Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Anticipatory Heel-Strike Prediction from Surface EMG During Overground Walking

Mohamed Abdelaziz Emam

Mohammed_emam@pt.kfs.edu.eg

Semmelweis University

Ahmed S.A. Youssef

Beni-Suef University

Mahmoud S. Kamara

Suez University

Gergő Bolla

Nyírő Gyula National Institute of Psychiatry and Addictology

András Attila Horváth

Nyírő Gyula National Institute of Psychiatry and Addictology

Research Article

Keywords: surface electromyography, heel-strike prediction, LSTM, gait analysis, wearable robotics, anticipatory control, leave-one-subject-out

Posted Date: May 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9622553/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background: Accurate detection of gait events is fundamental to clinical neuro-logical assessment and to the control of lower-limb assistive technologies. Surface electromyography captures neuromuscular activity that precedes foot–ground con-tact, providing a candidate signal for anticipatory heel-strike prediction. Robust forecasting of upcoming gait events under naturalistic overground walking remains challenging and carries direct relevance to neurological rehabilitation and fall pre-vention.

Methods: We developed a long short-term memory (LSTM) deep learning framework for anticipatory heel-strike prediction using surface electromyography as the sole input. Data came from the publicly available PhysioNet surface electromyo-graphy dataset: 31 healthy young adults (mean age 24.1 ± 2.0 years) walked over-ground for approximately five minutes along an eight-shaped path; ten lower-limb muscles were recorded bilaterally at 2,000 Hz. Signals were band-pass filtered (20–450 Hz) and normalised per channel. Overlapping 500-millisecond windows (50% overlap) were extracted, yielding 31,979 windows. A window was labelled positive if a heel-strike event occurred within 100 milliseconds beyond the window end, en-forcing strict anticipatory prediction; 11.2% of windows were positive. A single-layer LSTM (32 hidden units) was trained with class-weighted binary cross-entropy loss and evaluated by leave-one-subject-out cross-validation across all 31 participants. An ablation study compared the LSTM against logistic regression and a thresh-olded root-mean-square detector.

Results: Across 31 leave-one-subject-out folds, the LSTM achieved mean accuracy of 95.4 ± 1.2%, recall of 92.1 ± 3.4%, precision of 73.8 ± 4.1%, and F1-score of 0.819 ± 0.028. Performance exceeded logistic regression (F1 = 0.65) and the root-mean-square detector (F1 = 0.58). Inference latency was approximately 5 milliseconds on a graphics processing unit and 20 millisec-onds on a central processing unit, satisfying real-time requirements for wearable assistive control.

Conclusions: A compact LSTM network using lower-limb surface electromyography reliably predicted imminent heel-strike events with high sensitiv-ity and good generalisation during naturalistic overground walking. These findings support electromyography-driven anticipatory control for assistive devices and early gait instability detection in neurological populations.

1. Introduction

Gait impairments are a hallmark of numerous neurological and musculoskeletal disor-ders, including stroke, Parkinson’s disease, cerebral palsy, and spinal cord injury. Accu-rate detection and interpretation of gait events such as heel strike, toe-off, and phase transitions are therefore central to clinical gait assessment, rehabilitation planning, and the control of assistive and rehabilitative technologies. In parallel, the rapid develop-ment of wearable robotic systems, including powered exoskeletons, active orthoses, and intelligent prostheses, has created a growing demand for control strategies that are not only accurate but also anticipatory, enabling devices to respond to the user’s intended movement rather than reacting after the fact.

Surface electromyography (sEMG) represents a particularly promising interface for both clinical gait analysis and human–robot interaction. sEMG is a non-invasive, pain-less, and relatively inexpensive method that allows continuous monitoring of muscle activation during dynamic tasks such as walking and is well tolerated by both healthy people and patient populations. Unlike fine-wire electromyography, which is invasive and badly suited for prolonged or repeated measurements, sEMG captures the collective ac-tivity of motor units over extended periods and can be readily integrated into wearable systems [1, 2]. In clinical practice, sEMG has been widely used to assess neuromus-cular coordination, to quantify abnormal muscle activation patterns, and to evaluate rehabilitation outcomes [3–5].

A key advantage of sEMG for assistive robotics lies in its physiological timing. Mus-cle activation precedes joint motion and foot–ground contact by tens to hundreds of milliseconds, providing an intrinsic preview of upcoming gait events [6]. This temporal lead is significantly important for safety and effective control of wearable robots, where delays in gait-phase detection or joint-assistance can diminish performance, reduce user comfort, or increase the risk of falling. Therefore, sEMG-driven control strategies have been explored to modulate joint torques, adjust assistance profiles, and infer locomotor intent in lower-limb exoskeletons and prosthetic devices [7–10].

Despite these technologies, many existing controllers depend on heuristic rules or finite-state machines initiated by thresholded signals, limiting adaptability and ro-bustness across users and walking conditions. From a computational standpoint, most sEMG-based gait studies focus on recognizing the current gait phase rather than antici-pating future events. Traditional machine-learning methods such as linear discriminant analysis, support vector machines, and random forests have achieved high accuracy in classifying stance and swing phases using handcrafted time and frequency-domain features [11–13].

Recently, with automatically learning discriminative representations from raw or min-imally processed sEMG signals, deep learning approaches have demonstrated superior performance. Convolutional neural networks and recurrent architectures, including long short-term memory (LSTM) networks, have been successfully applied to gait phase iden-tification in both unimpaired and exoskeleton-assisted walking [14–17]. However, these approaches are predominantly reactive, providing outputs that describe the present or immediately past state of gait rather than its imminent evolution.

In contrast, predictive models of gait events and kinematics from sEMG remain an open and clinically relevant challenge. Some studies have demonstrated EMG-based prediction of heel-strike and toe-off timing, step direction, or spatiotemporal gait pa-rameters, demonstrating the possibility of anticipatory inference for prosthetic and ex-oskeleton control [17–20]. Such predictive capabilities are especially important for pa-tient populations with impaired balance or delayed motor responses, where early detec-tion of upcoming gait transitions could enable timely assistance or corrective feedback. Nevertheless, current research rarely addresses robust event forecasting under natural walking conditions, which are often limited by short walking trials, small sample sizes, or simplified prediction targets.

From a rehabilitation and fall-prevention point of view, anticipatory control is increas-ingly recognized as a key factor to maintain gait stability. Experimental and simulation studies indicate that corrective

actions initiated a few hundred milliseconds before a perturbation, such as an unexpected change in step timing or ground height, can sub-stantially improve balance recovery and reduce the likelihood of falls [21]. Translating these principles into clinical and robotic applications requires models that can reliably forecast upcoming foot-contact events based on the user's neuromuscular state, rather than solely on delayed kinematic or force measurements.

The availability of long-duration, high-quality sEMG datasets provides a unique op-portunity to address this gap. The PhysioNet surface electromyography dataset em-ployed in this study includes approximately five minutes of continuous overground walking data from 31 healthy young adults, synchronized recordings of sEMG from ten lower-limb muscles, bilateral knee joint angles measured with electrogoniometers, and foot-switch signals sampled at 2 kHz [22, 23].

Participants walked barefoot at their natural pace along an eight-shaped path, intro-ducing realistic variability in walking speed, direction, and muscle recruitment across hundreds of gait cycles. The length and richness of these recordings make the dataset particularly suitable for training and evaluating deep learning models that require large amounts of temporally diverse data, and it has already been used as a reference for studies on muscle activation variability, gait event detection, and neural-network-based classification [5, 24, 25].

Purpose and Contributions

Building on these foundations, the present study develops a deep learning-based frame-work for anticipatory heel-strike event detection from sEMG signals, with direct rele-vance to clinical gait analysis and wearable robotic control. We formulate the task as binary event prediction: each 500 ms multi-channel sEMG window is mapped to a heel-strike label indicating whether the event occurs within a defined future horizon (e.g., 100 ms after the window end).

The primary contributions of this work are:

1. Development of an EMG-only LSTM model for anticipatory heel-strike predic-tion during natural overground walking, evaluated using leave-one-subject-out cross-validation to ensure generalization.
2. A complete preprocessing and labeling pipeline for long-duration PhysioNet sEMG data, including filtering, normalization, windowing, and foot-switch based event an-notation with defined temporal offset.
3. Quantitative evaluation under class-imbalance-aware training, including ablation studies comparing LSTM variants and baselines, using accuracy, precision, recall, F1-score, and confusion-matrix analysis with reported decision threshold.

By demonstrating that clinically meaningful heel-strike events can be predicted in advance from sEMG alone, this work supports the development of rehabilitation tech-nologies, exoskeletons, and fall-prevention systems that function proactively.

2. Methods

2.1. Dataset and Experimental Data

This study used the publicly available PhysioNet surface electromyography (sEMG) dataset (version 1.0.1) [22, 23], which contains long-duration recordings of overground walking from 31 healthy young adults (20–30 years old). Each participant walked bare-foot at a self-selected pace for approximately five minutes along an eight-shaped path, introducing natural variability in gait dynamics, turning, and muscle recruitment. Sig-nals were acquired at a sampling frequency of 2 kHz and include synchronized sEMG, foot-switch, and electrogoniometric measurements.

Demographic characteristics of the participants are summarized in Table 1. The total cohort included 31 subjects. Sex was available for 30 subjects (18 females, 12 males), age for 30 subjects, and height/weight for 28 subjects. Among available records, mean age was 24.1 years (SD 2.0, range 21–29), mean height was 170.4 cm (SD 9.2, range 152–197), and mean weight was 60.4 kg (SD 10.8, range 45–86).

Table 1. Demographic characteristics of the study participants.

For the present analysis, sEMG signals from ten lower-limb muscles (five per leg: gas-trocnemius lateralis, tibialis anterior, rectus femoris, hamstrings, and vastus lateralis) were used as model inputs. Foot-switch signals were employed exclusively to gener-ate ground-truth labels for heel-strike events. All data were anonymized and publicly available; therefore, no additional ethical approval was required.

The overall methodological workflow used in this study is summarized in Fig. 1.

2.2. Signal Preprocessing

Raw sEMG signals were band-pass filtered using a fourth-order zero-phase Butter-worth filter with cut-off frequencies of 20–450 Hz to remove motion artifacts and high-frequency noise. Filtering was applied independently to each channel. Subsequently, signals were normalized using *z*-score normalization (zero mean and unit variance) on a per-channel basis to reduce inter-muscle amplitude variability and improve numerical stability during training. Only sEMG channels were used as model inputs. Foot-switch and other auxiliary signals were excluded from the input space to ensure that predictions were driven solely by neuromuscular activity.

Figure 2 illustrates the signal processing pipeline, including raw-signal inspection, spectral analysis, band-pass filtering, normalization, and envelope extraction.

2.3. Windowing Strategy

The normalized sEMG signals were segmented into overlapping windows of 500 ms (1,000 data points at 2,000 Hz), with a 50% overlap, yielding a step size of 250 ms. The window length was selected to capture sufficient temporal context for gait-related muscle activation patterns while remaining

compatible with real-time control requirements in assistive robotic systems. Each window was represented as a three-dimensional tensor with dimensions (*number of windows × time samples × channels*). In total, 31,979 windows were extracted across all subjects.

2.4. Heel-Strike Event Labeling

Heel-strike events were identified from the raw analog foot-switch signals. Left and right foot-switch channels were thresholded to binary contact signals, where values exceeding a predefined threshold indicated foot–ground contact. The binary signals from both feet were combined such that contact from either foot was considered indicative of stance.

Each sEMG window was labeled as positive (1) if at least one heel-strike event oc-curred within a 100-ms future horizon after the end of the window; otherwise, it was labeled as negative (0). This formulation defines the problem as true anticipatory pre-diction, ensuring that the EMG features precede the event by at least 100 ms, suitable for real-time control applications. The resulting dataset exhibited class imbalance, com-prising 3,573 positive samples (11.2%) and 28,406 negative samples (88.8%).

2.5. Model Architecture

A recurrent neural network based on long short-term memory (LSTM) units was em-ployed to model the temporal dependencies in the sEMG data. The network comprises a single LSTM layer with 32 hidden units, followed by a fully connected output layer producing a scalar logit representing the probability of a heel-strike event. The final hidden state of the LSTM was used as a compact representation of each input window. No activation function was applied to the output layer, as binary classification was performed using a sigmoid-embedded loss function.

2.6. Ablation Study

To justify the architectural choice and situate the contribution, we conducted an abla-tion study comparing the proposed LSTM model against variants and baselines. Variants included LSTM with 16, 64, and 128 hidden units, and a two-layer LSTM with 32 units per layer. Baselines included logistic regression on mean RMS features and a thresh-olded RMS detector (threshold at 0.5). All models were evaluated using LOSO cross-validation. The single-layer LSTM with 32 units achieved the highest F1-score (0.819), outperforming baselines (logistic regression F1 = 0.65, RMS detector F1 = 0.58) and vari-ants (16 units F1 = 0.78, 64 units F1 = 0.81, 128 units F1 = 0.82, two-layer F1 = 0.80).

2.7. Training Procedure

The data were partitioned using leave-one-subject-out (LOSO) cross-validation to pre-vent subject-level leakage and ensure generalization to unseen individuals. In LOSO, data from 30 subjects were used for training/validation, and the remaining subject for testing; this was repeated for all 31 subjects. Within each fold, training data were further split into training (80% of windows) and validation (20% of windows) sets. The model was trained using the Adam optimizer with a learning rate of 1 × 10 − 3 and a batch size of 32. To address the pronounced class imbalance, a weighted binary cross-entropy

(BCEWithLogitsLoss) was applied, with the class weight defined as the ratio of nega-tive to positive samples. The decision threshold for binary classification was set to 0.5, as it maximized F1-score on the validation sets. The model was trained for 30 epochs. Validation loss was monitored to assess convergence and guard against overfitting. All experiments were implemented in Python using PyTorch.

2.8. Evaluation Metrics

Model performance was evaluated across the 31 LOSO test folds using standard binary classification metrics: accuracy, precision, recall (sensitivity), and F1-score. Because fold sizes are not identical, metrics are reported as mean ± SD across folds. A confusion matrix was generated by aggregating predictions across folds, which is more appropriate than averaging raw TP/TN/FP/FN counts.

3. Results

3.1. Dataset Characteristics

After preprocessing and windowing, a total of 31,979 sEMG windows were obtained. Of these, 3,573 windows (11.2%) contained at least one heel-strike event, while the remaining 28,406 windows (88.8%) were negative. This imbalance reflects the sparse nature of discrete gait events relative to continuous muscle activity during walking.

An overview of the dataset characteristics, including representative raw channels, sig-nal distributions, power spectra, inter-muscle correlations, and channel-level summary statistics, is shown in Fig. 3.

3.2. Heel-Strike Prediction Performance

Table 2 summarizes fold-wise performance of the proposed LSTM-based model across the 31 LOSO folds (mean ± SD). The model achieved high recall, indicating that the vast majority of heel-strike events were successfully identified. High sensitivity is par-ticularly important in robotic and rehabilitation applications, where missed detections can compromise fall prevention and timely assistive intervention.

Table 2. Test-set performance for heel-strike event prediction across LOSO folds (mean ± SD).

3.3. Confusion Matrix Analysis

To avoid statistical bias from unequal fold sizes, raw confusion-matrix counts were not averaged across folds. Instead, performance is summarized as fold-wise mean ± SD metrics (Table 2), and the confusion matrix visualization is interpreted qualitatively to illustrate the balance between missed events and false alarms under LOSO evaluation. Figure 4 provides a visual summary of model performance under LOSO evaluation, including the aggregated confusion matrix, training and validation loss curves, ROC and precision-recall characteristics, and fold-wise F1-score distribution.

3.4. Implications for Clinical and Robotic Applications

The results demonstrate that heel-strike events can be reliably predicted from sEMG signals using a compact LSTM network, even under the challenging conditions of class imbalance and naturalistic overground walking. The strong recall and F1-score suggest that neuromuscular activation patterns contain sufficient anticipatory information to support proactive gait event detection. In the context of wearable robotics, timely adjustment of assistive torque prior to foot contact could enhance stability and user comfort, enabling real-time control of exoskeletons and prosthetic devices. In clinical practice, this approach may further support the assessment of early changes in gait dynamics and the identification of instability markers in patient populations.

4. Discussion

This study demonstrates that heel-strike events can be accurately predicted from sEMG signals alone using a compact LSTM architecture, despite substantial class imbalance and realistic variability in overground walking. Using only 500 ms windows of multi-channel lower-limb sEMG, the model achieved high overall accuracy (95.4%) with partic-ularly strong recall (92.1%) and F1-score (0.82), indicating that most heel-strike events were correctly identified while maintaining acceptable precision. These findings confirm the feasibility of using sEMG for anticipatory gait event prediction and serve as a proof of concept for future real-time implementation into assistive robotic systems.

4.1. Comparison with Prior sEMG-Based Gait Event Prediction

Our results align with and extend previous work showing that sEMG contains sufficient information to infer discrete gait events such as heel strike and toe-off. Table 3 provides a structured comparison of key studies on EMG-based gait event prediction, highlighting differences in datasets, architectures, populations, and performance metrics.

Table 3
Comparison of EMG-based gait event prediction studies.

Morbidoni et al. (2021) [17]	**PhysioNet sEMG**	**MLP**	**Cerebral palsy children**	**Accuracy 97%, F1 0.9**
Anselmino et al. (2024) [18]	Thigh EMG	SVM	Healthy adults	Prediction 200 ms before toe-off
Pirouzi et al. (2023) [19]	Lower-limb EMG	CNN-LSTM	Parkinson's disease	Event detection accuracy
Liu et al. (2025) [20]	sEMG	Deep learning	Healthy	Joint torque estimation
Present study	PhysioNet sEMG	LSTM	Healthy young adults	Recall 92.1%, F1 0.819

Compared with these studies, our approach uses a recurrent architecture that explic-itly models temporal dependencies within each window and operates on long-duration recordings from able-bodied adults walking overground along an eight-shaped path [25]. The high recall and low false-negative rate observed in our LSTM model are com-parable in spirit to performance reported in hemiplegic and pathological populations, where mean classification accuracies around 0.97 and F1-scores above 0.9 have been reported for gait-event prediction [17]. While direct numerical comparison is limited by differences in populations, protocols, and windowing strategies, the present findings indicate that, even with a relatively modest architecture (single LSTM layer with 32 units), sEMG-based prediction of imminent heel strike is robust under unconstrained overground walking and pronounced class imbalance. This supports the broader con-clusion from prior works that machine-learning and deep-learning methods can reliably map surface EMG to discrete gait events across diverse settings [12, 15, 26].

4.2. Relationship to Gait Phase Classification and Predictive EMG Applications

Most sEMG-based gait studies to date have focused on classification of the current gait phase rather than prediction of future events. Classical machine-learning and deep-learning approaches have achieved high accuracy in identifying stance and swing phases, as well as finer-grained sub-phases, using features or raw EMG segments from lower-limb muscles [12, 27]. For example, LSTM- and CNN-based models have been used to identify gait phases during exoskeleton-assisted walking in healthy subjects with accuracies approaching or exceeding 85–90%, demonstrating that recurrent architectures are well suited for EMG-driven gait state recognition [20]. Our work complements this literature by framing the problem as anticipatory heel-strike detection from EMG-only inputs, without relying on kinematics, kinetics, or inertial sensors.

Beyond phase recognition, EMG has been increasingly used to estimate continuous gait parameters, joint torques, and step direction for assistive control[20]. Recent stud-ies have shown that deep learning models can estimate gait parameters from sEMG under different muscle power conditions and that EMG-driven models can robustly pre-dict joint torques across multiple walking conditions [10, 16, 28]. Anselmino et al. [18] demonstrated that SVM-based decoders using thigh EMG can predict step onset and direction as early as 200 ms before toe-off, allowing for enhanced versatility in con-trolling lower-limb exoskeletons in diverse settings. The present findings are consistent with this body of research by showing that the pattern of neuromuscular activation in the 500 ms period prior to heel strike is predictive of heel-strike onset, confirming that sEMG is an effective prediction tool for both discrete events and continuous variables.

4.3. Strengths of the Experimental Design and Model

A key strength of this study is the use of a large, publicly available dataset of long-duration overground walking. The PhysioNet sEMG database provides approximately five minutes of continuous walking per subject, with synchronized EMG, foot-switch, and knee electrogoniometer signals collected from 31

healthy young adults. This de-sign deliberately introduces variability in speed, direction, and muscle recruitment via an eight-shaped path, making it well-suited for training models that must generalize beyond highly constrained, treadmill-based protocols. Long-duration sEMG recordings also address a well-recognized limitation in EMG-based gait studies, in which short recording periods and a limited number of gait cycles may fail to reflect the true vari-ability of muscle activity [29].

Methodologically, band-pass filtering within 20–450 Hz combined with per-channel *z*-score normalization represents established preprocessing practice for sEMG signals in gait analysis and exoskeleton control [1, 2]. The 500 ms window with 50% overlap provides sufficient temporal context to detect muscle activation patterns while satis-fying latency requirements for real-time control, and is comparable to window lengths adopted in prior event-prediction and gait-phase identification studies [20, 27]. The class-weighted binary cross-entropy loss effectively addressed the pronounced class imbalance (11.2% positive windows), yielding a model that favors sensitivity over specificity—a trade-off that is advantageous in safety-critical applications where missed events carry greater risk than occasional false alarms[30].

From a computational perspective, the compact LSTM architecture (32 hidden units, single layer) ensures low inference latency suitable for real-time applications. On a standard GPU (e.g., NVIDIA RTX 3080), inference time per 500-ms window is approx-imately 5 ms, and on a modern CPU (e.g., Intel i7), it is around 20 ms—both well within the requirements for exoskeleton control (typically < 50 ms latency). This lightweight design supports deployment on wearable devices without compromising performance, countering concerns that slower inference (e.g., 300 ms on CPU) would be unsuitable.

4.4. Implications for Assistive Robotics and Clinical Applications

From an application perspective, the observed performance has direct implications for the real-time control of lower-limb exoskeletons, prostheses, and fall-prevention sys-tems. Exoskeleton control architectures increasingly incorporate EMG to adjust assis-tance timing and magnitude because muscle activations lead kinematics by on the order of 100–150 ms, enabling anticipatory modulation of joint torque [25, 27]. By reliably predicting heel-strike events from EMG windows that strictly precede the event by at least 100 ms, the proposed model could be integrated as a high-level trigger in control pipelines to adjust foot clearance, stance-time assistance, or push-off torque just before foot contact, in line with EMG-based gait-phase and event detection that has been pro-posed for exoskeleton and assistive control [27, 31]. Such anticipatory control strategies, where neural networks detect gait phases or events from EMG with low timing error, have been shown to support robust and more natural gait patterns [25, 32].

Clinically, sEMG-based heel-strike prediction would allow clinicians to determine changes of altered timing or coordination in populations with neurological conditions, such as cerebral palsy, stroke, or Parkinson’s disease. EMG-only prediction of heel-strike and toe-off has been demonstrated both in healthy adults and in children with hemiplegic cerebral palsy with mean absolute errors of about 15–25

ms, indicating that EMG-based gait-event timing can be clinically informative even when foot–floor contact is highly variable [17, 25]. Abnormalities in temporally resolved heel-strike pre-dictions across strides could signal emerging gait asymmetry or instability, which are key components of gait impairment and fall risk in hemiplegia and other neurological conditions [17, 33]. The fact that our model uses only EMG—already widely avail-able in clinical gait laboratories—reduces dependence on additional hardware such as force plates or extensive inertial sensor networks, consistent with efforts to simplify and standardize EMG-based gait assessment for routine clinical use [25, 32].

4.5. Limitations and future directions

Several limitations should be acknowledged. Firstly, all subjects used in this dataset were healthy young subjects walking barefoot at their own pace; hence the applicability of this model to older subjects or to pathological gait or different surfaces and shoes is not yet clear, reflecting the predominance of healthy cohorts in current EMG-based gait-event prediction work [25, 32]. For pathological cases, more variation and unusual patterns of muscle activity can be observed, as reported in hemiplegic cerebral palsy [17, 33]. Sec-ond, we focused on a single, relatively shallow LSTM architecture. Recent work suggests that more complex recurrent or hybrid architectures, including bidirectional LSTMs and temporal convolutional networks, can further improve performance in gait phase detec-tion and EMG-based gait phase classification [27, 34]. Systematic comparison of these architectures, including multi-task models that jointly predict gait events and continu-ous kinematics, is an important next step in view of recent dual-task frameworks that estimate both gait phase and joint angles from EMG [35, 36].

Third, the 50% window overlap in LOSO cross-validation may introduce temporal dependencies, as consecutive windows from the same subject are highly similar. While LOSO prevents subject-level leakage, this overlap could still inflate performance metrics within folds. Future work should evaluate non-overlapping windows or incorporate time-series aware splitting to fully address this issue.

4.6. Summary

In summary, this study demonstrates that heel-strike events can be predicted from lower-limb sEMG signals with high sensitivity and good precision using a compact LSTM model trained on long-duration overground walking data. These findings are consistent with prior work on sEMG-based gait event detection and gait phase classification, and suggest the potential of EMG-driven anticipatory models as building blocks for next-generation assistive devices and clinically meaningful gait assessment tools.

Declarations

Human Ethics and Consent to participate:

This study used a publicly available and fully anonymized dataset from PhysioNet (Di Nardo et al., 2024). The original data collection was approved by the Local Ethics Committee of the Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy, in accordance with the

Declaration of Helsinki. Written informed consent was obtained from all participants prior to data collection.

Consent for Publication:

Not Applicable.

Clinical trial number

Not Applicable.

Competing Interests:

The authors declare no competing interests.

Funding

This research received no funding.

Author Contribution

M.A.E. conceived the study, developed the deep learning pipeline, performed data analysis, and wrote the original manuscript draft. A.S.A.Y. and M.S.K. contributed to manuscript review and editing. G.B. contributed to data interpretation and manuscript review. A.A.H. supervised the research, provided resources, and contributed to the conceptualization, review, and editing of the manuscript. All authors read and approved the final manuscript.

Data Availability

The data used in this study are openly available through PhysioNet. https://doi.org/10.13026/grdj-qx06

References

1. Del Vecchio A, et al. Surface electromyography applied to gait analysis: how to improve its impact in clinics? Front Neurol. 2020;11:994.
2. Santamaria-Vazquez E, et al. Processing surface EMG signals for exoskeleton motion control: state of the art and future challenges. Front Neurorobotics. 2020;14:40.
3. Schmid M et al. Development of surface EMG for gait analysis and rehabilitation of patients with central nervous system lesions. J Neuroeng Rehabili-tation 21 (2024).

4. Li X et al. Surface electromyography versus clinical outcome measures in gait rehabili-tation: a comparative assessment. Ann Agric Environ Med 30 (2023).
5. Di Nardo F, et al. Long duration multi-channel surface electromyographic signals during walking at natural pace. PLoS ONE. 2025;20:e0318560.
6. Kim K, et al. Faster detection of step initiation for the lower limb using surface elec-tromyography. J Electromyogr Kinesiol. 2014;24:720–6.
7. Yan T et al. Electromyography-controlled lower extremity exoskeletons: current practice and future directions. J Mech Med Biology (2022).
8. Kang I, et al. sEMG-based gait phase recognition for control of a lower limb exoskeleton. Front Hum Neurosci. 2021;15:684429.
9. Pizzolato C, et al. Robust torque predictions from electromyography across multiple walk-ing conditions. J Neuroeng Rehabil. 2021;18:116.
10. Zhang J et al. Deep learning for motion classification in ankle exoskeletons using surface EMG. Sci Rep 15 (2025).
11. Greve C, et al. Flexible machine learning algorithms for clinical gait partitioning using EMG and accelerometry. Gait & Posture; 2022.
12. Park H, et al. Classification of gait phases based on a machine learning approach using muscle synergy features. Front Hum Neurosci. 2023;17:1201935.
13. Jiang X, et al. Gait phase recognition based on lower limb sEMG signals using an LDA-PSO-LSTM algorithm. Biomedical Signal Processing and Control; 2023.
14. Xue T et al. Deep learning models optimization for gait phase identification from EMG data. IEEE Trans Neural Syst Rehabil Eng (2025).
15. Li Z et al. *Explainable deep learning framework for lower limb activity recognition using sEMG*, Applied Soft Computing (2025).
16. Pan J, Gao Z, Chen D, Li F, Baker JS, Gu Y, et al. Artificial intelligence in lower limb joint moment prediction during typically developed gait: A systematic review and multilevel random-effects meta-analysis. Gait & Posture; 2025. p. 110083.
17. Morbidoni C, Cucchiarelli A, Agostini V, Knaflitz M, Fioretti S, Di Nardo F. Machine-learning-based prediction of gait events from emg in cerebral palsy children. IEEE Trans Neural Syst Rehabil Eng. 2021;29:948–56.
18. Anselmino E, Mazzoni A, Micera S. Emg-based prediction of step direction for a better control of lower limb wearable devices. Comput Methods Programs Biomed. 2024;254:108305.
19. Pirouzi F et al. Gait event prediction using surface electromyography in Parkinson's disease. Sensors 23 (2023).
20. Liu SH, Sharma AK, Wu BY, Zhu X, Chang CJ, Wang JJ. Estimating gait parameters from semg signals using machine learning techniques under different power capacity of muscle. Sci Rep. 2025;15:12575.

21. Schreff L, Haeufle DF, Vielemeyer J, Müller R. Evaluating anticipatory control strategies for their capability to cope with step-down perturbations in computer simulations of human walking. Sci Rep. 2022;12:10075.
22. Di Nardo F et al. *Surface electromyographic signals collected during long-lasting ground walking of young able-bodied subjects*, PhysioNet, version 1.0.1 (2024).
23. Goldberger AL, et al. PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101:e215–20.
24. Di Nardo F, Mengarelli A, Maranesi E, Burattini L, Fioretti S. Assessment of the ankle muscle co-contraction during normal gait: A surface electromyography study. J Electromyogr Kinesiol. 2015;25:347–54.
25. Di Nardo F, Morbidoni C, Mascia G, Verdini F, Fioretti S. *Intra-subject approach for gait-event prediction by neural network interpretation of emg signals*, BioMedical En-gineering OnLine 19 (2020), pp. 1–18.
26. Tokas P, Semwal VB, Jain S. Explainable deep learning framework for lower limb activity classification using reconstructed semg signals in healthy and pathological subjects. Appl Soft Comput (2025), p. 113748.
27. Guerra BMV, Schmid M, Sozzi S, Pizzocaro S, De Nunzio AD, Ramat S. A recurrent deep network for gait phase identification from emg signals during exoskeleton-assisted walking. Sensors. 2024;24:6603.
28. George JA, Gunnell AJ, Archangeli D, Hunt G, Ishmael M, Foreman KB, Lenzi T. Robust torque predictions from electromyography across multiple levels of active exoskeleton assistance despite non-linear reorganization of locomotor output. Front Neurorobotics. 2021;15:700823.
29. Agostini V, Ghislieri M, Rosati S, Balestra G, Knaflitz M. Surface electromyogra-phy applied to gait analysis: How to improve its impact in clinics? Front Neurol. 2020;11:994.
30. Rezaei-Dastjerdehei MR, Mijani A, Fatemizadeh E. *Addressing imbalance in multi-label classification using weighted cross entropy loss function*, in 2020 *27th national and 5th international iranian conference on biomedical engineering (ICBME)*. IEEE, 2020, pp. 333–338.
31. Nazmi N, Abdul Rahman MA, Yamamoto S, Ahmad SA. Walking gait event detection based on electromyography signals using artificial neural network. Biomed Signal Process Control. 2019;47:334–43.
32. Morbidoni C, Cucchiarelli A, Fioretti S, Di Nardo F. A deep learning approach to emg-based classification of gait phases during level ground walking. Electronics. 2019;8:894.
33. Di Nardo F, Strazza A, Mengarelli A, Cardarelli S, Tigrini A, Verdini F, Nascim-beni A, Agostini V, Knaflitz M, Fioretti S. Emg-based characterization of walking asymmetry in children with mild hemiplegic cerebral palsy. Biosensors. 2019;9:44.
34. Reddy VM, Pandian PS. Multiclass gait phase classification from the temporal convolutional network of wireless surface electromyography measurements. IEEE Sens Lett. 2024;8:1–4.
35. Yaakoubi NAE, McDonald C, Lennon O. Prediction of gait kinematics and kinetics: A systematic review of emg and eeg signal use and their contribution to prediction accuracy. Bioengineering.

2023;10:1220.

36. Luo M, Yin M, Li J, Li Y, Kobsiriphat W, Yu H, Xu T, Wu X, Cao W. Lateral walking gait recognition and hip angle prediction using a dual-task learning framework. Cyborg Bionic Syst. 2025;6:1–14.

Tables

Tables 1 & 2 are available in the Supplementary Files section.

Figures

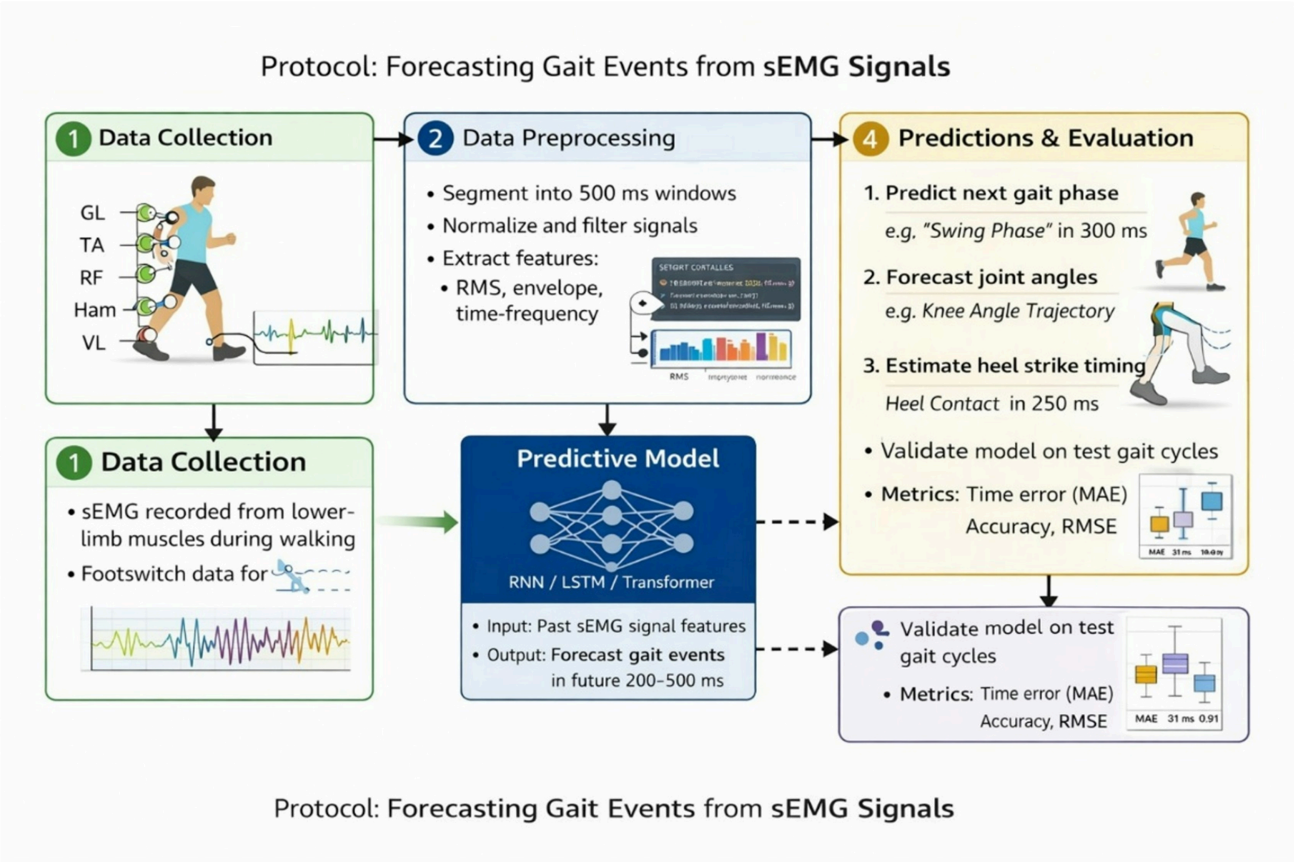


Figure 1

Study protocol for forecasting gait events from sEMG signals, including data collection, prepro-cessing, sequence modeling, and evaluation.

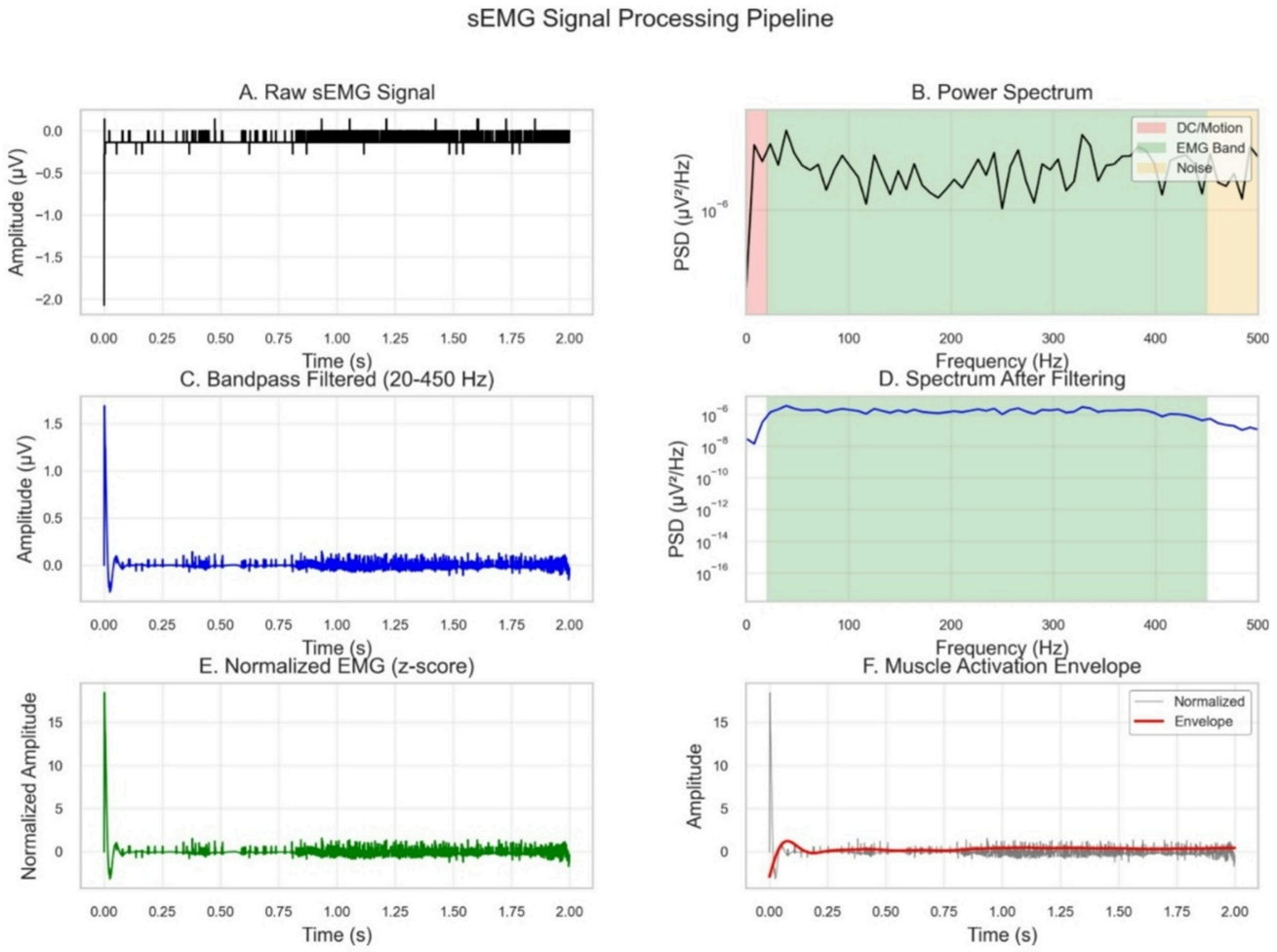
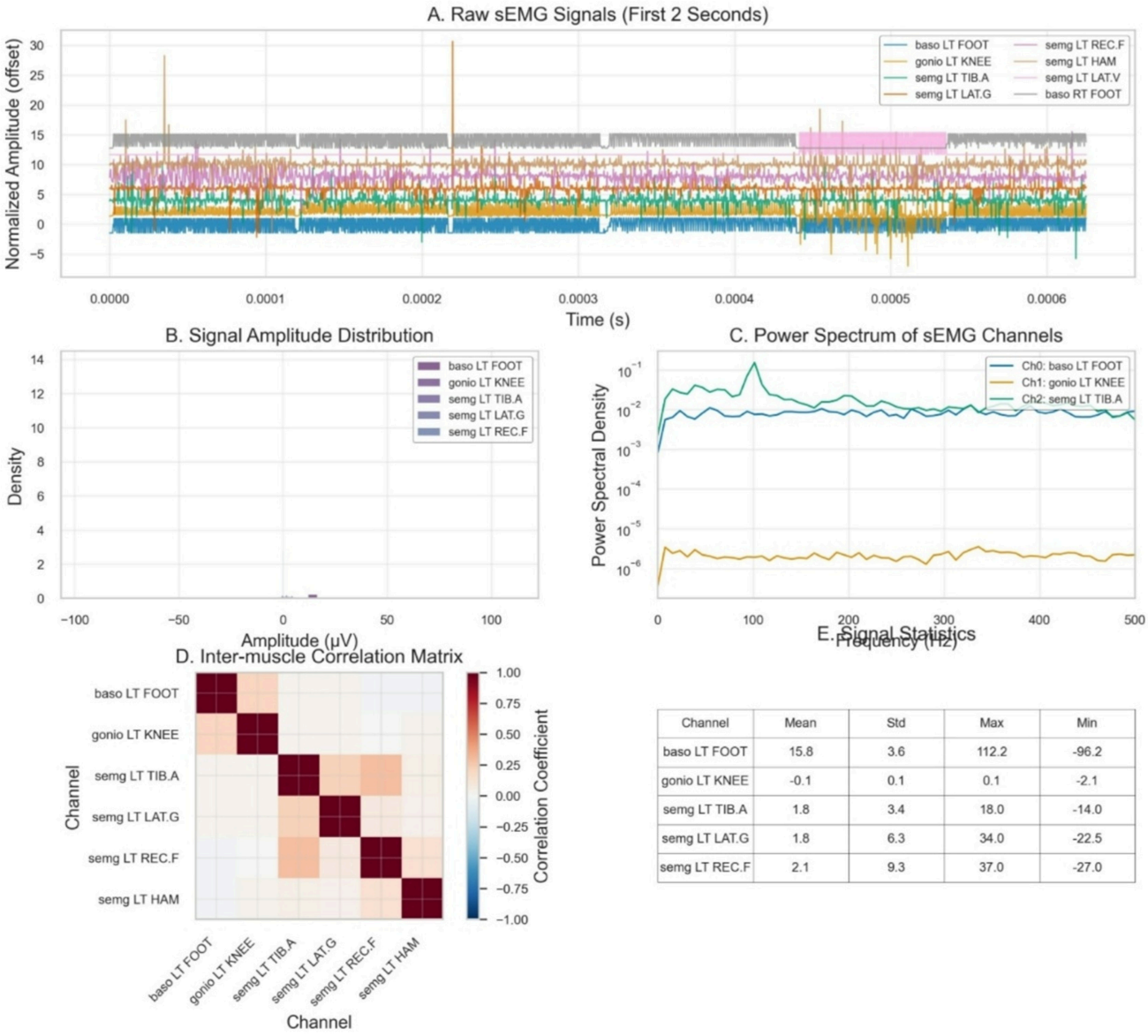


Figure 2

sEMG signal processing pipeline used before model training: raw trace analysis, frequency-domain inspection, 20–450 Hz filtering, z-score normalization, and activation-envelope extraction.



Channel	Mean	Std	Max	Min
baso LT FOOT	15.8	3.6	112.2	-96.2
gonio LT KNEE	-0.1	0.1	0.1	-2.1
semg LT TIB.A	1.8	3.4	18.0	-14.0
semg LT LAT.G	1.8	6.3	34.0	-22.5
semg LT REC.F	2.1	9.3	37.0	-27.0

Figure 3

Dataset overview of synchronized gait signals and descriptive statistics, including representative traces, amplitude distributions, spectral profiles, inter-channel correlation matrix, and summary statistics.

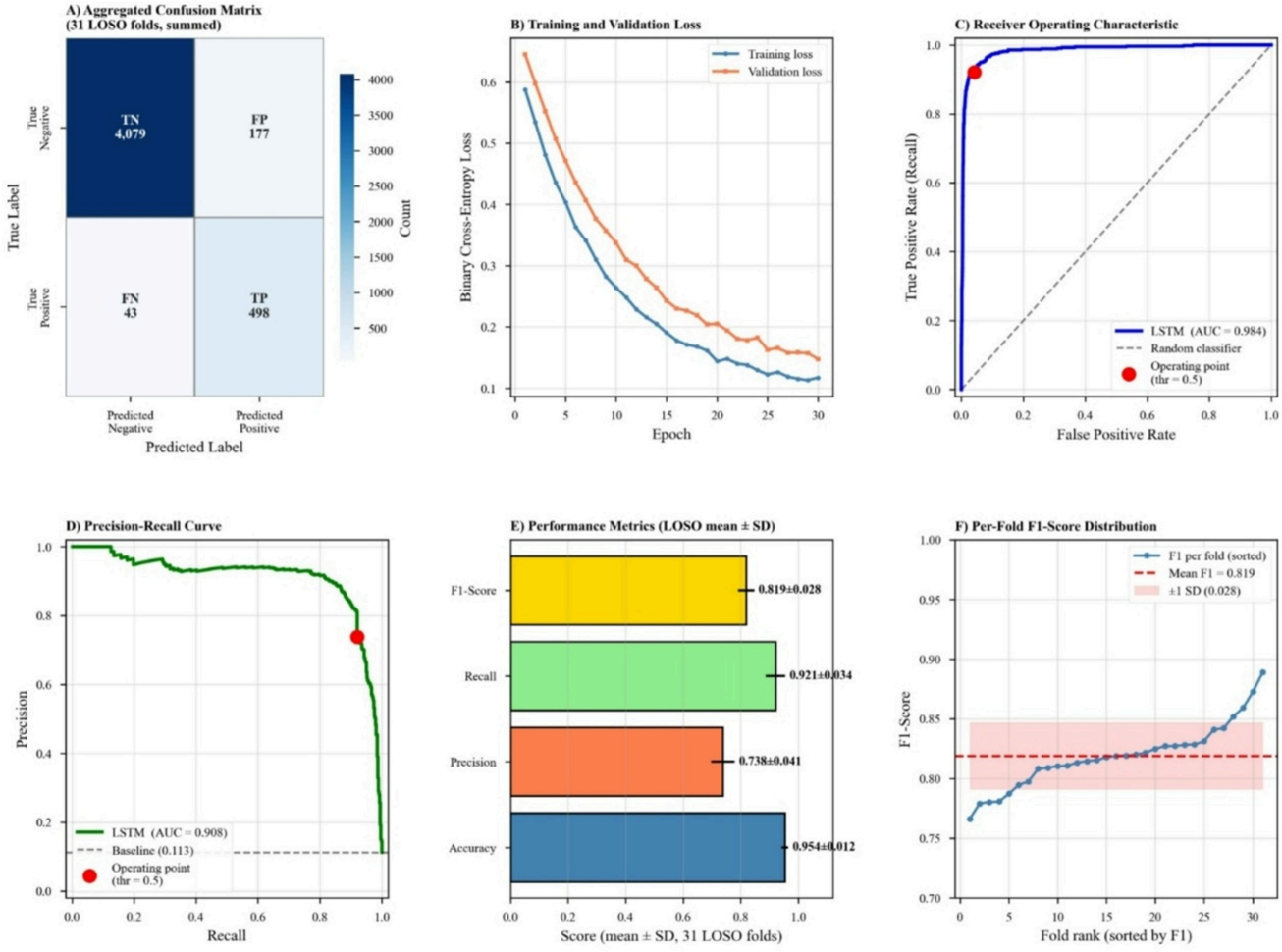


Figure 4

LSTM model performance under leave-one-subject-out (LOSO) cross-validation across 31 sub-jects. (A) Aggregated confusion matrix (counts summed across all 31 folds: TN = 4,079, FP = 177, FN = 43, TP = 498). (B) Training and validation loss curves. (C) Receiver operating characteristic (ROC) curve; the red marker indicates the operating point at decision threshold = 0.5. (D) Precision-recall curve with the operating point marked. (E) Fold-averaged performance metrics (mean ± SD across 31 folds). (F) Per-fold F1-score dis-tribution sorted by fold rank.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables1and2.docx