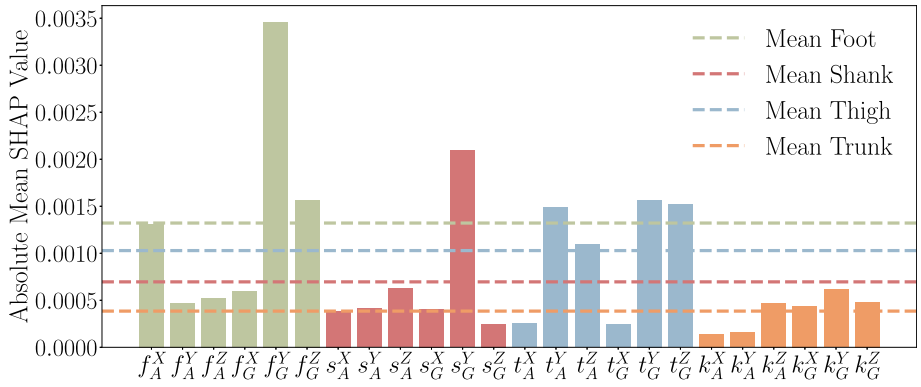


FIGURE 4
Feature importance for the IMU modality, derived by explaining the LSTM model used for the classification task (Level Ground, Stair Ascent, Stair Descent, Ramp Ascent, Ramp Descent) with the SHAP methodology. The labels x-axis are formatted as B_S^A, where B denotes the body part (foot: f, shank: s, thigh: t, trunk: k), S indicates the sensor type (accelerometer: A, gyroscope: G), and A specifies the axis (x-axis: X, y-axis: Y, z-axis: Z). The dotted horizontal lines represent the average sensor weights.

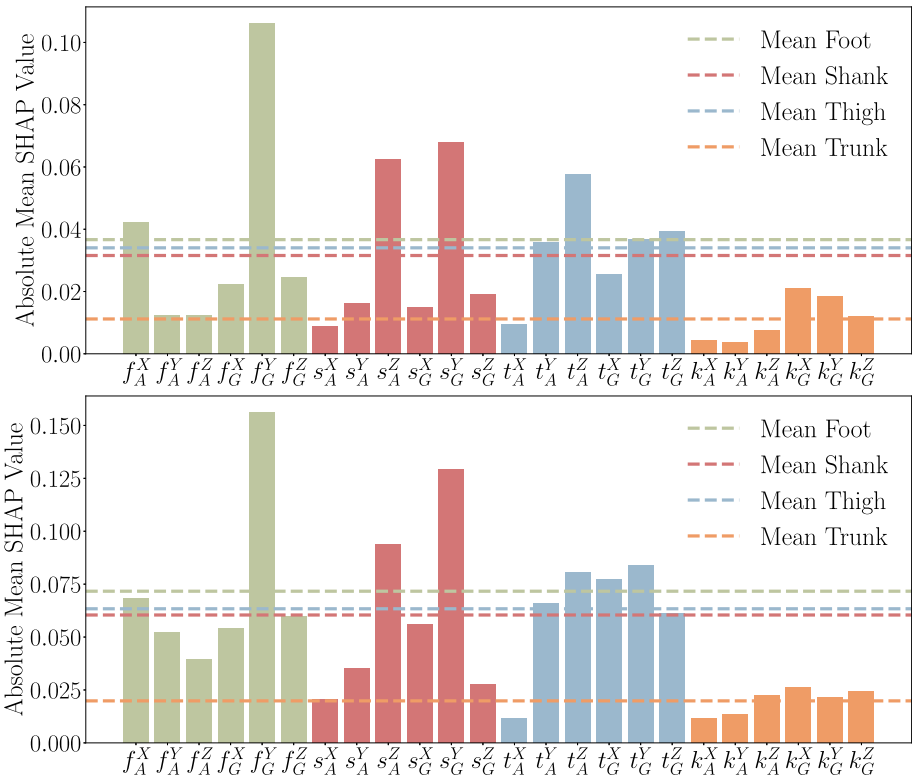


FIGURE 5
Feature importance for the IMU modality derived using the SHAP methodology, with the upper plot showing the LSTM model for slope prediction and the lower plot showing the CNN-LSTM model for stair height prediction, respectively. The features on *x*-axis are formatted as B_S^A, where *B* denotes the body part (foot: *f*, shank: *s*, thigh: *t*, trunk: *k*), *S* indicates the sensor type (accelerometer: *A*, gyroscope: *G*), and *A* specifies the axis (*x*-axis: *X*, *y*-axis: *Y*, *z*-axis: *Z*). The horizontal lines represent the average sensor weights.

than in direct terrain response, which leads to its lower contribution to ground prediction.

Focusing on the two regression tasks, Figure 5 adopts the same notation and shows the results of the SHAP analysis both for slope prediction using LSTM (upper panel) and for stair height prediction using the CNN-LSTM model (lower panel). Both plots confirm that foot sensors are the most influential of these two regressions, with the shank and thigh sensors following closely.

As expected, the trunk sensor consistently contributed the least. These insights underscore the critical role of foot sensors in capturing essential gyroscope and acceleration data during gait (Prasanth et al., 2021). Also in this case, analyzing each sensor's features, we see that the most influential is consistently the gyroscope on the *y*-axis of the foot. In second place is the thigh, where all sensors contribute equally except for the accelerometer on the *x*-axis. For what concerns the Shank the two most influential features are the acceleration on the *z*-axis and the gyroscope on the *y*-axis. The results indicate that the gyroscope on the *y*-axis of the foot is the most influential feature for predicting stair height and slope inclination, which is quite expected. This makes sense because the *y*-axis of the foot gyroscope captures rotational movement in the sagittal plane, which directly correlates with stair-climbing movements and slope changes. For the thigh, the fact that all sensors contribute equally—except for the accelerometer on the *x*-axis—is also reasonable. The thigh experiences both rotational and linear movements during stair ascent and descent, making all sensor modalities relevant. The lower influence of the *x*-axis accelerometer suggests that lateral movements of the thigh are less critical in determining stair height or slope. Regarding the shank, the dominance of the *z*-axis acceleration and the *y*-axis gyroscope aligns with expectations. The *z*-axis acceleration captures vertical displacement and impact forces, which are crucial for stair-related tasks. Meanwhile, the *y*-axis gyroscope reflects rotational motion in the sagittal plane, which is heavily involved in the adaptation to different stair heights and inclinations. Overall, these results align well with the kinematics of stair negotiation, where foot and shank dynamics play a critical role in adapting to slope and step height, while thigh movement remains more evenly distributed across sensors. The findings reinforce the importance of gyroscopic data, particularly along the *y*-axis, in capturing stair-climbing mechanics.

4.4 Minimal sensor setup

The previous analysis led us to reassess our sensor setup to simplify it without compromising performance. We evaluated four sensor combinations, starting with one IMU sensor and progressively adding sensors placed on different body districts based on their SHAP-derived importance. This iterative retraining process allowed us to assess the impact of each sensor combination on classification and regression performance.

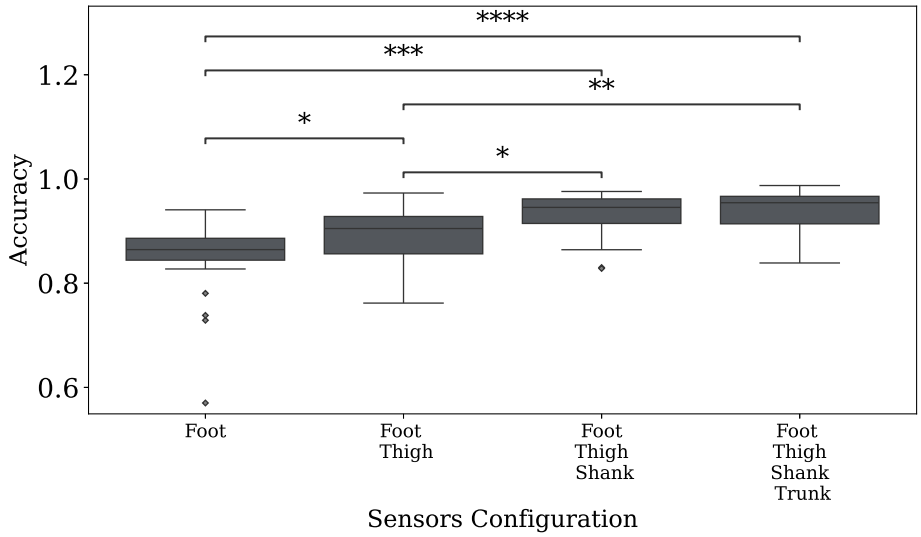


FIGURE 6
Accuracy of the LSTM model in the classification task as a function of the number of IMU sensors iteratively added based on their informativeness ranked by the SHAP methodology across the four body sectors. Statistical significance is denoted by asterisks, with *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, and ****$p \leq 0.0001$.

Figure 6 illustrates the classification accuracy of the LSTM model as a function of the IMU sensors considered. The results reveal a statistically significant difference between setups with four sensors and those with one or two sensors. However, there was no significant difference between using three sensors vs. four, suggesting that the trunk sensor does not substantially contribute to system functionality. Consequently, reducing the sensor setup to three sensors (foot, thigh, and shank) does not result in a significant performance drop, offering a more practical and cost-effective configuration.

Similarly, Figure 7 presents the MAE as a function of the IMU sensors, with sensors added iteratively based on their SHAP-derived importance for slope prediction (upper plot) and stair height prediction (lower plot). The findings indicate that excluding the trunk sensor did not significantly affect performance in either regression task, supporting the conclusion that we can exclude it from the setup.

5 Discussion

This study provides a comprehensive comparative analysis of deep learning models applied to human locomotion classification and terrain parameter estimation, particularly for lower-limb robotic exoskeletons. The results demonstrate that different architectures excel in locomotion modeling with and accuracy higher than 0.90, with LSTM emerging as the most effective model overall. However, a hybrid CNN-LSTM approach proves superior for stair height regression, highlighting the nuanced relationship between spatial and temporal dependencies in movement analysis.

LSTM's strength lies in its ability to capture long-term dependencies in sequential data, a fundamental requirement for understanding locomotion patterns. Unlike CNN, which is highly effective at learning spatial representations but lacks an inherent mechanism for handling temporal relationships, LSTM processes sequences by maintaining a memory of past information through its gated architecture. These gates regulate the flow of information, allowing the network to retain relevant features while discarding noise, which is particularly useful in gait analysis where movements evolve dynamically over time. For ground-level classification, LSTM demonstrates the highest performance due to its ability to recognize movement transitions and maintain continuity in gait sequences, making it well-suited for distinguishing between different locomotion modes.

In contrast, stair height regression benefits more from the CNN-LSTM hybrid model, which integrates the advantages of both convolutional and recurrent architectures. Stair climbing involves distinct movement phases, including lifting, pushing off, and landing, where CNN effectively extracts localized features related to these abrupt transitions. By combining CNN with LSTM, the model retains the ability to track the step-by-step progression while also refining the height estimation through convolutional feature extraction. This hybrid approach ensures that both spatial variations and long-term dependencies are adequately modeled, leading to more accurate predictions of stair height compared to an LSTM-only approach.

For slope regression, however, LSTM again outperforms other models, suggesting that gradual changes in movement mechanics are best captured by recurrent architectures. Unlike stair climbing, where discrete transitions between steps occur, slope walking involves a continuous evolution of movement patterns. The recurrent design of LSTM allows it to process these gradual shifts effectively, preserving the relationship between past and present gait states without the constraints imposed by convolutional layers. While transformers, such as the Time Series Transformer (TST) and MAMBA, offer alternative paradigms for sequential modeling through self-attention mechanisms, their advantage is more pronounced in datasets with extremely long dependencies.

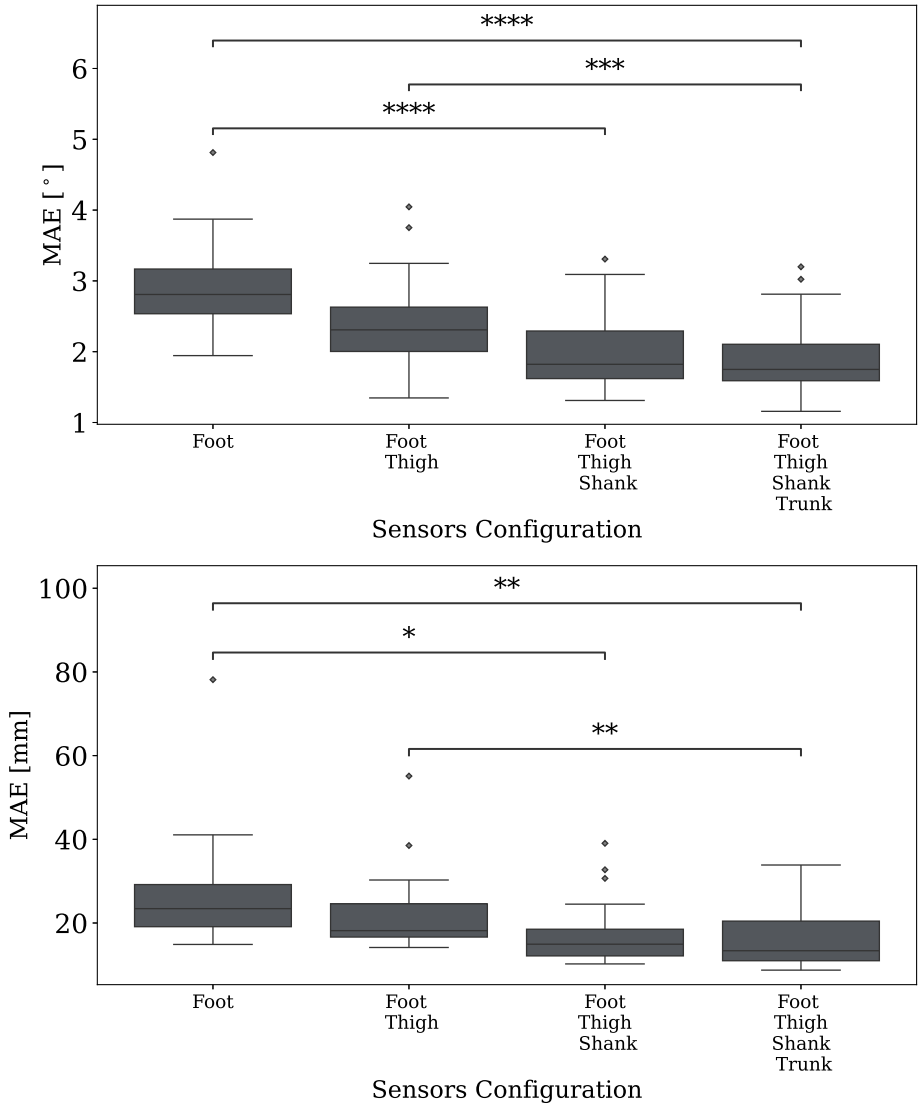


FIGURE 7
The MAE metric as a function of the number of IMU sensors iteratively added based on their informativeness ranked by the SHAP methodology across the four body sectors, with the upper plot showing slope prediction case and the lower plot showing stair height prediction case. Statistical significance is denoted by asterisks, with *$p \leq 0.05$, **$p \leq 0.01$ ***$p \leq 0.001$, and ****$p \leq 0.0001$.

In gait analysis, where temporal relationships are structured and periodic, LSTM remains a more efficient and reliable solution.

The study also examined the performance of MAMBA, XceptionTime, and GRU, which were found to be less effective in comparison to LSTM and CNN-LSTM. MAMBA, a state-space model designed to handle sequential data without recurrence, struggles with locomotion analysis due to its reliance on a different computational framework that may not be as finely tuned to the biomechanical constraints of human movement. While MAMBA excels in handling general sequential patterns, it lacks the ability to preserve structured periodic dependencies crucial for gait modeling. XceptionTime, a convolution-based model adapted for time-series tasks, faces inherent limitations due to its primary reliance on spatial feature extraction rather than the sequential nature of locomotion data. While it can learn feature hierarchies effectively, its inability to capture temporal transitions over time reduces its effectiveness in gait analysis. Similarly, GRU, despite being a recurrent model like LSTM, operates with a simpler gating mechanism that makes it computationally efficient but potentially less expressive in handling complex sequential patterns. GRU is effective at modeling short-term dependencies but may not retain sufficient context for more nuanced motion patterns that evolve over longer time spans, which explains why LSTM remains preferable in this scenario.

Beyond model selection, the study also emphasizes the critical role of sensor configuration. Although it was initially expected that a multimodal approach incorporating both IMU and EMG data would yield superior performance due to the additional physiological information from muscle activity, the findings indicate otherwise. IMU data alone proves sufficient and even more effective, suggesting that the richness of acceleration and gyroscope signals in capturing movement dynamics outweighs the benefits of

integrating EMG. The additional complexity introduced by EMG signals does not lead to a significant performance gain, reinforcing the idea that simpler sensor setups can be both practical and highly effective. Moreover, an analysis of feature importance using SHAP confirms that a minimal yet strategic sensor placement—covering the foot, shank, and thigh—achieves an optimal balance between accuracy and usability.

This study builds upon and extends the growing body of literature on terrain classification and gait parameter estimation using wearable sensor data. Compared to previous work summarized in Table 1, our approach offers several novel contributions and improved performance across multiple dimensions. First, unlike many prior studies, such as those by Negi et al. (2020) and Zhang et al. (2020) which primarily focus on unimodal approaches using either EMG or IMU signals, our work provides a comprehensive evaluation of both unimodal and multimodal setups. Importantly, we demonstrate that IMU-only models not only match but often outperform multimodal configurations (IMU+EMG), offering a more practical and cost-efficient solution for real-world robotic applications. Second, while previous efforts like (Zhao et al., 2022; Narayan et al., 2021) have explored CNN-based classification models, they generally do not extend their analysis to continuous terrain parameters such as ramp inclination or stair height. In contrast, our study uniquely integrates both classification and regression tasks—using LSTM for slope prediction and CNN-LSTM for stair height estimation—thus enhancing adaptability in real-time robotic control systems.

Third, our work improves upon methodological rigor. Several prior studies employ basic hold-out validation methods or do not clearly specify their data split strategies, as seen in Le et al. (2022). This lack of standardization limits reproducibility and complicates the evaluation of model generalizability. By employing leave-one-subject-out (LOSO) cross-validation on the publicly available CAMARGO dataset (Camargo et al., 2021), we ensure robust and interpretable performance evaluation, enabling more meaningful comparisons across studies. Moreover, our inclusion of Explainable AI (XAI) techniques via SHAP values enables a transparent evaluation of sensor importance, a feature rarely addressed in prior work. This allowed us to conduct an ablation study to identify a minimal sensor configuration—foot, shank, and thigh—without significant loss in performance. This contrasts with most of the literature, which either uses large sensor arrays or does not explore sensor minimization strategies. In terms of performance, our best classification accuracy (0.94 ± 0.04 with LSTM) is on par with or exceeds results from studies such as Kang et al. (2022) and Amer and Ji (2021), while our regression models achieve low MAE values for ramp slope ($1.93^\circ \pm 0.53$) and stair height (15.65 mm $\pm$ 6.33)—metrics that are rarely reported in the existing literature. Overall, our study provides a more holistic, efficient, and reproducible framework for terrain-aware control in wearable robotics, addressing many of the limitations observed in existing literature.

The findings of this study carry significant implications for wearable robotic exoskeletons working in unstructured environments. The ability to classify ground conditions and estimate locomotion parameters with high accuracy and minimal latency would directly influence the adaptability, safety, and usability of such systems. In real-world applications, exoskeletons should continuously interpret the environment and anticipate transitions to ensure seamless support, thereby preventing the user's destabilization or excessive cognitive burden. The ability to reliably distinguish between different locomotion contexts (e.g. level ground, ramps, stairs) would enable exoskeletons to preemptively adjust assistance strategies, optimizing the trajectories, the interaction control, and step timing. This is particularly crucial in rehabilitation, where precise adaptation to terrain variations can enhance training efficacy by encouraging more natural and effortful gait patterns. In the context of assistive applications for individuals with mobility impairments, real-time ground adaptation reduces the need for manual intervention, thereby fostering greater autonomy and reducing user fatigue.

Furthermore, the accuracy of parameter estimation (e.g., stair height, ramp inclination) impacts the granularity of control adjustments. A mismatch between perceived and actual locomotion demands could lead to either overcompensation, resulting in unnatural and inefficient movements, or undercompensation, which may compromise stability. The accuracy obtained by the proposed approach suggests that fine-grained terrain adaptation is feasible, paving the way for exoskeletons that can seamlessly modulate assistance not only across different terrains but also in response to subtle environmental variations.

The findings of this study serve to reinforce the critical role of real-time perception in the domain of wearable robotics. The reliable classification of terrain and the estimation of locomotion parameters are not merely technical benchmarks; they are fundamental enablers of more intuitive, responsive, and independent mobility assistance, thereby bridging the gap between robotic exoskeletons and natural human movement.

6 Conclusions

In this study, we have presented a neural network-based system for real-time ground condition analysis, finding that the integration of three specialized models may enhance adaptive robotic exoskeleton control. Specifically, an LSTM classifier identifies terrain type across five categories, a second LSTM estimates ramp slope, and a hybrid CNN-LSTM predicts stair height. A comparative analysis of literature models trained on the public CAMARGO dataset, together with SHAP explanations, informed the final architecture, which relies on three IMU sensors positioned on the foot, shank, and thigh to minimize sensor burden while maintaining high accuracy.

Furthermore, the use of SHAP for sensor relevance assessment revealed that the trunk sensor contributed minimally to both classification and regression tasks, enabling its removal without degrading performance. This streamlined setup maintained high accuracy in terrain classification (0.94 ± 0.04), as well as precise estimations of ramp slope ($1.95 \pm 0.58^\circ$) and stair height (15.65 ± 7.40 mm), corroborating findings in the literature. Notably, inference times of $\sim$1 ms make the system suitable for real-time applications, such as lower-limb exoskeleton control.

Contrary to the widespread assumption that multimodal sensor inputs improve performance, our experiments demonstrated that IMU data alone outperformed an IMU+EMG combination, reinforcing the efficiency of an IMU-only approach. These findings support a cost-effective and lightweight sensor configuration while maintaining robust classification and regression performance.

Future work will focus on validating the proposed system in real-world scenarios with healthy participants and developing control strategies for lower-limb exoskeletons that integrate the predicted locomotion parameters. In addition, a key future direction is the extension of this work to include individuals with neurological or musculoskeletal impairments. Studying pathological gait patterns will help determine whether the proposed models generalize to clinical populations or require adaptation to maintain effective terrain recognition and parameter estimation in assistive robotic applications. While this study focuses on data from healthy individuals, our methodology is intended for future extension to clinical populations with motor impairments. Future research should explore whether the models maintain performance on pathological gait data or require adaptation. This research contributes to the advancement of intelligent exoskeletons, offering accurate, low-latency terrain awareness for adaptive control in practical applications.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://www.sciencedirect.com/science/article/pii/S0021929021001007.

Author contributions

OC: Conceptualization, Data curation, Investigation, Methodology, Software, Visualization, Writing – original draft. CT: Conceptualization, Data curation, Investigation, Supervision, Visualization, Writing – original draft. MT: Conceptualization, Validation, Visualization, Writing – original draft. LF: Validation, Writing – original draft. RS: Conceptualization, Supervision, Validation, Writing – review & editing. LZ: Conceptualization, Formal analysis, Funding acquisition, Project administration, Supervision, Writing – review & editing. PS: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Italian Ministry of Research, under the complementary actions to the NRRP "Fit4MedRob - Fit for Medical Robotics" Grant (PNC0000007, CUP: B53C22006990001).

Acknowledgments

Coser Omar is a Ph.D. student enrolled in the National Ph.D. in Artificial Intelligence, XXXVIII cycle, course on Health and life sciences, organized by Università Campus Bio-Medico di Roma. Resources are partially provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Inffrastructure for Computing (SNIC) at Alvis @ C3SE, partially funded by the Swedish Research Council through grant agreement nos. 2022-06725 and 2018-05973. Authors thank prof. G. Iannello for his support in the research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

References

Ahamed, M. A., and Cheng, Q. (2024). TSCMAMBA: mamba meets multi-view learning for time series classification. *arXiv* [Preprint]. arXiv:2406.04419. doi: 10.48550/arXiv.2406.04419

Amer, A., and Ji, Z. (2021). Human locomotion activity recognition using spectral analysis and convolutional neural networks. *Int. J. Manuf. Res.* 16, 350–364. doi: 10.1504/IJMR.2021.119633

Anantrasirichai, N., Burn, J., and Bull, D. (2014). Terrain classification from body-mounted cameras during human locomotion. *IEEE Trans. Cybern.* 45, 2249–2260. doi: 10.1109/TCYB.2014.2368353

Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., Amirat, Y., et al. (2015). Physical human activity recognition using wearable sensors. *Sensors* 15, 31314–31338. doi: 10.3390/s151229858

Barshan, B., and Yüksek, M. C. (2014). Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. *Comput. J.* 57, 1649–1667. doi: 10.1093/comjnl/bxt075

Bartlett, H. L., and Goldfarb, M. (2017). A phase variable approach for IMU-based locomotion activity recognition. *IEEE Trans. Biomed. Eng.* 65, 1330–1338. doi: 10.1109/TBME.2017.2750139

Camargo, J., Ramanathan, A., Flanagan, W., and Young, A. (2021). A comprehensive, open-source dataset of lower limb biomechanics in multiple conditions of stairs, ramps, and level-ground ambulation and transitions. *J. Biomech.* 119:110320. doi: 10.1016/j.jbiomech.2021.110320

Caruana, R., Kangarloo, H., Dionisio, J. D., Sinha, U., and Johnson, D. (1999). "Case-based explanation of non-case-based learning methods," in *Proceedings of the AMIA Symposium* (Cambridge MA: MIT Press), 212.

Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* [Preprint]. arXiv:1412.3555. doi: 10.48550/arXiv.1412.3555

Coser, O., Tamantini, C., Soda, P., and Zollo, L. (2024). Ai-based methodologies for exoskeleton-assisted rehabilitation of the lower limb: a review. *Front. Robot. AI* 11:1341580. doi: 10.3389/frobt.2024.1341580

Díaz, I., Gil, J. J., and Sánchez, E. (2011). Lower-limb robotic rehabilitation: literature review and challenges. *J. Robot.* 2011, 1817–1820. doi: 10.1155/2011/759764

Doshi-Velez, F., and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv* [Preprint] arXiv:1702.08608. doi: 10.48550/arXiv.1702.08608

Gao, X., Zhang, P., Peng, X., Zhao, J., Liu, K., Miao, M., et al. (2023). Autonomous motion and control of lower limb exoskeleton rehabilitation robot. *Front. Bioeng. Biotechnol.* 11:1223831. doi: 10.3389/fbioe.2023.1223831

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* 52, 203–224. doi: 10.1093/biomet/52.1-2.203

Gu, A., and Dao, T. (2023). Mamba: linear-time sequence modeling with selective state spaces. *arXiv* [Preprint] arXiv:2312.00752. doi: 10.48550/arXiv.2312.00752

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D., et al. (2018). A survey of methods for explaining black box models. *ACM Comput. Surv.* 51, 1–42. doi: 10.1145/3236009

Gupta, R., and Agarwal, R. (2017). semg interface design for locomotion identification. *Int. J. Electr. Comput. Eng.* 11, 133–142. doi: 10.1007/s11062-019-09812-w

Hu, B., Rouse, E., and Hargrove, L. (2018). Benchmark datasets for bilateral lower-limb neuromechanical signals from wearable sensors during unassisted locomotion in able-bodied individuals. *Front. Robot. AI* 5:14. doi: 10.3389/frobt.2018.00014

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., and Muller, P.-A. (2019). Deep learning for time series classification: a review. *Data Min. Knowl. Discov.* 33, 917–963. doi: 10.1007/s10618-019-00619-1

Jiang, X., Liu, X., Fan, J., Ye, X., Dai, C., Clancy, E. A., et al. (2021). Open access dataset, toolbox and benchmark processing results of high-density surface electromyogram recordings. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 1035–1046. doi: 10.1109/TNSRE.2021.3082551

Jing, S., Huang, H.-Y., Vaidyanathan, R., and Farina, D. (2022). "Accurate and robust locomotion mode recognition using high-density emg recordings from a single muscle group," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (Glasgow: EEE), 686–689. doi: 10.1109/EMBC48229.2022.9871155

Kang, I., Molinaro, D. D., Choi, G., Camargo, J., and Young, A. J. (2022). Subject-independent continuous locomotion mode classification for robotic hip exoskeleton applications. *IEEE Trans. Biomed. Eng.* 69, 3234–3242. doi: 10.1109/TBME.2022.3165547

Le, D., Cheng, S., Gregg, R. D., and Ghaffari, M. (2022). Deep convolutional neural network and transfer learning for locomotion intent prediction. *arXiv* [Preprint]. arXiv:2209.12365. doi: 10.48550/arXiv.2209.12365

Lee, H., Ferguson, P. W., and Rosen, J. (2020). Lower limb exoskeleton systems—overview. *Wearable Robot.*2020, 207–229. doi: 10.1016/B978-0-12-814659-0.00011-4

Lee, T. H., Shair, E. F., Abdullah, A. R., Rahman, K. A., Ali, N. M., Saharuddin, N. Z., et al. (2025). Comparative analysis of 1d-CNN, GRU, and LSTM for classifying step duration in elderly and adolescents using computer vision. *Int. J. Robot. Control Syst.* 5, 426–439. doi: 10.31763/ijrcs.v5i1.1588

Liang, W., Wang, F., Fan, A., Zhao, W., Yao, W., Yang, P., et al. (2023). Deep-learning model for the prediction of lower-limb joint moments using single inertial measurement unit during different locomotive activities. *Biomed. Signal Process. Control* 86:105372. doi: 10.1016/j.bspc.2023.105372

Lines, J., Davis, L. M., Hills, J., and Bagnall, A. (2012). "A shapelet transform for time series classification," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: ACM), 289–297. doi: 10.1145/2339530.2339579

Lipton, Z. C. (2018). The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. *Queue* 16, 31–57. doi: 10.1145/3236386.3241340

Lundberg, S. M., and Lee, S.-I. (2017). "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems* (Red Hook, NY), 30.

Mekni, A., Narayan, J., and Gritli, H. (2025a). Multi-class classification of gait cycle phases using machine learning: a comprehensive study using two training methods. *Netw. Model. Anal. Health Inform. Bioinform.* 14, 1–20. doi: 10.1007/s13721-025-00522-4

Mekni, A., Narayan, J., and Gritli, H. (2025b). Quinary classification of human gait phases using machine learning: investigating the potential of different training methods and scaling techniques. *Big Data Cogn. Comput.* 9:89. doi: 10.3390/bdcc9040089

Molinaro, D. D., Kang, I., and Young, A. J. (2024). Estimating human joint moments unifies exoskeleton control, reducing user effort. *Sci. Robot.* 9:eadi8852. doi: 10.1126/scirobotics.adi8852

Molnar, C. (2020). *Interpretable Machine Learning*. Morrisville, NC: Lulu. com.

Müller, P. N., Müller, A. J., Achenbach, P., and Göbel, S. (2024). Imu-based fitness activity recognition using cnns for time series classification. *Sensors* 24:742. doi: 10.3390/s24030742

Napierala, M. A. (2012). What is the Bonferroni correction? *Aaos Now* (Poznań: Poznań University of Technology), 40–41.

Narayan, A., Reyes, F. A., Ren, M., and Haoyong, Y. (2021). Real-time hierarchical classification of time series data for locomotion mode detection. *IEEE J. Biomed. Health Inform.* 26, 1749–1760. doi: 10.1109/JBHI.2021.3106110

Negi, S., Negi, P. C., Sharma, S., and Sharma, N. (2020). Human locomotion classification for different terrains using machine learning techniques. *Crit. Rev. Biomed. Eng.* 48, 199–209. doi: 10.1615/CritRevBiomedEng.2020035013

O'Keeffe, B., and Rout, S. (2019). Prosthetic rehabilitation in the lower limb. *Indian J. Plast. Surg.* 52:134. doi: 10.1055/s-0039-1687919

Prasanth, H., Caban, M., Keller, U., Courtine, G., Ijspeert, A., Vallery, H., et al. (2021). Wearable sensor-based real-time gait detection: a systematic review. *Sensors* 21:2727. doi: 10.3390/s21082727

Rahimian, E., Zabihi, S., Atashzar, S. F., Asif, A., and Mohammadi, A. (2019). Xceptiontime: a novel deep architecture based on depthwise separable convolutions for hand gesture classification. *arXiv* [Preprint]. arXiv:1911.03803. doi: 10.48550/arXiv.1911.03803

Ramírez-Mena, A., Andrés-León, E., Alvarez-Cubero, M. J., Anguita-Ruiz, A., Martinez-Gonzalez, L. J., and Alcala-Fdez, J. (2023). Explainable artificial intelligence to predict and identify prostate cancer tissue by gene expression. *Comput. Methods Programs Biomed.* 240:107719. doi: 10.1016/j.cmpb.2023.107719

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: ACM), 1135–1144. doi: 10.18653/v1/N16-3020

Ribeiro, N. F., and Santos, C. P. (2017). "Inertial measurement units: a brief state of the art on gait analysis," in *2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG)* (Coimbra: IEEE), 1–4. doi: 10.1109/ENBENG.2017.7889458

Sadeghzadehyazdi, N., Batabyal, T., and Acton, S. T. (2021). Modeling spatiotemporal patterns of gait anomaly with a cnn-lstm deep neural network. *Expert Syst. Appl.* 185:115582. doi: 10.1016/j.eswa.2021.115582

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., et al. (2017). "Grad-cam: visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision* (Venice: IEEE), 618–626. doi: 10.1109/ICCV.2017.74

Shi, L.-F., Yan, X., Zhou, W., and Shi, Y. (2024). Simple and efficient step detection algorithm for foot-mounted imu. *Meas. Sci. Technol.* 36:016302. doi: 10.1088/1361-6501/ad7f78

Simonyan, K., Vedaldi, A., and Zisserman, A. (2013). Deep inside convolutional networks: visualising image classification models and saliency maps. *arXiv* [Preprint]. arXiv:1312.6034. doi: 10.48550/arXiv.1312.6034

Son, C.-S., and Kang, W.-S. (2023). Multivariate cnn model for human locomotion activity recognition with a wearable exoskeleton robot. *Bioengineering* 10:1082. doi: 10.3390/bioengineering10091082

Sorkhabadi, S. M. R., Chinimilli, P. T., Gaytan-Jenkins, D., and Zhang, W. (2019). "Human locomotion activity and speed recognition using electromyography based features," in *2019 Wearable Robotics Association Conference (WearRAcon)* (Scottsdale, AZ: IEEE), 80–85. doi: 10.1109/WEARRACON.2019.8719626

Su, B., and Gutierrez-Farewik, E. M. (2020). Gait trajectory and gait phase prediction based on an lstm network. *Sensors* 20:7127. doi: 10.3390/s20247127

Tortora, M., Cordelli, E., Sicilia, R., Nibid, L., Ippolito, E., Perrone, G., et al. (2023). Radiopathomics: multimodal learning in non-small cell lung cancer for adaptive radiotherapy. *IEEE Access* 11, 47563–47578. doi: 10.1109/ACCESS.2023.3275126

Valgeirsdóttir, V. V., Sigurardóttir, J. S., Lechler, K., Tronicke, L., Jóhannesson, Ó. I., Alexandersson, Á., et al. (2022). How do we measure success? A review of performance evaluations for lower-limb neuroprosthetics. *J. Prosthet. Orthot.* 34, e20–e36. doi: 10.1097/JPO.0000000000000355

Wang, B., Zheng, J., Gao, Y., Wang, Y., and Wu, G. (2022). "Locomotion mode recognition method based on inertial measurement units in exoskeleton robot," in *International Symposium on Robotics, Artificial Intelligence, and Information Engineering (RAIIE 2022)*, *Vol. 12454* (Piscataway, NJ: IEEE), 231–237. doi: 10.1117/12.2659268

Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., and Eickhoff, C. (2021). "A transformer-based framework for multivariate time series representation learning," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (New York, NY: ACM), 2114–2124. doi: 10.1145/3447548.3467401

Zhang, K., Wang, J., de Silva, C. W., and Fu, C. (2020). Unsupervised cross-subject adaptation for predicting human locomotion intent. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 646–657. doi: 10.1109/TNSRE.2020.2966749

Zhao, C., Lu, X., Zhang, T., Feng, Y., and Chen, W. (2022). "Multi-channel separated encoder based convolutional neural network for locomotion intention recognition," in *2022 34th Chinese Control and Decision Conference (CCDC)* (Hefei: IEEE), 2768–2773. doi: 10.1109/CCDC55256.2022.10034195

Zhao, H., Qiu, S., Wang, Z., Yang, N., Li, J., Wang, J., et al. (2019). "Applications of mems gyroscope for human gait analysis," in *Gyroscopes-Principles and Applications* (London: IntechOpen). doi: 10.5772/intechopen.86837

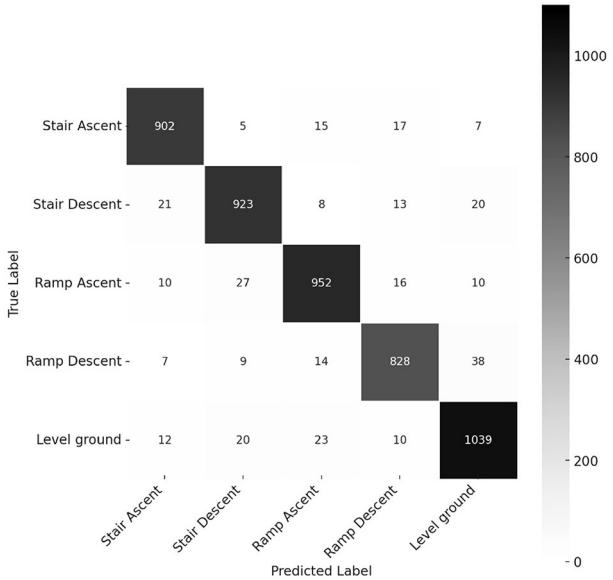
Zheng, J., Peng, M., Huang, L., Gao, Y., Li, Z., Wang, B., et al. (2022). A cnn-svm model using imu for locomotion mode recognition in lower extremity exoskeleton. *J. Mech. Med. Biol.* 22:2250043. doi: 10.1142/S0219519422500439

10.3389/fcomp.2025.1597143

Appendix

Confusion matrix for subjects: 10, 11, 12 that exhibits the mean accuracy (0.94) for the LSTM

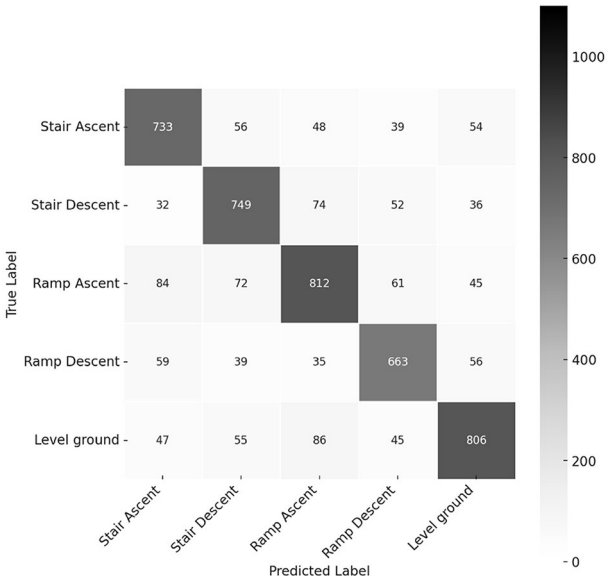
- **Total samples:** 4946
- **Class distributions (true labels):**
  - Class 1: 950
  - Class 2: 978
  - Class 3: 1010
  - Class 4: 908
  - Class 5: 1100



Rows represent **true classes**, columns represent **predicted classes**. Class order: Stair Acent, Stair Descent, Ramp Ascent, Ramp Descent, Level Ground.

Confusion matrix for subjects: 12, 13, 14 of the GRU that shows the mean accuracy (0.78)

- **Total samples:** 4,846
- **Class distributions (true labels):**
  - Class 1: 939
  - Class 2: 956
  - Class 3: 1,047
  - Class 4: 879
  - Class 5: 1,025



Rows = **True Classes**, Columns = **Predicted Classes**. Class order: Stair Acent, Stair Descent, Ramp Ascent, Ramp Descent, Level Ground.