Estimation of Lower Extremity Joint Moments and 3D Ground Reaction Forces Using IMU Sensors in Multiple Walking Conditions: A Deep Learning Approach

SUBMISSION DATE / POSTED DATE

13-08-2022 / 16-08-2022

CITATION

Hossain, Md Sanzid Bin; Guo, Zhishan; Choi, Hwan (2022): Estimation of Lower Extremity Joint Moments and 3D Ground Reaction Forces Using IMU Sensors in Multiple Walking Conditions: A Deep Learning Approach. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.20484201.v1

DOI

10.36227/techrxiv.20484201.v1

Estimation of Lower Extremity Joint Moments and 3D Ground Reaction Forces Using IMU Sensors in Multiple Walking Conditions: A Deep Learning Approach

Md Sanzid Bin Hossain, *Student Member, IEEE*, Zhishan Guo *Senior Member, IEEE*, and Hwan Choi, *Member, IEEE*

***Abstract*— Human kinetics, specifically joint moments and ground reaction forces (GRFs) can provide important clinical information and can be used to control assistive devices. Traditionally, collection of kinetics is mostly limited to the lab environment because it relies on data that measured from a motion capture system and floor-embedded force plates to calculate the dynamics via musculoskeletal models. This spatially limited method makes it extremely challenging to measure kinetics outside the laboratory in a variety of walking conditions due to the expensive device setup and large space required. Recently, employing machine learning with IMU sensors are suggested as an alternative method for biomechanical analyses. Although these methods enable estimating human kinetic data outside the laboratory by linking IMU sensor data with kinetics dataset, they were limited to show inaccurate kinetic estimates even in highly repeatable single walking conditions due to the employment of generic deep learning algorithms. Thus, this paper proposes a novel deep learning model, Kinetics-FM-DLR-Ensemble-Net to estimate the hip, knee, and ankle joint moments in the sagittal plane and 3 dimensional ground reaction forces (GRFs) using three IMU sensors on the thigh, shank, and foot under several representative walking conditions in daily living, such as treadmill, level-ground, stair, and ramp with different walking speeds. This is the first study that implements both joint moments and GRFs in multiple walking conditions using IMU sensors via deep learning. Our deep learning model is versatile and accurate for identifying human kinetics across diverse subjects and walking conditions and our model outperforms state-of-the-art deep learning model for kinetics estimation by a large margin.**

***Index Terms*— Kinetics Estimation; Wearable IMU Sensor; Joint Moment; Ground Reaction Force; Deep Learning; Machine Learning**

I. Introduction

Estimation of human kinetics, specifically ground reaction forces (GRFs) and joint moments, play an important role in providing insights and fundamental information in clinical decisions and controls of exoskeleton devices. For example, joint moments can be used to understand the impact of different joint arthritis in walking [1]–[5] and to provide assistive torque to the powered exoskeleton to minimize muscle efforts [6]. GRFs have also been used for the evaluation of pathological gait patterns [7], [8] or for analyzing the gait of amputees [9]–[12].

Traditionally, joint moments are calculated using infrared based motion capture cameras and ground reaction forceplates. Collected joint kinematics from motion capture cameras and GRFs from the floor embedded forceplates are then implemented to computational musculoskeletal modeling softwares such as OpenSim [13], Visual3D (C-Motion, MD), Nexus (Vicon, UK), or Anybody (Anybody Technology, Denmark) to calculate joint moment. Although using experimental data and musculoskeletal modeling software can provide reliable joint moment data, this method requires extensive manual post data processing of motion and GRF data to employ musculoskeletal modeling softwares, hindering prompt evaluations. In addition to this, there are major technical hurdles when collecting human kinetics in different walking conditions outside the laboratory such as walking on ramps and stairs due to the specific bulky and heavy equipment setups along with expertise. Thus, this method poses constraints on estimation outside the lab, especially in different walking environments that are mostly encountered in daily living.

To overcome these limitations imposed by traditional kinetics estimation methods, there is a trend in adapting wearable sensors with computational human dyanmic models [14]–[24], neuromusculoskeletal models [23], [25], or wearable forceplates [19]–[22]. However, wearable sensors along with human dynamic models or neuromusculoskeletal model based kinetics estimation requires a large number of sensors (e.g., 7 IMUs in [15], 17 IMUs in [17], and 15 IMUs in [18]). Wearable electromyography (EMG) sensors on specific muscle groups can also be used with neuromusculoskeletal modeling to calculate joint moment [23], [25]. However, EMG signals are sensitive to skin impedance and the location of muscle belly, making it challenging to acquire consistent and repeatable signals. Also, EMG signals are prone to additional noise due to the motion artifacts and electrical fields in environment. Thus, neuromusculoskeletal model based estimation of kinetics with wearable sensors and EMG sensors are subjected to multiple limitations including a large number of sensors, subject specific anthropometric information, and variability of EMG signals. Moreover, all of these studies are limited to the level-ground condition, which imposes the applicability to different walking conditions such as stairs and ramps that are commonly found in daily living. There is also widespread research to get GRFs data with portable forceplates. GRFs can be calculated by using a strain gauge transducer [26]–[28], piezoelectric sensors [29]–[31], and fiber-optic force sensors [32], [33]. However, these sensors are vulnerable to hysteresis, sensitive to temperature, and constrained by force ranges and deformation of the sensors' materials [19].

Md Sanzid Bin Hossain and Zhishan Guo are with the Electrical and Computer Engineering Department, University of Central Florida, Orlando, FL 32816, USA, (e-mail: sanzid@knights.ucf.edu and zsguo@ucf.edu)

Hwan Choi is with the Mechanical and Aerospace Engineering Department, University of Central Florida, Orlando, FL 32816, USA, (e-mail: hwan.choi@ucf.edu)

Wearable GRF plates are also heavy and stiff, which hinders the wearer's control of natural dynamic tasks. Moreover, these shoe-embedded force plates needs to be custom fabricated to fit into a specific subject's foot size and shape.

Recently, researchers are focusing on acquiring kinetics parameters through data-driven method [34]–[43], [43]–[46]. However, these studies are still relying on a large number of sensors [35], reversely acquired simulated IMU data from the retrospective motion capture data [34], [37], and highly repetitive treadmill/level-ground walking [36]–[43], [43]–[46]. Since experimentally collected IMU data contains noise induced from multi-frequncy vibration during the impact of the limb on the ground, the simulated IMU data that was acquired from retrospectively captured motion capture data may not provide a reliable kinetics estimation during real application. All these limitations are further aggravated by utilizing conventional deep learning models such as Temporal Convolutional Networks (TCN) [47], Feedforward Neural Network (FNN), and Long Short Term Memory (LSTM) [48] for kinetics estimation.

To address the limitations imposed by musculoskeletal or biomechanical modeling, shoe-embedded force plates, simulated IMU data, multi-modal sensor systems, a large number of wearable sensors, conventional deep learning model, proper validation method (leave-subject-out), and estimation of the simple repetitive level-ground / treadmill motion, we aim to predict the joint moment of the hip, knee, and ankle in the sagittal plane and 3D GRFs (anterior-posterior, vertical, and medio-lateral) using 3 IMU sensors on the foot, shank, and thigh using a novel deep learning model Kinetics-FM-DLR-Ensemble-Net in multiple walking environments and speeds. To the best of our knowledge, this is the first study that implements both fundamental kinetics parameters– 3D GRFs and lower extremity joint moments estimation in multiple walking condition using IMU sensors via deep learning. Our contribution is six-folds: (i) proposing an end-to-end trained model Kinetics-Net leveraging different deep learning layers to increase human kinetics prediction performance; (ii) presenting a FM to integrate output from three primary models in Kinetics-Net, creating Kinetics-FM-Net to further improve human kinetics prediction performance; (iii) introducing a novel technique to utilize two loss functions, which outperforms conventional loss design in deep learning models; (iv) further utilizing an existing ensemble technique, bagging to improve human kinetics prediction accuracy; (v) conducting extensive evaluation with ablation studies to show the effectiveness of our deep learning model; (vi) conducting experimental comparison with the state-of-the art deep learning model for human kinetics estimation, and our proposed method outperforms these models by a large margin.

The rest of this paper is organized as follows: Section II discusses the related work for IMU based kinetics estimation using musculoskeletal modeling and data driven methods. The problem statement and the detailed structure of Kinetics-FM-DLR-Ensemble-Net are discussed in Section III. Section IV describes the protocol of the dataset, dataset pre-processing, validation method, and implementation details of the deep learning model. Section V demonstrates the results. In Section VI, implications of these results, limitations and future works are discussed. We conclude our paper in Section VI.

II. Related Work

In this section, we will discuss related work to estimate kinetics using IMU sensors with musculoskeletal modeling or machine learning methods. First, we will discuss IMU and model based kinetics estimation methods. Later, we will discuss data driven methods for kinetics estimation and their limitations.

Yang et al. [15] used seven IMU sensors to estimate GRFs and moments during walking with a three-dimensional analytical model. Karatsidis et al. [17] also estimated GRFs during walking using kinematic data from 17 IMU sensors with a biomechanical model. They have also predicted joint moments along with GRFs with 17 IMU sensors with a musculoskeletal model based inverse dynamics method for three level-ground walking speeds. Aurbach et al. [18] used a musculoskeletal model to compute GRFs using the kinematics data from 15 IMU sensors during level-ground gait. All these methods are using a large number of wearable sensors for the estimation of kinetics.

Due to the limitation imposed by musculoskeletal model based kinetics estimation or shoe-embedded forceplate, researchers are focusing on estimating GRFs [43], [45], [46], [49] and joint moments [34]–[41], [43], [44] using data-driven method with IMU sensors. Dorschky et al. [44] used 4 IMU sensors to estimate hip, knee, and ankle joint moments, anterior-posterior GRF, and vertical GRF in the level-ground and treadmill walking and running utilizing a 2D Convolutional Neural Network (CNN) based deep learning model with the data augmentation techniques with musculoskeletal model simulation. Their approach is limited to the highly repeated level-ground or treadmill walking condition only. Leporace et al. [45] used a single accelerometer on the shank to predict the 3D GRFs during walking using a Multilayer Perceptron (MLP). However, the model was trained on limited samples (only 4 gait cycles per participant), and data was collected in level-ground walking condition only. Guo et al. [46] used only acceleration data from a single waist-mounted IMU sensor to predict the vertical GRF in self-selected walking speeds in outdoor level-ground settings using Orthogonal Forward Regression (OFR) algorithm . All these studies are implemented in simple repetitive motion such as level-ground or treadmill walking and the use of conventional machine learning models.

To utilize dataset without the presence of experimentally collected IMU sensors with motion capture and GRF data, Mundt et al. [37] placed virtual IMU sensors on the retrospectively collected walking model to predict hip, knee, and ankle joint moments using feedforward and LSTM based model. Molinaro et al. [34] also used a musculoskeleletal model to generate virtual IMU sensor signals on the trunk and thigh to estimate hip joint moment using TCN [47]. As these studies are implemented using virtual IMU data, it is unclear how reliable the outcome of their methods on real IMU application with noise introduced from skin surface movements.

To verify which deep learning model has better estimates on joint moments, there was a comparison between cur-

rent deep learning methods, LSTM, MLP, and pre-trained convolutional neural network [40]. This study showed MLP has better performance in joint moment prediction compared with other models and LSTM would be considered for real-time estimation. More recently, Camargo et al. [35] estimated hip, knee, and ankle joint moments in multiple locomotion modes (treadmill, stair, ramp) using extracted features from the cluster of electrogoniometer, EMG, and IMU sensors' data, performed feature selection, and then used the selected feature as input into the Artificial Neural Network (ANN) and XGBoost. However, the total number of 18 sensors and the training set that established from a single subject make doubt on practicality in real-world deployment and reproducibilty with unseen human subjects. In addition to this, they used handcrafted feature engineering, which adds complexity into their proposed method. To reduce the sensor count, Lim et al. [36] proposed a single sacrum mounted IMU to predict joint moments and GRFs with extracted features (acceleration, velocity, displacement, and time) as the input to ANN, but they are still limited to using conventional deep learning models and predicting treadmill walking condition.

Most of these studies are implemented for simple repetitive motion in level-ground and/or treadmill, simulated IMU data, and multi-modal sensors system, and they are limited to a specific kinetics component, either joint moment or GRFs. These limitations are compounded further with the use of conventional deep learning models such as LSTM, CNN, TCN, and ANN, which limit more accurate multi-variable human kinetics estimations. To address the limitations imposed by these discussed works, we propose a novel deep learning method to estimate kinetics during gait using 3 IMU sensors in multiple walking conditions.

III. Proposed Approach

A. Problem Statement

This paper estimates three dimensional GRFs and sagittal plane hip, knee, and ankle joint moment using 3 IMU sensors on thigh, shank, and foot via a novel deep learning model Kinetics-FM-DLR-Ensmeble-Net. If we have an IMU data of $\mathbb{I}_{\Delta T}$=[I_1,I_2,.....,$I_{\Delta T}$] $\in \mathbb{R}^{\Delta T \times D_{IMU} \times N}$ for a specific window length of ΔT, then the prediction from our model is $\mathbb{K}_{\Delta T}$=[K_1,K_2,.....,$K_{\Delta T}$] $\in \mathbb{R}^{\Delta T \times D_K}$. Here, ΔT represents the window length of data that will be input to the model, N is the number of IMU sensors, D_{IMU} is the dimension of IMU sensors, and D_k is the dimension of total kinetics parameters (anterior-posterior GRF, vertical GRF, medial-lateral GRF, sagittal plane hip, knee, and ankle joint moment). In this problem, we use D_{IMU}=6, $N = 3$, ΔT=100, D_k=6. Mathematically, $\mathbb{I}_{\Delta T} \rightarrow \mathbb{K}_{\Delta T}$

B. Kinetics-FM-DLR-Ensemble-Net

Kinetics-FM-DLR-Ensemble-Net is built mainly with the model Kinetics-FM-DLR-Net (Fig. 2). We implement bagging [50] techniques using Kinetics-FM-DLR-Net to create Kinetics-FM-DLR-Ensemble-Net, which is our final proposed model. Kinetics-FM-DLR-Net mainly consists of two Kinetics-FM-Nets (Fig. 2), where each model will be trained

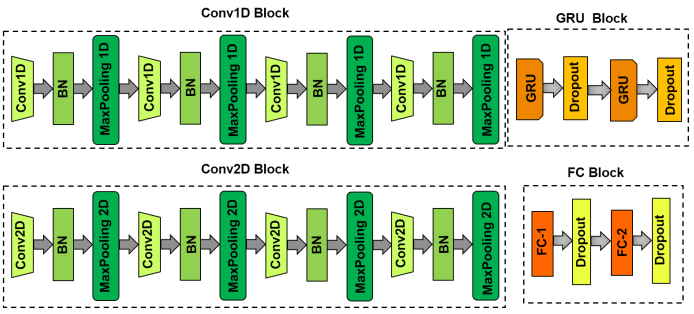

Fig. 1. Fundamental Blocks of Kinetics-FM-DLR-Ensemble-Net

using two different loss functions and combined using a novel technique, Double Loss Regression (DLR). Kinetics-FM-Net is built with Kinetics-Net and a FM. We build Kinetics-Net using different deep learning layers–GRU, Conv1D, Conv2D, and fully connected dense layers. All the components of Kinetics-FM-DLR-Ensemble-Net will be described in this section.

1) Kinetics-Net: Kinetics-Net mainly consists of three primary models– GRU-Net, GRU-Conv1D-Net, and GRU-Conv2D-Net. Predictions from these three models will be different because of the difference in architecture. The combination of these primary models may increase prediction performance. Typically, three models are trained separately, then combined with late fusion by taking the average to get the final prediction. However, this approach can cause complexity due to the separate training process of multiple primary models along with late fusion of the prediction. To address this issue, we train all the three primary models simultaneously along with the final prediction from the combined model to create an end-to-end trained kinetics-Net. We minimize the loss function for four output values of Kinetics-Net, which results in good predictive performance of the final model Kinetics-Net. If the predictions from GRU-Net, GRU-Conv2D-Net, and GRU-Conv1D-Net are $\mathbb{K}_{\Delta T}^{Output-1}$, $\mathbb{K}_{\Delta T}^{Output-2}$, and $\mathbb{K}_{\Delta T}^{Output-3}$ respectively, then output from Kinetics-Net

$$\mathbb{K}_{\Delta T}^{Output} = \frac{1}{3}(\mathbb{K}_{\Delta T}^{Output-1} + \mathbb{K}_{\Delta T}^{Output-2} + \mathbb{K}_{\Delta T}^{Output-3}) \quad (1)$$

a) **Primary Models:**

- **GRU-Net.** GRU-Net consists of an input layer followed by a Batch Normalization (BN) layer, a GRU block, a flatten layer, and an output layer. The BN layer is applied after the input signal to address heterogeneity of the source data [51]. After BN, a GRU block is added, and then output from the GRU block is flattened to add with the output layer.
- **GRU-Conv2D-Net.** In GRU-Conv2D-Net, we have two branches using the features from the flatten layer of GRU-Net and Conv2D-Net. For the second branch, we use a Conv2D block followed by a FC block. The output from the fully connected block is flattened and then concatenated with the features from the flatten layer of GRU-Net. The concatenated features are then connected to the Output-2 layer to make the prediction. The rationale

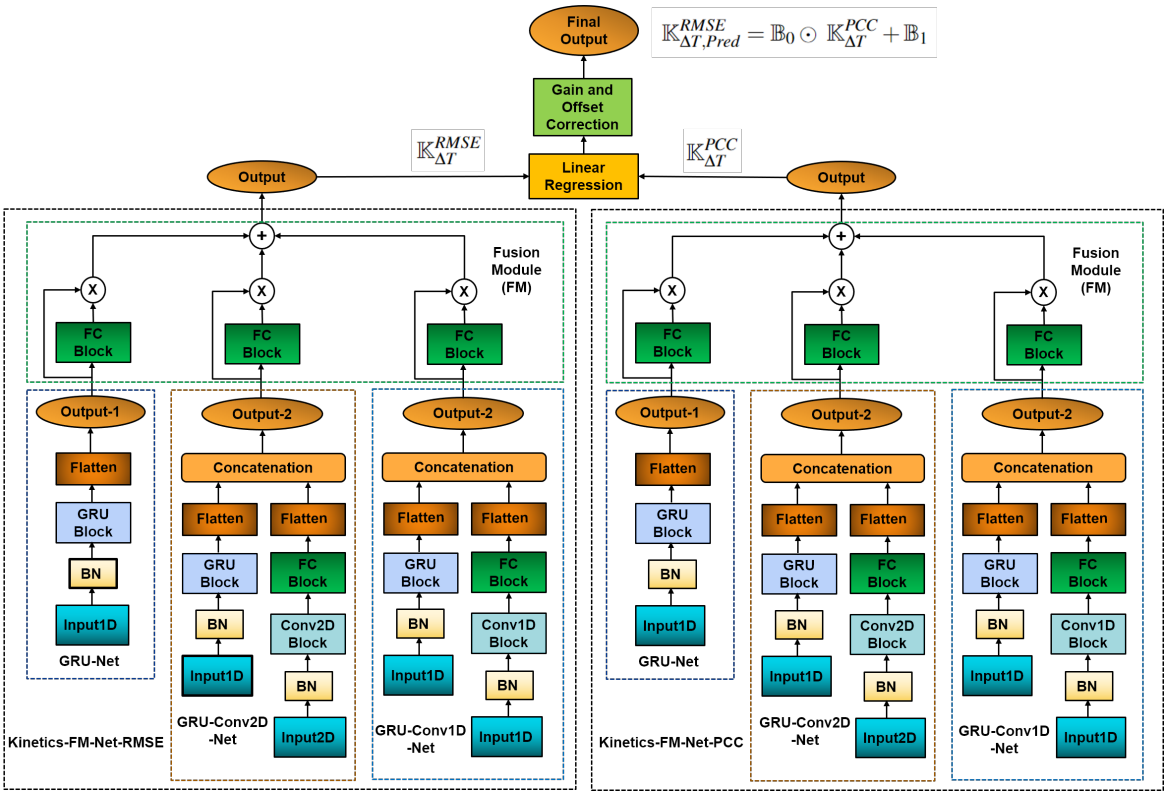

Fig. 2. Entire structure of Kinetics-FM-DLR-Net

for using two branches is to improve the prediction by integrating more diversified features from different types of deep learning layers.

- **GRU-Conv1D-Net.** In GRU-Conv1D-Net, we follow the same architecture as GRU-Conv2D-Net, but only replace the Conv2D block with Conv1D block.

b) **Fundamental Blocks:** All the fundamental blocks to create primary models are provided in this section:

- **GRU Block.** GRU block uses two GRU layers. A dropout layer is added after the GRU layer to avoid overfitting during the training of the model.
- **Conv2D Block.** In Conv2D block, a conv2D layer is used followed by a BN layer. BN layer helps to reduce internal co-variance shift. Then, a max-pooling2D layer is applied to reduce the feature space, which helps reduce the model's complexity and select dominant features. A conv2D, batch normalization, and max-pooling layer create the main unit of the Conv2D block. Four such units are added sequentially to create the Conv2D block.
- **Conv1D Block.** In Conv1D block, a conv1D layer is used followed by a BN layer. Then, a max-pooling1D layer is applied. A conv1D, BN, and max-pooling layer create the main unit of the Conv1D block. Four such units are added sequentially to create the Conv1D block.

2) Fusion Module (FM): In Kinetics-Net, we initially take the simple average of the output from three primary models. However, this may not ensure optimal performance from three models as we assign equal weights to all the models. Since the performance of each model will be different, proper weight needs to be assigned to ensure best performance gain from these three models. To do this, we design a FM creating Kinetics-FM-Net using two fully connected layers. Suppose Output-1, Output-2, Output-3 (Fig. 2) from GRU-Net, GRU-Conv2D-Net, GRU-Conv1D-Net are $\mathbb{K}^{GRU-Net}_{\Delta T} \in \mathbb{R}^{\Delta T \times D_K}$, $\mathbb{K}^{GRU-Conv2D-Net}_{\Delta T} \in \mathbb{R}^{\Delta T \times D_K}$, $\mathbb{K}^{GRU-Conv1D-Net}_{\Delta T} \in \mathbb{R}^{\Delta T \times D_K}$ respectively. These three outputs are passed to two Fully Connected (FC) layers. All the FC layers are dense layers, where rectified linear unit activation is applied between the first two layers and a sigmoid activation is in the last layer to enforce the output between 0 to 1. If output of each primary model after passing through the the dense layers are $FC(\mathbb{K}^{GRU-Net}_{\Delta T}) \in \mathbb{R}^{\Delta T \times D_K}$, $FC(\mathbb{K}^{GRU-Conv2D-Net}_{\Delta T}) \in \mathbb{R}^{\Delta T \times D_K}$, and $FC(\mathbb{K}^{GRU-Conv1D-Net}_{\Delta T}) \in \mathbb{R}^{\Delta T \times D_K}$ respectively, then output of the FM will be

$$\begin{aligned} \mathbb{K}^{Output}_{\Delta T} = FC(\mathbb{K}^{GRU-Net}_{\Delta T}) \odot \mathbb{K}^{GRU-Net}_{\Delta T} + \\ FC(\mathbb{K}^{GRU-Conv2D-Net}_{\Delta T}) \odot \mathbb{K}^{GRU-Conv2D-Net}_{\Delta T} + \\ FC(\mathbb{K}^{GRU-Conv1D-Net}_{\Delta T}) \odot \mathbb{K}^{GRU-Conv1D-Net}_{\Delta T} \end{aligned} \quad (2)$$

3) Double Loss Regression (DLR): Previously for kinetics estimation, the deep learning model was trained using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) as the loss function [41], [44]. While measuring the performance, both Normalized RMSE(NRMSE)– proportional to RMSE and Pearson Correlation Coefficient (PCC) were used [42]. Typically, when multiple loss functions are available, a single loss function is derived using the weighted sum of those loss functions. This approach of combining multiple loss functions (RMSE and PCC for our case) may not ensure

proper performance for kinetics estimation as the optimizer is minimizing the combined loss without understanding the proper relation between these loss functions. To use different loss function properly, we devise a novel strategy that employs two Kinetics-FM-Nets (Kinetics-FM-Net-RMSE, Kinetics-FM-Net-PCC), which are trained with two loss functions (RMSE and PCC) separately (Figure 2). In Kinetics-FM-Net-RMSE, the optimizer will try to minimize RMSE between ground truth and prediction, whereas in Kinetics-FM-Net-PCC, the optimizer will maximize the PCC between ground truth and prediction. As a result, from Kinetics-FM-Net-PCC, we will have predictions with similar profiles of joint moments and GRFs with ground truth but different baseline and range. In Figure 3, we show a plot of a gait cycle for different kinetics components when RMSE and PCC loss are used to train the model.

As the Kinetics-FM-Net-PCC is mainly focused on increasing PCC, this will have a higher PCC than the Kinetics-FM-Net-RMSE. To maintain that high PCC while acquiring actual baseline and range of the graph, we need the actual baseline and range of the prediction. To do this, we acquire baseline and range information from the prediction of Kinetics-FM-Net-RMSE by minimizing the offset between ground truth and prediction. Then, we use the results from Kinetics-FM-Net-RMSE to correct gain and offset of the Kinetics-FM-Net-PCC results while preserving the PCC obtained from the model. If the prediction from Kinetics-FM-Net-PCC is $\mathbb{K}_{\Delta T}^{PCC} \in \mathbb{R}^{\Delta T \times D_K}$, gain and the offset correction are $\mathbb{B}_0$=[B_{00},B_{01},.....,B_{0D_k}] $\in \mathbb{R}^{D_K}$,$\mathbb{B}_1$=[B_{10},B_{11},.....,B_{1D_k}] $\in \mathbb{R}^{D_K}$ respectively, then the corrected prediction from Kinetics-FM-Net-PCC can be considered as the element wise multiplication of gain and addition of offset corrected matrix.

$$\mathbb{K}_{\Delta T}^{RMSE} = \mathbb{B}_0 \odot \mathbb{K}_{\Delta T}^{PCC} + \mathbb{B}_1 \tag{3}$$

In Equation 3, $\mathbb{K}_{\Delta T}^{RMSE} \in \mathbb{R}^{\Delta T \times D_K}$ is the prediction from Kinetics-FM-Net-RMSE as it has closely similar range and offset of the ground truth. As we can estimate $\mathbb{K}_{\Delta T}^{RMSE}$ and $\mathbb{K}_{\Delta T}^{PCC}$ from the two Kinetics-FM-Net models, we can calculate the coefficient matrix of $\mathbb{B}_0$,$\mathbb{B}_1$ using D_K number of linear regression for each component. After calculating the gain and offset correction matrix of $\mathbb{B}_0$,$\mathbb{B}_1$, the final prediction after gain and offset correction:

$$\mathbb{K}_{\Delta T,Pred}^{RMSE} = \mathbb{B}_0 \odot \mathbb{K}_{\Delta T}^{PCC} + \mathbb{B}_1 \tag{4}$$

In Figure 3, we demonstrate the qualitative and quantitative impact on our DLR loss design for performance improvement.

4) Ensemble (Bagging): Finally, we apply bagging on Kinetics-FM-DLR-Net to create Kinetics-FM-DLR-Ensemble-Net. At first, from the training dataset, we create bootstrap samples (random sampling with replacement) from our training dataset. Suppose, we create K number of bootstrap samples from the whole training dataset. Each bootstrap sample will be used to train Kinetics-FM-DLR-Net. The output from each Kinetics-FM-DLR-Net will be $\mathbb{K}_{\Delta T,1}^{Bag}$, $\mathbb{K}_{\Delta T,K}^{Bag}$, ..., $\mathbb{K}_{\Delta T,2}^{Bag}$. Then, the final output of Kinetics-FM-DLR-Ensemble-Net:

$$\mathbb{K}_{\Delta T,Pred}^{Ensemble} = \frac{1}{K}\sum_{k=1}^{K} \mathbb{K}_{\Delta T,k}^{Bag} \tag{5}$$

TABLE I
HYPERPARAMETERS OF DIFFERENT LAYERS. VALUE OF DROPOUT IN THE BRACKET ARE FOR THE MODEL TRAINED WITH PCC AS LOSS FUNCTION.

Unit	GRU-Net	GRU-Conv2D-Net	GRU-Conv1D-Net	Fusion Module
GRU-1	Unit: 512 Dropout: 0.1 (0.5)	Unit: 512 Dropout: 0.1 (0.5)	Unit: 512 Dropout: 0.1 (0.5)	NA
GRU-2	Unit: 256 Dropout: 0.1 (0.5)	Unit: 256 Dropout: 0.1 (0.5)	Unit: 256 Dropout: 0.1 (0.5)	NA
Convolutional-1	NA	Filter: 256 Kernel: 5X3 Pool: 2X2	Filter: 256 Kernel: 3 Pool: 2	NA
Convolutional-2	NA	Filter: 256 Kernel: 5X3 Pool: 2X2	Filter: 256 Kernel: 3 Pool: 2	NA
Convolutional-3	NA	Filter: 512 Kernel: 5X3 Pool: 2X2	Filter: 512 Kernel: 3 Pool: 2	NA
Convolutional-4	NA	Filter: 512 Kernel: 5X3 Pool: 2X1	Filter: 512 Kernel: 3 Pool: 2	NA
FC-1	NA	Units: 64 Dropout: 0.25	Units: 64 Dropout: 0.25	Units: 128 Dropout: NA (0.4)
FC-2	NA	Units: 32 Dropout: 0.25	Units: 32 Dropout: 0.25	Units: 6 Dropout: NA (0.4)

IV. EXPERIMENTS

A. Dataset Description

This paper uses a publicly available dataset [52] to build Kinetics-FM-DLR-Ensemble-Net for the joint moment and GRFs estimation. We will briefly discuss the protocol of data collection in this subsection. This dataset is comprised with four locomotion modes, i.e., treadmill walking, level-ground walking, ramp ascent/descent, and stair ascent/descent. Treadmill walking was collected for 28 different speeds ranging from 0.5 to 1.85 m/s in 0.05 m/s increments across seven trials (four speeds per trial). Level-ground walking data was collected for 30 circuits–5 clockwise and 5 counterclockwise for each speed, which includes both straight walking and turning with self-selected slow, normal, and fast speed. In stair walking, the subject was walking on a six-step stair-case with four different stair heights. All the subjects completed a total of 40 trials with a set of five trials starting with their instrumented leg. Four IMU sensors (thigh, shank, and foot), three electronic goniometers (hip, knee, and ankle joint), and 11 EMG sensors were attached on the instrumented leg. A set of five trials starting with their non-instrumented leg for each stair height (4in, 5in, 6in, and 7in). Sixty ramp trials were performed with six different inclination angle (5.2°, 7.8°, 9.2°, 11°, 12.4°, and 18°) and five for each starting leg on six

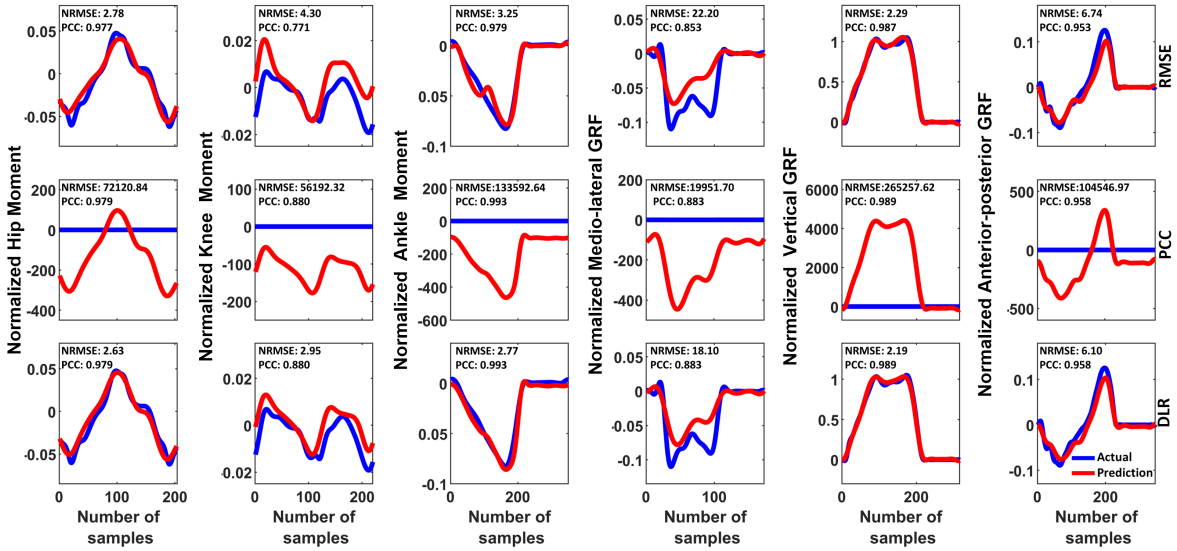

Fig. 3. A sample of joint moments and 3D GRFs in a single gait cycle to demonstrate how DLR improves the performance from RMSE and PCC loss function of deep learning model. Blue lines represent experimentally collected and calculated normalized joint moments and GRFs, and red lines present estimated results using the deep learning model.

inclination. Other detailed information of the protocol of the experiment is described in [52].

B. Dataset Pre-processing

Twenty subjects' data (8 female, age: 21.7 ± 3.65 years, height: 1.70 ± 0.07 m, weight: 68.21 ± 11.52 Kg) are used for model training, discarding two subjects (AB06, AB20) because of the absence of GRFs for level-ground condition (AB20) and IMU data mismatch (AB06) comparing to rest of the subjects. IMU data and marker trajectory data from the motion capture system were collected at 200 Hz. GRFs data were measured at 1000 Hz, then they are re-sampled to 200 Hz to synchronize with IMU and motion capture data. Joint moment is calculated using Opensim [13] inverse dynamics tool with motion capture and GRFs data. For level-ground, stair, and ramp walking, IMUs, joint moments, and GRFs data are segmented for a full gait cycle based on the availability of GRFs from the installed force plates on the instrumented right leg. On the other hand, data segmentation is not performed for the treadmill walking as GRFs are present for all the gait cycle. We normalize joint moment data with body weight and height of the participants. We also normalize GRF data with body weight of the participants. IMU data are collected from the foot, shank, thigh, and torso. IMU data from the foot, shank, and thigh are taken to build the deep learning model while discarding torso data, as we are primarily focused on predicting lower body joint kinetics. We provide details of segmentation of forceplate for level-ground, ramp, and stair condition in the supplementary materials.

C. Implementation Details

We train all of our models in Keras with a TITAN Xp GPU (NVIDIA, CA). The Kinetics-FM-DLR-Ensemble-Net is trained with a run time of 20 hours (2 hours per bootstrap sample). Adam [53] is used as the optimizer, and all the models are run for 40 epochs with a batch size of 64. We mainly use two loss functions, i.e. $\mathcal{L}_{RMSE}$, $\mathcal{L}_{PCC}$ for training. Joint Loss ($\mathcal{L}_{JL}$) is derived using these two loss functions with the following equation.

$$\mathcal{L}_{JL} = \mathcal{L}_{RMSE} + W \times \mathcal{L}_{PCC} \tag{6}$$

In Table I, we present all the hyperparameters of different layers of Kinetics-FM-DLR-Ensemble-Net.

D. Evaluation Procedures

Leave-one-subject out cross validation is implemented to assess the performance of the all the models. Excluding test subject from training data ensures proper validation as including test subject data in the training set can yield better model performance as the model can understand the relation between IMU and kinetics for that specific subject. For the performance measurement, we use two metrics, Normalized RMSE (NRMSE) and PCC. To get NRMSE, we normalize RMSE by the range (difference between maximum and minimum value) of corresponding experimentally measured GRF and directly calculated joint moments.

V. Results

In this section, we will present the results of model ablation, FM, and Double Loss Regression (DLR). We will also provide results for varying numbers of bootstrap samples and parameter sweeps of W of Equation 6. Additionally, we will also show the result comparison of our proposed model with the state-of-the-art deep learning algorithm for kinetics estimation.

A. Model Ablation

In Table II, we show the ablation study on our model and report the NRMSE and PCC values for kinetics estimation

TABLE II

MEAN AND STANDARD DEVIATION OF NRMSE AND PCC VALUES OF JOINT MOMENTS AND 3D GRFS FOR DIFFERENT MODELS IN ENTIRE WALKING CONDITIONS. BOLD NUMBERS REPRESENT THE RESULT OF KINETICS-FM-NET THAT HAS THE BEST PERFORMANCE IN NRMSE AND PCC.

Model	NRMSE			PCC		
	RMSE	JL	DLR	RMSE	JL	DLR
GRU-Net	4.84 ± 1.00	4.80 ± 1.12	4.57 ± 0.85	0.905 ± 0.036	0.904 ± 0.045	0.918 ± 0.028
Conv2D-Net	5.46 ± 0.99	5.45 ± 0.96	5.20 ± 0.91	0.867 ± 0.027	0.866 ± 0.040	0.890 ± 0.027
Conv1D-Net	5.38 ± 0.88	5.34 ± 0.87	5.19 ± 0.81	0.868 ± 0.026	0.872 ± 0.026	0.890 ± 0.026
GRU-Conv2D-Net	4.71 ± 0.91	4.74 ± 0.94	4.45 ± 0.83	0.907 ± 0.035	0.906 ± 0.033	0.920 ± 0.032
GRU-Conv1D-Net	4.79 ± 0.98	4.79 ± 0.93	4.39 ± 0.83	0.904 ± 0.040	0.907 ± 0.033	0.922 ± 0.029
Conv2D-Conv1D-Net	5.16 ± 0.75	4.89 ± 1.24	4.88 ± 0.79	0.884 ± 0.027	0.902 ± 0.047	0.906 ± 0.027
Kinetics-Sub-Net-1	4.60 ± 0.93	4.64 ± 1.13	4.44 ± 0.88	0.913 ± 0.035	0.913 ± 0.038	0.922 ± 0.030
Kinetics-Sub-Net-2	4.60 ± 0.90	4.51 ± 0.89	4.38 ± 0.86	0.913 ± 0.032	0.917 ± 0.032	0.923 ± 0.029
Kinetics-Sub-Net-3	4.59 ± 1.00	4.56 ± 0.93	4.28 ± 0.84	0.912 ± 0.036	0.915 ± 0.031	0.925 ± 0.029
Kinetics-Net	4.56 ± 0.93	4.52 ± 0.98	4.31 ± 0.88	0.915 ± 0.034	0.916 ± 0.034	0.926 ± 0.030
Kinetics-FM-Net	**4.38 ± 0.83**	**4.37 ± 0.83**	**4.21 ± 0.86**	**0.920 ± 0.030**	**0.921 ± 0.030**	**0.928 ± 0.028**

TABLE III

MEAN AND STANDARD DEVIATION OF NRMSE AND PCC VALUES OF JOINT MOMENTS AND 3D GRFS OF DIFFERENT WALKING ENVIRONMENTS FOR FINAL PROPOSED MODEL, KINETICS-FM-DLR-ENSEMBLE-NET.

Metrics	Scene	Hip Moment	Knee Moment	Ankle Moment	Medio-lateral GRF	Vertical GRF	Anterior-posterior GRF	Mean
NRMSE	**Treadmill**	5.31 ± 1.44	6.24 ± 1.77	6.00 ± 2.75	8.14 ± 3.55	5.36 ± 3.20	3.42 ± 1.05	5.74 ± 1.53
	Level-ground	9.16 ± 3.54	9.45 ± 3.00	9.23 ± 3.48	8.85 ± 1.77	7.40 ± 6.45	6.18 ± 3.87	8.38 ± 2.71
	Ramp	6.31 ± 1.61	5.37 ± 2.30	4.94 ± 2.64	8.32 ± 1.74	4.60 ± 2.80	3.99 ± 1.03	5.59 ± 1.16
	Stair	8.44 ± 2.52	7.49 ± 2.18	7.68 ± 2.18	8.16 ± 1.56	4.51 ± 2.24	4.64 ± 0.87	6.82 ± 1.20
PCC	**Treadmill**	0.948 ± 0.031	0.915 ± 0.041	0.966 ± 0.037	0.913 ± 0.045	0.976 ± 0.037	0.962 ± 0.031	0.947 ± 0.032
	Level-ground	0.764 ± 0.187	0.710 ± 0.207	0.869 ± 0.123	0.833 ± 0.101	0.957 ± 0.087	0.883 ± 0.177	0.836 ± 0.113
	Ramp	0.893 ± 0.058	0.908 ± 0.060	0.929 ± 0.062	0.847 ± 0.066	0.978 ± 0.043	0.951 ± 0.040	0.918 ± 0.043
	Stair	0.867 ± 0.076	0.926 ± 0.049	0.944 ± 0.041	0.885 ± 0.042	0.983 ± 0.024	0.926 ± 0.030	0.922 ± 0.032

TABLE IV

EFFECT OF VARYING NUMBER OF BOOTSTRAP SAMPLE ON THE MEAN AND STANDARD DEVIATION OF NRMSE FOR KINETICS ESTIMATION IN ENTIRE WALKING CONDITIONS. BOLD NUMBERS REPRESENT THE BEST PERFORMANCE IN EACH JOINT MOMENT AND GRF.

No	Hip Moment	Knee Moment	Ankle Moment	Medio-lateral GRF	Vertical GRF	Anterior-posterior GRF	Mean
1	4.75 ± 1.34	4.47 ± 1.83	4.21 ± 2.32	4.36 ± 1.26	5.07 ± 2.32	3.50 ± 0.71	4.39 ±0.88
2	4.58 ± 1.24	4.28 ± 1.69	4.05 ± 2.16	4.26 ± 1.30	4.97 ± 2.34	3.42 ± 0.72	4.26 ± 0.87
3	4.56 ± 1.23	4.29 ± 1.71	4.01 ± 2.17	4.26 ± 1.38	4.92 ± 2.37	3.39 ± 0.71	4.24 ± 0.89
4	4.51 ± 1.24	4.20 ± 1.60	3.98 ± 2.13	4.22 ± 1.37	4.88 ± 2.35	3.39 ± 0.71	4.20 ± 0.87
5	4.49 ± 1.24	4.19 ± 1.60	3.97 ± 2.14	4.19 ± 1.38	4.87 ± 2.31	3.37 ± 0.71	4.18 ± 0.88
6	4.48 ± 1.23	4.20 ± 1.60	3.93 ± 2.12	4.17 ± 1.35	4.86 ± 2.31	3.37 ± 0.71	4.17 ± 0.87
7	4.48 ± 1.24	4.15 ± 1.56	3.93 ± 2.12	4.18 ± 1.35	4.86 ± 2.31	3.36 ± 0.71	4.16 ± 0.87
8	4.46 ± 1.23	4.14 ± 1.55	3.92 ± 2.10	4.16 ± 1.33	4.85 ± 2.31	3.35 ± 0.71	4.15 ± 0.86
9	4.45 ± 1.22	4.13 ± 1.52	3.91 ± 2.10	4.14 ± 1.31	4.85 ± 2.31	3.35 ± 0.71	4.14 ± 0.86
10	**4.45 ± 1.21**	**4.11 ± 1.51**	**3.90 ± 2.08**	**4.14 ± 1.32**	**4.85 ± 2.31**	**3.34 ± 0.71**	**4.13 ± 0.86**

with different loss designs. As different layers are added to the model, the performance of estimation of both GRFs and joint moments are improved.

B. Fusion Module (FM)

Our designed fusion model that assigns proper weight to the prediction of primary models (GRU-Net, GRU-Conv2D-Net, and GRU-Conv1D-Net) increases joint moments and GRFs prediction in all walking conditions. Table II demonstrates that integrating the FM in Kinetics-Net decreases the mean NRMSE from 4.56, 4.52, 4.31 to 4.38, 4.37, 4.21 and increases mean PCC from 0.915, 0.916, 0.926 to 0.920, 0.921, 0.928 for RMSE, PCC, and DLR loss design respectively.

C. Double Loss Regression (DLR)

We find that our proposed DLR has the superiority in performance over RMSE and JL design for all the models (Table II). Although we initially design DLR to increase PCC of kinetics prediction, we have 'Double Reward' with DLR as it improves both NRMSE and PCC. More specifically, mean NRMSE decreases from 4.38 (RMSE), 4.37 (JL) to 4.21 (DLR) and mean PCC increases from 0.920 (RMSE), 0.921 (JL) to 0.928 (DLR) for Kinetics-FM-Net.

D. Bagging (Ensemble)

Kinetics-FM-DLR-Ensemble-Net that is created from Kinetics-FM-DLR-Net with application of bagging techniques

TABLE V

EFFECT OF VARYING NUMBER OF BOOTSTRAP SAMPLE ON THE MEAN AND STANDARD DEVIATION OF PCC FOR KINETICS ESTIMATION IN ENTIRE WALKING CONDITIONS. BOLD NUMBERS REPRESENT THE BEST PERFORMANCE IN EACH JOINT MOMENT AND GRF WITH 10 BOOTSTRAP SAMPLE.

No	Hip Moment	Knee Moment	Ankle Moment	Medio-lateral GRF	Vertical GRF	Anterior-posterior GRF	Mean
1	0.907 ± 0.033	0.896 ± 0.041	0.944 ± 0.035	0.882 ± 0.039	0.975 ± 0.030	0.950 ± 0.027	0.926 ± 0.028
2	0.911 ± 0.036	0.902 ± 0.042	0.946 ± 0.035	0.885 ± 0.039	0.975 ± 0.030	0.952 ± 0.027	0.928 ± 0.028
3	0.912 ± 0.036	0.904 ± 0.042	0.947 ± 0.035	0.887 ± 0.038	0.976 ± 0.030	0.953 ± 0.027	0.929 ± 0.028
4	0.913 ± 0.036	0.905 ± 0.040	0.947 ± 0.034	0.888 ± 0.037	0.976 ± 0.029	0.953 ± 0.027	0.930 ± 0.028
5	0.912 ± 0.037	0.905 ± 0.041	0.947 ± 0.034	0.889 ± 0.038	0.976 ± 0.029	0.953 ± 0.027	0.931 ± 0.028
6	0.912 ± 0.037	0.905 ± 0.041	0.947 ± 0.034	0.890 ± 0.038	0.976 ± 0.029	0.953 ± 0.027	0.931 ± 0.027
7	0.912 ± 0.036	0.906 ± 0.041	0.948 ± 0.034	0.890 ± 0.037	0.976 ± 0.029	0.953 ± 0.027	0.931 ± 0.027
8	0.912 ± 0.036	0.906 ± 0.041	0.948 ± 0.034	0.890 ± 0.038	0.976 ± 0.029	0.953 ± 0.027	0.931 ± 0.027
9	0.913 ± 0.036	0.906 ± 0.041	0.948 ± 0.034	0.890 ± 0.038	0.976 ± 0.029	0.953 ± 0.027	0.931 ± 0.027
10	**0.913 ± 0.036**	**0.906 ± 0.041**	**0.948 ± 0.034**	**0.890 ± 0.038**	**0.976 ± 0.029**	**0.953 ± 0.027**	**0.931 ± 0.027**

reduces mean NRMSE from 4.21 to 4.13 and increases mean PCC from 0.928 to 0.931 compared with Kinetics-FM-DLR-Net. Table IV, V, demonstrates the performance of adding different numbers of bagging samples to the Kinetics-FM-DLR-Net. After adding five bootstrap samples, results almost get saturated for both NRMSE and PCC. We have little improvement in NRMSE (0.01 increase after each bootstrap sample) over the next five bootstrap samples. Since adding more bootstrap samples will increase computational complexity, we avoid adding more bootstrap samples and stop our experiment at ten.

E. Parameter Sweep (W)

Typically, when we combine two loss function to create joint loss, weight of the loss is varied to ensure best performance gain. In Table II, we only show the results of JL with $W = 1$ (Equation 6), which may not give the best performance from the JL. To demonstrate how varying W impacts performance compared with our DLR loss design, we change the W in $\mathcal{L}_{JL}$ (Equation 6) from 1 to 10 and summarize the results in Table VI,VII. The rationale of choosing W from 1 to 10 is to make the actual $\mathcal{L}_{PCC}$ to be one-third to three times of $\mathcal{L}_{RMSE}$ to have proper weight on both $\mathcal{L}_{RMSE}$ and $\mathcal{L}_{PCC}$ to ensure valid comparison for joint loss. The effect of varying W in $\mathcal{L}_{JL}$ (Equation 6) did not provide distinct performance improvement. This result further demonstrates our superiority of loss design over $\mathcal{L}_{JL}$.

F. Comparison with State-Of-The-Art

In Table VIII, we show the comparison of our results with state-of-the-art deep learning algorithm for kinetics estimation. All these architectures that are used in these studies are optimized based on this dataset to ensure better performance and valid comparison. Here, FFN network uses raw IMU data as inputs. We extract 17 Handcrafted Features (HF) for each of the 18 components of IMU signals (total=306) and used those features as input to the FFN model (FFN-HF). Detailed architecture of all the models and feature extraction process are provided in the supplementary materials.

Table III shows the prediction outcomes of Kinetics-FM-DLR-Ensemble-Net in each walking condition. As an example, we plotted one gait cycle of each kinetics component representative of all walking conditions to demonstrate a sample qualitative comparison of ground truth and prediction from the model.

VI. DISCUSSION

This paper estimates sagittal plane hip, knee, and ankle joint moments as well as anterior-posterior, vertical, and medial-lateral GRFs in treadmill, level-ground, stairs, and ramp walking conditions using three IMU sensors on the thigh, shank, and foot using our Kinetics-FM-DLR-Ensemble-Net. This is the first study that implemented both GRFs and joint moments estimation on multiple walking conditions and speeds using IMU sensors via deep learning. We apply our algorithm to a comprehensive dataset with a large number of subjects, multiple walking environments, multiple walking speeds, multi variable ramp angles, and stair heights. We also perform validation of our model on an unseen subject (leave-one-subject out cross validation), which also makes sure that our model is not just memorizing a specific subject's IMU and kinetics relation. This extensive training set, generalized and versatile capacity of our novel algorithm, as well as rigorous validations will enable the most accurate estimates of joint moments and GRFs for new testing subjects in any walking speeds and various walking conditions compared with the other state-of-the-art deep learning algorithms for kinetics estimation. In addition to this, as we establish and evaluate our model on the public dataset, other researchers can easily validate our model with their algorithm and also have the potential to improve our contribution in this field further.

By leveraging different conventional deep learning layers, (i.e., 1D, 2D convolutional, GRU, and dense layers) this paper proposes an end-to-end model Kinetics-Net. In Table II, we demonstrate extensive model ablation with different loss function designs to validate the use of different layers in our model. Firstly, we build three simple models GRU-Net, Conv1D-Net, and Conv2D-Net by utilizing convolutional, GRU, and dense layers. Then, we concatenate features of two models from the combination of these three models to build GRU-Conv2D-Net, GRU-Conv1D-Net, and Conv2D-Conv1D-Net. From the results, we have an improvement in kinetics estimation performance when a combination of features of

TABLE VI

MEAN AND STANDARD DEVIATION OF ENTIRE WALKING NRMSE RESULTS FOR PARAMETER SWEEP (W) OF $\mathcal{L}_{JL}$ FOR KINETICS-FM-DLR-NET IN ENTIRE WALKING CONDITIONS. BOLD NUMBERS REPRESENT THE BEST PERFORMANCE IN EACH JOINT MOMENT AND GRF.

W	Hip Moment	Knee Moment	Ankle Moment	Medio-lateral GRF	Vertical GRF	Anterior-posterior GRF	Mean
1	4.78 ± 1.15	4.30 ± 1.28	4.15 ± 2.29	4.66 ± 1.56	4.98 ± 2.37	3.33 ± 0.66	4.37 ± 0.83
2	4.75 ± 1.25	**4.25 ± 1.22**	4.33 ± 2.29	4.59 ± 1.50	4.98 ± 2.36	3.28 ± 0.68	4.36 ± 0.82
3	**4.62 ± 1.28**	4.28 ± 1.20	4.29 ± 2.24	4.69 ± 1.50	4.91 ± 2.38	3.28 ± 0.71	4.34 ± 0.85
4	4.74 ± 1.29	4.33 ± 1.27	4.26 ± 2.11	4.68 ± 1.53	4.92 ± 2.42	3.25 ± 0.68	4.36 ± 0.83
5	4.67 ± 1.37	4.32 ± 1.20	4.26 ± 2.14	4.54 ± 1.41	**4.86 ± 2.30**	**3.23 ± 0.64**	**4.31 ± 0.82**
6	4.73 ± 1.29	4.33 ± 1.22	4.29 ± 2.23	4.63 ± 1.50	4.91 ± 2.33	3.27 ± 0.71	4.36 ± 0.82
7	4.80 ± 1.16	4.39 ± 1.19	4.20 ± 2.32	4.65 ± 1.69	5.05 ± 2.36	3.31 ± 0.67	4.40 ± 0.85
8	4.82 ± 1.23	4.43 ± 1.29	4.16 ± 2.21	**4.44 ± 1.27**	4.98 ± 2.22	3.30 ± 0 .66	4.35 ± 0.78
9	4.83 ± 1.23	4.47 ± 1.23	**4.13 ± 2.23**	4.54 ± 1.41	4.92 ± 2.21	3.30 ± 0.68	4.37 ± 0.80
10	4.86 ± 1.30	4.37 ± 1.17	4.14 ± 2.12	4.60 ± 1.46	4.89 ± 2.19	3.31 ± 0.68	4.36 ± 0.79

TABLE VII

MEAN AND STANDARD DEVIATION OF ENTIRE WALKING PCC RESULTS FOR PARAMETER SWEEP (W) OF $\mathcal{L}_{JL}$ FOR KINETICS-FM-DLR-NET IN ENTIRE WALKING CONDITIONS. BOLD NUMBERS REPRESENT THE BEST PERFORMANCE IN EACH JOINT MOMENT AND GRF.

W	Hip Moment	Knee Moment	Ankle Moment	Medio-lateral GRF	Vertical GRF	Anterior-posterior GRF	Mean
1	0.902 ± 0.030	0.891 ± 0.042	0.941 ± 0.039	0.866 ± 0.050	0.975 ± 0.029	0.952 ± 0.027	0.921 ± 0.030
2	0.901 ± 0.028	**0.892 ± 0.042**	**0.943 ± 0.037**	0.861 ± 0.083	0.975 ± 0.029	0.953 ± 0.027	0.921 ± 0.031
3	**0.904 ± 0.030**	0.890 ± 0.041	0.942 ± 0.038	0.850 ± 0.112	0.975 ± 0.030	0.952 ± 0.028	0.919 ± 0.035
4	0.901 ± 0.029	0.891 ± 0.041	0.943 ± 0.036	0.868 ± 0.052	0.975 ± 0.030	**0.953 ± 0.027**	**0.922 ± 0.029**
5	0.892 ± 0.055	0.877 ± 0.078	0.941 ± 0.038	**0.873 ± 0.047**	**0.976 ± 0.029**	0.952 ± 0.027	0.918 ± 0.037
6	0.901 ± 0.028	0.887 ± 0.043	0.941 ± 0.039	0.871 ± 0.049	0.975 ± 0.029	0.952 ± 0.028	0.921 ± 0.030
7	0.897 ± 0.033	0.888 ± 0.042	0.941 ± 0.037	0.863 ± 0.062	0.975 ± 0.030	0.953 ± 0.027	0.920 ± 0.030
8	0.901 ± 0.030	0.886 ± 0.043	0.941 ± 0.034	0.874 ± 0.040	0.976 ± 0.028	0.952 ± 0.026	0.922 ± 0.026
9	0.899 ± 0.036	0.884 ± 0.051	0.940 ± 0.037	0.845 ± 0.119	0.974 ± 0.030	0.952 ± 0.028	0.916 ± 0.038
10	0.898 ± 0.030	0.887 ± 0.048	0.940 ± 0.039	0.867 ± 0.048	0.976 ± 0.028	0.952 ± 0.027	0.920 ± 0.028

TABLE VIII

COMPARISON OF THE KINETICS ESTIMATION PERFORMANCE OF OUR ALGORITHM WITH STATE-OF-THE-ART DEEP LEARNING METHOD OF KINETICS ESTIMATION. OUR KINETICS-FM-NET SHOWS THE BEST PREDICTION ACCURACY IN NRMSE AND PCC COMPARED WITH THE OTHER STATE-OF-ART DEEP LEARNING MODELS.

Metric	Models	Hip Moment	Knee Moment	Ankle Moment	Medio-Lateral GRF	Vertical GRF	Anterior-posterior GRF	Mean
NRMSE	FFN (HF) [35]	5.71 ± 1.37	5.97 ± 1.74	5.16 ± 2.66	5.08 ± 1.26	6.63 ± 2.15	4.13 ± 0.59	5.45 ± 0.85
	TCN [34]	5.48 ± 1.25	5.05 ± 1.48	5.01 ± 2.82	4.92 ± 1.81	5.64 ± 2.26	3.70 ± 0.64	4.97 ± 0.91
	FFN [45], [37]	5.48 ± 1.49	5.49 ± 2.04	4.71 ± 2.50	5.19 ± 1.89	6.06 ± 2.43	3.90 ± 0.81	5.14 ± 1.19
	LSTM [37], [40]	5.29 ± 1.24	4.66 ± 1.58	4.77 ± 2.81	4.69 ± 1.39	5.23 ± 2.25	3.57 ± 0.68	4.70 ± 0.91
	Conv2D [44]	6.11 ± 1.46	6.32 ± 1.85	5.42 ± 3.05	5.11 ± 1.43	5.77 ± 2.27	4.05 ± 0.57	5.46 ± 0.99
	Ours	**4.45 ± 1.21**	**4.11 ± 1.51**	**3.90 ± 2.08**	**4.14 ± 1.32**	**4.85 ± 2.31**	**3.34 ± 0.71**	**4.13 ± 0.86**
PCC	FFN (HF) [35]	0.847 ± 0.036	0.778 ± 0.050	0.912 ± 0.039	0.819 ± 0.042	0.959 ± 0.033	0.923 ± 0.028	0.873 ± 0.027
	TCN [34]	0.866 ± 0.030	0.847 ± 0.041	0.924 ± 0.038	0.840 ± 0.046	0.972 ± 0.030	0.943 ± 0.026	0.899 ± 0.027
	FFN [45], [37]	0.861 ± 0.053	0.824 ± 0.110	0.926 ± 0.041	0.815 ± 0.097	0.966 ± 0.035	0.933 ± 0.037	0.888 ± 0.059
	LSTM [37], [40]	0.877 ± 0.027	0.868 ± 0.046	0.931 ± 0.034	0.857 ± 0.040	0.974 ± 0.029	0.943 ± 0.027	0.908 ± 0.026
	Conv2D [44]	0.821 ± 0.041	0.747 ± 0.070	0.911 ± 0.037	0.821 ± 0.045	0.970 ± 0.030	0.933 ± 0.026	0.867 ± 0.027
	Ours	**0.913 ± 0.036**	**0.906 ± 0.041**	**0.948 ± 0.034**	**0.890 ± 0.038**	**0.976 ± 0.029**	**0.953 ± 0.027**	**0.931 ± 0.027**

two models is used. This validates our approach of adding features from two models to improve the prediction. From the first six models in Table II, we dismiss Conv1D-Net, Conv2D-Net, and Conv2D-Conv1D-Net for further model development due to their poor performance. Later, we use GRU-Net, GRU-Conv2D-Net, and GRU-Conv1D-Net to build three Kinetics-Sub-Nets using the average prediction of two models from these three models in an end-to-end manner. We see a further improvement in results by creating those Sub-Nets. Finally, we use GRU-Net, GRU-Conv2D-Net, and GRU-Conv1D-Net and take the average of these three models to create end-to-end model Kinetics-Net, which outperforms all the previous models. This validates our approach of adding different layers to create a more complex model for performance improvement.

Although many studies have applied deep learning algorithms to estimate kinetics [34]–[43], [43]–[46], a direct comparison of their results with ours cannot be valid due to the different sensor modalities, number of sensors, walking environments, and number of subjects. Moreover, those datasets are not publicly available, which makes it challenging to apply

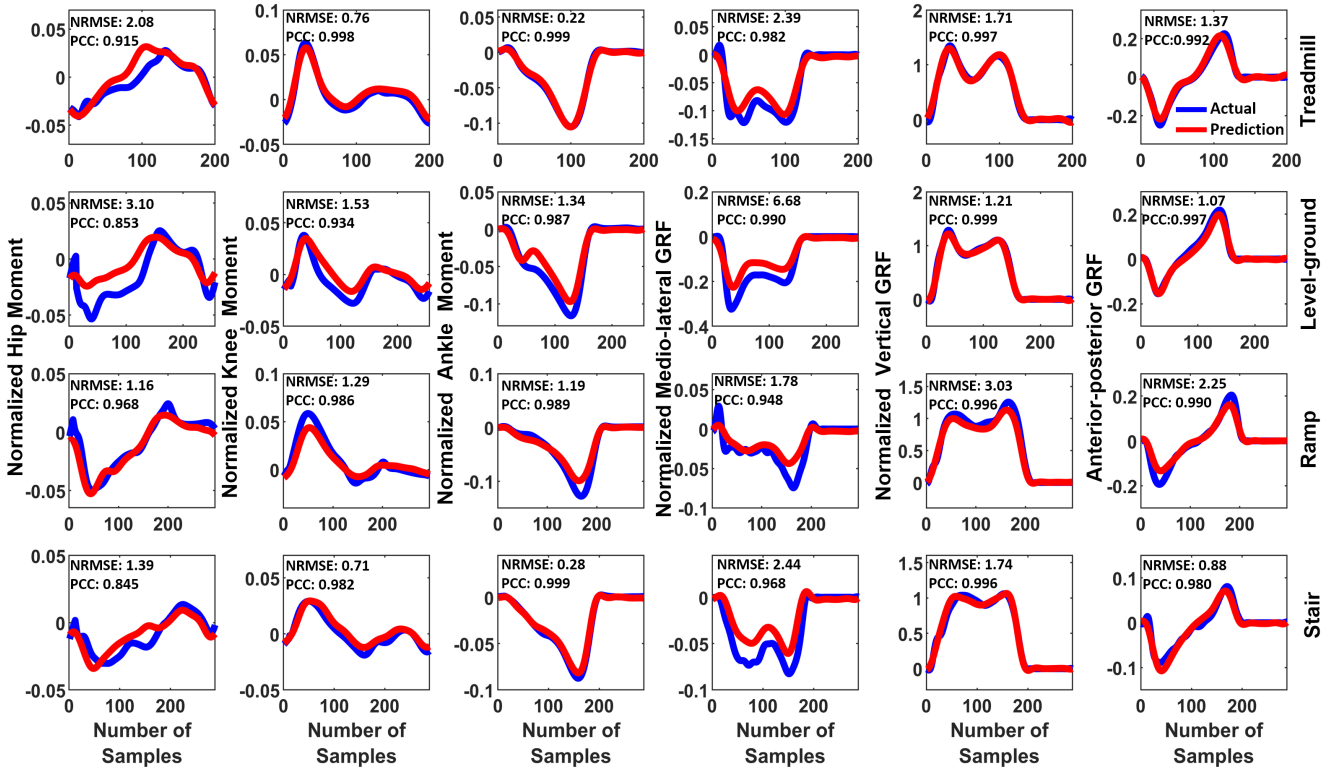

Fig. 4. A sample plot for a single gait cycle for different walking condition along with the corresponding NRMSE and PCC of subject AB24.

our algorithm to make the comparisons. To address this issue, we adopt their machine learning algorithms and applied to the dataset that we used. However, directly applying their specific machine learning architecture to this dataset may not be a valid method, as those models are specifically built for their dataset and inputs (variable number and multi-modal sensors). For this reason, we use the same types of layers such as LSTM, TCN, Conv2D, FFN, etc, and optimize these models for our dataset for achieving high performance to ensure valid comparison. In Table VIII, we show the comparison of the performance of the state-of-the-art deep learning algorithm with ours. Our algorithm Kinetics-FM-DLR-Ensemble-Net outperforms those algorithms by a large margin, which proves the effectiveness of our model over the conventional deep learning method for kinetics estimation.

Although we provide the most accurate prediction of joint moments and GRFs with extensive walking conditions, speeds and number of subjects, there are several limitations on this study. We use a relatively large number of IMU sensors (three) to estimate kinetics. However, to ensure practicality and users' comfort, we need to minimize the number of IMU sensors. More specifically, if we can implement only shoe-mounted sensors similar to kinematics estimation in [54], it would be more helpful to maintain sensors. Since single limb joint moments and GRFs are affected by contralateral limb during early stance phase and terminal stance phase, incorporating IMU information on both limbs would gather more meaningful knowledge of walking dynamics, and this can improve the prediction further. Ensemble learning–bagging is added to improve the prediction. However, it increases the computational complexity as the model is repeated ten times. If we want to make a trade-off between accuracy and computation complexity, we can avoid using bagging techniques in our model. The different training data set impacts the accuracy. From the results, level-ground condition has the largest error compared with the other walking conditions. The main reason for this is the lack of training data during the level-ground condition. As we segment the dataset for level-ground, we have only limited right leg strikes for the walking circuit, which results in fewer data (4% of the total dataset) compared to other conditions. A future research direction to improve the level-ground kinetics prediction may be to augment the level-ground dataset by repetition and make it roughly equal to other walking environments.

VII. Conclusion

This study proposes a novel deep learning model to estimate kinetics in multiple walking conditions and speeds. From our extensive evaluation of our developed model, we justify our design choices. This accurate estimation will enable tracking of kinetics parameters outside the lab, removing the limitation of traditional motion capture cameras and floor-embedded force plate based kinetics estimation.

References

[1] J. L. Astephen, K. J. Deluzio, G. E. Caldwell, and M. J. Dunbar, "Biomechanical changes at the hip, knee, and ankle joints during gait are associated with knee osteoarthritis severity," *Journal of orthopaedic research*, vol. 26, no. 3, pp. 332–341, 2008.

[2] L. Woolnough, A. Pomputius, and H. K. Vincent, "Juvenile idiopathic arthritis, gait characteristics and relation to function," *Gait & Posture*, vol. 85, pp. 38–54, 2021.

[3] M. B. Simonsen, R. P. Hirata, K. Næsborg-Andersen, P. D. C. Leutscher, K. Hørslev-Petersen, J. Woodburn, and M. S. Andersen, "Different types of foot orthoses effect on gait mechanics in patients with rheumatoid arthritis," *Journal of Biomechanics*, p. 110496, 2021.

[4] P. Aleixo, J. Vaz-Patto, H. Moreira, and J. Abrantes, "Dynamic joint stiffness of the ankle in healthy and rheumatoid arthritis post-menopausal women," *Gait & posture*, vol. 60, pp. 225–234, 2018.

[5] M. Hall, S. Chabra, N. Shakoor, S. E. Leurgans, H. Demirtas, and K. C. Foucher, "Hip joint moments in symptomatic vs. asymptomatic people with mild radiographic hip osteoarthritis," *Journal of Biomechanics*, vol. 96, p. 109347, 2019.

[6] T. Lenzi, M. C. Carrozza, and S. K. Agrawal, "Powered hip exoskeletons can reduce the user's hip and ankle muscle activations during walking," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 6, pp. 938–948, 2013.

[7] M. Alaqtash, T. Sarkodie-Gyan, H. Yu, O. Fuentes, R. Brower, and A. Abdelgawad, "Automatic classification of pathological gait patterns using ground reaction forces and machine learning algorithms," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 453–457.

[8] C. Nüesch, V. Valderrabano, C. Huber, V. von Tscharner, and G. Pagenstert, "Gait patterns of asymmetric ankle osteoarthritis patients," *Clinical Biomechanics*, vol. 27, no. 6, pp. 613–618, 2012.

[9] D. P. Soares, M. P. de Castro, E. A. Mendes, and L. Machado, "Principal component analysis in ground reaction forces and center of pressure gait waveforms of people with transfemoral amputation," *Prosthetics and orthotics international*, vol. 40, no. 6, pp. 729–738, 2016.

[10] A. M. Cárdenas, J. Uribe, J. M. Font-Llagunes, A. M. Hernández, and J. A. Plata, "The effect of prosthetic alignment on the stump temperature and ground reaction forces during gait in transfemoral amputees," *Gait & Posture*, vol. 95, pp. 76–83, 2022.

[11] T. Kobayashi, M. Hu, R. Amma, G. Hisano, H. Murata, D. Ichimura, and H. Hobara, "Effects of walking speed on magnitude and symmetry of ground reaction forces in individuals with transfemoral prosthesis," *Journal of biomechanics*, vol. 130, p. 110845, 2022.

[12] N. A. Hashim, N. A. A. Razak, and N. A. A. Osman, "Ground reaction force of trilateral amputee during walking with and without upper limb prosthesis: Case report," in *Kuala Lumpur International Conference on Biomedical Engineering*. Springer, 2022, pp. 187–195.

[13] S. L. Delp, F. C. Anderson, A. S. Arnold, P. Loan, A. Habib, C. T. John, E. Guendelman, and D. G. Thelen, "Opensim: Open-source software to create and analyze dynamic simulations of movement," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 11, pp. 1940–1950, 2007.

[14] L. Ren, R. K. Jones, and D. Howard, "Whole body inverse dynamics over a complete gait cycle based only on measured kinematics," *Journal of biomechanics*, vol. 41, no. 12, pp. 2750–2759, 2008.

[15] E. C.-Y. Yang and M.-H. Mao, "3d analysis system for estimating intersegmental forces and moments exerted on human lower limbs during walking motion," *Measurement*, vol. 73, pp. 171–179, 2015.

[16] A. Ancillao, S. Tedesco, J. Barton, and B. O'Flynn, "Indirect measurement of ground reaction forces and moments by means of wearable inertial sensors: A systematic review," *Sensors*, vol. 18, no. 8, p. 2564, 2018.

[17] A. Karatsidis, G. Bellusci, H. M. Schepers, M. De Zee, M. S. Andersen, and P. H. Veltink, "Estimation of ground reaction forces and moments during gait using only inertial motion capture," *Sensors*, vol. 17, no. 1, p. 75, 2016.

[18] M. Aurbach, K. Wagner, F. Süß, and S. Dendorfer, "Implementation and validation of human kinematics measured using imus for musculoskeletal simulations by the evaluation of joint reaction forces," in *CMBEBIH 2017*. Springer, 2017, pp. 205–211.

[19] A. H. Abdul Razak, A. Zayegh, R. K. Begg, and Y. Wahab, "Foot plantar pressure measurement system: A review," *Sensors*, vol. 12, no. 7, pp. 9884–9912, 2012.

[20] E. Shahabpoor and A. Pavic, "Measurement of walking ground reactions in real-life environments: a systematic review of techniques and technologies," *Sensors*, vol. 17, no. 9, p. 2085, 2017.

[21] S. J. Kim, G. M. Gu, Y. Na, J. Park, Y. Kim, and J. Kim, "Wireless ground reaction force sensing system using a mechanically decoupled two-dimensional force sensor," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 1, pp. 66–75, 2019.

[22] T. Liu, Y. Inoue, and K. Shibata, "A wearable ground reaction force sensor system and its application to the measurement of extrinsic gait variability," *Sensors*, vol. 10, no. 11, pp. 10 240–10 255, 2010.

[23] D. G. Lloyd and T. F. Besier, "An emg-driven musculoskeletal model to estimate muscle forces and knee joint moments in vivo," *Journal of biomechanics*, vol. 36, no. 6, pp. 765–776, 2003.

[24] T. S. Buchanan, D. G. Lloyd, K. Manal, and T. F. Besier, "Neuromusculoskeletal modeling: estimation of muscle forces and joint moments and movements from measurements of neural command," *Journal of applied biomechanics*, vol. 20, no. 4, p. 367, 2004.

[25] G. Durandau, D. Farina, and M. Sartori, "Robust real-time musculoskeletal modeling driven by electromyograms," *IEEE transactions on biomedical engineering*, vol. 65, no. 3, pp. 556–564, 2017.

[26] B. L. Davis, J. E. Perry, D. C. Neth, and K. C. Waters, "A device for simultaneous measurement of pressure and shear force distribution on the plantar surface of the foot," *Journal of Applied Biomechanics*, vol. 14, no. 1, pp. 93–104, 1998.

[27] A. Faivre, M. Dahan, B. Parratte, and G. Monnier, "Instrumented shoes for pathological gait assessment," *Mechanics research communications*, vol. 31, no. 5, pp. 627–632, 2004.

[28] R. Riener, M. Rabuffetti, C. Frigo, J. Quintern, and G. Schmidt, "Instrumented staircase for ground reaction measurement," *Medical & biological engineering & computing*, vol. 37, no. 4, pp. 526–529, 1999.

[29] A. Pedotti, R. Assente, G. Fusi, D. De Rossi, P. Dario, and C. Domenici, "Multisensor piezoelectric polymer insole for pedobarography," *Ferroelectrics*, vol. 60, no. 1, pp. 163–174, 1984.

[30] J. Hidler, "Robotic-assessment of walking in individuals with gait disorders," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2. IEEE, 2004, pp. 4829–4831.

[31] M. A. Razian and M. G. Pepper, "Design, development, and characteristics of an in-shoe triaxial pressure measurement transducer utilizing a single element of piezoelectric copolymer film," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 3, pp. 288–293, 2003.

[32] G. Bakalidis, E. Glavas, N. Voglis, and P. Tsalides, "A low-cost fiber optic force sensor," *IEEE transactions on instrumentation and measurement*, vol. 45, no. 1, pp. 328–331, 1996.

[33] W.-C. Wang, W. R. Ledoux, B. J. Sangeorzan, and P. G. Reinhall, "A shear and plantar pressure sensor based on fiber-optic bend loss." *Journal of Rehabilitation Research & Development*, vol. 42, no. 3, 2005.

[34] D. D. Molinaro, I. Kang, J. Camargo, M. C. Gombolay, and A. J. Young, "Subject-independent, biological hip moment estimation during multimodal overground ambulation using deep learning," *IEEE Transactions on Medical Robotics and Bionics*, vol. 4, no. 1, pp. 219–229, 2022.

[35] J. Camargo, D. Molinaro, and A. Young, "Predicting biological joint moment during multiple ambulation tasks," *Journal of Biomechanics*, vol. 134, p. 111020, 2022.

[36] H. Lim, B. Kim, and S. Park, "Prediction of lower limb kinetics and kinematics during walking by a single imu on the lower back using machine learning," *Sensors*, vol. 20, no. 1, 2020.

[37] M. Mundt, W. Thomsen, T. Witter, A. Koeppe, S. David, F. Bamer, W. Potthast, and B. Markert, "Prediction of lower limb joint angles and moments during gait using artificial neural networks," *Medical & biological engineering & computing*, vol. 58, no. 1, pp. 211–225, 2020.

[38] M. Mundt, A. Koeppe, S. David, F. Bamer, W. Potthast, and B. Markert, "Prediction of ground reaction force and joint moments based on optical motion capture data during gait," *Medical Engineering & Physics*, vol. 86, pp. 29–34, 2020.

[39] M. Mundt, A. Koeppe, F. Bamer, S. David, and B. Markert, "Artificial neural networks in motion analysis—applications of unsupervised and heuristic feature selection techniques," *Sensors*, vol. 20, no. 16, p. 4581, 2020.

[40] M. Mundt, W. R. Johnson, W. Potthast, B. Markert, A. Mian, and J. Alderson, "A comparison of three neural network approaches for estimating joint angles and moments from inertial measurement units," *Sensors*, vol. 21, no. 13, p. 4535, 2021.

[41] M. Mundt, A. Koeppe, S. David, T. Witter, F. Bamer, W. Potthast, and B. Markert, "Estimation of gait mechanics based on simulated and measured imu data using an artificial neural network," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 41, 2020.

[42] M. Mundt, A. Koeppe, F. Bamer, W. Potthast, and B. Markert, "Prediction of joint kinetics based on joint kinematics using artificial neural networks," *ISBS Proceedings Archive*, vol. 36, no. 1, p. 794, 2018.

[43] M. Mundt, A. Koeppe, S. David, T. Witter, F. Bamer, W. Potthast, and B. Markert, "Estimation of gait mechanics based on simulated

and measured imu data using an artificial neural network," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 41, 2020.
[44] E. Dorschky, M. Nitschke, C. F. Martindale, A. J. Van den Bogert, A. D. Koelewijn, and B. M. Eskofier, "Cnn-based estimation of sagittal plane walking and running biomechanics from measured and simulated inertial sensor data," *Frontiers in bioengineering and biotechnology*, p. 604, 2020.
[45] G. Leporace, L. A. Batista, and J. Nadal, "Prediction of 3d ground reaction forces during gait based on accelerometer data," *Research on Biomedical Engineering*, vol. 34, pp. 211–216, 2018.
[46] Y. Guo, F. Storm, Y. Zhao, S. A. Billings, A. Pavic, C. Mazzà, and L.-Z. Guo, "A new proxy measurement algorithm with application to the estimation of vertical ground reaction forces using wearable sensors," *Sensors*, vol. 17, no. 10, p. 2181, 2017.
[47] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.
[48] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[49] H. Lim, B. Kim, and S. Park, "Prediction of lower limb kinetics and kinematics during walking by a single imu on the lower back using machine learning," *Sensors*, vol. 20, no. 1, p. 130, 2019.
[50] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, p. 123–140, Aug. 1996. [Online]. Available: https://doi.org/10.1023/A:1018054314350
[51] F. Li, K. Shirahama, M. A. Nisar, X. Huang, and M. Grzegorzek, "Deep transfer learning for time series data based on sensor modality classification," *Sensors*, vol. 20, no. 15, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/15/4271
[52] J. Camargo, A. Ramanathan, W. Flanagan, and A. Young, "A comprehensive, open-source dataset of lower limb biomechanics in multiple conditions of stairs, ramps, and level-ground ambulation and transitions," *Journal of Biomechanics*, vol. 119, p. 110320, 2021.
[53] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.
[54] M. S. B. Hossain, J. Dranetz, H. Choi, and Z. Guo, "Deepbbwae-net: A cnn-rnn based deep superlearner for estimating lower extremity sagittal plane joint kinematics using shoe-mounted imu sensors in daily living," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2022.

VIII. Supplementary Materials

IX. Data Segmentation

level-ground. For clockwise level-ground walking, we segment FP2, Combined, and FP5 forceplate data and for anti-clockwise level-ground walking, we segment FP5 data.
Stair. For stair trials with starting instrumented leg (right leg), we segment forceplate data of FP1 and FP5 for right leg. For the trials starting with non-instrumented leg (left leg), we segment FP2, FP3, and FP4.
Ramp. For ramp trials with starting non-instrumented leg (left leg), we segment forceplate data of FP2 and FP4 for right leg. For the trials starting with instrumented leg (right leg), we segment FP1, FP3, and FP5.

During the segmentation, we notice that some of the participants did not successfully step their entire right foot on the forceplates for all the trials. This incomplete force plate data set cannot derive complete joint moment calculations and misleading GRFs data. Thus, we identify these incomplete dataset and manually exclude from our model training dataset.

X. State-Of-The-Art model description

FFN (HF). In FFN-HF, we extract 17 features for all the component of IMU data. We extract those features for the window length of 100 (0.5s). Extracted features are mean, RMS, max, min, mean absolute value, standard deviation, mean absolute difference, mean difference, median difference, median absolute difference, interquartile range, kurtosis, skewness, median, variance, median absolute deviation, and mean absolute deviation. We extract total 306 features from 3 IMUs. Then, we use those extracted features as the input to a FFN neural network with five dense layers. Each layer is followed by a dropout to avoid overfitting during the training of the model. We flatten the output of the last dropout layer and connect it to the last layer for kinetics prediction. The units of the five layers are 1024, 512, 256, 128, and 64. For all the drouput layer, we use a dropout rate of 0.05.
TCN. For TCN, we use a single stacks of residual block with a kernel size 3, 128 filters in the convolutional layers, dilations of (1, 2, 4, 8, 16), dropout rate of 0.05, and weight normalization. TCN layer is followed by two dense layers and a dropout layer after each of them. The units of the dense layer is 64, 32 and dropout rate was 0.1. We flatten the features from the last dropout layer and connect it to the prediction layer.
FFN. For FFN, we use raw IMU data as input. Two dense layers with a dropout after each one was used. The units of the dense layers are 128, 64 respectively. The dropout rate was 0.1 for both dropout layers. Featurs from the last dropout layer was flattened and connected to the last prediction layer.
LSTM. We use two bidirectional LSTM with units 128 and 64 respectively. Dropout layer with a dropout rate 0.5 is used after each bidirectional LSTM. After the dropout layer, two dense layer each one followed by a dropout was used, flattened and connected to the prediction layer. The units of the dense layer are 128 and 64 respectively and dropout rate was 0.25 for both dropout layer.
Conv2D. We use the same architecture as Conv2D-Net.