

# A Deep Learning-based Approach for Foot Placement Prediction

Sung-Wook Lee

Thesis submitted to the Faculty of the  
Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements for the degree of

Master of Science  
in  
Mechanical Engineering

Alan Asbeck, Chair

Alexander Leonessa

Jason Xuan

April 21, 2023

Blacksburg, Virginia

Keywords: Sensor fusion, IMU sensors, Deep learning, Foot placement, Gait phase

Copyright 2023, Sung-Wook Lee

# A Deep Learning-based Approach for Foot Placement Prediction

Sung-Wook Lee

(ABSTRACT)

Foot placement prediction can be important for exoskeleton and prosthesis controllers, human-robot interaction, or body-worn systems to prevent slips or trips. Previous studies investigating foot placement prediction have been limited to predicting foot placement during the swing phase, and do not fully consider contextual information such as the preceding step or the stance phase before push-off. In this study, a deep learning-based foot placement prediction approach was proposed, where the deep learning models were designed to sequentially process data from three IMU sensors mounted on pelvis and feet. The raw sensor data are pre-processed to generate multi-variable time-series data for training two deep learning models, where the first model estimates the gait progression and the second model subsequently predicts the next foot placement. The ground truth gait phase data and foot placement data are acquired from a motion capture system. Ten healthy subjects were invited to walk naturally at different speeds on a treadmill. In cross-subject learning, the trained models had a mean distance error of 5.93 cm for foot placement prediction. In single-subject learning, the prediction accuracy improved with additional training data, and a mean distance error of 2.60 cm was achieved by fine-tuning the cross-subject validated models with the target subject data. Even from 25-81% in the gait cycle, mean distance errors were only 6.99 cm and 3.22 cm for cross-subject learning and single-subject learning, respectively.

# A Deep Learning-based Approach for Foot Placement Prediction

Sung-Wook Lee

(GENERAL AUDIENCE ABSTRACT)

This study proposes a new approach for predicting where a person's foot will land during walking, which could be useful in controlling robots and wearable devices that work with humans to prevent events such as slips and falls and allow for more smooth human-robot interactions. Although foot placement prediction has great potential in various domains, current works in this area are limited in terms of practicality and accuracy. The proposed approach uses data from inertial sensors attached to the pelvis and feet, and two deep learning models are trained to estimate the person's walking pattern and predict their next foot placement. The approach was tested on ten healthy individuals walking at different speeds on a treadmill, and achieved state-of-the-arts results. The results suggest that this approach could be a promising method when sufficient data from multiple people are available.

# Dedication

*To my parents and friends.*

# Acknowledgments

I would like to thank my father, Dr. Jeong Ho Lee, and my mother, Dr. Hee Ok Jeon, for their endless love and the patient support they have given me throughout my academic journey. I know I would not be where I am without your sacrifice.

I would also like to thank my supervisor and committee chair, Dr. Alan Asbeck, for giving me the research opportunity that I was truly interested in. I learned a lot during the process and the advises and the guidance you gave me were invaluable.

I would like to thank the respected committee members, Dr, Jason Xuan and Dr. Alexander Leonessa, for their commitments to their departments and the prolific research contributions to the research community. I also took courses taught by both professors and I was able to build a strong theoretical foundation for the research.

# Contents

<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Organization . . . . .	2
<b>2 Review of Literature</b>	<b>5</b>
2.1 Human gait estimation . . . . .	5
2.2 Foot placement prediction . . . . .	6
2.3 Limitations of the prior works . . . . .	7
<b>3 Background</b>	<b>8</b>
3.1 Vector/Matrix Notations . . . . .	8
3.2 Sensor system . . . . .	8
3.3 Sensor Placement-Free Orientation . . . . .	10
3.4 Foot Placement . . . . .	11
3.5 Recurrent Neural Networks for IMU Data . . . . .	12
<b>4 Methods</b>	<b>14</b>

4.1	Gait Phase Estimation . . . . .	15
4.1.1	Data Processing . . . . .	15
4.1.2	Deep Learning Model Architecture . . . . .	17
4.2	Foot Placement Prediction . . . . .	18
4.2.1	Gait Phase-based ZUPT (Zero Velocity Potential Update) Correction . . . . .	18
4.3	Data Processing . . . . .	19
4.4	Deep Learning Model . . . . .	21
4.5	Pre-Training and Fine-Tuning . . . . .	21
4.6	Human Subjects Data Collection . . . . .	22
4.7	Model Training . . . . .	23
4.8	Foot Placement Prediction Evaluation . . . . .	24
4.9	Interpretable AI/SHAP (SHapley Additive exPlanation) . . . . .	25
<b>5</b>	<b>Results</b>	<b>27</b>
5.1	Dataset Metrics . . . . .	27
5.2	Gait Phase Estimation and Foot Placement Prediction Accuracy . . . . .	27
5.3	Early Gait Phase Foot Placement Prediction . . . . .	30
5.4	Interpretation using SHAP values . . . . .	31
<b>6</b>	<b>Discussion</b>	<b>35</b>
<b>7</b>	<b>Conclusions</b>	<b>39</b>

<b>Bibliography</b>	<b>40</b>
<b>Appendices</b>	<b>45</b>
<b>Appendix A Discussion of Methods and Findings Not Included in the Main Body</b>	<b>46</b>
A.1 Deep Learning Architecture . . . . .	46
A.2 Temporal Rescaling . . . . .	47
A.3 Additional Features . . . . .	49
A.4 SHAP-based Feature Selection . . . . .	50
A.5 Multiple Foot Placements Prediction . . . . .	51
A.6 Different Sensor Placements . . . . .	51

# List of Figures

3.1	The sensor system. The figure on the left shows the fully-body MVN link system. The true placement of the foot mounted sensors is inside the shoes. The figure on the right is the MTx tracker. . . . .	9
3.2	Illustration of sensor-fixed coordinate systems and foot placement vector. Legend: x-axis: blue, y-axis: red, z-axis: green. Foot placement is defined as the displacement of right heel from the initial position when $GP = n_R\%$ to the next foot placement when $GP = 100+n_R\%$ , expressed in the coordinate system of the pelvis. . . . .	12
4.1	A visual illustration of the proposed approach. The proposed foot placement prediction method consists of two deep learning models: a GRU (Gated Recurrent Unit) model for gait phase estimation and a GRU model for foot placement prediction. For the foot placement prediction model, position and velocity features are corrected using ZUPT (Zero-Velocity Potential Update). The ground truth gait cycle data and 3D foot placement data are generated by processing the full-body recording data using the MVN Analyze software. . . . .	14
4.2	The gait phase label data and the 2d transformation. Note that for both right foot (RF) and left foot (LF), heel strike (HS) information provided by the MVN Analyze software is used to identify the foot placement (PM) instance independently. $n_R$ and $n_L$ indicate the gait phase progression percentages for right and left foot placements, respectively. Linear gait progression is assumed for linear extrapolation. . . . .	17

4.3	The ZUPT interval is parameterized by the start and end of the interval expressed in terms of gait phase percentage. The boxed green cell on the left indicates the interval with the least position drift MSE error. On the right, the foot velocity expressed in the sensor coordinate system is shown with the gait cycle progression. The foot velocity is set to zero during the ZUPT interval (grey area). . . . .	19
-----	--	----

5.1	Results of foot placement prediction with deep learning models. (a) The right gait phase RMSE for training datasets of different sizes. (b) and (c) Normalized histogram for right foot gait phase error for the base (orange) and fine-tuning (green) case. Note that this error is calculated by reconstructing the gait phase value from the output of the gait phase evaluation model. (d) Foot placement error when trained with datasets of different sizes. (e) and (f) Normalized histogram for foot placement error for the base (orange) and fine-tuning (green) case. . . . .	28
-----	--	----

5.2	(a) Foot placement error with gait progression. Gait phase values denote an interval of gait progression after the last foot placement, for example, 20 means 20~21% gait progression after the last foot placement, 40 means 40~41%, and so on. (b) Foot placement error for different step lengths. The step length refers to the distance between the last foot placement to the succeeding one. Note that the Gait Phase percentage is the percentage after the foot placement; this corresponds to approximately 25% in the traditional gait cycle after heel strike, so 87% on the graph is approximately the next heel strike. (c) X,Y, and Z components of foot placement error. (d) The foot placement prediction results for each participant using the last foot placement predictor, the base models, the pre-trained models, and the fine-tuned models. The shaded areas and the error bars indicate 95% confidence intervals. . . . .	29
5.3	The average SHAP values for all 480 timestamps using the fine-tuned models.	32
5.4	The average SHAP values computed for all 33 features using the fine-tuned models. . . . .	33
5.5	The average SHAP values for each feature using the fine-tuned models across 10 participants. . . . .	34
A.1	Different deep learning architectures . . . . .	47
A.2	Raw time series data vs temporally re-scaled series data for each gait. Green color indicates fast walks, and blue color indicates slow walks. . . . .	48

# List of Tables

4.1 Training Parameters .....	24
-------------------------------	----

# List of Abbreviations

a	Acceleration
CPU	Central Processing Unit
GPU	Graphics Processing Unit
GRU	Gated Recurrent Units
HS	Heel strike
IMU	Inertial Measurement Unit
IRB	Institutional Review Boards
LF	Left foot
LIPM	Linear Inverted Pendulum Model
MDE	Mean Distance Error
MSE	Mean Squared Error
NLP	Natural Language Processing
P, p	Position
PM	Placement
PV	Pelvis
q	Quaternion

R     Rotation matrix

RF    Right foot

RFE   Recursive Feature Elimination

RMSE  Root Mean Squared Error

RNN   Recurrent Neural Network

STD   Standard deviation

v     Velocity

ZUPT  Zero Velocity Potential Update

# Chapter 1

## Introduction

The placement of the foot is crucial for maintaining balance and stability, and it has a significant impact on gait behavior [1]. Foot placement influences the length and width of the stride, as well as the amount of force applied to the ground during each step [2]. It also plays a crucial role in indicating if a person is trying to restore balance, accelerate or decelerate, or if a person is in danger of losing balance and about to fall. Additionally, it is a great indicator of the trajectory of the body's center of mass and its orientation relative to the ground as well as a meaningful predictor of the future motions [3]. Overall, analyzing foot placement can provide insights into the underlying mechanisms of human gait, help identify factors that may lead to gait abnormalities or instability, and predict future events.

While the importance of foot placement is well-established, predicting it accurately can be challenging. The placement of the feet during locomotion is determined by a complex interplay between internal and external factors. External factors such as the nature of the walking surface can influence foot placement. Meanwhile, internal factors like changes in intention, physical and cognitive conditions also play a significant role in foot placement. [1]. This complex interplay of the various factors makes it very difficult to accurately predict foot placement.

Despite the challenges of foot placement prediction, it presents with high potential across various domains, such as fall prevention, rehabilitation, human-machine interaction, and bipedal robot control. It can be employed to identify risk factors for falls in elderly or at-risk

populations, and to design interventions to mitigate such risks [1]. In the context of rehabilitation, it can be leveraged to create personalized rehabilitation programs for individuals with mobility impairments, such as those caused by stroke or spinal cord injury. Additionally, it can enable natural and intuitive control of devices, such as virtual reality systems, through foot gestures in human-machine interaction [2]. In the field of robotics, anticipating the next foot placement in advance can help enhance the stability and robustness of bipedal robot control in changing environments [3].

Until recently, limitations in sensor technology and computational capacity of processing units have hindered the analysis and prediction of human kinematics/motions. However, with the advent of advanced sensors with higher accuracy and frequency, deep learning models capable of processing millions of data points, and the parallel computation units that allow for rapid learning, it has become possible to implement a large-scale deep learning model using big data. As a result, deep learning has recently emerged as a viable option for analyzing real-time data and generating predictive conclusions in various domains, and the interest for deep learning-based methods for sensor fusion has been growing rapidly in the last few years.

## 1.1 Organization

This thesis is organized as follows:

### Review of Literature

This section examines relevant studies and research papers related to foot placement prediction. It focuses on identifying key findings and insights from previous research and discussing

their implications for our study. Additionally, it highlights the limitation of the existing works and makes a strong case for the proposed approach.

## **Background**

This section includes the background information needed to understand the study, such as the sensor system, deep learning models, and the mathematical notations used in the Methods section. This section also defines the objective of the study.

## **Methods**

In this section, the methods used to conduct our study on the prediction of foot placement are explained in detail, including the sensor data pre-processing, deep learning model structures, data collection procedures, and the training scheme.

## **Results**

In this section, the results of the study are presented and analyzed. The findings are visualized in a number of figures, and any relevant statistical data is provided.

## **Discussion**

In this section, the results are compared to the existing literature, highlighting any differences between the current study and previous research. The implications of the findings are also discussed, and suggestions for future research are presented.

## Conclusions

In this section, the main findings of the study are summarized, and the implications of the results are discussed. The conclusions drawn from the study are presented.

# Chapter 2

## Review of Literature

### 2.1 Human gait estimation

The majority of previous research on foot placement has been focused on measuring foot placement. In [4], a single IMU sensor was used to measure the step length, achieving error smaller than 3% of the traversed distance. In [? ], a method for measuring the step length and the step width using four IMUs mounted on both feet achieved a mean stride length and duration within 1% of the ground truth.

Another very important gait estimation task is the continuous gait phase estimation. In the context of walking, gait phase refers to the specific stages of a person's walking cycle, and heel strike is usually defined as the start of a gait cycle. The aim of the continuous gait phase estimation is to process sensor information in real time to estimate the current gait phase with high frequency. In [5], Lee continuously estimated the gait phase during walking and reported an average error of  $1.67 \pm 1.36\%$  and  $1.45 \pm 1.47\%$  for walking speeds of 0.5 m/s and 1.5 m/s, respectively, using two IMU sensors. In [? ], Zhang used an IMU sensor to estimate the thigh angle and continuously estimate the gait phase with the root-mean-square error (RMSE) of  $4.14 \pm 1.68\%$  for steady walking.

## 2.2 Foot placement prediction

Predicting the trajectory of human motion has been extensively studied for a variety of applications, including robotics, autonomous vehicles, and human-robot interaction [6]. More recently, many lower body joint angle prediction studies have been proposed [7, 8]. However, in the context of foot position prediction, the existing literature is limited. [9] analyzed the single footstep for recovery after an external perturbation and reported  $R^2$  value of 0.858 between the perturbation momentum and the step position. [10] expanded to multiple steps for recovery, and found a mean error of 15 cm and 18 cm for the first foot step and the second foot step, respectively, using a passive walking model and a cost optimization method. For continued walking, [11] concluded that 80% of the next foot step can be explained using the pelvis states during mid-stance phase. Also, Zhang *et al.* [12] used a vision sensor for 3D gaze estimation to predict foot placements on rough terrain, achieving a mean error of 18 cm. They improved the resultsto 8.6 cm by fusing it with a pre-defined environment context. Chen *et al.* [13] manually extracted features from a single IMU on the foot and used a Bayesian inference algorithm to achieve 10 cm RMSE error along the forward axis and horizontal axis early in the swing phase. In a separate study, Chen *et al.* [14] used foot position data from a motion capture system and compared three machine learning algorithms, achieving an RMSE error of 4.4 cm in the forward direction and 4.2 cm in the sideways direction.ving an RMSE error of 4.4 cm in the forward direction and 4.2 cm in the sideways direction.

## 2.3 Limitations of the prior works

The previous methods for predicting foot placement during human walking have been limited by many assumptions that make them impractical for realistic use cases. For example, [14] uses motion capture data during inference, which requires offline calibration, and is limited to a laboratory setting. [13] requires manual extraction of features and is evaluated on a discrete set of walking speeds picked by the researchers. None of the above studies consider the temporal dependencies between the sensor readings of different timestamps and between neighboring steps, and the prediction time window is limited to the first half of the swing phase. These methods also rely on having the ground truth start of a gait, which is impractical in realistic use cases without gait phase estimation. Additionally, the prediction frequency for all the existing methods is once per step, which means that they are not capable of updating its prediction location using the most recent sensor data within a gait cycle. Lastly, none of these methods predict the full 3d information of the foot placement location, and relies on an absolute coordinate system for evaluation. The above shortcomings contribute to why they struggle with accuracy, particularly when predicting early or across subjects.

# Chapter 3

## Background

### 3.1 Vector/Matrix Notations

We use the following notation for representing vectors:  ${}^A_B v$  is a vector  $v$  relative to frame B, expressed in frame A.  $R_{B,A}(t)$  is the rotation matrix describing the rotation from frame B to frame A, and  $q_{B,A}(t)$  is the quaternion representation of the identical rotation matrix. Lastly,  ${}^C p_{B,A}(t)$  is the distance vector describing the translation from frame B to frame A, expressed in frame C.

### 3.2 Sensor system

To experiment with an IMU-based foot placement prediction approach and obtain accurate foot placement data, we utilize a motion capture system known as the MVN Link system. The MVN Link system is a wireless motion capture system that uses miniature inertial measurement units (IMUs) to track the motion of the human body in real-time.

Each MTx and MTx-STR tracker contains a suite of sensors, including 3D linear accelerometers, 3D rate gyroscopes, 3D magnetometers, and a barometer to measure atmospheric pressure. These sensors work together to provide highly accurate measurements of the motion of each body segment.



Figure 3.1: The sensor system. The figure on the left shows the fully-body MVN link system. The true placement of the foot mounted sensors is inside the shoes. The figure on the right is the MTx tracker.

The MTx trackers are fixed to the pelvis, sternum, hands, and head using a suit or harness, while the MTx-STR trackers are used to chain the upper and lower legs and feet, as well as the upper body. The MTx-STR trackers are connected in series, with one tracker placed on the upper leg, one on the lower leg, and one on the foot.

Each sensor unit fuses the inertial and magnetic sensor signals to estimate orientation, and calculates the gravity-compensated acceleration. The raw sensor data are used as inputs to our prediction algorithms. The raw sensor outputs of interest are as follows: 1) the gravity-compensated acceleration of the pelvis (PV), right foot (RF), and left foot (LF) sensor

module with respect to the global origin (G), expressed in the global coordinate system:  ${}^G_G a_{PV}$ ,  ${}^G_G a_{LF}$ ,  ${}^G_G a_{RF}$ ; and 2) the estimated orientation of the pelvis, right foot, and left foot sensor module with respect to the global coordinate system:  $q_{G,PV}$ ,  $q_{G,LF}$ ,  $q_{G,RF}$ .

### 3.3 Sensor Placement-Free Orientation

The orientation of sensors can be different between recording sessions even with careful measures, due to the variability of people's physical characteristics such as foot shape. The varying sensor orientation across different recording sessions introduces a misalignment between sensor data. To address this, we propose a methodology for calculating the sensor placement-free orientation.

For a given session, the desired orientation  $R_{G\_Target}$  is determined using the distance vector  $d_{fwd}$ , which is measured while the participant is commanded to walk forward on level ground. The corresponding unit vector  $u_{travel}$  is calculated using the following equations.

$$d_{travel} = {}^G P_{G,PV}(t_{end}) \quad (3.1)$$

$$u_{fwd} = \frac{d_{travel}}{\|d_{travel}\|} \quad (3.2)$$

$$R_{Target} = \begin{bmatrix} u_{fwd,X} & -u_{fwd,Y} & 0 \\ u_{fwd,Y} & u_{fwd,X} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.3)$$

The transformation matrix from the sensor orientation at each instance to the desired orientation  $R_{G\_Target}$  is averaged based on [15], and the transformation matrix for each sensor unit is multiplied with the corresponding sensor orientation to calculate the sensor placement-free

orientation:

$$\begin{aligned}
 R_{calibrated}(t) &= R_{sensor}(t) \cdot \\
 &avg(R_{sensor}(t_i)^{-1} \cdot R_{Target}, \forall t_i \in \{0, \dots, t_{end}\})
 \end{aligned}
 \tag{3.4}$$

## 3.4 Foot Placement

The next foot placement is calculated as the displacement of the right heel (RH) from its previous foot placement to its next foot placement in the calibrated coordinate system of the pelvis. The foot is assumed to be 'placed' at the gait cycle percentage with the lowest magnitude of the ground truth mean foot velocity. The next foot placement is defined mathematically as:

$$\begin{aligned}
 F(t) &= R_{G,PV}(t)^{-1} \cdot ({}^G P_{G,RH}(t_{GP=100+n_R\%}) - \\
 &{}^G P_{G,RH}(t_{GP=n_R\%})), \forall t \in \{t_{GP=n_R\%}, \dots, t_{GP=100+n_R\%}\},
 \end{aligned}
 \tag{3.5}$$

where  $n_R$  is the gait phase percentage for foot placement;  $n_R$  is found separately for each training dataset.

The post-processed foot position data and the heel strike ( $t_{GP=0\%}$ ) is provided by MVN Analyze 2021. A visual illustration of foot placement is shown in Fig. 3.2. The objective of this study is to predict the right foot placement using the raw sensor data from 3 IMUs mounted on pelvis and feet, before the foot placement occurs.

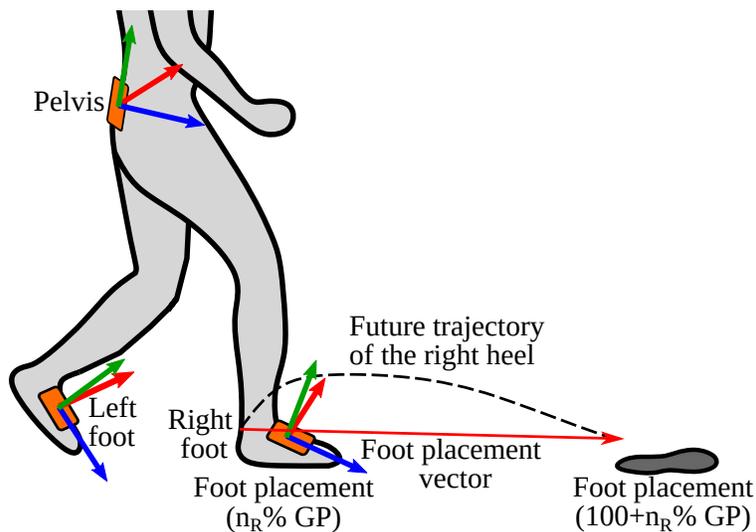


Figure 3.2: Illustration of sensor-fixed coordinate systems and foot placement vector. Legend: x-axis: blue, y-axis: red, z-axis: green. Foot placement is defined as the displacement of right heel from the initial position when  $GP = n_R\%$  to the next foot placement when  $GP = 100+n_R\%$ , expressed in the coordinate system of the pelvis.

### 3.5 Recurrent Neural Networks for IMU Data

Recurrent Neural Networks (RNNs) are well-suited for processing IMU data as they can handle temporal dependencies and high-dimensional and highly frequency data [16, 17, 18]. RNNs can also learn non-linear relationships and store information from previous time steps.

In this study, we choose GRUs (Gated Recurrent Units) for sequential data processing, a computationally efficient and low-memory variation of RNNs. A GRU network is composed of a series of interconnected cells, with each cell having a hidden state that is updated at each time step. The update process in a GRU network is controlled by a set of gating mechanisms that regulate the flow of information through the network. These mechanisms include an update gate, which determines how much of the previous hidden state should be retained, and a reset gate, which determines how much of the new input should be incorporated. GRUs have been shown to effectively model long-term dependencies while mitigating the vanishing gradient problem that is common in traditional RNNs. In this study, we use GRUs to process

pre-processed multi-variable time-series data obtained from raw sensor data.

# Chapter 4

## Methods

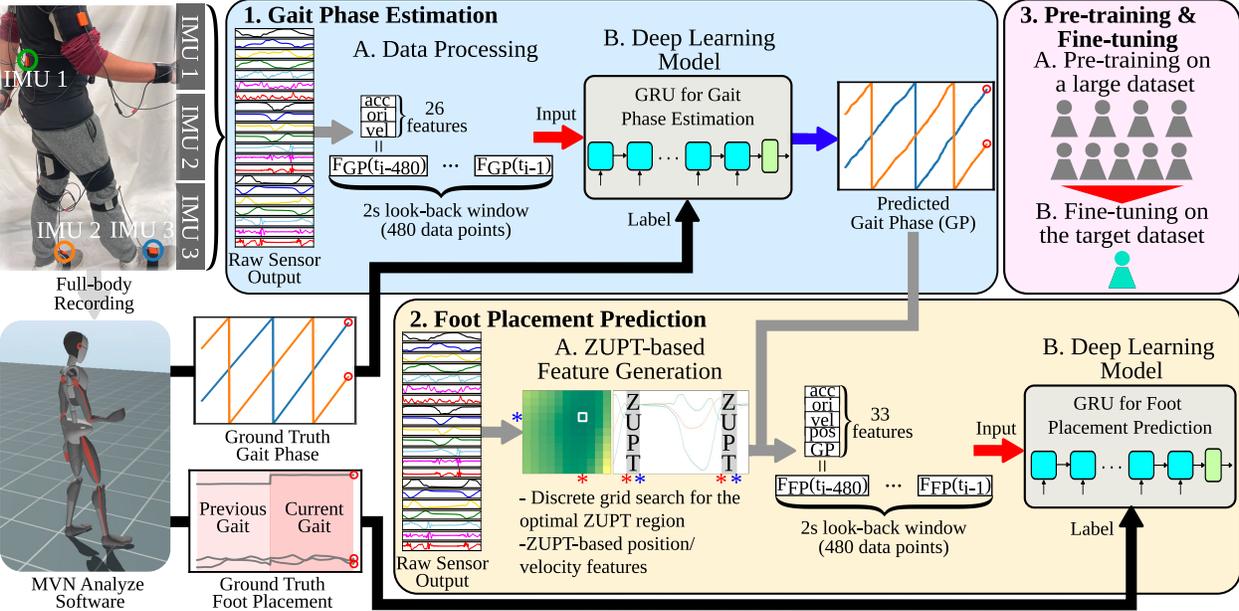


Figure 4.1: A visual illustration of the proposed approach. The proposed foot placement prediction method consists of two deep learning models: a GRU (Gated Recurrent Unit) model for gait phase estimation and a GRU model for foot placement prediction. For the foot placement prediction model, position and velocity features are corrected using ZUPT (Zero-Velocity Potential Update). The ground truth gait cycle data and 3D foot placement data are generated by processing the full-body recording data using the MVN Analyze software.

Figure 4.1 shows the overview of the approach.

## 4.1 Gait Phase Estimation

In this section, the methodologies for developing a recurrent neural network for gait phase estimation are discussed.

### 4.1.1 Data Processing

The raw sensor data are pre-processed to build the dataset for training the deep learning model. The input features are:

*Acceleration:* The pelvis, right foot, and left foot acceleration data are expressed in the orientation of the calibrated pelvis coordinate system, denoted as  ${}^G_{PV}\hat{a}_{PV}$ ,  ${}^G_{RF}\hat{a}_{RF}$ , and  ${}^G_{LF}\hat{a}_{LF}$ , respectively.

*Orientation:* The quaternion representation of the right foot and left foot orientation with respect to that of the pelvis are expressed in the orientation of the calibrated pelvis coordinate system, denoted as  $\hat{q}_{PV,RF}$  and  $\hat{q}_{PV,LF}$ , respectively.

*Velocity:* The pelvis, right foot, and left foot velocity are calculated by integrating the high-pass filtered (5th order Butterworth filter with 0.1 Hz cutoff frequency) acceleration data. The velocity features are normalized to the pelvis coordinate system, and are denoted as  ${}^G_{PV}\hat{v}_{PV}(t_i)$ ,  ${}^G_{RF}\hat{v}_{RF}(t_i)$ , and  ${}^G_{LF}\hat{v}_{LF}(t_i)$ .

The acceleration, orientation, and velocity data are each scaled to zero mean and unit variance then concatenated along the feature axis, which is denoted as  $F_{GP}(t)$ : To better handle different sensor placements and local peaks that can affect the scaling especially when the training data is small, the zero mean and the unit variance scaling is done per feature category, instead of per feature. For example, 6 velocity features (3 from right foot and 3 from left foot) are scaled by the same scaler, instead of 6 individual scalers.

The input time series data are generated by vertically concatenating the combined sensor output with a look-back window of 2 seconds (480 data points), as shown below. The 2 second look-back window is chosen because it is sufficient enough to capture at least a single full gait cycle.

$$F_{GP}(t_i) = \left\{ \begin{array}{l} \frac{PV}{G} \hat{a}_{PV}(t_i), \frac{PV}{G} \hat{a}_{RF}(t_i), \frac{PV}{G} \hat{a}_{LF}(t_i), \\ \hat{q}_{PV,RF}(t_i), \hat{q}_{PV,LF}(t_i), \frac{PV}{G} \hat{v}_{PV}(t_i), \\ \frac{PV}{G} \hat{v}_{RF}(t_i), \frac{PV}{G} \hat{v}_{LF}(t_i) \end{array} \right\} \in \mathbb{R}^{26} \quad (4.1)$$

$$X_{GP}(t_i) = \left\{ \begin{array}{l} F_{GP}(t_{i-480}) \\ F_{GP}(t_{i-479}) \\ \vdots \\ F_{GP}(t_{i-1}) \end{array} \right\} \in \mathbb{R}^{480 \times 26} \quad (4.2)$$

In order to accurately label gait phase data, we define a continuous vector  $Y_{GP,R}$  which is 0 at 0% gait phase of each gait cycle (heel strike) and linearly increases to 1 at 100% before resetting back to 0. To eliminate discontinuities at the reset points, we transform the this vector into a two-dimensional continuous vector  $(Y_{GP1,R}, Y_{GP2,R})$ , similar to [5, 19, 20].

$$Y_{GP1,R} = \sin(2\pi \cdot Y_{GP,R}) \quad (4.3)$$

$$Y_{GP2,R} = \cos(2\pi \cdot Y_{GP,R}) \quad (4.4)$$

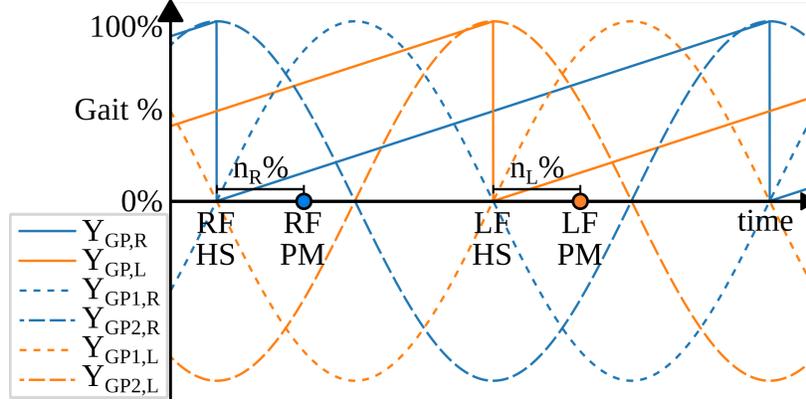


Figure 4.2: The gait phase label data and the 2d transformation. Note that for both right foot (RF) and left foot (LF), heel strike (HS) information provided by the MVN Analyze software is used to identify the foot placement (PM) instance independently.  $n_R$  and  $n_L$  indicate the gait phase progression percentages for right and left foot placements, respectively. Linear gait progression is assumed for linear extrapolation.

The inverse transform equation to derive  $Y_{GP}$  given  $(Y_{GP1}, Y_{GP2})$  is:

$$Y_{GP,R} = \frac{\text{atan2}(-Y_{GP1,R}, -Y_{GP2,R}) + \pi}{2\pi} \quad (4.5)$$

Similarly,  $Y_{GP,L}$  refers to the left foot gait phase data. The resultant gait phase labeled data  $Y_{GP}(t)$  is  $(Y_{GP,R1}(t); Y_{GP,R2}(t); Y_{GP,L1}(t); Y_{GP,L2}(t)) \in \mathbb{R}^4$ , as shown in Fig. 4.2.

### 4.1.2 Deep Learning Model Architecture

Each training sample  $X \sim X_{GP} \in \mathbb{R}^{l \times m}$  is a multi-variable time-series of length  $l$  (480) with  $m$  (26) features.

The uni-directional and single-layered GRU (Gated Recurrent Unit) with  $k$ -dimensional hidden states sequentially processes the time-series data and outputs the hidden states from each unit. The hidden state of the final layer of the GRU goes through a feedforward layer to output the estimated gait phase vector:

$$o, h = GRU(X), \quad (4.6)$$

$$\hat{Y}_{GP} = W_w h + b_w, \quad (4.7)$$

$$L_{MSE} = \sum_{i=1}^d (\hat{Y}_{GPi} - Y_{GPi})^2, \quad (4.8)$$

where  $o$  and  $h$  are the output and the hidden state of the GRU,  $W_w$  and  $b_w$  are the learnable parameters of the final feedforward network.  $Y_{GP}$  is the ground truth gait phase data of  $d$  dimensions, which is 4. The mean square error (MSE) loss is used for learning.

## 4.2 Foot Placement Prediction

In this section, the methodologies for developing a recurrent neural network for foot placement prediction are discussed.

### 4.2.1 Gait Phase-based ZUPT (Zero Velocity Potential Update) Correction

To provide more information to the deep learning model, position features are generated by double-integrating the acceleration data from the foot sensors. To correct the position drift error, we employ a ZUPT correction method, similar to [21]. We assume that the movement of the feet is significantly correlated to gait phase, and that it is possible to identify a range of percentages in the gait cycle during which the foot velocity is zero. This is particularly advantageous because it allows for automatic labeling of the zero velocity region by utilizing the gait phase estimates. It is also a more direct and interpretable way of identifying the zero-velocity interval compared to other ZUPT methods, which usually consider factors such as sensor characteristics, detection threshold, and gait type [21, 22, 23]. Since the exact

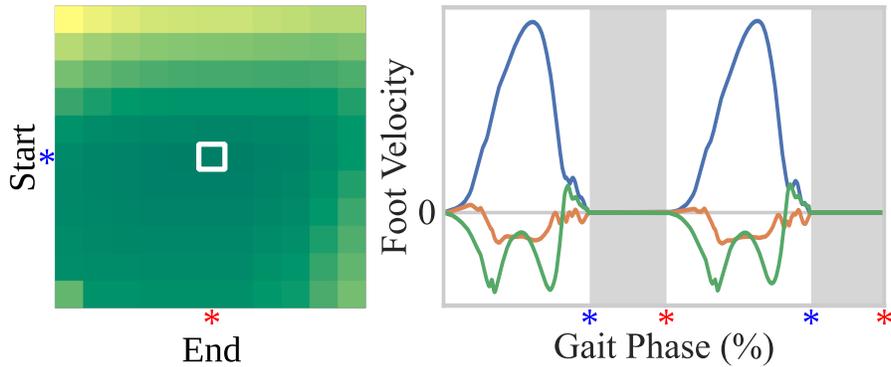


Figure 4.3: The ZUPT interval is parameterized by the start and end of the interval expressed in terms of gait phase percentage. The boxed green cell on the left indicates the interval with the least position drift MSE error. On the right, the foot velocity expressed in the sensor coordinate system is shown with the gait cycle progression. The foot velocity is set to zero during the ZUPT interval (grey area).

range of this interval is subject-dependent, we perform a two-dimensional grid search for the start and end of the interval to find the optimal range for each subject. The optimal interval minimizes the mean squared error (MSE) of the total position drift error. During the zero-velocity interval, the foot velocity is set to zero (Fig. 4.3).

### 4.3 Data Processing

The raw sensor data are pre-processed again to build the dataset for training the deep learning model. The input features are:

*Acceleration and Orientation:* The pelvis, right foot, and left foot acceleration data are expressed in the orientation of the calibrated pelvis coordinate system, denoted as  ${}^G{}^{PV}\hat{a}_{PV}$ ,  ${}^G{}^{PV}\hat{a}_{RF}$ , and  ${}^G{}^{PV}\hat{a}_{LF}$ , respectively.

*Velocity:* After each ZUPT interval (where the velocity is set to zero), the right foot, and left foot velocity are calculated by integrating the high-pass filtered (5th order Butterworth filter with 0.01 Hz cutoff frequency) acceleration data. The velocity features are normalized

to the pelvis coordinate system, and are denoted as  ${}^G{}^{PV}\hat{v}_{RF}(t_i)$ , and  ${}^G{}^{PV}\hat{v}_{LF}(t_i)$ , respectively.

*Displacement:* The estimated displacement of right foot and left foot is calculated by integrating the ZUPT-corrected velocity data to estimate the feet positions and subtracting by the previous foot placement. The estimated right foot and left foot positions at time stamp  $t_i$  are denoted as  $\hat{p}_{G,RF}(t_i)$  and  $\hat{p}_{G,LF}(t_i)$ , respectively, and the estimated last foot placement is denoted as  $\hat{p}_{G,RF}(t_{RFLP})$  and  $\hat{p}_{G,LF}(t_{LFLP})$ .

*Estimated Gait Phase:* The estimated gait phase for both right and left foot from the gait phase estimation model are used as input features for foot placement prediction (Fig. 4.1).

The features listed above are each scaled to zero mean and unit variance and concatenated along the feature axis, denoted as  $F_{FP}(t)$ . The input time series data are generated by vertically concatenating the combined sensor output with a look-back window of 2 seconds (480 data points):

$$F_{FP}(t_i) = \left\{ \begin{array}{l} {}^G{}^{PV}\hat{a}_{PV}(t_i), {}^G{}^{PV}\hat{a}_{RF}(t_i), {}^G{}^{PV}\hat{a}_{LF}(t_i), \hat{q}_{PV,RF}(t_i), \\ \hat{q}_{PV,LF}(t_i), {}^G{}^{PV}\hat{v}_{RF}(t_i), {}^G{}^{PV}\hat{v}_{LF}(t_i), \\ {}^{PV}(\hat{p}_{G,RF}(t_i) - \hat{p}_{G,RF}(t_{RFLP})), \\ {}^{PV}(\hat{p}_{G,LF}(t_i) - \hat{p}_{G,LF}(t_{LFLP})), \end{array} \right. \quad (4.9)$$

$$X_{FP}(t_i) = \left\{ \begin{array}{l} F_{FP}(t_{i-480}) \\ F_{FP}(t_{i-479}) \\ \vdots \\ F_{FP}(t_{i-1}) \end{array} \right\} \in \mathbb{R}^{480 \times 33} \quad (4.10)$$

## 4.4 Deep Learning Model

Each training sample  $X \sim X_{FP} \in \mathbb{R}^{l \times m}$  is a multi-variable time-series of length  $l$  (480) with  $m$  (33) features. The uni-directional and single-layered GRU (Gated Recurrent Unit) with  $k$ -dimensional hidden states and a final feedforward layer to output the predicted value are selected as the deep learning model. The structure of the model is similar to that of the gait phase estimation model.

$$o, h = GRU(X), \quad (4.11)$$

$$\hat{Y}_{FP} = W_w h + b_w, \quad (4.12)$$

$$L_{MSE} = \sum_{i=1}^d (\hat{Y}_{FPi} - Y_{FPi})^2, \quad (4.13)$$

where  $o$  and  $h$  are the output and the hidden state of the GRU,  $W_w$  and  $b_w$  are the learnable parameters of the final feedforward network.  $Y_{FP}$  is the ground truth foot placement data of  $d$  dimensions, which is 3. The mean square error (MSE) loss is used for learning.

## 4.5 Pre-Training and Fine-Tuning

In this study, we conduct a comparison between three different training cases for predicting foot placement. The first, referred to as the pre-training case, involves cross-subject training on a large, diverse dataset of walking data from multiple people excluding the target subject. Pre-training is a widely used machine learning technique in fields such as Natural Language Processing, image processing, and speech recognition, where a model is first trained on a larger, more diverse dataset before being fine-tuned on a specific task or dataset [24, 25, 26]. In this study, through cross-validating with other people's walking data, the models are

trained to generalize to the walking patterns of an unseen person. These cross-validated models are also used for cross-subject evaluation, which measures the models' effectiveness to learn without any supervision (ground truth) from the target subject.

The second case, referred to as the base case, involves training the models on the target subject dataset without pre-training.

Finally, the third case, referred to as the fine-tuning case, uses the pre-trained models as the initial models and subsequently trains them with the target subject datasets of different sizes. This fine-tuning process allows the models to identify important parameters of the target's gait such as the sensor orientation transformation matrix, foot placement timing, and the zero-velocity interval, and to directly learn the gait patterns of the target subject.

## 4.6 Human Subjects Data Collection

The IMU data and the ground truth foot placement data are collected using an XSens MVN Link suit. The locations of the sensors are shown in Fig. 3.2. Ten (8 male and 2 female) healthy subjects participated in the experiment, which was approved by the Virginia Tech Institutional Review Board (IRB #22-665). The average age and height is 25.6 and 173.2 cm. All participants had never suffered from musculoskeletal diseases. During the experiment, subjects were asked to walk naturally on a treadmill. During each walking session, the treadmill speed increased from 1.0 mph ( $\sim 0.45$  m/s) speed to 3.0 mph ( $\sim 1.34$  m/s) with increments of 0.1 mph then slowed back down to 1.0 mph with 0.1 mph decrements. The treadmill stayed at each speed for different periods of time, depending on the session. The intervals used for each of the sessions were 1 second, 2 second, 5 second, 10 second, 15 second, and 30 second. Using these different intervals in different sessions allowed us to investigate how the size of the training dataset affects the performance of a foot placement prediction

model. For each participant, the dataset with the 15-second interval was selected as the evaluation set. For cross-subject training and evaluation, all the available intervals were used.

## 4.7 Model Training

The deep learning models for gait phase estimation and foot placement prediction are trained successively, as the evaluation of ZUPT intervals depends on the output of the gait phase estimation model. Except for the fine-tuning case, the learnable parameters of the models are initialized from a normal distribution. Also, Adam optimizer and cross validation scheme are used [27, 28]. For pre-training, group-fold cross-validation for 9 datasets from 9 other people is performed. Hyperparameter search is done using the TPE (Tree-structured Parzen Estimator) algorithm from Optuna, which is an open-source Python library for hyperparameter optimization [29]. For each training scenario, at least 100 hyperparameter searches are made. The data-processing and training are carried out using AMD EPYC 7742 CPUs and NVIDIA Tesla A100 GPUs. In Table 4.1, decay factor is the multiplicative factor for the learning rate after every epoch for the base case, and initialization std refers to the standard deviation of the normally distributed initial weights of deep learning models. A single-epoch with a batch size of 1024 and a batch-wise learning rate decay are used for pre-training due to the large size of the dataset.

Table 4.1: Training Parameters

Parameters	Values		
	Base	Pre-training	Fine-tuning
Epoch	5	1	5
Batch	256	1024	256
Cross-validation	5 fold	4 group-fold	5 fold
Learning rate	1e-5~1e-3	1e-5~1e-3	1e-7~1e-5
Decay factor	0.05~0.5	0.9~1	0.5~1
Initialization std	0.01~0.1	0.01~0.1	pre-trained

## 4.8 Foot Placement Prediction Evaluation

In this experiment, we evaluate the trained models for each participant using the following metrics:

$$\text{Mean Squared Error (RMSE)} = \frac{1}{N} \sum_{i=1}^N (\hat{Y}_{i,k} - Y_{i,k})^2, \quad (4.14)$$

$$\text{Mean Distance Error (MDE)} = \frac{1}{N} \sum_{i=1}^N \sqrt{\sum_{j=1}^d (\hat{Y}_{i,j} - Y_{i,j})^2}, \quad (4.15)$$

where  $N$  is the number of evaluation data points,  $k$  is the index of the output, and  $d$  is the output dimension. To summarize the results from 10 participants, the mean and the standard error are calculated across the 10 cases. Additionally, to compare the results, we perform statistical analyses, including paired sample t-tests and F-tests. We also investigate the linear relationships between the performance metrics of different prediction methods using Pearson’s coefficient.

Since the collected data are on a treadmill, some correlation is expected between successive foot placements. To show that the trained models make inference from the previous gait cycles better than simply referring to the previous foot placement, we compare the trained

models to a simple predictor that uses the ground truth last foot placement vector as the next foot placement.

## 4.9 Interpretable AI/SHAP (SHapley Additive exPlanation)

Deep learning models are typically views as black-box models that does not provide insights on how it makes the prediction. This becomes a problem in applications where the explainability is crucial.

SHAP (SHapley Additive exPlanation) is a recently developed game-theory method that is used to explain the output of a black box model[30]. The method is based on Shapley values, a concept used in game theory that analyze the contribution of multiple agents working in coalition. The SHAP values are model-agnostic and provide a local interpretability of a predictive model, meaning that it is able to quantify the level of contribution given sample input-output pairs.

In this study, the trained deep learning models are analyzed using SHAP values to assess the contribution of the input data and analyze the spatio-temporal importance in making the foot placement prediction. In order to analyze the impact of input data on prediction accuracy in a uni-variate context, the Mean Squared Error (MSE) is used as the output metric instead of the 3D foot placement prediction vector.

There are two dimensions along which the SHAP values are analyzed. First, to explain how much each feature is contributing to the prediction accuracy, the calculated absolute SHAP values are added along the temporal dimension. Similarly, to explain the temporal contribution, the calculated absolute SHAP values are added along the spatial dimension.

The average summation of the absolute SHAP values for sample data points are used as the metric for determining the contribution level. To perform the evaluation, a sample of data consisting of 512 points is selected from the test dataset. Additionally, a set of 2048 data points is chosen as background data.

# Chapter 5

## Results

### 5.1 Dataset Metrics

The mean and standard deviation of the number of right foot steps for the intervals of 1 second, 2 second, 5 second, 10 second, 15 second, and 30 second from 10 participants are:  $24.7 \pm 2.5$ ,  $53.8 \pm 3.2$ ,  $146 \pm 9.4$ ,  $301.5 \pm 12.3$ ,  $473.9 \pm 27.0$ , and  $910.3 \pm 40.5$  steps, respectively. The corresponding dataset sizes are:  $7153 \pm 558$ ,  $16928 \pm 168$ ,  $46898 \pm 1169$ ,  $98702 \pm 1641$ ,  $154069 \pm 8197$ , and  $297862 \pm 11984$  points (data is collected at 240 Hz). Overall, the  $n_R$  value where the foot placement occurred was an average of 25.2% in the ground truth gait cycle.

### 5.2 Gait Phase Estimation and Foot Placement Prediction Accuracy

As shown in Fig. 5.2(a) and (d), gait phase detection accuracy as well as foot placement prediction accuracy are consistently improved when the models are trained with larger datasets. For example, the models trained with all the available data have significantly better performance than the foot placement prediction models trained with the 30 s interval data for both the base case ( $P \leq 0.024$ ) and the fine-tuning case ( $P \leq 0.007$ ).

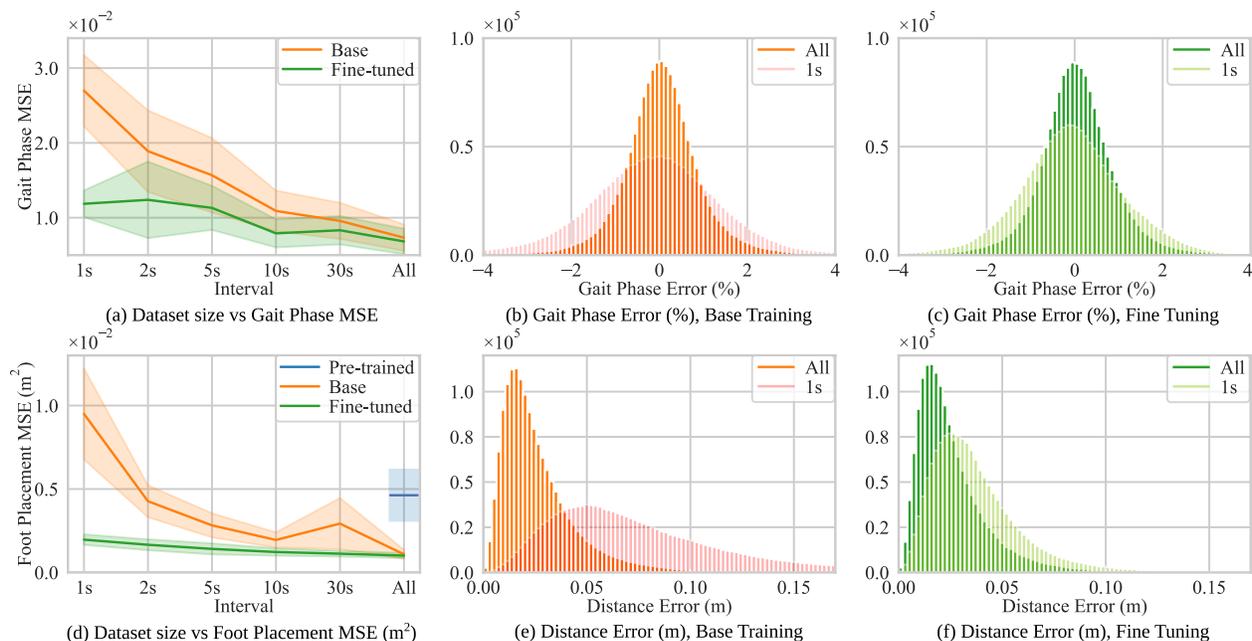


Figure 5.1: Results of foot placement prediction with deep learning models. (a) The right gait phase RMSE for training datasets of different sizes. (b) and (c) Normalized histogram for right foot gait phase error for the base (orange) and fine-tuning (green) case. Note that this error is calculated by reconstructing the gait phase value from the output of the gait phase evaluation model. (d) Foot placement error when trained with datasets of different sizes. (e) and (f) Normalized histogram for foot placement error for the base (orange) and fine-tuning (green) case.

It is also observed that the fine-tuned models consistently outperform the models from the base case across all the dataset sizes (see Fig. 5.2(a) and (d)). The largest improvement is made when the models are trained with the 1 second interval datasets; when trained with the 1 second interval datasets, the base models have 3.0% and 1.69 cm or less for the gait phase estimation error and the foot placement prediction, 95% of the time, respectively, whereas the fine-tuned models have them lower than 2.3% and 7.57 cm 95% of the time (see Fig. 5.2(b), (c), (e), and (f)). When trained with all the available data, the difference becomes smaller; the corresponding values are 1.8% and 5.81 cm for the base models and 1.7% and 5.58 cm for the fine-tuned models. In addition, the fine-tuning process generally reduces cross-subject variance in terms of prediction accuracy, compared to the base models.

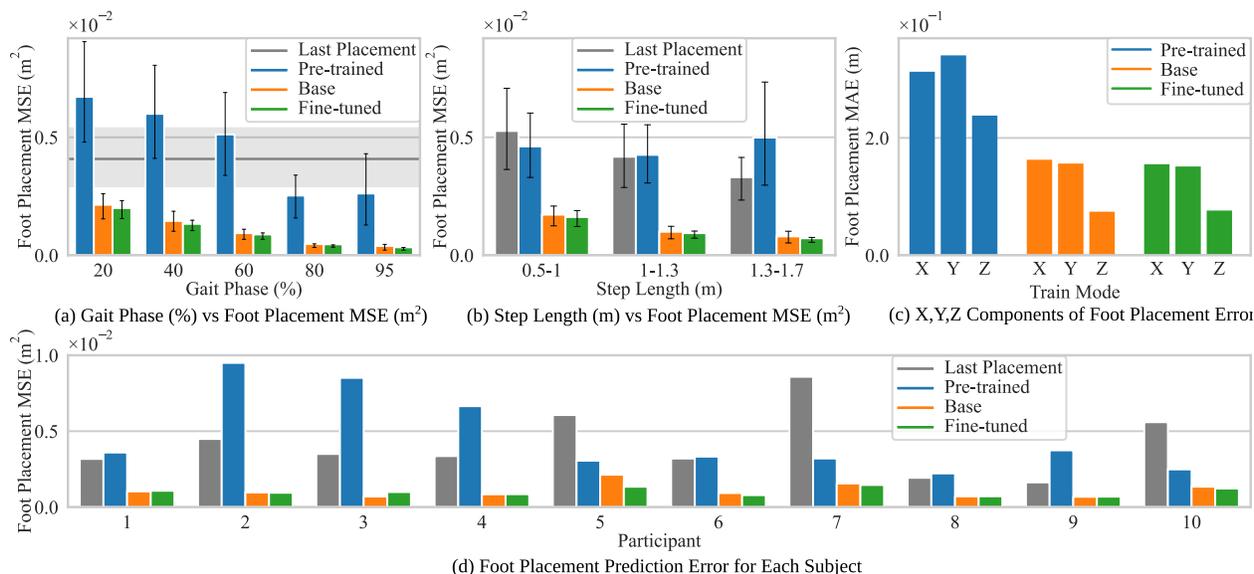


Figure 5.2: (a) Foot placement error with gait progression. Gait phase values denote an interval of gait progression after the last foot placement, for example, 20 means 20~21% gait progression after the last foot placement, 40 means 40~41%, and so on. (b) Foot placement error for different step lengths. The step length refers to the distance between the last foot placement to the succeeding one. Note that the Gait Phase percentage is the percentage after the foot placement; this corresponds to approximately 25% in the traditional gait cycle after heel strike, so 87% on the graph is approximately the next heel strike. (c) X,Y, and Z components of foot placement error. (d) The foot placement prediction results for each participant using the last foot placement predictor, the base models, the pre-trained models, and the fine-tuned models. The shaded areas and the error bars indicate 95% confidence intervals.

Overall, the best foot placement prediction accuracies are achieved by the models trained on all the available data: 2.69 cm MDE for the base case and 2.60 cm MDE for the fine-tuning case. Additionally, for cross-subject evaluation, the pre-trained models have a 5.93 cm MDE (Fig. 5.2g). The overall RMSEs of foot placement prediction in the  $X$ ,  $Y$ , and  $Z$  axis of the pelvis coordinate system are, respectively, 4.04 cm, 4.46 cm, and 3.21 cm for the pre-trained models, 2.25 cm, 2.21 cm, and 0.94 cm for the base models, and 2.15 cm, 2.12 cm, and 0.97 cm for the fine-tuned models, when trained by all the available data. Note that the performance improvement is statistically insignificant for the all datasets case ( $P \geq 0.46$ ), and the cross-subject variance reduction is statistically insignificant for the 5 second interval

case ( $P \geq 0.21$ ).

### 5.3 Early Gait Phase Foot Placement Prediction

Fig. 5.2(g) shows the foot placement prediction error given different ranges for gait progression when the models were trained using all the available data. The base and fine-tuned models achieve competitive performance even at very early in the gait phase. For example, the base models ( $P \leq 0.001$ ) and the fine-tuned models ( $P \leq 0.001$ ) trained by all the available data have the foot placement prediction MDE of 4.60 cm and 4.70 cm with gait progression less than 1% after the previous foot placement, respectively, which are significantly lower than the prediction MSE error of the last foot placement predictor at 5.34 cm. The pre-trained models have the corresponding MDE of 6.88 cm, which is higher than that of the last foot placement predictor. With gait progression less than 56% after the previous foot placement (half way through swing), the foot placement prediction MDEs are 6.99 cm, 3.32 cm, and 3.22 cm from the pre-trained, base, and fine-tuned models, respectively. The intervals that achieve the best accuracy are 81-82%, 84-85%, and 82-83% after the last foot placement for the pre-trained, base, and fine-tuned models, respectively, with the corresponding MDEs of 3.95 cm, 1.69 cm, and 1.59 cm within the specified intervals (see Fig. 5.2(g)). This is after the time of heel strike for the foot.

Throughout the experiment, a positive correlation between the performance of the last foot placement predictor and the trained models is observed: For the base case and the fine-tuning case trained by all the data, the Pearson correlation coefficients are  $r = 0.76$  and  $0.91$ ,  $N = 10$ , respectively, indicating strong linear relationships (see Fig. 5.2(j)). In a similar context, the best foot placement prediction accuracy is achieved when the step length ranges from 1.3 m to 1.7 m for both the base case and the fine-tuning case, where the highest correlation exists

between the previous foot placement and the subsequent foot placement (Fig. 5.2(h)).

## 5.4 Interpretation using SHAP values

As the fine-tuned models have demonstrated the highest level of foot placement accuracy, it is conjectured that the SHAP values derived from these models can provide a more precise understanding of the impact of individual features on foot placement prediction. This assumption is motivated by the fact that the models achieving the best accuracy are more likely to be capturing the most relevant information in the data, and thus the SHAP values computed from these models can be indicative of the most influential features in the prediction process. This section seeks to leverage the insights gained from SHAP values to identify the features that contribute the most to foot placement prediction.

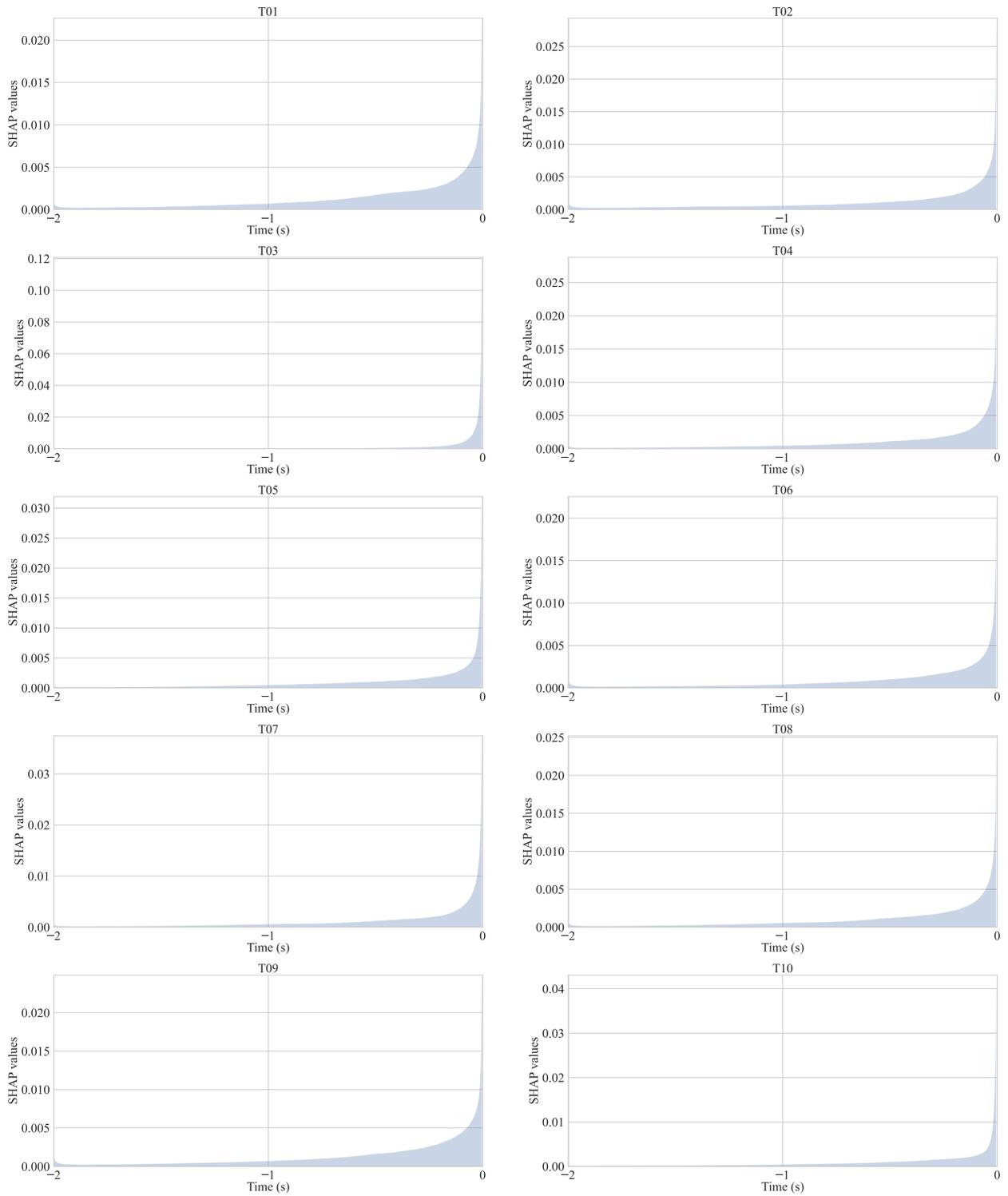


Figure 5.3: The average SHAP values for all 480 timestamps using the fine-tuned models.

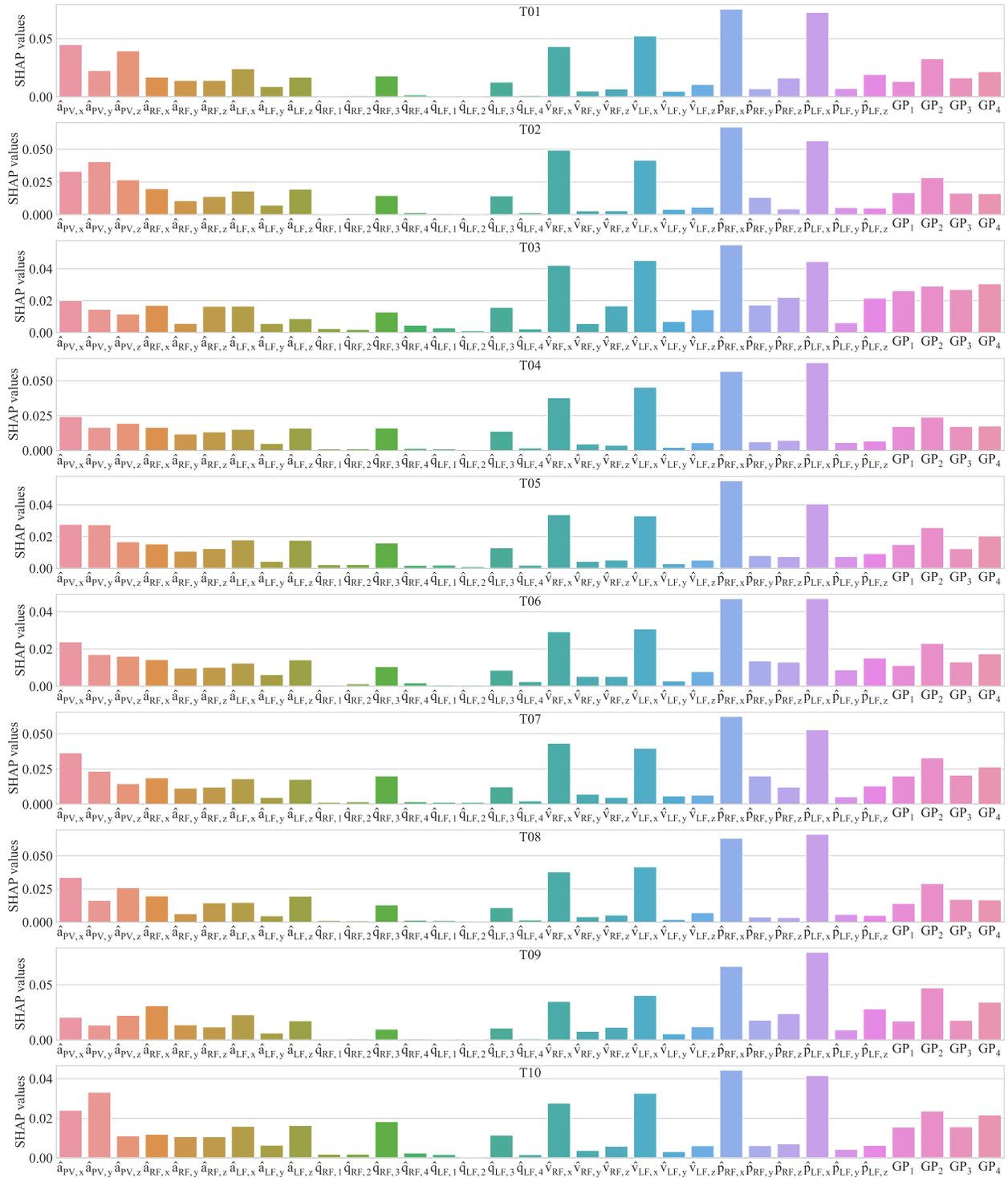


Figure 5.4: The average SHAP values computed for all 33 features using the fine-tuned models.

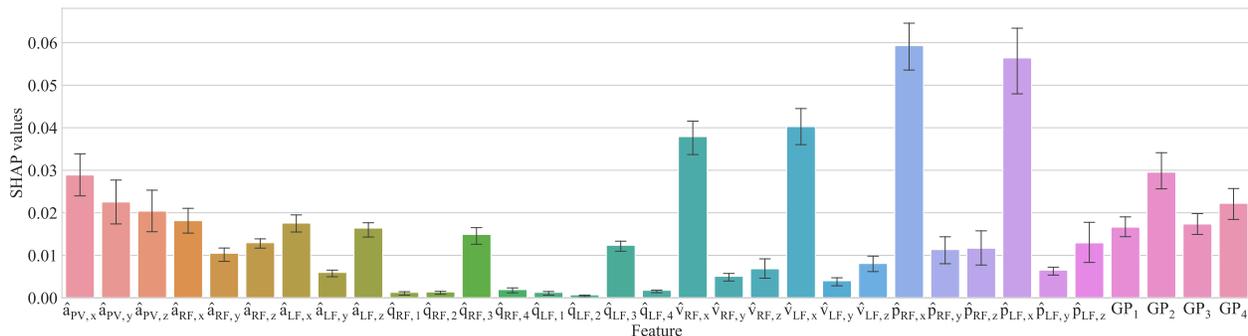


Figure 5.5: The average SHAP values for each feature using the fine-tuned models across 10 participants.

Fig. 5.3 shows the relationship between contribution and time progression. The plot demonstrates that the temporal contribution increases exponentially over time, indicating that the most recent data points carry greater weight in predicting the foot placement. This trend is consistent with the underlying assumption that recent events are more relevant than earlier events in time-series data.

Fig. 5.4 and Fig. 5.5 present the average contribution of each feature. The descending order of the most influential features, based on their average contribution, includes  $\hat{p}_{RF,x}$ ,  $\hat{p}_{LF,x}$ ,  $\hat{v}_{LF,x}$ ,  $\hat{p}_{RF,x}$ ,  $GP_2$ , and  $\hat{a}_{PV,x}$ . It is observed that features that represent the x-component (forward direction) have the highest contribution among the same feature category. Position features in average have the highest impact on the prediction, followed by velocity features. Interestingly, pelvis acceleration features have higher contribution than feet acceleration features.

# Chapter 6

## Discussion

In this study, we present a new approach for predicting foot placement using three IMU sensors and a deep learning algorithm (Fig. 4.1). Our approach employs a time-series processing deep learning model that predicts foot placement in three dimensions. Unlike any of the existing methods, the proposed method relies completely on local sensor information for inference and is thus a mobile solution once implemented. It is also relatively free of manual engineering such as feature extraction or data labeling, and is capable of cross-subject learning and small dataset learning through cross-validated training and fine-tuning. To the best of our knowledge, we are the first to use recurrent neural networks for foot placement prediction, and our approach achieves state-of-the-art foot placement prediction accuracy in comparison to other studies.

Our algorithm produces state-of-the-art results in comparison to prior wearable sensor-based works. In [5], Lee continuously estimated the gait phase during walking and reported an average error of  $1.67 \pm 1.36\%$  and  $1.45 \pm 1.47\%$  for walking speeds of 0.5 m/s and 1.5 m/s, respectively. This means that the corresponding upper bounds for the 95% confidence interval are approximately 4.34% and 4.32%, respectively, which are higher than our best reported upper bound of 1.7%.

For foot placement prediction, Chen [14] found an RMSE of 4.4 cm and 4.2 cm in the forward and sideways directions, respectively, using a motion capture system to make predictions after the first 0.2 s of the swing phase. Also, Chen [13] used a foot-mounted IMU and found

an RMSE of 12.2 cm and 10.6 cm in the forward and sideways directions in single-subject learning and 13.3 cm and 13.7 cm in cross-subject learning, respectively, with a prediction window of the first 50% of the swing phase. Using the same prediction window, Zhang found a best MDE of 8.6 cm in single-subject learning when subjects walked through rough terrain [12]. In both [13] and [12], the earliest a prediction can be made is at least 50% in gait phase percentage after the last foot placement, which corresponds to the mid-swing phase. In comparison, our method achieves the foot placement prediction MDEs of 6.99 cm, 3.32 cm, and 3.22 cm from the pre-trained, base, and fine-tuned models, respectively, with gait progression up to a similar point (around 56% in Fig. 5.2(e)), which are significant improvements from the above studies in accuracy.

Additionally, our models begin making predictions immediately after the previous foot placement (around  $n_R = 25\%$  in the gait cycle), with increasing accuracy over time. This is in contrast to the previous works that only consider the foot motion during the swing phase. Interestingly, early in the gait cycle the pre-trained models are less accurate than an estimate based on the person’s last footstep until the mid-swing phase. However, if the models are trained on target subject data (base-trained and fine-tuned), then they are always more accurate than the previous foot placement, even early in the gait cycle. Also, increasing the amount of training data leads to better accuracy in both gait estimation and foot placement prediction.

The competitive early gait-phase foot placement prediction accuracy indicates that the deep learning models are capable of extracting valuable information from the pelvis and feet kinematics before the swing phase to create a more informed prediction. The deep learning models look back the previous 2 seconds of sensor data, which is sufficient to capture at least a single gait cycle. This approach differs from other studies, such as [13] and [14], which only consider the foot motion during the swing phase. Our deep learning models also

update predictions at the IMU sensor sampling rate, which enhances the interpretability and reliability of the predictions.

Our method also offers several notable other advantages over the existing methods for practical implementation. First, it does not need a reference coordinate system and relies on internally normalized sensor data for inference, making it a viable mobile solution. Additionally, our method is scalable to the dimensionality of foot placement, whereas other distribution-based studies can suffer from exponentially higher computational loads when an additional dimension is introduced.

One interesting finding is that the foot placement prediction accuracy increases with longer steps, which is contrary to the suggestion made by many step length evaluation studies that scale the distance error by the true step length [4, 31]. A possible explanation is that the walking speed that the participants found natural is close to the maximum allowed speed during the experiment: when asked to walk in low speeds, the participants had to make conscious adjustments to slow down, potentially introducing additional patterns that the trained models had a harder time generalizing. This could also be a reason why there was higher correlation between neighboring foot placements for step length range of 1.3~1.7 m than that of 0.5~1.0 m.

Also, in many studies, gender differences in walking are reported [32, 33]. However, in terms of cross-subject foot placement prediction, no gender-specific discrepancy is found between the female participants (7 and 8) and the male participants in this study. Female participants showed lower prediction error than the average during cross-subject learning, although the models were trained on a dataset consisting mainly of walking data from male participants. Thus, it appears it is more beneficial to build a pre-training dataset with diverse walking patterns from the general population rather than to limit it to specific target attributes.

Based on the computed SHAP values, several observations are made. Firstly, among the acceleration features, the pelvis features are the most significant ones. This aligns with the bipedal walking models, including LIPM (Linear Inverted Pendulum Model), which assert that the kinematics of the center of mass (COM), which can be approximated as the pelvis, is the primary factor for foot placements. Secondly, most of the orientation features are found to be least relevant for foot placement prediction, except for the third quaternion components for both feet. Lastly, the gait phase estimation model’s output plays a meaningful role in predicting foot placements.

In future work, we will extend our method to uncontrolled walking settings and implement online learning methods. Since the current prediction models were trained and tested on a computer with multiple GPUs, we will improve this system by making it a fully wearable solution. From the deep learning perspective, future work includes using Bayesian deep learning networks to capture a distribution of potential foot placements, better explaining model with interpretable AI techniques, and making the learning completely end-to-end with better feature extraction ability and a shorter latency period.

# Chapter 7

## Conclusions

In this study, we present a new approach for foot placement prediction using three IMU sensors and deep learning models with state-of-the-art results for gait phase estimation and for foot placement prediction during human walking. In cross-subject learning, the trained deep learning models had the mean distance error as low as 5.93 cm. In single-subject learning, when the models were initialized without pre-training prior to learning, a mean distance error as low as 2.69 cm was achieved. When using the cross-subject trained models as the initial models instead, the prediction accuracy of the trained models were improved especially with smaller training datasets, and cross-subject variances in performance were reduced, with the best mean distance error of 2.60 cm.

The experimental results demonstrate that deep learning models are able to capture temporal dependencies by sequentially processing time-series sensor data to make informed predictions even early in the gait cycle. The prediction accuracy increased by training on the target individual, and by increased data set sizes. The proposed foot placement prediction method can help lower-body wearable robots to provide better safety to users by anticipating the next foot placement and performing risk-aversion techniques such as fall prevention or obstacle avoidance, or by improving control systems to be more in tune with the body. This study is also expected to open up more deep learning-based works for practical human motion anticipation.

# Bibliography

- [1] Yueng Santiago Delahoz and Miguel Angel Labrador. Survey on fall detection and fall prevention using wearable and external sensors. *Sensors*, 14(10):19806–19842, 2014.
- [2] Nisal Menuka Gamage, Deepana Ishtaweera, Martin Weigel, and Anusha Withana. So predictable! Continuous 3D hand trajectory prediction in virtual reality. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pages 332–343, 2021.
- [3] Juan Alejandro Castano, Zhibin Li, Chengxu Zhou, Nikos Tsagarakis, and Darwin Caldwell. Dynamic and reactive walking for humanoid robots based on foot placement control. *International Journal of Humanoid Robotics*, 13(02):1550041, 2016.
- [4] Alper Köse, Andrea Cereatti, and Ugo Della Croce. Bilateral step length estimation using a single inertial measurement unit attached to the pelvis. *Journal of neuroengineering and rehabilitation*, 9:1–10, 2012.
- [5] Jinwon Lee, Woolim Hong, and Pilwon Hur. Continuous gait phase estimation using LSTM for robotic transfemoral prosthesis across walking speeds. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:1470–1477, 2021.
- [6] Andrey Rudenko, Luigi Palmieri, Achim J Lilienthal, and Kai O Arras. Human motion prediction under social grouping constraints. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3358–3364. IEEE, 2018.
- [7] Kouros Darvish, Serena Ivaldi, and Daniele Pucci. Simultaneous action recognition and human whole-body motion and dynamics prediction from wearable sensors. In *2022*

- IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*, pages 488–495. IEEE, 2022.
- [8] Zachary Choffin, Nathan Jeong, Michael Callihan, Edward Sazonov, and Seongcheol Jeong. Lower body joint angle prediction using machine learning and applied biomechanical inverse dynamics. *Sensors*, 23(1):228, 2022.
- [9] Lei Zhang and Chenglong Fu. Predicting foot placement for balance through a simple model with swing leg dynamics. *Journal of Biomechanics*, 77:155–162, 2018.
- [10] Zohaib Aftab, Thomas Robert, and Pierre-Brice Wieber. Predicting multiple step placements for human balance recovery tasks. *Journal of biomechanics*, 45(16):2804–2809, 2012.
- [11] Yang Wang and Manoj Srinivasan. Stepping in the direction of the fall: the next foot placement can be predicted from current upper body state in steady-state walking. *Biology letters*, 10(9):20140405, 2014.
- [12] Kuangen Zhang, Haiyuan Liu, Zixuan Fan, Xinxing Chen, Yuquan Leng, Clarence W de Silva, and Chenglong Fu. Foot placement prediction for assistive walking by fusing sequential 3d gaze and environmental context. *IEEE Robotics and Automation Letters*, 6(2):2509–2516, 2021.
- [13] Xinxing Chen, Kuangen Zhang, Haiyuan Liu, Yuquan Leng, and Chenglong Fu. A probability distribution model-based approach for foot placement prediction in the early swing phase with a wearable IMU sensor. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:2595–2604, 2021.
- [14] Xinxing Chen, Zijian Liu, Jiale Zhu, Kuangen Zhang, Yuquan Leng, and Chenglong Fu. Comparison of machine learning regression algorithms for foot placement prediction. In

- 2021 27th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, pages 169–174. IEEE, 2021.
- [15] F Landis Markley, Yang Cheng, John L Crassidis, and Yaakov Oshman. Averaging quaternions. *Journal of Guidance, Control, and Dynamics*, 30(4):1193–1197, 2007.
- [16] Mustafa Sarshar, Sasanka Polturi, and Lutz Schega. Gait phase estimation by using LSTM in IMU-based gait analysis—proof of concept. *Sensors*, 21(17):5749, 2021.
- [17] Changhui Jiang, Shuai Chen, Yuwei Chen, Boya Zhang, Ziyi Feng, Hui Zhou, and Yuming Bo. A MEMS IMU de-noising method using long short term memory recurrent neural networks (LSTM-RNN). *Sensors*, 18(10):3470, 2018.
- [18] Ming Zhang, Mingming Zhang, Yiming Chen, and Mingyang Li. IMU data processing for inertial aided navigation: A recurrent neural network based approach. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3992–3998. IEEE, 2021.
- [19] Keehong Seo, Young Jin Park, Jusuk Lee, Seungyong Hyung, Minhyung Lee, Jeonghun Kim, Hyundo Choi, and Youngbo Shim. Rnn-based on-line continuous gait phase estimation from shank-mounted IMUs to control ankle exoskeletons. In *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*, pages 809–815. IEEE, 2019.
- [20] Inseung Kang, Pratik Kunapuli, and Aaron J Young. Real-time neural network-based gait phase estimation using a robotic hip exoskeleton. *IEEE Transactions on Medical Robotics and Bionics*, 2(1):28–37, 2019.
- [21] Zhelong Wang, Hongyu Zhao, Sen Qiu, and Qin Gao. Stance-phase detection for

- ZUPT-aided foot-mounted pedestrian navigation system. *IEEE/ASME Transactions On Mechatronics*, 20(6):3170–3181, 2015.
- [22] Yusheng Wang and Andrei M Shkel. Adaptive threshold for zero-velocity detector in ZUPT-aided pedestrian inertial navigation. *IEEE Sensors Letters*, 3(11):1–4, 2019.
- [23] Tianyi Zhao and Mohammed Jalal Ahamed. Pseudo-zero velocity re-detection double threshold zero-velocity update (ZUPT) for inertial sensor-based pedestrian navigation. *IEEE Sensors Journal*, 21(12):13772–13785, 2021.
- [24] Hangbo Bao, Li Dong, Songhao Piao, and Furu Wei. Beit: Bert pre-training of image transformers. *arXiv preprint arXiv:2106.08254*, 2021.
- [25] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [26] Alexei Baevski, Michael Auli, and Abdelrahman Mohamed. Effectiveness of self-supervised pre-training for speech recognition. *arXiv preprint arXiv:1911.03912*, 2019.
- [27] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [28] Christoph Bergmeir and José M Benítez. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213, 2012.
- [29] James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*, 24, 2011.
- [30] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

- [31] Luis Enrique Díez, Alfonso Bahillo, Jon Otegui, and Timothy Otim. Step length estimation methods based on inertial sensors: A review. *IEEE Sensors Journal*, 18(17):6908–6926, 2018.
- [32] Tessa M Pollard and Janelle M Wagnild. Gender differences in walking (for leisure, transport and in total) across adult life: a systematic review. *BMC public health*, 17(1):1–11, 2017.
- [33] D Casey Kerrigan, Mary K Todd, and Ugo Delia Croce. Gender differences in joint biomechanics during walking normative study in young adults, 1998.
- [34] Emre Aksan, Manuel Kaufmann, Peng Cao, and Otmar Hilliges. A spatio-temporal transformer for 3D human motion prediction. In *2021 International Conference on 3D Vision (3DV)*, pages 565–574. IEEE, 2021.
- [35] Cheolhwan Oh, Seungmin Han, and Jongpil Jeong. Time-series data augmentation based on interpolation. *Procedia Computer Science*, 175:64–71, 2020.
- [36] Satya Narayan Shukla and Benjamin M Marlin. Interpolation-prediction networks for irregularly sampled time series. *arXiv preprint arXiv:1909.07782*, 2019.
- [37] Vadim Romanuke. Time series smoothing improving forecasting. *Applied Computer Systems*, 26(1):60–70, 2021.

# Appendices

# Appendix A

## Discussion of Methods and Findings Not Included in the Main Body

The main body of this thesis presents the results and conclusions of our research. However, as with any research project, there were many aspects that were explored and attempted that did not make it into the final analysis. In this bonus chapter, we aim to provide a comprehensive discussion of the methods and findings that were not included in the main paper. This chapter covers the approaches that were explored but ultimately found to be unfeasible or not as effective as initially hypothesized. Additionally, we discuss other relevant findings that did not make it into the journal paper, including preliminary results, but may be of interest to readers seeking a more comprehensive understanding of the research process.

### A.1 Deep Learning Architecture

Given the same input dataset, many different deep learning architectures were compared in terms of the foot placement prediction accuracy. With some preliminary testing, following conclusions were made: First, transformer models with spatio-temporal self-attention mechanisms, similar to [34], performed noticeably worse than GRU models with attention mechanisms. Additionally, models with attention mechanism, including the GRU models, not only suffered from longer training and inference time, but also did not provide mean-

ingful advantages compared to the vanilla GRU model used in the method section in terms of performance. The structure of the GRU models with and without attention mechanism is illustrated in Fig. A.1. Because the vanilla GRU model without attention led to faster training and inference time and the best prediction accuracy, it was selected as the optimal model. The possible explanation for why the vanilla GRU model worked better than the other models is, as shown in Fig 5.3, that GRU model without attention inherently places higher importance on the more recent data points, which aligns with the task of predicting the foot placement, where the more recent sensor data are expected to have higher impact on the predictions made.

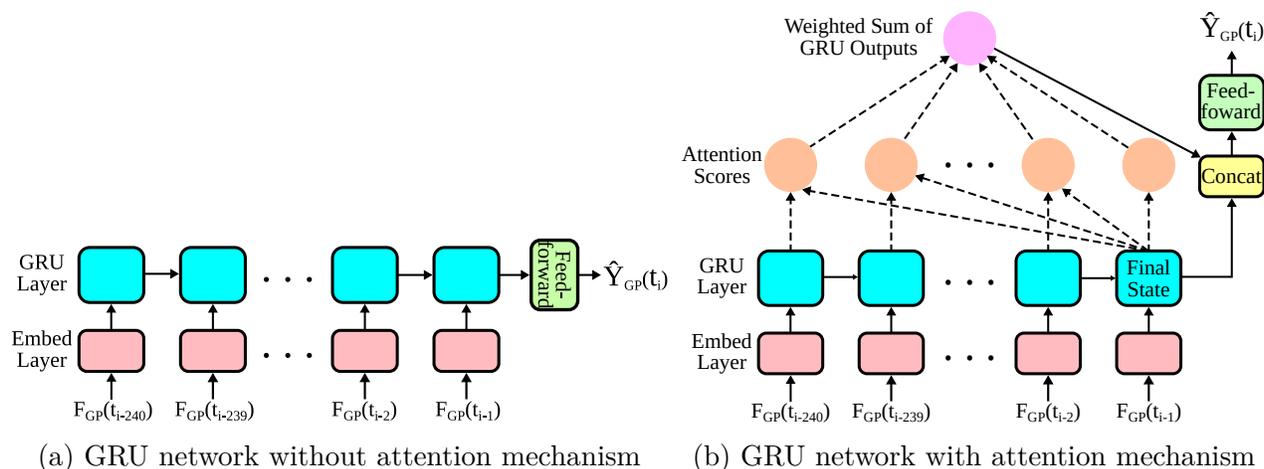


Figure A.1: Different deep learning architectures

## A.2 Temporal Rescaling

It is commonly observed that many deep learning models designed for time series data employ a fixed look-back window. This is because some models, such as Transformers, cannot process variable length inputs without zero-padding as their architecture requires a fixed input size. Other models like RNNs, LSTMs, and GRUs are capable of handling variable-length inputs

but often prefer fixed-length inputs for efficient batch learning.

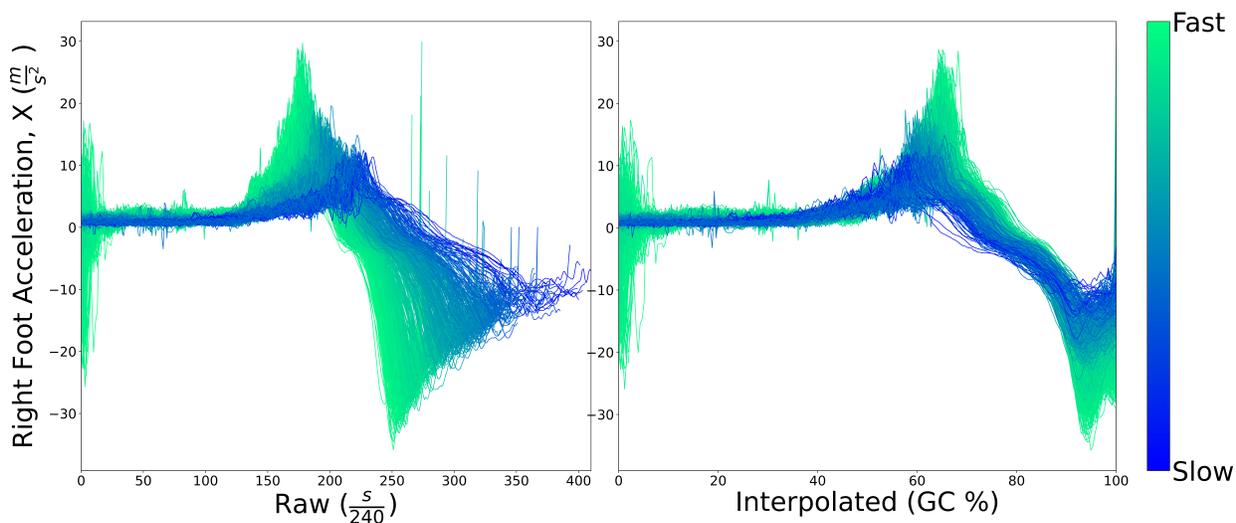


Figure A.2: Raw time series data vs temporally re-scaled series data for each gait. Green color indicates fast walks, and blue color indicates slow walks.

The time window within which the sensor readings carry relevant information for the next foot placement prediction is affected by the current gait cycle progression as well as the gait frequency. This observation was used to hypothesize that it leads to a high variability of the optimal look-back window between different input data. Also, as suggested by [35, 36, 37], the time series data interpolated to a certain temporal scale can lead to for better deep learning performances. The time-series input data for foot placement prediction were re-scaled according to the gait progression instead of time. As a preceding step, the gait progression was estimated at each time instance for proper rescaling.

However, preliminary results showed that the temporal rescaling does not lead to better prediction accuracy. It also introduces an additional hyperparameter, the re-scaling ratio, and is affected by the current gait progression estimation. The selected deep learning models were able to capture information from the input data using data without re-scaling regardless of the rate of the gait progression.

### A.3 Additional Features

Additional features were generated in an attempt to achieve better accuracy, but omitted in the final work because there was no observed improvement. The additional features are:

*Time*: the time progression data from the last right foot flat instance in time-series were used as follows:

$$\text{Time progression features} = \left\{ t - t_{GP=n_R\%}, t - t_{GP=n_L\%} \right\}$$

*LIPM-based COM position*: According to the LIPM model, the center of mass position of the system is:

$$P_{COM}(t) = -x_0 \cosh\left(\frac{t}{T_c}\right) + T_c v_0 \sinh\left(\frac{t}{T_c}\right), \quad (\text{A.1})$$

where  $T_c$  is the time constant computed as  $\sqrt{\frac{l}{g}}$ ,  $l$  is the length of the inverted pendulum,  $g$  is the gravity constant, and  $v_0$  is the initial COM velocity. The second term of the equation is used as the input feature, where  $T_c$  is replaced by the estimated step completion duration  $\frac{t}{2\pi Y_{GP,R}(t)}$ , and  $v_0$  is estimated by averaging the previous swinging foot velocity until the heel strike. Note that in the SLIP model, the z-position of the COM is assumed to be constant, so only the x and y component are used. We define the x and y component of the estimated COM position as  $SLIP_x$  and  $SLIP_y$ , respectively. Mathematically,

$$\begin{aligned} SLIP_x &= \frac{t}{2\pi Y_{GP,swf}(t)} \hat{v}_{0,x} \sinh(2\pi Y_{GP,swf}(t)) \\ SLIP_y &= \frac{t}{2\pi Y_{GP,swf}(t)} \hat{v}_{0,y} \sinh(2\pi Y_{GP,swf}(t)) \\ \hat{v}_0 &\propto \int_{sw,i}^{sw,e} v_{swf,prev}(t) dt \end{aligned} \quad (\text{A.2})$$

The estimated initial COM velocity  $\hat{v}_0$  is replaced by the integral of the estimated previous

swinging foot velocity during the swinging phase. Note that the swinging foot (swf) switches every half of the gait cycle: From  $GP = 0\%$  to  $GP = 50\%$ , the left foot is the swinging foot, and from  $GP = 50\%$  to  $GP = 100\%$ , the right foot is the swinging foot, and so on.

The inclusion of both time progression and SLIP features in the model did not result in any improvement, according to the preliminary results. The reasons behind this are twofold: Firstly, the time progression feature may not have provided any additional information that was not already provided by the output of the gait estimation model (GP in Fig. 4.1). Secondly, the LIPM model, being a passive model with pre-determined walk trajectory based on initial conditions such as the initial COM velocity, differs from actual human walking, thus rendering the SLIP features less useful in improving prediction accuracy.

## A.4 SHAP-based Feature Selection

To more accurately predict foot placements and to enhance the autonomy of the learning process, a SHAP-based feature selection method was proposed. Recursive Feature Elimination (RFE) is a feature selection methodology which iteratively eliminates the least contributing feature from the current feature subset until a certain condition is met. In the proposed SHAP-based feature selection method, at each iteration, SHAP values are assessed to calculate the feature contribution and eliminate the least contributing feature to. When there is no improvement in the performance after an iteration, the process stops and the final feature subset is selected.

Although preliminary results were positive, the practical application of this method was hindered by the training time required for GRU models with large datasets at each iteration, and the need to customize models for the selected features in each scenario (e.g., different dataset size and participant). Therefore, it was deemed practically impossible to rigorously

apply this method.

## A.5 Multiple Foot Placements Prediction

An additional aspect of the study was to test if the model is able to make predictions for the multiple future foot placements. Although this is a valuable capability for long-term motion predictions, it was not included in the paper due to the controlled nature of the data collection process, where the speed and direction of the walks were managed. Consequently, the prediction accuracy achieved in this study could differ significantly from the case where the speed and direction of the walk are not controlled, and therefore, the results were not reported in the paper. The future works include collecting uncontrolled overground walking data and use the deep learning models to perform multiple foot placement predictions.

## A.6 Different Sensor Placements

In our study, each participant did not try multiple sensor placements during data collection. This means that, during the single-subject learning, the training dataset and the test dataset had the same sensor orientation for each participant, assuming no slippage of the sensors. In contrast, during the cross-subject learning, each subjects had different sensor placements. It is worthwhile to mention that if each participant had to don and doff the system after each session to test different sensor placements, the single-subject models would have yielded results comparable to that of the pre-trained models, which learned across different subjects and sensor placements.