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 are 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 are 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 for single limb prediction of hip, knee, and ankle joint moments and 3 dimensional Ground Reaction Forces (GRFs) using three IMU sensors on the thigh, shank, and foot under several representatives walking conditions in daily living, such as treadmill, level-ground, stair, and ramp. 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]–[3] and to provide assistive torque to the powered exoskeleton to minimize muscle efforts [4]. Moreover, knee abduction moment (KAM) and knee flexion moment (KFM) are critical biomechanical measurements for assessing knee joint loading [5], which is considered to contribute to knee osteoarthritis [6]. GRFs have also been used

for the evaluation of pathological gait patterns [7]–[9] or for analyzing the gait of amputees [10], [11].

Traditionally, joint moments are calculated using infrared-based motion capture cameras and ground reaction force plates. Collected joint kinematics from motion capture cameras and GRFs from the floor-embedded force plates are then implemented to computational musculoskeletal modeling softwares such as OpenSim [12], 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. 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 dynamic models [13]–[23], neuromusculoskeletal models [22], [24], or wearable force plates [18]–[21]. 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 [14], 17 IMUs in [16], and 15 IMUs in [17]). Wearable electromyography (EMG) sensors on specific muscle groups can also be used with neuromusculoskeletal modeling to calculate joint moment [22], [24]. 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 force plates. GRFs can be calculated by using a strain gauge transducer [25], piezoelectric sensors

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

Zhishan Guo is with the Computer Science Department, North Carolina State University, Raleigh, North Carolina, 27695, USA, (email: zguo32@ncsu.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)

[26], and fiber-optic force sensors [27]. However, these sensors are vulnerable to hysteresis, sensitive to temperature, and constrained by force ranges and deformation of the sensors' materials [18]. 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 [28]–[37], [37]–[40]. However, these studies are still relying on a large number of sensors [29], reversely acquired simulated IMU data from the retrospective motion capture data [28], [31], and highly repetitive treadmill/level-ground walking [30]–[37], [37]–[40]. Since experimentally collected IMU data contains noise induced from multi-frequency 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) [41], Feedforward Neural Networks (FNN), and Long Short Term Memory (LSTM) [42] 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 kinetics in the simple repetitive level-ground / treadmill motion, we aim to predict the single limb sagittal plane joint moment of the hip, knee, and ankle, the frontal plane joint moment of the hip and knee, and 3D GRFs (anterior-posterior, vertical, and medio-lateral) using three IMU sensors on the foot, shank, and thigh using a novel deep learning model Kinetics-FM-DLR-Ensemble-Net in multiple walking environments. To implement the algorithm, we utilize two publicly available datasets– Dataset A [43] (sagittal plane hip, knee, ankle moments, hip abduction moment, and 3D GRFs) and Dataset B [44] (KAM, KFM, and 3D GRFs). 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 conditions using IMU sensors via deep learning. Our contribution is seven-fold: (i) proposing an end-to-end trained model Kinetics-Net leveraging different deep learning layers to increase human kinetics prediction performance; (ii) presenting a Fusion Module (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; (vii) demonstrating generalization capabilities of the model on two publicly available datasets.

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 VII.

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 musculoskeletal or analytical model-based kinetics estimation methods. Later, we will discuss data-driven methods for kinetics estimation and their limitations.

Yang et al. [14] used seven IMU sensors to estimate GRFs and moments during walking with a three-dimensional analytical model. Karatsidis et al. [16] 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. [17] 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 force plate, researchers are focusing on estimating GRFs [37], [39], [40], [45] and joint moments [28]–[35], [37], [38] using data-driven methods with IMU sensors. Dorschky et al. [38] 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. [39] 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. [40] 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 a dataset without the presence of experimentally collected IMU sensors with motion capture and GRF

data, Mundt et al. [31] placed virtual IMU sensors on the retrospectively collected walking model to predict hip, knee, and ankle joint moments using feedforward and LSTM based models. Molinaro et al. [28] also used a musculoskeleletal model to generate virtual IMU sensor signals on the trunk and thigh to estimate hip joint moment using TCN [41]. As these studies are implemented using virtual IMU data, it is unclear how reliable the outcome of their methods is on real IMU applications with noise introduced from skin surface movements.

To verify which deep learning model has better estimates on joint moments, there was a comparison between current deep learning methods, LSTM, MLP, and pre-trained convolutional neural network [34]. 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. [29] 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 make doubt on practicality in real-world deployment. In addition to this, they used handcrafted feature engineering, which adds complexity into their proposed method. To reduce the sensor count, Lim et al. [30] 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 of joint kinetics estimation 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 three IMU sensors in multiple walking conditions.

III. PROPOSED APPROACH

A. Problem Statement

This paper estimates three-dimensional GRFs and hip, knee, and ankle joint moment using three IMU sensors on the 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 of the single limb (anterior-posterior GRF, vertical GRF, mediolateral GRF, sagittal plane hip, knee, and ankle joint moment, and frontal plane hip joint moment for Dataset

A and 3D GRFs, KAM, KFM for Dataset B). In this problem, we use D_{IMU}=6; $N=3$; ΔT=100 (Dataset A), 50 (Dataset B); D_k=7 (Dataset A), 5 (Dataset B). Mathematically, $\mathbb{I}_{\Delta T} \rightarrow \mathbb{K}_{\Delta T}$.
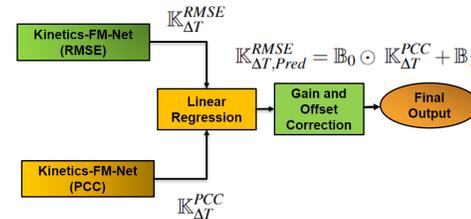
B. Kinetics-FM-DLR-Ensemble-Net

Fig. 1. Entire structure of Kinetics-FM-DLR-Net

Kinetics-FM-DLR-Ensemble-Net is built mainly with the model Kinetics-FM-DLR-Net (Fig. 1). We implement bagging [46] 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 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.
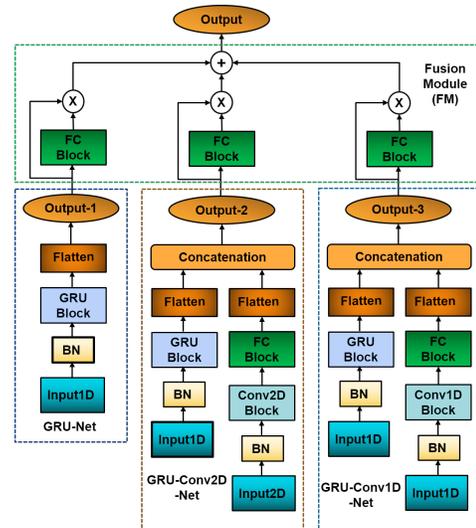
Fig. 2. Entire structure of Kinetics-FM-Net

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. In this paper, we combine the output of the three primary models by taking the average of the predictions. We train all three primary models simultaneously along with the final prediction from the combined model to create an

end-to-end trained model 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 perform an operation similar to standard normalization of the input data [47]. 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 Fully Connected (FC) block. The output from the FC 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 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 (Fig. 3) to create primary models are provided in this section:
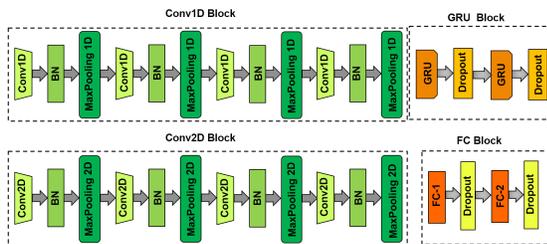

Fig. 3. Fundamental Blocks of Kinetics-FM-DLR-Ensemble-Net

- **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. 1) from GRU-Net, GRU-Conv2D-Net, GRU-Conv1D-Net are $\mathbb{K}_{\Delta T}^{GRU-Net} \in \mathbb{R}^{\Delta T \times D_K}$, $\mathbb{K}_{\Delta T}^{GRU-Conv2D-Net} \in \mathbb{R}^{\Delta T \times D_K}$, $\mathbb{K}_{\Delta T}^{GRU-Conv1D-Net} \in \mathbb{R}^{\Delta T \times D_K}$ 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 $FC(\mathbb{K}_{\Delta T}^{GRU-Net}) \in \mathbb{R}^{\Delta T \times D_K}$, $FC(\mathbb{K}_{\Delta T}^{GRU-Conv2D-Net}) \in \mathbb{R}^{\Delta T \times D_K}$, and $FC(\mathbb{K}_{\Delta T}^{GRU-Conv1D-Net}) \in \mathbb{R}^{\Delta T \times D_K}$ respectively, then output of the FM will be

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

   Where, $\odot$ is element-wise multiplication.

   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 [35], [38]. While evaluating the performance, both Normalized RMSE (NRMSE)– proportional to RMSE and Pearson Correlation Coefficient (PCC) were used [36]. Typically, when multiple loss functions are available, a single loss function is derived from the weighted sum of the 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 relationship between these loss functions. To use different characteristics of loss functions 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 (Fig. 1). 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 experimentally collected ground truth and prediction. As a result, from Kinetics-FM-Net(PCC), we will have similar profiles of joint moments and GRFs with ground truth but shifted and scaled (right Y-axis of Kinetic-FM-Net(PCC) in Fig. 4).

   As the Kinetics-FM-Net(PCC) is mainly focused on increasing PCC, this will result in a higher PCC than the Kinetics-FM-Net(RMSE). To maintain that high PCC while correcting the shifting and the scaling of the Kinetics-FM-Net(PCC) prediction, we need to acquire the actual offset and

range. In Fig. 4, the prediction of Kinetics-FM-Net(RMSE) has closely similar range and offset as the ground truth. As a result, we can leverage that range and offset information from the prediction of Kinetics-FM-Net(RMSE) to correct the shifted and scaled prediction of Kinetics-FM-Net(PCC). 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 a 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 Fig. 4, we demonstrate the qualitative and quantitative impact on our DLR loss design for performance improvement.
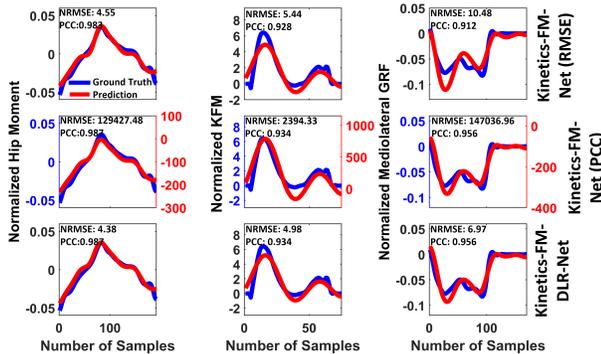

Fig. 4.   A sample plot of joint moments and GRF to demonstrate how DLR improves the performance from RMSE and PCC loss function.

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 $\mathbb{K}_{\Delta T,1}^{Bag}$, $\mathbb{K}_{\Delta T,2}^{Bag}$, ..., $\mathbb{K}_{\Delta T,K}^{Bag}$. Then, the final output of Kinetics-FM-DLR-Ensemble-Net:

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

5) State-Of-The-Art (SOTA) model description: For comparison with our proposed model Kinetics-FM-DLR-Ensemble-Net, we adopt multiple deep learning models from the literatures [28], [29], [31], [34], [38], [39]. Description of those models are given below:

a) **FFN (HF):** In the FeedForward Neural network with Handcrafted Features (FFN-HF) from input data, we extract 17 features for each of the axis (3D accelerometer and 3D gyroscope) of IMU data. We extract those features for the window length of 100 samples (0.5s) for Dataset A [43] and 50 samples (0.5s) for Dataset B [44]. The 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 a total of 102 features from each IMU. Then, we use those extracted features as the input to an FFN neural network with multiple 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.

b) **TCN:** For TCN, we use a single stack of the residual block with weight normalization. The TCN layer is followed by two dense layers and a dropout layer after each of them. We flatten the features from the last dropout layer and then connect it to the prediction layer.

c) **FFN:** For FeedForward Neural network(FFN), we use raw IMU data as input. Two dense layers with a dropout after each one is used. Features from the last dropout layer are flattened and connected to the prediction layer.

d) **Bi-LSTM:** We use two bidirectional LSTM layers, where each one is followed by a dropout layer. After the dropout layer, two dense layers, each one followed by a dropout, are used, flattened, and connected to the prediction layer.

e) **Conv2D:** We derive Conv2D architecture from GRU-Conv2D-Net by discarding the branch of GRU-Net (Fig. 2).

IV. EXPERIMENTS

A. Dataset Description

This paper uses two publicly available datasets [43], [44] to build Kinetics-FM-DLR-Ensemble-Net for the joint moment and GRFs estimation.

1) Dataset A [43]: Twenty subjects' data (8 females and 12 males, 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. 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 self-selected slow, normal, and fast speed, which includes both straight walking and turning. 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 (right leg). Four IMU sensors (thigh, shank,

TABLE I
HYPERPARAMETERS OF DIFFERENT LAYERS FOR DATASET A, DATASET B SEPARATED BY A COMMA.

Unit	GRU-Net	GRU-Conv2D-Net	GRU-Conv1D-Net	Fusion Module
GRU-1	Unit: 512, 512 Dropout: 0.1 (0.5), 0.3 (0.25)	Unit: 512, 512 Dropout: 0.1 (0.5), 0.3 (0.3)	Unit: 512, 512 Dropout: 0.1 (0.5), 0.3 (0.3)	NA
GRU-2	Unit: 256, 256 Dropout: 0.1 (0.5), 0.3 (0.25)	Unit: 256, 256 Dropout: 0.1 (0.5), 0.3 (0.3)	Unit: 256, 256 Dropout: 0.1 (0.5), 0.3 (0.3)	NA
Conv-olutional-1	NA	Filter: 256, 256 Kernel: 5X3, 5X3 Pool: 2X2, 2X2 Strides: 1X1, 1X1	Filter: 256, 256 Kernel: 3, 3 Pool: 2, 2 Strides: 1, 1	NA
Conv-olutional-2	NA	Filter: 256, 256 Kernel: 5X3, 5X3 Pool: 2X2, 2X2 Strides: 1X1, 1X1	Filter: 256, 256 Kernel: 3, 3 Pool: 2, 2 Strides: 1, 1	NA
Conv-olutional-3	NA	Filter: 512, 512 Kernel: 5X3, 5X3 Pool: 2X1, 2X1 Strides: 1X1, 1X1	Filter: 512, 512 Kernel: 3, 3 Pool: 2, 2 Strides: 1, 1	NA
Conv-olutional-4	NA	Filter: 512, 512 Kernel: 5X3, 5X3 Pool: 2X1, 2X1 Strides: 1X1, 1X1	Filter: 512, 512 Kernel: 3, 3 Pool: 2, 2 Strides: 1, 1	NA
FC-1	NA	Unit: 64, 64 Dropout: 0.25, 0.05 (NA)	Unit: 64, 64 Dropout: 0.25, NA(NA)	Unit: 128, 128 Dropout: NA (0.4), NA (0.4)
FC-2	NA	Unit: 32, 32 Dropout: 0.25, 0.05 (NA)	Unit: 32, 32 Dropout: 0.25, NA (NA)	Unit: 7, 5 Dropout: NA (0.4), NA (0.4)

The value of dropout in the first Parentheses is for the model trained with PCC as a loss function. Other parameters are same for all three loss designs (RMSE, JL, DLR).

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 (4 in, 5 in, 6 in, and 7 in). Sixty ramp trials were performed with six different inclination angles (5.2°, 7.8°, 9.2°, 11°, 12.4°, and 18°) with a set of five trials for each starting leg (left and right) on each inclination. Other detailed information on the protocol of the experiment is described in [43].

Dataset Pre-processing: IMU data and marker trajectory data from the motion capture system are collected at 200 Hz. GRFs data are 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 the Opensim [12] 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 entire gait cycles. We normalize joint moment data with the body weight and height of the participants. We also normalize GRF data with the 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 use 100 samples (0.5 s) for feature extraction of the deep learning model, as it provides satisfactory results for kinetics estimation.

Details of the segmentation of force plates for level-ground, ramp, and stair conditions are given below. Force plate number and their position in the laboratory can be found in [43].

a) **level-ground:** For clockwise level-ground walking, we segment FP2, Combined, and FP5 force plate data, and for anti-clockwise level-ground walking, we segment FP5 data.

b) **Stair:** For stair trials starting with the instrumented leg (right leg), we segment force plate data of FP1 and FP5 for the right leg. For the trials starting with the non-instrumented leg (left leg), we segment FP2, FP3, and FP4.

c) **Ramp:** For ramp trials starting with the non-instrumented leg (left leg), we segment force plate data of FP2 and FP4 for the right leg. For the trials starting with the instrumented leg (right leg), we segment FP1, FP3, and FP5.

2) Dataset B [44]: Seventeen subjects (all male; age: 23.2 ± 1.1 years; height: 1.76 ± 0.06 m; mass: 67.3 ± 8.3 Kg) participated in that study. Different walking speeds, foot progression angles, step widths, and trunk sway angles were used as testing conditions. Subjects initially performed a 2-minute normal walking trial with a self-selected speed (1.16 ± 0.04 m/s), which was used as a baseline to determine foot progression angle and step width. Three foot progression angles (baseline–15°, baseline, and baseline+15°) with three different speeds (self-selected–0.2 m/s, self-selected, and self-selected+0.2 m/s) were used in combinations to perform nine trials. In the same way, three step widths (baseline–0.054 m, baseline, and baseline+0.070 m) and trunk sway angles (4°, 8°, and 12°) trials with three different speeds were also performed.

Dataset Pre-processing: Infrared-based optical motion capture data was recorded at 100 Hz with synchronized ground reaction force data at 1000 Hz. Eight IMU data were collected with a sampling rate of 100 Hz. In their pre-processed dataset, they have provided 3D GRFs with the same number of samples as IMU and joint moments. As they have only provided KAM, KFM, and 3D GRFs, we use these variables to build our algorithm. In the dataset, zeros are appended at the end of shorter steps to synchronize with the longest steps with a length of 152. We remove those extra zeros from the shorter steps and use the rest of the dataset to build our algorithm.

B. 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 (two hours per bootstrap sample) for Dataset A and of 4 hours (24 minutes per bootstrap sample) for Dataset B. Adam [48] is used as the optimizer. All the models with RMSE loss are run for 40 epochs, and models with PCC loss are run for 55 epochs with early

TABLE II

MEAN AND STANDARD DEVIATION OF NRMSE AND PCC VALUES OF JOINT MOMENTS AND 3D GRFs FOR DIFFERENT MODELS WITH THREE DIFFERENT LOSS DESIGN APPROACHES (RMSE, JL, DLR) IN ENTIRE WALKING CONDITIONS FOR **DATASET A**.

Models	NRMSE			PCC		
	RMSE	JL	DLR	RMSE	JL	DLR
GRU-Net	$5.18 \pm 0.97^*$	$5.26 \pm 0.78^*$	$4.78 \pm 1.08^*$	$0.897 \pm 0.039^*$	$0.896 \pm 0.030^*$	$0.916 \pm 0.043^*$
Conv2D-Net	$5.74 \pm 0.95^*$	$5.75 \pm 0.92^*$	$5.46 \pm 0.97^*$	$0.860 \pm 0.027^*$	$0.861 \pm 0.029^*$	$0.890 \pm 0.028^*$
Conv1D-Net	$5.72 \pm 0.85^*$	$5.70 \pm 0.89^*$	$5.37 \pm 1.04^*$	$0.856 \pm 0.027^*$	$0.859 \pm 0.028^*$	$0.890 \pm 0.027^*$
GRU-Conv2D-Net	$5.29 \pm 1.08^*$	$5.32 \pm 0.79^*$	$4.72 \pm 1.08^*$	$0.891 \pm 0.040^*$	$0.892 \pm 0.029^*$	$0.918 \pm 0.043^*$
GRU-Conv1D-Net	$5.18 \pm 0.93^*$	$5.16 \pm 0.84^*$	$4.78 \pm 0.99^*$	$0.898 \pm 0.036^*$	$0.895 \pm 0.028^*$	$0.917 \pm 0.040^*$
Conv2D-Conv1D-Net	$5.67 \pm 0.91^*$	$5.28 \pm 0.88^*$	$5.19 \pm 0.88^*$	$0.864 \pm 0.028^*$	$0.895 \pm 0.028^*$	$0.906 \pm 0.040^*$
Kinetics-Sub-Net-1	$4.91 \pm 0.86^*$	$5.31 \pm 0.84^*$	$4.68 \pm 1.40^*$	$0.909 \pm 0.032^*$	$0.893 \pm 0.030^*$	$0.919 \pm 0.052^*$
Kinetics-Sub-Net-2	$4.99 \pm 0.87^*$	$4.98 \pm 0.76^*$	$4.71 \pm 0.81^*$	$0.905 \pm 0.040^*$	$0.905 \pm 0.029^*$	$0.919 \pm 0.034^*$
Kinetics-Sub-Net-3	$5.02 \pm 0.94^*$	$4.99 \pm 0.83^*$	$4.68 \pm 0.93^*$	$0.906 \pm 0.036^*$	$0.904 \pm 0.029^*$	$0.921 \pm 0.036^*$
Kinetics-Net	$4.88 \pm 0.90^*$	$4.96 \pm 0.82^*$	$4.64 \pm 1.07^*$	$0.910 \pm 0.034^*$	$0.906 \pm 0.029^*$	0.922 ± 0.041
Kinetics-FM-Net	**$4.72 \pm 0.92^*$**	**$4.70 \pm 0.92^*$**	**4.49 ± 0.82**	**$0.915 \pm 0.033^*$**	**$0.915 \pm 0.031^*$**	**0.923 ± 0.030**

Bold numbers represent the result of Kinetics-FM-Net which has the best performance in NRMSE and PCC. * denotes a significant difference of NRMSE and PCC between Kinetic-FM-DLR-Net with the rest of the models.

TABLE III

HYPERPARAMETERS OF DIFFERENT LAYERS OF SOTA MODELS

Model	Hyperparameters	Dataset A	Dataset B
FFN-HF	Number of Layers	5	3
	Units	1024, 512, 256, 128, 64	512, 256, 128
	Dropout rate	0.05, 0.05, 0.05, 0.05, 0.05	0.05, 0.05, 0.05
TCN	Kernel Size	3	3
	Filter	128	64
	Dilations	1, 2, 4, 8, 16	1, 2, 4, 8
	Dropout rate	0.05	0.05
	Number of stacks	1	1
	Number of Dense layer	2	2
	Unit	64, 32	64, 32
	Dropout rate	0.10, 0.10	0.05, 0.05
FFN	Number of Layers	2	2
	Units	128, 64	128, 64
	Dropout rate	0.10, 0.10	0.05, 0.05
Bi-LSTM	Number of Bi-LSTM layers	2	2
	Units	128, 64	128, 64
	Dropout rate	0.50, 0.50	0.50, 0.50
	Number of Dense layer	2	2
	Unit	128, 64	128, 64
	Dropout rate	0.25, 0.25	0.10, 0.10

stopping callbacks (patience of 15 epochs). The best model's weights are restored based on the best values (minimum loss) of patience epochs (15) on validation data. All the models are trained 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.

$$\mathcal{L}_{JL} = \mathcal{L}_{RMSE} + W \times \mathcal{L}_{PCC} \quad (6)$$

In Table I, III, we present all the hyperparameters of different layers of Kinetics-FM-DLR-Net and SOTA models respectively.

C. Evaluation Procedures

Leave-one-subject out cross validation is implemented to assess the performance of all the models. Excluding the test subject from the 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. We conduct Repeated-Measures Analysis of Variance (ANOVA) with a least significant difference correction post hoc test for NRMSE and PCC separately with a *p-value* less than 0.05 to determine if our proposed method significantly improves the kinetics estimation compared to other methods.

V. RESULTS

A. Model Ablation

In Table II and IV, we show the ablation study on our model and report the NRMSE and PCC values for kinetics estimation with different loss designs. 11 models trained with three loss designs result in a total of 33 models. We perform statistical analysis to validate whether our final proposed model Kinetics-FM-DLR-Net (Kinetics-FM-Net with DLR loss design) significantly improves kinetics estimation accuracy compared to the rest of the 32 models (Table II and IV). For both datasets and evaluation metrics (NRMSE, PCC), overall Kinetics-FM-DLR-Net significantly outperforms rest of the models.

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 and IV demonstrates that integrating the FM in Kinetics-Net decreases the mean NRMSE and PCC for both datasets.

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 and IV). Although we initially design DLR to increase PCC of kinetics prediction, we have 'Double Reward' with DLR as it improves both NRMSE and PCC.

D. Sensor Combination

We use different combinations of sensors and apply Kinetics-FM-DLR-Net to see how performance changes due to different sensor locations. We implement statistical analysis to

TABLE IV

MEAN AND STANDARD DEVIATION OF NRMSE AND PCC VALUES OF JOINT MOMENTS AND 3D GRFs FOR DIFFERENT MODELS WITH THREE DIFFERENT LOSS DESIGN APPROACHES (RMSE, JL, DLR) IN ENTIRE WALKING CONDITIONS FOR **DATASET B**.

Models	NRMSE			PCC		
	RMSE	JL	DLR	RMSE	JL	DLR
GRU-Net	$8.14 \pm 0.75^*$	$8.08 \pm 0.71^*$	$8.05 \pm 0.71^*$	$0.864 \pm 0.031^*$	$0.865 \pm 0.029^*$	$0.866 \pm 0.032^*$
Conv2D-Net	$8.16 \pm 0.64^*$	$8.20 \pm 0.72^*$	$8.09 \pm 0.61^*$	$0.862 \pm 0.031^*$	$0.861 \pm 0.031^*$	$0.865 \pm 0.033^*$
Conv1D-Net	$7.98 \pm 0.66^*$	$8.02 \pm 0.72^*$	$7.96 \pm 0.66^*$	$0.868 \pm 0.029^*$	$0.869 \pm 0.029^*$	$0.869 \pm 0.031^*$
GRU-Conv2D-Net	$7.88 \pm 0.70^*$	$7.84 \pm 0.62^*$	$7.80 \pm 0.68^*$	$0.869 \pm 0.031^*$	$0.869 \pm 0.030^*$	$0.872 \pm 0.030^*$
GRU-Conv1D-Net	$7.92 \pm 0.75^*$	$7.87 \pm 0.70^*$	$7.87 \pm 0.69^*$	$0.871 \pm 0.030^*$	$0.870 \pm 0.030^*$	$0.874 \pm 0.029^*$
Conv2D-Conv1D-Net	$8.01 \pm 0.75^*$	$7.93 \pm 0.74^*$	$7.94 \pm 0.73^*$	$0.869 \pm 0.030^*$	$0.870 \pm 0.029^*$	$0.871 \pm 0.029^*$
Kinetics-Sub-Net-1	$7.75 \pm 0.68^*$	$7.74 \pm 0.70^*$	$7.70 \pm 0.72^*$	$0.874 \pm 0.029^*$	$0.875 \pm 0.029^*$	$0.876 \pm 0.031^*$
Kinetics-Sub-Net-2	$7.80 \pm 0.65^*$	$7.79 \pm 0.69^*$	$7.75 \pm 0.67^*$	$0.874 \pm 0.029^*$	$0.875 \pm 0.028^*$	0.877 ± 0.029
Kinetics-Sub-Net-3	$7.71 \pm 0.67^*$	$7.71 \pm 0.74^*$	$7.69 \pm 0.64^*$	$0.877 \pm 0.028^*$	$0.876 \pm 0.029^*$	$0.878 \pm 0.029^*$
Kinetics-Net	$7.69 \pm 0.65^*$	$7.79 \pm 0.72^*$	7.62 ± 0.67	$0.877 \pm 0.029^*$	$0.877 \pm 0.028^*$	$0.881 \pm 0.029^*$
Kinetics-FM-Net	$\mathbf{7.60 \pm 0.73}^*$	$\mathbf{7.64 \pm 0.71}$	$\mathbf{7.54 \pm 0.70}$	$\mathbf{0.881 \pm 0.029}^*$	$\mathbf{0.881 \pm 0.029}^*$	$\mathbf{0.884 \pm 0.029}$

Bold numbers represent the result of Kinetics-FM-Net that has the best performance in NRMSE and PCC. * denotes a significant difference of NRMSE and PCC between Kinetic-FM-DLR-Net with the rest of the models.

TABLE V

EFFECT OF VARYING NUMBER OF BOOTSTRAP SAMPLE ON THE MEAN AND STANDARD DEVIATION OF NRMSE FOR KINETICS ESTIMATION.

No	Dataset A		Dataset B	
	NRMSE	PCC	NRMSE	PCC
1	4.55 ± 0.84	0.922 ± 0.032	7.54 ± 0.70	0.882 ± 0.029
2	4.42 ± 0.82	0.925 ± 0.031	7.50 ± 0.71	0.884 ± 0.030
3	4.38 ± 0.82	$0.927 \pm 0.030^*$	7.49 ± 0.72	$0.885 \pm 0.030^*$
4	$4.35 \pm 0.82^*$	$0.927 \pm 0.030^*$	$7.46 \pm 0.72^*$	$0.886 \pm 0.030^*$
5	$4.33 \pm 0.82^*$	$0.928 \pm 0.029^*$	$7.45 \pm 0.71^*$	$0.886 \pm 0.030^*$
6	$4.32 \pm 0.82^*$	$0.928 \pm 0.029^*$	$7.44 \pm 0.71^*$	$0.886 \pm 0.030^*$
7	$4.32 \pm 0.81^*$	$0.928 \pm 0.029^*$	$7.44 \pm 0.71^*$	$0.886 \pm 0.030^*$
8	$4.32 \pm 0.81^*$	$0.928 \pm 0.029^*$	$7.44 \pm 0.71^*$	$0.886 \pm 0.030^*$
9	$4.32 \pm 0.82^*$	$0.928 \pm 0.029^*$	$7.43 \pm 0.71^*$	$0.886 \pm 0.030^*$
10	$\mathbf{4.32 \pm 0.81}^*$	$\mathbf{0.929 \pm 0.029}^*$	$\mathbf{7.43 \pm 0.71}^*$	$\mathbf{0.886 \pm 0.030}^*$

Bold numbers represent the result of Kinetics-FM-Net which has the best performance in NRMSE and PCC. * denotes a significant difference of NRMSE and PCC between Kinetic-FM-DLR-Net with rest of the models.

TABLE VI

EFFECT OF VARYING NUMBER OF SENSORS ON THE MEAN AND STANDARD DEVIATION OF NRMSE FOR KINETICS ESTIMATION.

Sensor	Dataset A		Dataset B	
	NRMSE	PCC	NRMSE	PCC
Foot	$5.06 \pm 1.06^*$	$0.904 \pm 0.036^*$	$7.93 \pm 0.59^*$	$0.867 \pm 0.034^*$
Shank	$4.65 \pm 0.82^*$	0.920 ± 0.029	7.66 ± 0.63	$0.876 \pm 0.033^*$
Thigh	$4.86 \pm 0.92^*$	$0.913 \pm 0.028^*$	$7.82 \pm 0.78^*$	$0.872 \pm 0.032^*$
Foot-Shank	4.56 ± 0.78	0.922 ± 0.029	7.55 ± 0.71	$0.880 \pm 0.030^*$
Shank-Thigh	4.50 ± 0.85	0.922 ± 0.027	7.59 ± 0.73	$0.880 \pm 0.031^*$
Foot-Thigh	4.61 ± 0.93	$0.919 \pm 0.030^*$	7.65 ± 0.83	$0.880 \pm 0.030^*$
Shank-Foot-Thigh	$\mathbf{4.49 \pm 0.82}$	$\mathbf{0.923 \pm 0.030}$	$\mathbf{7.54 \pm 0.70}$	$\mathbf{0.884 \pm 0.029}$

Bold numbers represent the result of Kinetics-FM-Net which has the best performance in NRMSE and PCC. * denotes a significant difference of NRMSE and PCC between Shank-Foot-Thigh with the rest of the sensor combinations.

see if the combination of all three sensors (Shank-Foot-Thigh) can significantly improve the result compared to other sensor combinations. Compared to a single sensor, the Shank-Foot-Thigh sensor significantly improves the kinetics prediction. We have a marginal improvement compared to the combination of the two sensors, which is statistically insignificant. However, as Shank-Foot-Thigh sensors have the lowest NRMSE and highest PCC, we proceed with the combination of the Shank-Foot-Thigh sensors for the implementation of bagging.

E. Bagging (Ensemble Learning)

Kinetics-FM-DLR-Ensemble-Net, created from Kinetics-FM-DLR-Net with the application of bagging techniques, reduces mean NRMSE from 4.49 to 4.32 for Dataset A, from 7.54 to 7.43 for Dataset B and increases mean PCC from 0.923 to 0.929 for Dataset A, from 0.884 to 0.886 for Dataset B compared with Kinetics-FM-DLR-Net. Table V demonstrates the performance of adding different numbers of bagging samples to the Kinetics-FM-DLR-Net. Statistical analysis is performed to validate in which bootstrap sample number we achieve significant performance improvement over Kinetic-FM-DLR-Net. After adding four bootstrap samples in bagging techniques, there are significant improvements in both datasets. We continue to add samples of up to ten and eventually stopped the experiment due to increasing complexity and saturation in performance.

F. Parameter Sweep (W)

Typically, when we combine two loss functions to create joint loss, the weight of the loss is varied to ensure the best performance gain. In Table II and IV, 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) and summarize the results in Table VII. The rationale of choosing the specific value of W is to make the actual value of $\mathcal{L}_{PCC}$ to be one-fifth, one-fourth, one-third, half, equal, two-times, three-times, four-times, and five times of $\mathcal{L}_{RMSE}$, which will ensure a wide range of weight on $\mathcal{L}_{PCC}$ to make sure valid comparison of joint loss. We perform statistical analysis to determine if our model Kinetics-FM-DLR-Net significantly performs better than Kinetics-FM-Net with weighted joint loss. For Dataset A, we do not achieve significant improvement when $W = 8$ and $W = 4$ for PCC. For the rest of the cases of Dataset A including Dataset B, our model significantly outperforms the model with different joint losses.

G. Comparison with State-Of-The-Art

In Table VIII, we show the comparison of our results with the state-of-the-art deep learning algorithm for kinetics estimation. For Dataset A, we have an improvement of up to 24.74% in NRMSE and 8% in PCC, and for Dataset B, an improvement of up to 11.54% in NRMSE and 3.38% in PCC are observed.

Table IX and X show the prediction outcomes of Kinetics-FM-DLR-Ensemble-Net in each walking condition for Dataset A and Dataset B respectively. As an example, we plot one gait

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3262164

AUTHOR *sanzid et al.*: ESTIMATION OF LOWER EXTREMITY JOINT MOMENTS AND 3D GROUND REACTION FORCES USING IMU SENSORS IN MULTIPLE WALKING CONDITIONS: A DEEP LEARNING APPROACH
9

TABLE VII
MEAN AND STANDARD DEVIATION OF ENTIRE WALKING NRMSE RESULTS FOR PARAMETER SWEEP (W) OF $\mathcal{L}_{JL}$ FOR KINETICS-FM-DLR-NET.

Dataset A			Dataset B		
W	NRMSE	PCC	W	NRMSE	PCC
4/5	4.71 ± 0.86*	0.914 ± 0.032*	3	8.40 ± 0.80*	0.858 ± 0.034*
1	4.72 ± 0.82*	0.915 ± 0.032*	15/4	8.34 ± 0.79*	0.857 ± 0.030*
4/3	4.72 ± 0.78*	0.914 ± 0.031*	5	8.30 ± 0.66*	0.858 ± 0.030*
2	4.67 ± 0.75*	0.916 ± 0.031*	15/2	8.21 ± 0.77*	0.861 ± 0.033*
4	4.68 ± 0.83	0.916 ± 0.033*	15	8.20 ± 0.72*	0.862 ± 0.035*
8	4.65 ± 0.90	0.916 ± 0.031	30	8.24 ± 0.88*	0.863 ± 0.033*
12	4.67 ± 0.84*	0.917 ± 0.031*	45	8.27 ± 0.73*	0.864 ± 0.029*
16	4.74 ± 0.86*	0.915 ± 0.032*	60	8.21 ± 0.85*	0.866 ± 0.031*
20	4.72 ± 0.81*	0.917 ± 0.031*	75	8.16 ± 0.79*	0.867 ± 0.032*

Bold numbers represent the result of Kinetics-FM-Net which has the best performance in NRMSE and PCC. * denotes a significant difference of NRMSE and PCC between Kinetic-FM-DLR-Net with the rest of the models.

TABLE VIII
COMPARISON OF THE KINETICS ESTIMATION PERFORMANCE OF OUR ALGORITHM WITH STATE-OF-THE-ART DEEP LEARNING METHOD OF KINETICS ESTIMATION.

Model	Dataset A		Dataset B	
	NRMSE	PCC	NRMSE	PCC
FFN (HF) [29]	5.59 ± 0.85*	0.873 ± 0.029*	8.40 ± 0.80*	0.858 ± 0.034*
TCN [28]	5.29 ± 0.98*	0.892 ± 0.032*	8.34 ± 0.79*	0.857 ± 0.030*
FFN [39], [31]	5.38 ± 0.89*	0.888 ± 0.034*	8.30 ± 0.66*	0.858 ± 0.030*
Bi-LSTM [31], [34]	5.05 ± 0.94*	0.901 ± 0.033*	8.21 ± 0.77*	0.861 ± 0.033*
Conv2D [38]	5.74 ± 0.95*	0.860 ± 0.027*	8.20 ± 0.72*	0.862 ± 0.035*
Kinetics-FM-DLR Ensemble-Net	**4.32 ± 0.81**	**0.929 ± 0.029**	**7.43 ± 0.71**	**0.886 ± 0.03**

Bold numbers represent the result of Kinetics-FM-Net which has the best performance in NRMSE and PCC. * denotes a significant difference of NRMSE and PCC between Kinetic-FM-DLR-Ensemble-Net with the rest of the models.

cycle of each kinetics component for Dataset A and stance phase of each kinetics component of Dataset B in each walking condition to demonstrate a sample qualitative comparison of ground truth and prediction from the model (Fig. 5).

VI. DISCUSSION

This study estimates lower extremity joint moments as well as anterior-posterior, vertical, and mediolateral 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 to implement GRFs and joint moments estimation on multiple walking conditions using IMU sensors via deep learning. We apply our algorithm to two datasets; Dataset A [43] with a good number of subjects, multiple walking environments, multiple walking speeds, multi-variable ramp angles, and stair heights, and Dataset B [44] with different treadmill conditions. 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. The extensive training sets, generalized and versatile capacity of our novel algorithm, and rigorous validations enable the most accurate estimates of joint moments and GRFs for new testing subjects compared with the other state-of-the-art deep learning algorithms. In addition to this, as we establish and evaluate our model on the public datasets, other researchers can 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 and IV, we demonstrate extensive model ablation with different loss function designs to validate the use of different layers in our model in two publicly available datasets. 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 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 and IV, 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 an end-to-end model Kinetics-Net, which outperforms all the previous models. We further add FM and DLR modules, which significantly improve the kinetics estimation compared to other 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 [28]–[37], [37]–[40], 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 our algorithm to make the comparisons. To address this issue, we adopt their machine-learning algorithms and applied them to the Dataset A and Dataset B. 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 the hyperparameters of those models manually for our dataset to achieve 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 for kinetics with ours. Our algorithm Kinetics-FM-DLR-Ensemble-Net significantly outperforms those algorithms by a large margin, which proves the effectiveness of our model over the conventional deep learning method for kinetics estimation.

From Table IX, the performance of the level-ground condition is inferior to the other walking condition. The main reason for this is the lack of training data for the level-ground condition. As we segment the dataset of level-ground conditions, we have limited numbers of right leg strikes from the walking circuit, which results in fewer data (4% of the total dataset) compared to other walking environments. Prediction from the treadmill trials is highly correlated with the ground truth with a mean PCC of 0.945 compared to trials of other

TABLE IX

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 FOR **DATASET A**.

Mertric	Scene	Hip Flexion Moment	Hip Abduction Moment	Knee Moment	Ankle Moment	Medio-lateral GRF	Vertical GRF	Anterior-Posterior GRF	Mean
NRMSE	Treadmill	3.67 ± 1.40	6.03 ± 2.28	3.25 ± 1.31	3.20 ± 2.01	4.12 ± 1.94	4.42 ± 2.91	2.86 ± 1.01	6.23 ± 1.08
	Level-ground	6.06 ± 2.27	8.16 ± 2.70	5.48 ± 1.97	5.59 ± 2.49	8.65 ± 1.34	5.49 ± 4.36	3.77 ± 1.30	8.52 ± 1.57
	Ramp	5.64 ± 1.86	5.69 ± 2.09	5.05 ± 2.05	4.35 ± 2.46	3.12 ± 0.99	4.34 ± 2.81	3.46 ± 0.94	5.69 ± 1.22
	Stair	4.06 ± 1.43	4.51 ± 1.55	4.35 ± 1.86	3.96 ± 2.70	3.07 ± 0.88	4.05 ± 1.88	2.56 ± 0.43	6.95 ± 1.03
PCC	Treadmill	0.948 ± 0.031	0.925 ± 0.046	0.917 ± 0.040	0.966 ± 0.038	0.917 ± 0.044	0.977 ± 0.037	0.967 ± 0.031	0.945 ± 0.032
	Level-ground	0.765 ± 0.177	0.810 ± 0.131	0.718 ± 0.191	0.868 ± 0.117	0.826 ± 0.095	0.957 ± 0.088	0.922 ± 0.078	0.838 ± 0.110
	Ramp	0.895 ± 0.051	0.899 ± 0.061	0.912 ± 0.054	0.930 ± 0.061	0.856 ± 0.061	0.979 ± 0.042	0.952 ± 0.040	0.918 ± 0.043
	Stair	0.870 ± 0.075	0.922 ± 0.038	0.928 ± 0.046	0.944 ± 0.042	0.884 ± 0.043	0.984 ± 0.024	0.927 ± 0.029	0.923 ± 0.032

TABLE X

MEAN AND STANDARD DEVIATION OF NRMSE AND PCC VALUES OF JOINT MOMENTS AND 3D GRFs OF TREADMILL WALKING FOR FINAL PROPOSED MODEL KINETICS-FM-DLR-ENSEMBLE-NET FOR **DATASET B**.

Mertric	KFM	KAM	Medio-lateral GRF	Vertical GRF	Anterior-Posterior GRF	Mean
NRMSE	7.11 ± 1.47	8.27 ± 1.72	7.49 ± 1.09	7.86 ± 1.1	6.42 ± 0.5	7.43 ± 0.71
PCC	0.865 ± 0.09	0.805 ± 0.055	0.885 ± 0.02	0.959 ± 0.011	0.917 ± 0.014	0.886 ± 0.03

scenes, which is due to the collection of data in a repetitive manner with strictly controlled walking speeds for each of the participants.

We use different configurations of sensor placement in our model to identify how kinetics estimation performance varies. We find that IMU on the shank has a better performance compared to the sensor on thigh and foot location on both datasets when a single sensor is used for algorithm construction. One plausible explanation for inaccuracy while using only the foot sensor is due to the lack of movement of the foot (resulting in a lack of IMU signal) during the stance phase when the foot is firmly constrained to the ground. In addition to this, the thigh segment may have greater IMU signal noise due to the compliant surface and movement of muscle whereas the frontal area of shank may have less noise due to the tibia's boney surface. Moreover, we achieve more accurate results when the IMU sensor on the shank is combined with the thigh and foot sensors for both datasets, which implies that our model is able to perform effectively with multiple sensor configurations. Overall, there are no significant differences between Shank-Foot-Thigh and Foot-Shank, Shank-Thigh. This will help us to reduce the number of sensors without sacrificing performance for kinetics estimation.

Although we provide the most accurate prediction of joint moments and GRFs with extensive walking conditions, speeds, good number of subjects, and two independent public datasets, there are several limitations on this study. We use a relatively large number of IMU sensors (three) to estimate single limb 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 or foot mounted sensors similar to kinematics estimation in [49] or using insoles such as in [50], it would be more helpful to maintain those sensors. Since single limb joint moments and GRFs are affected by the contralateral limb during the 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. Additionally, adding six sensors on both feet would help to acquire both limbs' kinetics information. 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.

As the level-ground condition has the highest error compared to other conditions, a direction for future improvement may be to augment the level-ground dataset by repetition and make it roughly equal to other walking environments. Although we have decent accuracy for healthy individuals, patients with musculoskeletal issues may not have an accurate outcome from the algorithm due to the absence of patients' training data. Moreover, our proposed approach relies only on IMU data while ignoring static and anthropometric properties of the individual. As this can provide essential information for representing the dynamic model of human kinetics, fusing this information into the model has the potential to improve the performance of the prediction. Thus, a probable future direction to address the limitation is to fuse deep learning-based features of static variables such as segment length, mass matrix of the segments, etc. to the model using any dense fully connected layers. IMU placement errors such as orientation and position errors can affect the accuracy of moment estimation [44] or significantly reduce GRF accuracy [51] when machine learning based data-driven methods are used. Our model may suffer the same performance degradation due to sensor placement errors. A future research direction to tackle the problem is to train the model with a simulated dataset of different placement and orientations of the IMU data with corresponding kinetics information.

As we use extensive datasets with our novel algorithm, our outcome provides reliable estimates of joint moment and ground reaction forces compared with state-of-the-art studies. Thus, our algorithm can be used to measure the joint kinetics and 3D ground reaction forces in the clinic or other research lab where they cannot accommodate prohibitive
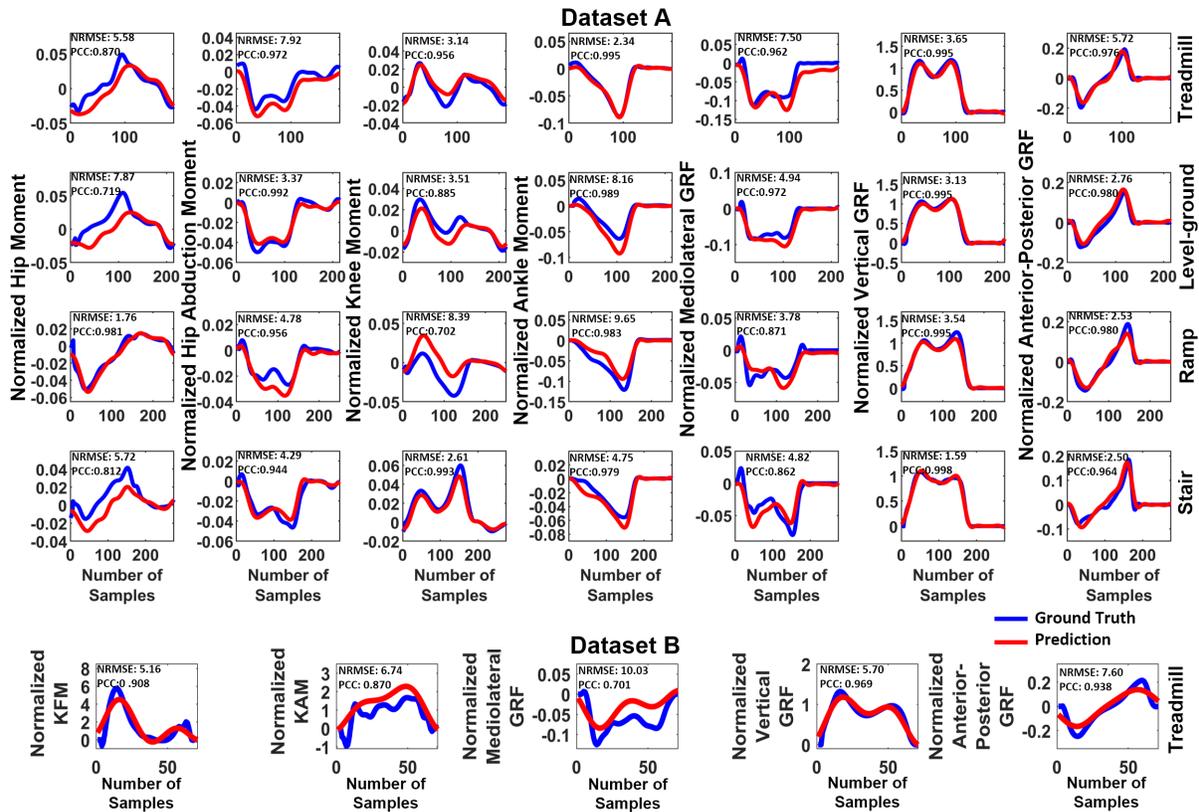
This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3262164

AUTHOR *sanzid et al.*: ESTIMATION OF LOWER EXTREMITY JOINT MOMENTS AND 3D GROUND REACTION FORCES USING IMU SENSORS IN MULTIPLE WALKING CONDITIONS: A DEEP LEARNING APPROACH   11

Fig. 5.   A sample plot for a single gait cycle for different walking conditions along with the corresponding NRMSE and PCC of subject AB25 in Dataset A and sample plot of stance phase for kinetics components in Dataset B. Blue lines represent experimentally collected and normalized joint moments and GRFs. Red lines present estimated results using Kinetic-FM-DLR-Ensemble-Net.

measurement modalities such as motion capture cameras, and overground force plates. To guarantee reliable performance, users need to set up the IMU with the protocol mentioned in [43], [44]. Our model also has the potential to serve as a platform that enables updates or modifications via retraining of the model to accommodate kinetics estimation of different subject populations who have deviated kinetics due to musculoskeletal issues or aging in the future. Our model can be found at Github (https://github.com/Sanzid-Priam/Estimation-of-Kinetics-using-Kinetics-FM-DLR-Ensemble-Net).

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] 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.

[2] 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.

[3] A. Chang, K. Hayes, D. Dunlop, J. Song, D. Hurwitz, S. Cahue, and L. Sharma, "Hip abduction moment and protection against medial tibiofemoral osteoarthritis progression," *Arthritis & Rheumatism*, vol. 52, no. 11, pp. 3515–3519, 2005.

[4] 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.

[5] J. P. Walter, D. D. D'Lima, C. W. Colwell Jr, and B. J. Fregly, "Decreased knee adduction moment does not guarantee decreased medial contact force during gait," *Journal of orthopaedic research*, vol. 28, no. 10, pp. 1348–1354, 2010.

[6] T. P. Andriacchi and A. Mündermann, "The role of ambulatory mechanics in the initiation and progression of knee osteoarthritis," *Current opinion in rheumatology*, vol. 18, no. 5, pp. 514–518, 2006.

[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. Slijepcevic, F. Horst, S. Lapuschkin, B. Horsak, A.-M. Raberger, A. Kranzl, W. Samek, C. Breiteneder, W. I. Schöllhorn, and M. Zeppelzauer, "Explaining machine learning models for clinical gait analysis," *ACM Transactions on Computing for Healthcare (HEALTH)*, vol. 3, no. 2, pp. 1–27, 2021.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] T. S. Buchanan, D. G. Lloyd, K. Manal, and T. F. Besier, "Neuromuscoskeletal 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

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

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[43] 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.

[44] T. Tan, D. Wang, P. B. Shull, and E. Halilaj, "Imu and smartphone camera fusion for knee adduction and knee flexion moment estimation during walking," *IEEE Transactions on Industrial Informatics*, 2022.

[45] 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.

[46] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, p. 123–140, Aug. 1996. [Online]. Available: https://doi.org/10.1023/A:1018054314350

[47] 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

[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[49] 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.

[50] M. Iber, B. Dumphart, V.-A. de Jesus Oliveira, S. Ferstl, J. M. Reis, D. Slijepčević, M. Heller, A.-M. Raberger, and B. Horsak, "Mind the steps: Towards auditory feedback in tele-rehabilitation based on automated gait classification," in *Audio Mostly 2021*, 2021, pp. 139–146.

[51] T. Tan, D. P. Chiasson, H. Hu, and P. B. Shull, "Influence of imu position and orientation placement errors on ground reaction force estimation," *Journal of biomechanics*, vol. 97, p. 109416, 2019.