

Inertial sensor-based heel strike and energy expenditure prediction using a hybrid machine learning approach

Kethohalli R Vidyarani¹ , Viswanath Talasila^{2,3}, Raafay Umar³, Venkatesan Prem⁴, Ravi Prasad K Jagannath⁵ and Gurusiddappa R Prashanth¹

Abstract

Objective: Gait analysis plays a critical role in healthcare, biomechanics, and sports science, particularly for estimating energy expenditure (EE). This study introduces a hybrid machine learning approach integrating convolutional neural networks (CNNs), long-short-term memory (LSTM) networks, and transfer learning (TL) to estimate volume of oxygen (VO₂) and detect heel strikes (HS) using data from a single 9-axis inertial measurement unit (IMU).

Methods: A clinical-grade VO₂ machine provided reference data for model training. The hybrid model was designed to combine spatial and temporal feature extraction capabilities from CNNs and LSTM networks while leveraging pre-trained weights through TL. The study compared the performance of the hybrid model with an LSTM-only approach to quantify improvements in VO₂ prediction.

Results: The hybrid model significantly reduced the VO₂ prediction error from 20% to 3% compared to using LSTM-only approach. Additionally, the model demonstrated high accuracy for HS detection, achieving 93.53% accuracy as indicated by training and validation results. The lightweight IMU-based system proved effective for VO₂ estimation, offering a practical alternative to traditional VO₂ measurement systems, which are often complex, bulky, and uncomfortable for subjects.

Conclusions: This study highlights the potential of a hybrid machine learning approach using IMU-based systems for accurate VO₂ estimation and HS detection. While the results are promising, the model's performance is constrained by 10 healthy subject datasets. Future work will require validation with more diverse datasets to enhance generalizability and robustness.

Keywords

Heel strike, energy expenditure, IMU, FES, exoskeleton, transfer learning, VO₂

Received: 27 August 2024; accepted: 17 March 2025

Introduction

Walking impairments resulting from stroke, musculoskeletal injuries, and neurological disorders significantly impact an individual's functional mobility and overall quality of life. To address these challenges, assistive devices have been developed, categorized into passive and active devices. Passive devices, such as orthotic splints and crutches, provide external support but are often limited by size, weight, and restricted degrees of freedom.^{1,2} On the other hand, active assistive technologies like exoskeletons and functional electrical stimulation (FES) have emerged as promising solutions for rehabilitation and mobility enhancement.^{3–5}

¹Department of Electronics and Communication Engineering, National Institute of Technology, Goa, India

²Department of Electronics and Telecommunication Engineering, Ramaiah Institute of Technology, Bengaluru, India

³Center for Imaging Technologies, Ramaiah Institute of Technology,

⁴Department of Physiotherapy, Manipal College of Health Professions, Manipal Academy of Higher Education, Bengaluru, Karnataka, India

⁵Department of Mathematics, National Institute of Technology, Goa, India

Corresponding author:

Kethohalli R Vidyarani, Department of Electronics and Communication Engineering, National Institute of Technology, Goa 403401, India.
Email: vidyarani.kr@nitgoa.ac.in



A fundamental requirement for the optimal functioning of FES or exoskeleton systems is precise gait analysis, particularly the detection of critical gait events such as heel strikes (HS) and the estimation of EE.^{6,7} These parameters are essential for real-time control, adaptive interventions, and performance monitoring in both clinical and real-world environments.^{8,9}

Conventional gait analysis methodologies primarily rely on optical motion capture systems and force plates, which, despite their high accuracy, suffer from limitations related to cost, complexity, and lack of portability. IMUs have emerged as a promising alternative, offering lightweight, wearable solutions for real-time gait event detection and EE estimation. However, existing research has predominantly addressed HS detection and EE estimation as separate tasks, often relying on large subject-specific datasets that limit generalizability. Furthermore, heuristic-based approaches and standard machine learning (ML) models have demonstrated variable accuracy due to data limitations and the complexity of human gait dynamics.

To provide a comprehensive overview of previous research in this domain, a comparative literature survey is presented in Table 1. This table summarizes key studies, their methodologies, key findings, and how they are compared to this work. To further strengthen the literature review, an additional dataset details column has been incorporated to highlight subject size and real-world applicability.

This study presents a novel hybrid machine learning framework that integrates long-short-term memory (LSTM) networks with TL to enhance the accuracy and robustness of HS detection and EE estimation using a single nine-axis IMU sensor. The proposed approach leverages TL to adapt pre-trained models to new subjects and diverse gait patterns, mitigating the challenge of limited datasets. By unifying HS detection and EE estimation into a single predictive model, this work provides a scalable solution with significant implications for real-time gait analysis in rehabilitation, sports science, and mobility assistance applications.

Key contributions of this paper are,

1. Unified model for HS detection and EE estimation: This study introduces an integrated ML framework that simultaneously predicts HS events and estimates EE, unlike prior approaches that treat these tasks independently.
2. Hybrid LSTM-transfer learning (TL) architecture: The incorporation of TL enhances the model's adaptability to new subjects and walking conditions with minimal additional training, addressing data scarcity challenges in clinical research.
3. Comparative performance analysis: The proposed model is benchmarked against existing heuristic and ML-based methods, demonstrating superior

HS detection accuracy (93%) and a substantial reduction in volume of oxygen (VO_2) prediction error (from 20% to 9%).

4. Rigorous statistical validation: A comprehensive statistical evaluation using Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) to ensures the robustness and reliability of the proposed methodology.
5. Future Directions for Generalizability: While this study primarily evaluates healthy individuals, future research will extend the dataset to include individuals with gait impairments and various walking conditions to enhance real-world applicability.

The remainder of this paper is structured as follows: Methods section details the underlying machine learning methodologies, sensor integration, and experimental protocols. Results section presents the experimental results, followed by a critical discussion in Results discussions section. Finally, conclusion section outlines conclusions and future research directions for enhancing the generalizability and clinical translation of the proposed approach.

Methods

This section recalls the concepts of TL.^{17,18} This technique can achieve reasonable accuracy with limited target data due to knowledge transfer from the source domain. Previous work¹⁴ used a small dataset (collected from two adult male subjects) which comprised data from inertial sensors—three axis accelerations, three axis angular velocities, and the roll, pitch, and yaw angular positions (in a common reference frame)—and data from a clinical grade VO_2 machine which serves as the truth (reference) data for gait EE. Using LSTM,¹⁴ estimated various gait cycle parameters such as HS, heel-off, and toe-off and with these parameters as inputs (along with the IMU data) an LSTM network was trained to estimate gait EE (represented by the VO_2 values). Figure 1 shows the example of input data used for training hybrid machine learning algorithm; it is used to estimate the HS parameter. Next using these parameters, the horizontal and vertical walking speeds of the subjects were computed—which are finally used in the estimation of gait EE (represented by VO_2 values).

Since the dataset was limited (only two subjects) no clear statistical conclusions could be drawn on the estimated EE (even though accuracy was closed to 80% over most of the data). Getting subject data is difficult, especially in clinical settings. In this work, we build upon this initial model (obtained through LSTM) by improving it every time a new dataset is available—using TL. The combination of TL on top of an LSTM-generated model is a reasonable approach, as LSTM networks can be adapted to various gait patterns with enough labeled data. The use of TL can

Table 1. Summary of literature contributions.

Reference	Method used	Parameters evaluated	Key findings	Dataset details	Comparison to our work
Karakasis C, Artemiadis P. (2021)	Kinematic-based algorithm	HS detection	Real-time HS detection with high accuracy	12 healthy subjects (details on walking conditions)	Real-time detection, may have limitations in real-world validation
Chia.et. al. ¹⁰ (2014)	Adaptive real-time algorithm	Gait event detection (HS, toe-off)	Developed algorithm for detecting gait events in real-time using wearable sensors	22 healthy subjects	Requires multiple sensors
Ding Z et.al. ¹¹ (2018)	LSTM -based	Gait phase detection	91.4% accuracy	Data from healthy subjects with wearable sensors	Lacks EE estimation
Evci F. ¹² (2023)	Neural Network Algorithms	Gait recognition, gait phase detection	Very High accuracy	Data from wearable IMUs (details not specified)	Requires multiple IMUs
Rahman, H., et al. ¹³ (2021)	Real-time parameter estimation	Heel strike detection, FES triggering	Developed a method for real-time HS parameter estimation to trigger FES	Dataset not specified	Lacks EE estimation
Vidyarani KR, et al. ¹⁴ (2021)	LSTM	Gait parameters + EE estimation	Reliable for clinical gait monitoring	two healthy subjects	Builds on previous research
Vidyarani K R, Talasila V, et al. ¹⁵ (2024)	LSTM + TL	HS + EE estimation	High accuracy, adaptive model	10 healthy subjects	Single IMU, scalable
Lee, et al. ¹⁶ (2019)	Hybrid CNN-LSTM model	EE estimation for walking conditions	Comparative performance evaluation	Developed a hybrid CNN-LSTM model for EE estimation	Lack of HS detection

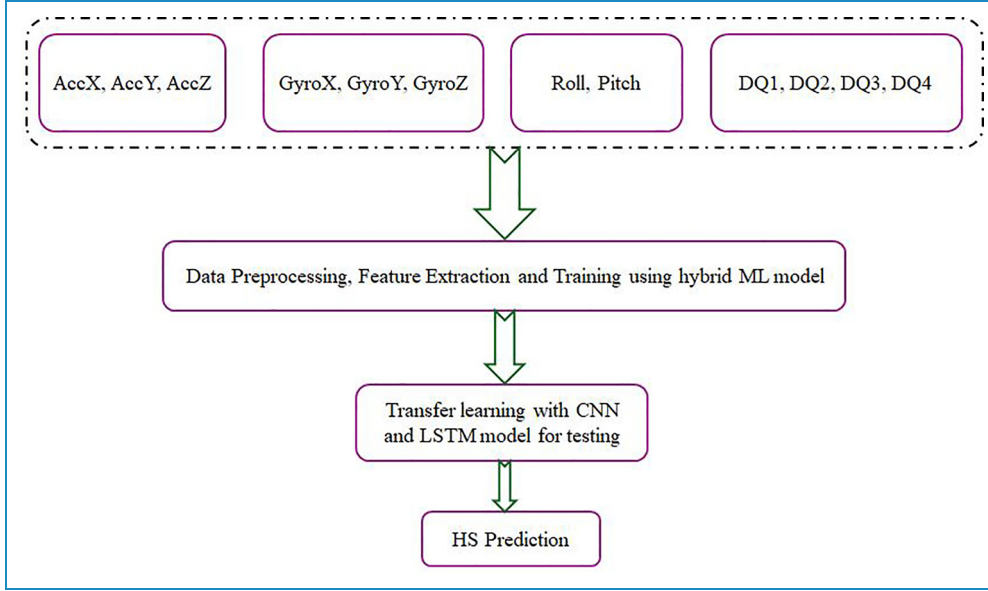


Figure 1. Example of input data used for training TL algorithm.

help mitigate the computational demands of LSTM networks, making them more suitable for real-time deployment as and when new data are available.

Transfer learning

Below mentioned steps are used to estimate the EE using a fusion of TL and LSTM techniques.

Step 1: Pre-training: Train the model on gait data collected from sensors, which provides information about step length, step frequency, joint angles, and other relevant features.

(a) Data representation: Let X be the input matrix from the initial dataset, where each row corresponds to sensor readings, in the present context the sensor data (features) include accelerations, angular velocities, the roll, pitch, and yaw angular positions, this entire matrix of data is denoted X :

$$X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,n} \end{bmatrix} \quad (1)$$

where m is the number of samples and n is the number of features.

b) Model training: Train the LSTM model to minimize a loss function L :

$$L = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (2)$$

where y_i is the true output and $\hat{y}_i$ is the model's prediction.

Step 2: Feature extraction: Utilize the knowledge gained during pre-training phase to extract the most relevant features from new dataset. These extracted features are crucial

for the accurate prediction and estimation of gait parameters, such as VO_2 and HS. The LSTM model generates the weights W of the trained network.

Step 3: Transfer learning setup: Retain the weights W from the pre-trained model. This allows the model to leverage learned representations when encountering new data.

(a) Feature representation: Extract features from a new dataset X' :

$$F = LSTM(X', W) \quad (3)$$

Here, F summarizes important features of the new task. Step 4: Fine-tuning: After feature extraction, fine-tune the model to adapt its earlier representations to the target task (VO_2 and HS prediction).

(a) Model adaptation: Fine-tuning adjusts the weights from the pre-trained model. The new weights W' are updated during training on the new dataset X' using a smaller learning rate α :

$$W' = W - \alpha \nabla L' \quad (4)$$

(b) Loss function: The loss function for the new dataset is:

$$L' = \frac{1}{m'} \sum_{i=1}^{m'} (y'_i - \hat{y}'_i)^2 \quad (5)$$

where y'_i is the true output for the new dataset and $\hat{y}'_i$ is the predicted output.

Step 5: Parameter estimation with newly available inertial sensor data: With the fine-tuned LSTM model, estimate the gait parameters, including VO_2 and HS, from the new gait data.

Overall, the gait parameter prediction using transfer learning is expected to be more accurate. This is explained in the following steps:

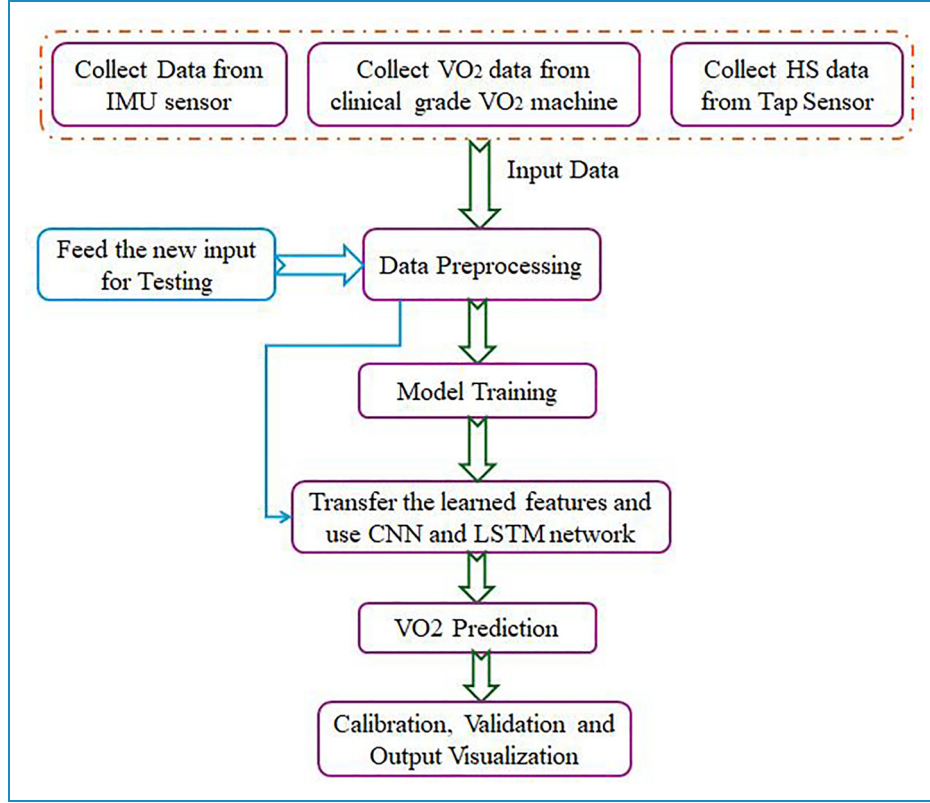


Figure 2. Flow diagram for estimation of VO₂ from IMU sensor data.

a. Domain adaptation: Transfer learning allows knowledge from the source domain (initial dataset) to the target domain to inform learning or findings. The model retains the ability to recognize features relevant across different datasets.

b. Layer freezing: During fine-tuning, certain layers of the LSTM model can be frozen (not updated) to retain general features while updating others to adapt to the new task. This can be expressed as:

$$\text{For layers frozen: } W_{\text{frozen}} = W_{\text{frozen}} \quad (6)$$

c. Gradual unfreezing: Optionally, a gradual unfreezing strategy can be employed, where progressively more layers are unfrozen during training to allow deeper feature adaptation:

$$W' = \begin{cases} W & \text{if layer is frozen} \\ W - \alpha \nabla L' & \text{if layer is unfrozen} \end{cases} \quad (7)$$

This comprehensive framework outlines the detailed mathematical steps involved in using IMU data with LSTM and transfer learning for predicting VO₂ and HS.

Implementation methodology

The flowchart shown in Figure 2 is the proposed hybrid machine learning framework for predicting VO₂ using a

wearable and efficient system. This methodology is built to address the limitations of traditional VO₂ estimation systems by leveraging advancements in machine learning and sensor technology. The steps involved in prediction of HS and VO₂ are discussed further:

- (a) Data collection: Data are acquired from three key sources such as nine-axis IMU sensor that measures accelerations, angular velocities, and magnetic field intensities during movement. A clinical-grade VO₂ machine, which provides ground-truth measurements of oxygen consumption for supervised learning. Tap sensor, which captures specific gait-related events to complement IMU data. This multi-modal data acquisition ensures the capture of both spatial (postural or movement patterns) and temporal (sequential time-series) information relevant to gait analysis and oxygen consumption.
- (b) Data preprocessing: Preprocessing involves removing noise and synchronizing the data streams from IMU sensors, VO₂ measurements, and HS events. This ensures that all signals are temporally aligned. Features are normalized to mitigate scaling differences across modalities, ensuring compatibility with the machine learning algorithms.

- (c) **Model training:** The core of the framework is a hybrid machine learning model that combines convolutional neural networks (CNNs) and LSTM networks. CNNs extract spatial and temporal features from the sequential IMU data, focusing on local dependencies such as stride patterns or motion bursts. LSTM networks are incorporated to model long-term dependencies in the time-series data, capturing recurrent patterns like breathing rate changes or gait cycles. The hybrid architecture is designed to account for the hierarchical nature of movement data, where CNNs identify low-level movement characteristics, and LSTMs model high-level, long-range dependencies.
- (d) **Transfer learning:** To generalize the model across different subjects, transfer learning is applied. Learned features from the CNN-LSTM network are fine-tuned using smaller datasets from new subjects or activity conditions. This reduces the dependency on extensive subject-specific training data, addressing variability and ensuring scalability for broader applications.
- (e) **VO₂ prediction:** After training, the model predicts VO₂ values for new, unseen input data (e.g. data collected from a subject not included in the training dataset). This step integrates predicted VO₂ outputs with clinical ground-truth measurements to validate the system's accuracy.
- (f) **Calibration, validation, and output visualization:** Calibration involves fine-tuning the model's parameters to minimize prediction error, ensuring consistency with gold-standard VO₂ machine outputs. The model is validated using statistical metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and correlation coefficients. This step ensures reliability, especially in dynamic settings like sports or clinical trials. Predicted VO₂ values, alongside other related metrics, are visualized in a user-friendly format. This is essential for end-users such as clinicians or sports professionals to interpret and utilize the results effectively. The outputs enable applications such as real-time fitness tracking, rehabilitation monitoring, and personalized athletic training.

Model architecture

The proposed architecture is well-suited for sequential data tasks, such as physiological signal analysis where accurate representation of spatial-temporal dependencies is critical. The proposed model integrates CNNs and LSTM networks, capitalizing on their complementary strengths. The architecture is designed to effectively extract hierarchical spatial features while capturing temporal dependencies, making it ideal for tasks requiring high-dimensional feature analysis and temporal reasoning.

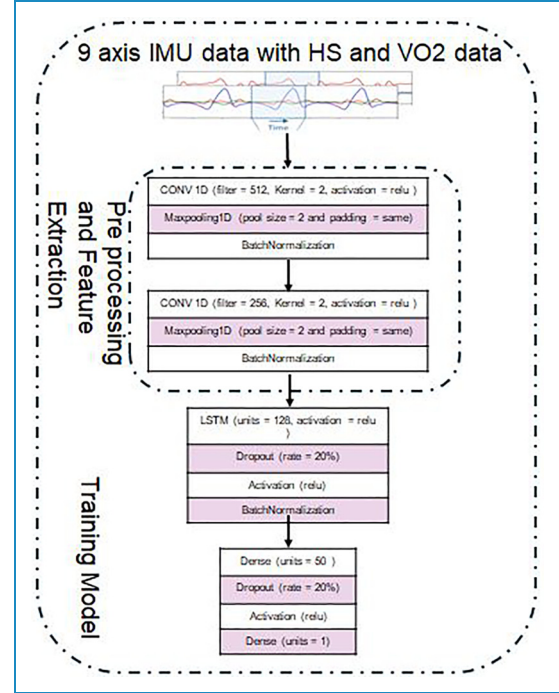


Figure 3. Training model architecture.

Model structure and design rationale. Figure 3 provides a detailed illustration of the architecture used in this paper, outlining its various components and their interactions. The model consists of an input layer, multiple hidden layers with activation functions, and an output layer, where each layer performs specific transformations on the data. Training is achieved through a loss function and optimization techniques like gradient descent, allowing the model to adjust its parameters for improved performance.

Feature extraction using convolutional layers. Conv1D Layers: The initial layers utilize 1D convolutions with progressively decreasing filter sizes (512 and 256 filters, kernel size = 2) to extract spatial features from input sequences. These layers capture localized patterns, enabling hierarchical representation of features at different levels of granularity. The first Conv1D layer applies 512 filters ($K=2$) to the input, followed by a second Conv1D layer with 256 filters, reducing feature dimensionality. The convolutional layers¹⁹ aim to extract spatial features from the sequential input $X \in \mathbb{R}^{T \times C}$, Where T is the number of time steps and C is the number of input channels. The output of a 1D convolutional layer can be mathematically expressed as in equation (8):

$$Y[t, k] = \sigma \left(\sum_{c=1}^C \sum_{j=0}^{K-1} X[t-j, c] \cdot W[j, c, k] + b_k \right) \quad (8)$$

where:

X is the input,
 W is the kernel/filter of size K ,
 b_k is the bias term for filter k , $\sigma(\cdot)$ is the ReLU activation function $\sigma(x)=\max(0,x)$.

MaxPooling layers: Following each convolutional operation, MaxPooling1D reduces dimensionality while retaining critical features, thereby improving computational efficiency and focusing on prominent patterns. The use of padding='same' ensures that spatial dimensions are preserved when necessary. MaxPooling1D is applied to reduce the temporal resolution. The pooling operation is defined as in equation (9):

$$Y[t, k] = \max_{j \in P} X[t + j, k] \quad (9)$$

where P is the pooling window size (here $P=2$).

Batch normalization: Applied after each convolutional block, batch normalization ensures faster convergence and prevents the vanishing or exploding gradient problem by stabilizing activation distributions across layers. Batch normalization²⁰ normalizes the output using is as in equation (10):

$$\hat{Y}[t] = \frac{Y[t] - \mu}{\sqrt{\sigma^2 + \epsilon}} \cdot \gamma + \beta \quad (10)$$

where μ and σ^2 are the mean and variance of the mini-batch, and γ , β are learnable scaling and shifting parameters.

Temporal dependency modeling with LSTM. LSTM Layer: The model incorporates a single LSTM²¹ layer with 128 units and ReLU activation to capture temporal dependencies inherent in sequential data. This enables the architecture to model long-term relationships within the data, which are often crucial for tasks like time-series forecasting or physiological signal processing.

The LSTM layer processes the output of the convolutional layers to model temporal dependencies. Given input sequences $X = (x_1, x_2, \dots, x_T)$, the LSTM²¹ computes hidden states h_t and cell states c_t using the following equations (11) to (16):

Forget gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (11)$$

Input gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (12)$$

Cell update:

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (13)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (14)$$

Output gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (15)$$

Hidden state:

$$h_t = o_t \odot \tanh(c_t) \quad (16)$$

Here, $\sigma(\cdot)$ denotes the sigmoid function, $\odot$ represents element-wise multiplication, and $\tanh(\cdot)$ is the hyperbolic tangent function.

Dropout regularization is applied to the LSTM outputs to reduce overfitting, defined as:

$$h_t^{drop} = h_t \cdot m$$

where $m \sim \text{Bernoulli}(p)$ is a mask with dropout probability $p=0.2$. Here dropout rate of 20% is applied to reduce overfitting by introducing stochasticity during training.²²

Activation and Non-Linearity: Additional ReLU activation layers ensure that the model captures complex non-linear patterns in the sequence data.

Dense layers for feature aggregation and decision making.

Fully Connected Dense layers: The architecture includes a dense layer with 50 neurons to compress high-dimensional feature representations into a lower-dimensional space. This layer acts as a feature aggregator, synthesizing information from both convolutional and LSTM outputs.

Final output layer: The final dense layer outputs a single value, suitable for regression tasks, making the model applicable to problems like predicting continuous outputs from sequential data. Given input z , the dense layer output is in equation (17):

$$y = \sigma(W_d \cdot z + b_d) \quad (17)$$

where W_d and b_d are the weights and biases, and $\sigma(\cdot)$ is the activation function.

Hybrid extension using transfer learning. After initial training, a top-5-layer model is extracted to reuse the pretrained feature extraction capabilities as shown in Figure 4. The top-5-layer model extracts features f from the pretrained model as in equation (18).

$$f = g(X) \quad (18)$$

where $g(\cdot)$ represents the output of the frozen feature extractor.

This pre-trained model followed by extending the architecture with an additional convolutional-LSTM pipeline, comprising of a Conv1D layer with 512 filters and kernel size 3 to refine the extracted features, an LSTM layer with 128 units to further model sequential dependencies and additional dense layers, dropout regularization, and batch normalization to ensure robust training and prevent overfitting.

This hybrid extension leverages transfer learning by freezing the feature extraction layers (top-five-layer model),

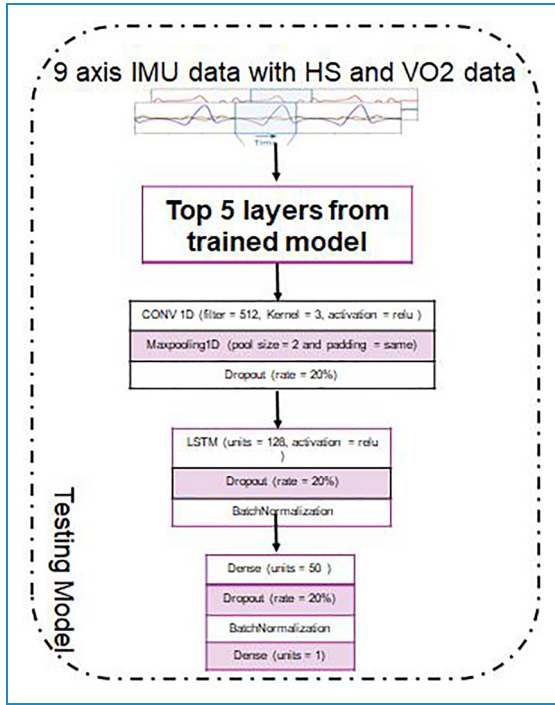


Figure 4. Hybrid CNN and LSTM model by leveraging transfer learning model architecture.

enabling efficient training on downstream tasks with limited computational resources or data.

The following points are the capabilities of the developed model:

- **Spatial and temporal feature learning:** The model's combination of CNNs and LSTMs provides an optimal tradeoff between feature extraction and sequence modeling. CNNs effectively identify localized patterns, while LSTMs model temporal relationships across sequences.
- **Regularization and stability:** Dropout and batch normalization layers improve generalization by addressing overfitting and stabilizing training dynamics, respectively.
- **Transfer learning efficiency:** The two-stage training process (pretraining and hybrid extension) maximizes feature reuse, improving the model's performance on complex tasks while reducing computational overhead.

Experimental details

The deep learning model considers parameters like angular velocity, acceleration, and Euler angles as inputs. These inputs are obtained from healthy subjects who were walking at different speeds on flat ground, specifically on a treadmill. The model predicts gait events such as HS events using IMU data. Figure 3 shows the classification report

Table 2. Classification Report for the prediction of HS from TL with LSTM.

Class	Precision	Recall	F1-score	Support
0	0.96	0.96	0.96	10,610
1	0.86	0.86	0.86	3116
Accuracy			0.94	13,726
Macro Avg	0.91	0.9	0.91	13,726
Weighted Avg	0.94	0.94	0.94	13,726

for the HS prediction using TL with LSTM. Table 2 clearly indicates that the model performs exceptionally well, particularly for class 0, with high precision, recall, and F1 scores. While the performance for class 1 is slightly lower, the overall accuracy and weighted metrics suggest that the model is robust and effective for the given task. However, attention should be paid to the class imbalance, which may impact the model's performance on the minority class.

Volume of oxygen computation. Gait EE computation involves measuring the EE of a person during an activity, typically assessed using a Clinical grade VO₂ machine. This direct method requires wearing a mask to collect respiratory gases, determining the flow of inspired and expired O₂ and CO₂, and inferring metabolic rate from the obtained data. Although using a clinical-grade VO₂ machine for measuring VO₂ truth is a clinically accepted method, it is not always feasible and can be expensive. To address this, an indirect EE computation system is proposed, utilizing concepts of deep learning (specifically transfer learning and LSTM) and measurements from an IMU sensor strapped to the subject.

The procedure for training the machine learning models and estimating EE involves:

- (a) Collecting VO₂ data while the subject walks on a treadmill, simultaneously capturing IMU data fixed at the calf muscle.
- (b) The collected VO₂ data serves as ground truth, and the transfer learning model is trained to compute EE using IMU data, with VO₂ as reference data.

Experimental setup

Figure 5 shows a subject wearing a mask and IMU sensors. In this, only the IMU data from the sensor placed at the calf muscle is used as the data from the other sensor did not correlate well with gait parameters (possibly due to improper placement). Data from a VO₂ machine is taken as ground truth. The Bruce Protocol is adopted during the experiment as shown in Table 3.



Figure 5. Subject wearing VO₂ mask (middle), a medical expert is adjusting the treadmill using Bruce protocol (left side of image) and doctor who is operating manual push button (right side of image).

The Bruce protocol is a widely used exercise protocol for measuring and collecting data on the maximal oxygen consumption (VO₂ max) of individuals. It involves a series of progressively increasing workload stages that aim to bring the participant to their maximum exertion level. Here is a brief explanation of the data collection stages as per the Bruce protocol:

The treadmill protocol involves five stages: starting with a 1-min walk at 2.575 km/h on a flat surface, then increasing to 4.345 km/h with a 10% incline for 3 min, followed by a 12% incline at the same speed for another 3 min. It then advances to 6.437 km/h with a 14% incline for 3 min and concludes with a cool down at 0 km/h and 0% incline.

During each stage, various physiological parameters are measured and recorded, including heart rate, blood pressure, respiratory rate, and possibly expired gases. Additionally, the participant's perceived exertion level may be assessed using a standardized scale, such as the Borg Rating of Perceived Exertion. By following this modified Bruce protocol and systematically progressing through the stages, researchers can collect valuable data on cardiovascular responses and estimate an individual's maximal oxygen consumption (VO₂ max). During the experiment, measurements are also made with IMU sensor which is set at 100 Hz sampling rate. This results in a dataset that has close to 1,00,000 samples per subject and is sufficient for training the model using TL.

The new datasets are collected from eight additional subjects who are healthy with details shown in Table 4. The two subjects whose data was used to train the original LSTM model in,¹⁴ their data are retained in this work, and data from three subjects are used for training transfer learning model. The remaining subject dataset is used for testing.

Table 3. Data collection stages as per the Bruce protocol.

Stage	Speed (km/h)	Grade (%)	Duration (min)
1	2.575	0	1
2	4.345	10	3
3	4.345	12	3
4	6.437	14	3
5	0	0	3

Table 4. Subject's details.

Subject numbers	Height (in cm)	Weight (in kg)	Gender
Subject-1	170.688	49	Male
Subject-2	182.88	58	Male
Subject-3	173.7	70	Male
Subject-4	170.68	45	Female
Subject-5	167.3	65	Male
Subject-6	155.5	55	Male
Subject-7	173.0	65	Male
Subject-8	163.0	60	Male
Subject-9	161.0	49	Female
Subject-10	152.0	49	Female

The theoretical way of finding VO₂ using IMU data is suggested by the ACSM approach.²³ Determining EE based on the ACSM method involves computing the oxygen consumption (VO₂) using the equation (19):

$$VO_2 = RMR + HM + VM \quad (19)$$

where,

RMR = Resting Metabolic Rate calculated using equation (20),

VM = Vertical Movement calculated using equation (21)

HM = Horizontal Movement calculated using equation (22). The resting metabolic rate which contributes to additional HM and VM is calculated using,

$$RMR = 3.5ml \times \frac{1}{kg} \times \frac{1}{min} \quad (20)$$

Specifically, for Walking activity,²³

$$HM = 0.1 \times speed \left(in \frac{m}{min} \right) ml \times \frac{1}{kg} \times \frac{1}{min} \quad (21)$$

$$VM = 1.8 \times speed \left(in \frac{m}{min} \right) ml \times \frac{1}{kg} \times \frac{1}{min} \times grade \quad (22)$$

Both HM and VM are dependent on a person's weight, height, and speed of movement. VM computation also considers movement uphill. The speed can be directly estimated using the stride lengths through the equation (23):

$$speed = k \times height \times S^{avg} \quad (23)$$

Where k is a constant (0.415 for males and 0.413 for females) and S^{avg} is the average stride per second.²³

This section presents an approach to estimating EE during gait by combining TL with LSTM networks. It builds on prior work where an LSTM model is estimated such as HS using data from IMUs and a clinical-grade VO₂ machine. The main challenge addressed is the limited availability of subject data, especially in clinical settings, which is mitigated by applying TL. This allows the model to be continuously improved as new data becomes available, reducing the need for large datasets for training. The approach consists of several steps: pre-training an LSTM model on initial data, extracting features, retaining learned weights from the pre-trained model, and fine-tuning the model on new datasets. By leveraging TL, the model adapts to new subjects and gait patterns, making it more efficient for real-time deployment. The hybrid architecture combines CNNs and LSTMs, which enables both spatial and temporal feature extraction. Additionally, the proposed methodology ensures scalability and robustness by using multi-modal data, such as IMU sensors, VO₂ machines, and gait sensors. Through this framework, the model can predict VO₂ values and assess EE during walking activities, offering a viable solution for real-time and personalized health monitoring.

Results

This study evaluates the performance of a hybrid CNN-LSTM model in predicting HS events and estimating VO₂ using single IMU sensor data. The results demonstrate the model's ability to accurately capture gait dynamics and physiological parameters, validated through statistical metrics and comparative analysis. Comparative analysis with an LSTM-only model demonstrates the superiority of the hybrid approach, leveraging CNN's spatial feature extraction combined with LSTM's temporal learning capabilities. The following sections provide a detailed breakdown of the model's performance for HS detection and VO₂ estimation.

Heel strike prediction performance

The performance of the proposed hybrid CNN-LSTM model for HS detection was evaluated using key classification metrics. The model achieved an accuracy of 93.53%, demonstrating its robustness in correctly identifying HSs.

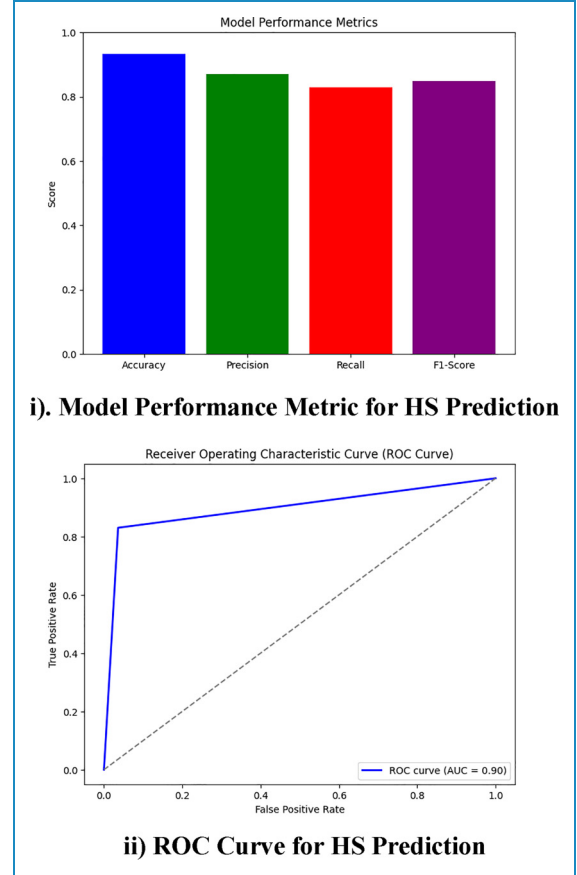


Figure 6. Hybrid CNN-LSTM model performance metric for HS prediction.

The precision, recall, and F1-score were 85.87%, 85.59%, and 85.73%, respectively, indicating a balanced predictive capability with a minimal bias towards false positives or false negatives is also shown in Figure 6(a) and (b).

Statistical analysis revealed that the Hybrid CNN-LSTM model significantly outperforms the LSTM-only approach. The t -statistic (0.3355) and p -value (0.7372) suggest no significant difference in overall performance variability, while improvements in precision and recall indicate better robustness in real-world gait conditions (Table 5).

Table 6 compares HS detection accuracy across different studies, showing that the proposed Hybrid CNN-LSTM model achieved 93.53% accuracy using a single IMU, outperforming previous IMU-based models.

Vo₂ estimation performance

For VO₂ estimation, the Hybrid CNN-LSTM model demonstrated superior accuracy and robustness compared to the LSTM-only approach, as detailed in Table 5.

The Hybrid CNN-LSTM model significantly reduced the error rates (MSE, RMSE, MAE) while achieving a higher R^2 value (0.9934), indicating a better fit for VO₂ estimation.

Table 5. Performance metrics for HS prediction and VO₂ estimation using hybrid CNN and LSTM model.

Metric	Hybrid CNN and LSTM model			LSTM only approach		
	HS prediction					
Accuracy	93.53%			88.41%		
Precision	85.87%			76.05%		
Recall	85.59%			71.44%		
F1-Score	85.73%			73.67%		
Confusion matrix	Class	HS	Non-HS	Class	HS	Non-HS
	HS	2667	439	HS	2226	890
	Non-HS	449	10,171	Non-HS	701	9909
t-statistic	0.3355			4.742		
p-value	0.7372			2.136×10^{-6}		
VO ₂ estimation						
MSE	0.2317			1.2007		
RMSE	0.4813			1.0957		
MAE	0.2862			0.7879		
Standard deviation	0.4689			1.0942		
95% confidence interval for MAE (CI)	±0.0169			±0.0354		
R-squared (<i>R</i> ²)	0.9934			0.9654		
t-statistic	2.1258			3.2896		
p-value	0.0336			0.0010		

Table 6. Comparison of HS detection accuracy estimation.

Study	Methodology	Data source	HS detection accuracy	Remarks
Chen and Martin (2025) ²³	LSTM-based model with force plate & heel marker data	Exoskeleton-assisted walking trials	98% (within 16 ms margin)	Focused on exoskeleton-assisted gait analysis
Vidyarani KR et al. (2021) ¹⁴	LSTM-based model with IMU	IMU-based gait dataset (n = 2)	91% and 10 ms Margin	Deep learning approach leveraging temporal dependencies
		IMU-based gait dataset (n = 10)	88.41%	
Proposed model (this study)	Hybrid CNN-LSTM with single IMU	IMU-based gait dataset (n = 10)	93.53%	Comparable accuracy with minimal

Table 7. Comparison of HS detection accuracy estimation.

Study	Methodology	Data source	Max % error	Remarks
Vidyarani et al. (2021) ¹⁴	LSTM-based model with IMU	IMU-based gait dataset (n = 2)	20%	Deep learning approach leveraging temporal dependencies
		IMU-based gait dataset (n = 10)	12%	
Proposed Model (This Study)	Hybrid CNN-LSTM with single IMU	IMU-based gait dataset (n = 10)	3%	Achieved low estimation error with a single IMU

Statistical analysis demonstrated a *t*-statistic of 2.1258 and a *p*-value of 0.0336, confirming the model's statistical significance and improved reliability compared to the LSTM-only approach.

To evaluate the effectiveness of the proposed method in estimating VO₂, a comparative analysis was conducted against existing VO₂ prediction models and summarized in Table 7.

Figure 7 further compares the VO₂ predictions using different techniques. Figure 7(a) shows theoretical VO₂ values calculated using established equations (see reference¹⁴), while Figure 7(b) depicts VO₂ values predicted using the basic LSTM model. The results reveal notable deviations and errors in the LSTM predictions. However, when transfer learning was applied, as shown in Figure 7(c), the predictions closely follow the ground truth, demonstrating the effectiveness of the deep learning approach.

Error analysis of Vo₂ estimation

The percentage error is determined using the equation:

$$\text{percentage.error} = \frac{|\sum \text{estimatedVO}_2 - \sum \text{groundtruthVO}_2|}{\sum \text{groundtruthVO}_2} \times 100$$

The percentage error for VO₂ prediction is illustrated in Figure 8(a) and (b). The basic LSTM model produced a maximum error of 11% with average errors around 3%–4%, while the percentage of error further, from 20% (see reference¹⁴) to 11%, demonstrating the importance of training on a larger dataset for improving predictive accuracy. The error graph of the Hybrid CNN-LSTM model demonstrates significant advantages over the LSTM-only approach. This proposed model is particularly used in minimizing prediction errors and improving model generalization and it provides enhanced stability and robustness across different VO₂ estimation scenarios.

Results discussions

A statistical comparison between the two models was conducted using *t*-tests to determine whether the observed differences were statistically significant.

Heel Strike Prediction: The *t*-statistic (0.3355, *p* = 0.7372) suggests no significant difference in overall variability, though improvements in precision and recall indicate better performance in detecting true positive HS events.

VO₂ Estimation: The Hybrid CNN-LSTM model exhibited a statistically significant improvement in VO₂ prediction, with a *t*-statistic of 2.1258 and a *p*-value of 0.0336, indicating enhanced model reliability.

The results of this study emphasize the advantages of using transfer learning with LSTM for predicting VO₂ values and detecting HS, particularly when dealing with limited datasets. While the LSTM model demonstrated potential for VO₂ prediction, it exhibited significant errors due to the complexity of the task and the limited data available. LSTM networks are powerful tools for analyzing sequential data, but they often require large amounts of training data to generalize effectively. When data are insufficient, as in this study, LSTM models may struggle to capture the intricate patterns in the dataset, resulting in higher prediction errors. However the percentage error reduction from 20% to 11%, achieved through increasing the dataset size, further underscores the significance of data availability for improving model accuracy.

The application of transfer learning proved to be more effective, as it allowed the model to generalize better despite the small dataset size. By utilizing pre-trained weights from the top layers of the base LSTM model, transfer learning improved prediction accuracy and reduced errors. This is particularly advantageous when working with datasets containing limited subject variability, as transfer learning leverages prior knowledge to enhance model performance on new, unseen data.

The reduction in percentage error from 20% in previous studies¹⁴ to 11% in this study, achieved using LSTM, and further to a maximum error of 3% from the integration of customized transfer learning techniques with LSTM is a key contribution of this work. This customized method is driving significant improvements in accuracy and performance. Furthermore, the increase in dataset size played a

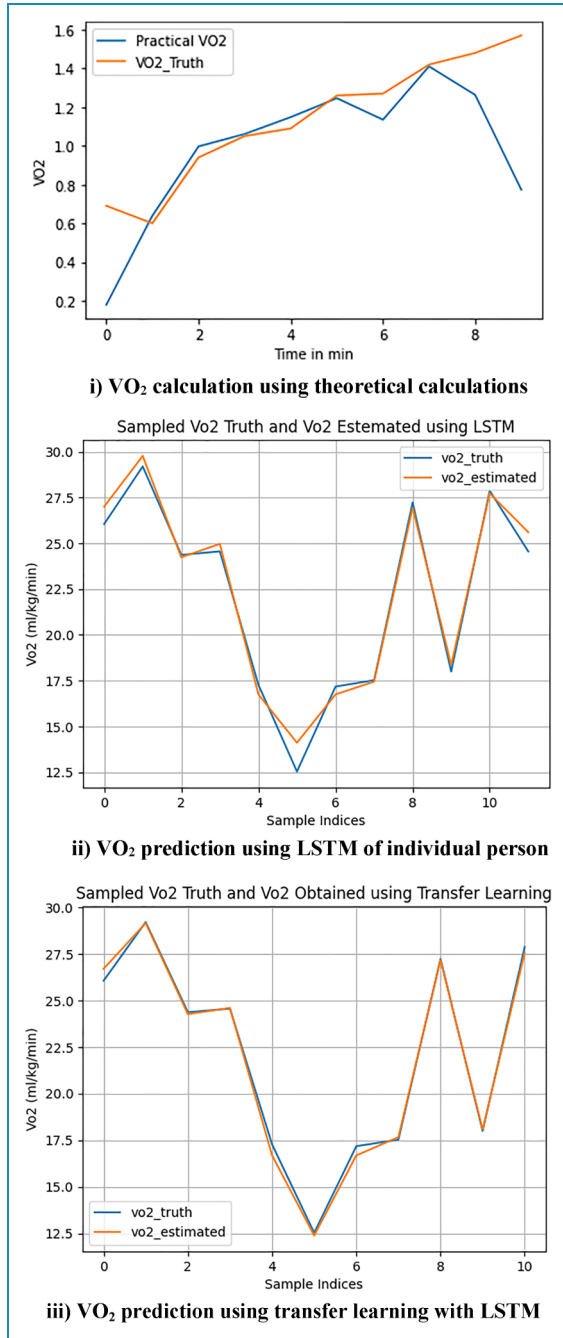


Figure 7. Prediction VO₂ using theoretical calculation, LSTM, and transfer learning with LSTM.

critical role by enabling the model to learn from a broader range of samples, thereby capturing the underlying data patterns more effectively.

In summary, The Hybrid CNN-LSTM model outperforms the LSTM-only approach in both HS prediction and VO₂ estimation, offering enhanced accuracy, precision, and generalizability. These findings underscore the efficacy of integrating deep learning models for

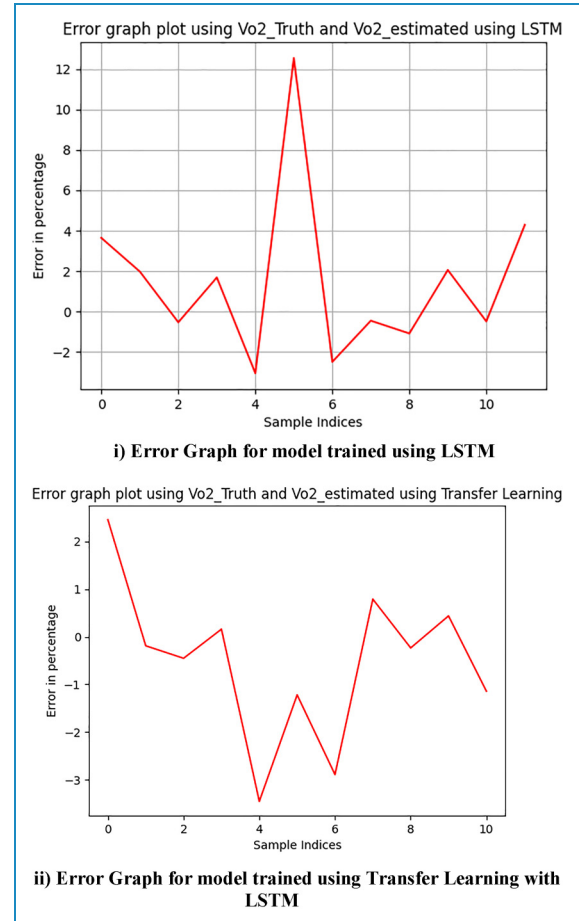


Figure 8. Percentage error plots for the predicted value of VO₂.

real-time gait analysis and metabolic estimation, supporting their application in wearable and rehabilitation technologies.

Limitations

Despite these promising results, the study has limitations that must be acknowledged. The dataset used consists of only ten subjects, with nine subjects for training and one for testing. This small sample size restricts the model's ability to capture the variability inherent in larger, more diverse populations, potentially limiting its generalizability to unseen data. The subject-specific evaluation approach may also impact the model's robustness, particularly when applied to datasets with significant inter-subject variability.

Conclusion

This study demonstrates the potential of a hybrid machine learning model combining CNNs, LSTM networks, and Transfer Learning for estimating HS and EE (VO₂) using data from a single nine-axis. By leveraging transfer learning,

the model achieved a 3% reduction in prediction error compared to traditional systems, highlighting its ability to address the challenges of accuracy and practical application in clinical and sports settings. Despite these advancements, the study's limited dataset size and subject-specific evaluation constrain its generalizability to broader populations, warranting further validation with larger, more diverse datasets.

Future work should focus on integrating advanced techniques like Generative Adversarial Networks (GANs) for synthetic data generation and fusion with LSTM networks offers exciting possibilities to overcome data scarcity and enhance prediction accuracy.

Acknowledgements

The authors would like to thank Dr Karthikbabu S (KMCH, Coimbatore) for his valuable insights into the experimental set up and Manipal Hospital (Bangalore) for providing the facility for data collection.

ORCID iD

Kethohalli R Vidyarani  <https://orcid.org/0009-0007-6112-7629>

Statements and declarations

Ethical considerations

This study was conducted using data collected from healthy participants, and no patient data were involved. Informed consent was obtained from all participants prior to data collection. As the study did not involve invasive procedures, sensitive personal information, or clinical interventions, formal ethical approval from an institutional ethics committee was not sought. The research adhered to the ethical principles outlined in the Declaration of Helsinki to ensure participant rights and welfare was respected throughout.

Author contributions/CRedit

All authors have made substantial contributions to the conception and design of the study, data acquisition, analysis, and interpretation of data, and have drafted the manuscript or revised it critically for important intellectual content. All authors have read and approved the final version of the manuscript, and agree to be accountable for all aspects of the work.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Perry J and Burnfield JM. *Gait analysis: normal and pathological function*. 2nd ed. Thorofare, NJ: Slack Incorporated, 2010.
2. Winter DA. *Biomechanics and motor control of human movement*. 4th ed. Hoboken, NJ: John Wiley & Sons, 2009.
3. Dollar AM and Herr H. Lower extremity exoskeletons and active orthoses: challenges and state-of-the-art. *IEEE Trans Robot* 2008; 24: 144–158.
4. Zhang T, Xing Y, Yang X, et al. Human-in-the-loop optimization of assistive strategies for exoskeletons using reinforcement learning. *IEEE Sens J* 2024; 24: 24723–24736.
5. Del-Ama AJ, Gil-Agudo Á, Pons JL, et al. Hybrid FES-robot cooperative control of ambulatory gait rehabilitation exoskeleton. *J Neuroeng Rehabil* 2014; 11: 27.
6. Tan HX, Aung NN, Tian J, et al. Time series classification using a modified LSTM approach from accelerometer-based data: a comparative study for gait cycle detection. *Gait Posture* 2019; 74: 128–134.
7. Mannini A and Sabatini AM. Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors (Basel)* 2010; 10: 1154–1175.
8. Chen S, Lach J, Lo B, et al. Toward pervasive gait analysis with wearable sensors: a systematic review. *IEEE J Biomed Health Inform* 2016; 20: 1521–1537.
9. Karakasis C and Artemiadis P. F-VESPA: a kinematic-based algorithm for real-time heel-strike detection during walking. In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 27 September 2021 - 01 October 2021. pp.1234–1240.
10. Chia Bejarano N, Ambrosini E, Pedrocchi A, et al. A novel adaptive, real-time algorithm to detect gait events from wearable sensors. *IEEE Trans Neural Syst Rehabil Eng* 2014; 23: 413–422.
11. Ding Z, Yang C, Xing K, et al. The real-time gait phase detection based on long short-term memory. In *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, 18–21 June 2018, pp. 270–275. 2018.
12. Evci F, Saroglu Y and Konukseven EI. Gait recognition and phase detection using wearable IMU sensors and neural network algorithms. In *2023 The 7th International Conference on Advances in Artificial Intelligence (ICAII)*, Istanbul, Turkey, October 2023. <https://doi.org/10.1145/3633598.3633620>. 2023.
13. Rahman H, Kumbala A, Megharjun VN, et al. (2022). Real-time heel strike parameter estimation for FES triggering. In *Distributed computing and optimization techniques*, Lecture notes in electrical engineering, Volume 903. Springer Nature Singapore, pp.749–760.
14. Vidyarani KR, Talasila V, Megharjun N, et al. An inertial sensing mechanism for measuring gait parameters and energy expenditure. *Biomed Signal Process Control* 2021; 70: 103056.
15. Vidyarani KR, Talasila V, et al. Inertial sensor-based heel strike and energy expenditure prediction using a hybrid machine learning approach. Under review in *Digital Health* 2024.
16. Lee CJ and Lee JK. IMU-based energy expenditure estimation for various walking conditions using a hybrid CNN–LSTM model. *Sensors* 2024; 24: 14.
17. Pan SJ and Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010; 22: 1345–1359.

18. Weiss K, Khoshgoftaar TM and Wang D. A survey of transfer learning. *J Big Data* 2016; 3: 9.
19. Krizhevsky A, Sutskever I and Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017; 60: 84–90 DOI: 10.1145/3065386.
20. Ioffe S and Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the 32nd International Conference on Machine Learning (ICML)* 2015; 37: 448–456.
21. Hochreiter S and Schmidhuber J. Long short-term memory. *Neural Comput* 1997; 9: 1735–1780.
22. Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014; 15: 1929–1958.
23. Lester J, Hartung C, Pina L, et al. Validated caloric expenditure estimation using a single body-worn sensor. *11th Int Conf UbiComp*; 2009 Sep 30–Oct 3; Orlando, FL.s.