Ankle Kinematics Estimation using Artificial Neural Network and Multimodal IMU Data

Lefan Wang, Pingfan Song, Thomas Stone, Adrian Weller, and Sebastian Pattinson

***Abstract*—Inertial measurement units (IMUs) have become attractive for monitoring joint kinematics due to their portability and versatility. However, their limited accuracy, inability to analyze data in real-time, and complex data fusion algorithms requiring precise sensor-to-segment calibrations hinder their clinical and daily use. This paper introduces KEEN (KinEmatics Estimation Network), an innovative framework that exploits lightweight artificial neural networks (ANNs) to provide real-time predictions of multi-plane ankle kinematics using a minimal number of IMUs, without calibration requirements. Five ANN algorithms were developed and evaluated using 42 inputs derived from four IMUs in both intra-subject and inter-subject tasks. Extensive experimental results yielded exciting findings: even a single IMU located at the heel can provide clinically acceptable estimations of ankle kinematics, implying significant potential for cost and energy savings. Statistical analysis demonstrated the superiority of the developed Long Short-Term Memory (LSTM) network over the other models in intra-subject tasks, achieving impressive accuracy (RMSE: 1.88°±0.02°, MAE: 1.41°±0.01°, and r2 score: 0.93±0.01), indicating strong generalization within the same subject. In inter-subject tasks, the convolutional neural network (CNN) and the CNN-LSTM models showed comparable performance but statistically outperformed the other models in terms of estimation accuracy across various inputs. When using a single IMU, the CNN model achieved the lowest error (RMSE: 4.13°±0.55°, MAE: 3.33°±0.48°, and r2 score: 0.50±0.21), showcasing its effective generalization to new subjects. Furthermore, deploying the CNN into a microcontroller, with a sinlge IMU at the heel, resulted in promising real-time ankle kinematics estimations (RMSE: 3.34°±0.48°, MAE: 2.68°±0.46° and r2 score: 0.63±0.07). Overall, this research highlights the potential of combining IMUs with ANNs as reliable and practical tools for early prevention and rehabilitation of ankle injuries.**

***Index Terms*—Ankle Kinematics, CNN, CNN-LSTM Model, Feedforward Network, Inertial Measurement Units, LSTM, Transformer.**

I. Introduction

ANKLE joint is an essential part of the body's musculoskeletal system, playing an essential role in locomotion and balance. Unfortunately, it is commonly injured, often subject to sprains or fractures. In the UK, ankle sprains account for 3%-5% of all Emergency Department visits, resulting in around 5600 incidences per day [1]. In the US, approximately 2 million ankle sprains occur annually, with an incidence rate of 2-7 per 1000 people [2]. In Australia, up to 70% of the general population reported having an ankle injury during their lifetime [3]. Studies reveal that 3%-34% of people with ankle sprain experience persistent complains of pain, swelling, muscular weakness, re-sprain, or limitations in activities and sports participation [4], [5]. Ankle injuries also impose high social and economic costs, with the cost of inpatient hospital care for ankle fractures in England exceeding £63.1 million in 2016/17 [6] and over $367 in the US [7]. Therefore, preventing ankle injuries is essential to mitigate their negative impacts on individuals and society.

Ankle kinematics involves analyzing the motion of ankle joint across three planes: sagittal, frontal, and transverse. However, the ankle primarily functions in the sagittal and frontal planes. In the sagittal plane, movements include plantar-flexion (toes pointing downward) and dorsi-flexion (toes lifting upward), while the frontal plane consists of inversion (sole turning inward) and eversion (sole turning outward). Understanding these motions is crucial for assessing ankle function, stability, and lower limb biomechanics, particularly in the context of ankle injury prevention.

Optical motion capture systems (MoCap), such as Vicon and Optitrack, which uses multiple cameras to track reflective markers on the body, are considered the gold standard for non-invasive ankle kinematics measurements [8]. However, they are expensive, vulnerable to occlusion and lighting conditions, and have a restricted capture volume. In contrast, inertial measurement units (IMUs), which typically combine accelerometers, gyroscopes, and magnetometers, offer a viable alternative for motion tracking due to their low cost, light weight, and applicability outside of the laboratory environment. Accelerometers measure linear acceleration to provide insights into position and motion, while gyroscopes capture angular velocity to detail rotational movements. Magnetometers provide directional information for orientation tracking but are sensitive to local magnetic disturbances, limiting their use in joint kinematics estimations.

Various data processing algorithms, such as integration of linear acceleration and/or angular velocity data [9]–[11], extended Kalman filters [12]–[14], and complementary filters [15], [16] are commonly studied for converting IMU data into lower limb kinematics. However, these algorithms typically require accurate sensor placement or calibration strategies to identify sensor-to-segment orientations [17], which can be challenging or even impossible to achieve, particularly for simultaneous estimations of joint angles in multiple planes.

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC grant EP/V004654/1 and EP/V025279/1).

L.Wang and S. Pattinson are with the Institute for Manufacturing, Department of Engineering, University of Cambridge, Cambridge CB3 0FS, UK. (e-mail: lw683@cam.ac.uk and swp29@cam.ac.uk).

P. Song and A. Weller are with the Information Engineering, Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK. (e-mail: ps898@cam.ac.uk and aw665@cam.ac.uk).

T. Stone is with the Cambridge Clinical Movement Laboratory, Cambridge University Hospital, Cambridge CB2 0QQ, UK. (e-mail: thomas.stone@nhs.net).

Recently, artificial neural network (ANN) based models have been studied for joint kinematics estimation [18]–[23], mainly because of their ability to accurately learn complicated non-linear relationships from experimental datasets. For instance, Senanayake et.al [18] developed a generative adversarial network (GAN) to convert data streams from three IMUs into 3D anatomical ankle joint angles. Sung et al. [19] proposed a long short-term memory (LSTM) recurrent neural network model with a single layer to estimate hip, knee, and ankle motion in the sagittal plane during walking using a shank-mounted IMU. Mundt et al. [20] used a fully-connected feedforward neural network to map the experimental and the simulated IMU data into 3D joint angles of the lower limbs. Later, they developed a LSTM model and convolutional neural network (CNN) for comparison [21]. Hossain et al. [23] proposed a complicated framework, DeepBBWAE-Net, using a CNN and Gated Recurrent Unit as the base learners to estimate lower-limb sagittal kinematics in different walking speeds. Although most of these ANN models presented promising results, researchers rarely consider the possibility of deploying them into a microcontroller for data processing and decision making in real-time, which limits their potential applications in real-world scenarios, such as injury prevention. Furthermore, the majority of studies have focused solely on estimating ankle motion in the sagittal plane (plantar- and dorsi-flexion), with inadequate research on measuring ankle inversion/eversion measurements in the frontal plane. This can result in a limited understanding of ankle joint function and hinder ankle injury prevention efforts, as ankle inversion/eversion is crucial to the occurrence of ankle injuries.

Apart from modeling methods, the number of IMUs and their location can also significantly affect the accuracy of ankle kinematics prediction. Many studies employ multiple IMUs (2-7 sensors) for lower-limb kinematics estimation [18], [23]–[27]. Although this can bring benefits in terms of improved accuracy and robustness, it comes with additional complexity and cost, making it impractical for real-life use. To address this, some research groups [19], [28] have started using a reduced set of IMUs to estimate joint kinematics, but the impact on the prediction accuracy and reliability remains unclear. Currently, there is no standard to place IMUs for optimal estimations of ankle kinematics, with common positions including dorsal foot, heel, and/or lateral shank [29].

This research aims to provide an ANN-based framework to accurately estimate ankle motion in both the sagittal and frontal planes using a minimal number of IMUs, while avoiding the requirement for precise sensor-to-segment alignment and/or calibration. The proposed ANN models are designed with the requirements for deploying the algorithm into a microcontroller for real-time measurements of ankle kinematics in mind. In this paper, we outline the data collection methodology and the proposed framework, along with the model optimization in Section II. Section III presents and compares model performances and evaluates the system's real-time estimation capabilities. Section IV discusses the findings, and Section V draws conclusions.

II. METHODOLOGY

A. Data Collection and Pre-processing

Optimizing the location of IMUs is essential for accurately and reliably measuring ankle kinematics. In this study, a system was developed, consisting of four IMUs (BNO055 breakout board, Adafruit) that combines accelerometer, gyroscope, and magnetometer in a single package. However, given the susceptibility of magnetometers to local disturbances in the magnetic field, such as interference from nearby metal objects or electrical equipment, only the tri-axis accelerometer and the tri-axis gyroscope were utilized for gait data collection in this study. The system is based on an open-source Arduino-compatible microcontroller, which utilizes an ARM Cortex-M4 processor (Teensy 3.6, PJRC, USA) and provides the necessary connectivity for integrated microSD port and I2C ports for the IMUs. A LiPo battery, with a rated voltage of 3.7V and a rated capacity of 2 AH, was used to power the system. As shown in Fig. 1, the four IMUs were placed on the participant's right lateral shank (S1), right lateral thigh (S2), right superior heel (S3), and right lateral calcaneus (S4). These locations were chosen to capture comprehensive and varied motion data across both the lower leg and the foot. Importantly, when placed onto the participants, the IMUs were only roughly aligned with the body segments. This approach reflects a more realistic, real-world scenario where prefect alignment is often impractical, yet reliable measurements can still be obtained.

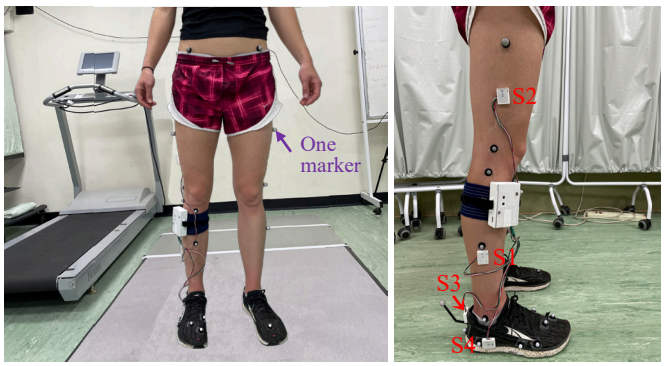

Fig. 1. Measurement system for gait data collection.

To acquire ground truth values of ankle motion at the same time, we used a MoCap system with 12 cameras (VICON, Oxford Metrics, UK) to capture the 3D trajectories of retro-reflective markers placed on the lower body of the same participant. As presented in Fig. 1, the Oxford Foot Model combined with a lower-body VICON Plug-in Gait model (based on Newington-Helen Hayes gait model) was adopted to measure the ankle motion in multiple planes. The VICON Nexus software, specifically designed for capturing and analyzing motion data from VICON cameras, automatically generated joint kinematics and kinetics, providing an accurate ground truth reference. Prior to data collection, the MoCap system was calibrated according to the manufacturer's specifications. To ensure that the optical markers were placed accurately and consistently across all participants, a clinical scientist took responsibility for placing the markers directly onto each participant's skin.

The gait data collection has obtained ethical approval, with reference number #256, from the Ethics Committee of the Department of Engineering at the University of Cambridge on 05/06/2019. 11 healthy participants (4 females and 7 males, age: 30.9 ± 4.6 years, height: 176.6 ± 8.0 cm, and body mass: 69.4 ± 11.9 kg), labeled P01 to P11, were involved in this study, where P01 referes to Participant 1. All participants provided written informed consent prior to participation. Gait data was collected in the Cambridge Movement Laboratory at Addenbrookes Hospital, Cambridge, UK. Participants wore their own trainers and walked on a treadmill at four different speeds each for about 80 seconds, a fixed speed 3.5 km/h, self-selected comfortable gait speed, self-selected comfortable gait speed + 1 km/h (fast), and self-selected comfortable gait speed – 1 km/h (slow). These speeds were chosen to reflect a range of typical walking conditions encountered in daily life, covering both natural and slightly modified gait patterns to simulate different real-world scenarios [30]. The marker trajectories were sampled at a rate of 100 Hz using the VICON system, and then modelled and processed using VICON Nexus software to obtain the ground truth values of ankle kinematics. In the meantime, IMU data were acquired at 50 Hz, stored into an on-board SD card, and finally synchronized with the VICON data in a manual way. During gait data collection, IMUs were placed next to relevant markers, instead of being directly attached to the markers, in order to reflect real-world scenarios more faithfully. In total, 193,435 time-step IMU data were collected, covering various walking speeds and movement patterns. This rich dataset enhanced the robustness of the models developed in this study by capturing a wide range of ankle kinematic behaviours, contributing to the generalizability of the findings.

B. Dataset splitting Methods

To quantitatively train and test the ANN models, the IMU dataset was divided into training, validation, and testing sets using different methods based on two types of ankle kinematics estimation tasks: intra-subject task and inter-subject task.

As illustrated in Fig. 2(a), for the intra-subject tasks, every participant's data were involved in the model training and validation process. At each walking speed of each participant, 20% of the time steps starting from a random index were selected as the testing set, the following 20% as the validation set, and the remaining data was used to train ANN models. This resulted in approximately 116,061 time-step IMU data being used for model training and 38,687 time-step data for testing. For the inter-subject tasks, as shown in Fig. 2(b), a single participant's data were chosen only for model testing while the remaining participants' data were used for model training and validation. Since the number of samples varied slightly among participants, the testing samples ranged from approximately 17,267 to 20,254. Consequently, the remaining data used for model training and validation comprised approximately 173,181 to 176,168 time-step IMU data.

C. Artificial Neural Network Based Framework

In this research, we proposed an innovative framework called KinEmatics Estimation Network (KEEN) to predict ankle kinematics utilizing lightweight ANNs and multimodal IMU data, as illustrated in Fig. 3. To enhance prediction performance, a time window sliding technique was applied to the training, testing, and validation sets of the IMU data before feeding them into the ANN models. This technique allows the models to leverage historical information when making predictions.

The scale of the features affects the convergence speed and final training results. To facilitate training, we performed standardization on the features. The scaling parameters were calculated from the training set and then applied to both validation and testing sets to ensure consistency. After comparing 5 scaler approaches: Standard Scaler, Min Max Scaler, Max Abs Scaler, Robust Scaler, and Normalizer, we noticed that Max Abs Scaler, which scaled each feature individually by its maximum absolute value, demonstrated the best empirical performance for our tasks. This might be in part due to the scaler's inherent characteristic that transforms input features to a similar scale while preserving the sparsity of the original data.

To estimate multi-plane ankle kinematics, five different ANNs were developed and evaluated: Fully Connected (FC) network, CNN, LSTM, Transformer (TX), and CNN-LSTM

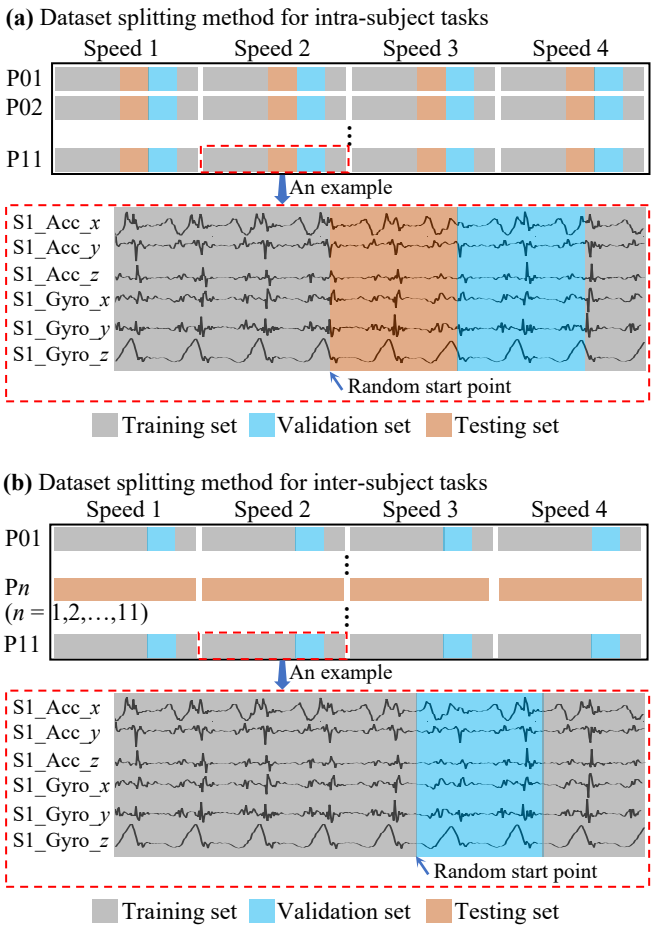

Fig. 2. Dataset splitting methods for (a) intra-subject tasks and (b) inter-subject tasks. Participants are labeled as P01 to P11, where P01 refers to Participant 1.

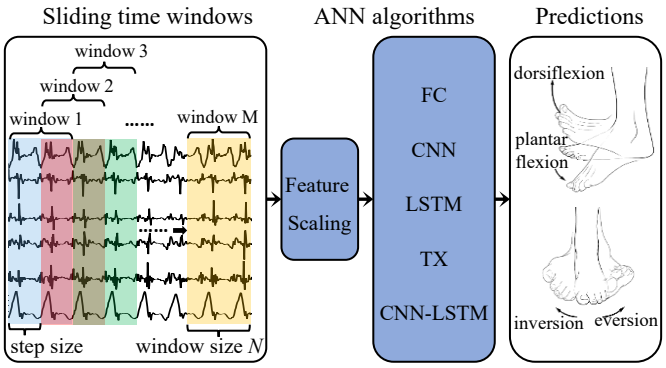

Fig. 3. The proposed IMU data processing framework, KEEN, for real-time ankle kinematics predictions.

networks, each chosen to address distinct challenges posed by the IMU data. The FC network served as a baseline due to its simplicity, offering a reference point for comparison. The CNN model was chosen for its ability to extract spatial patterns from the IMU data, which was crucial for analyzing ankle joint motion across space. The LSTM network was specifically chosen to capture the temporal dependencies inherent in time-series IMU data, beneficial for modelling long-term dependencies and sequential information in ankle kinematics over time. The TX model was selected as it can leverage the attention mechanism to efficiently capture long-term dependencies in a parallel manner, making it increasingly popular for handling complex time-series data. Lastly, the CNN-LSTM network combined the strengths of CNNs for spatial feature extraction and LSTMs for temporal sequence modelling by feeding the CNN output into the LSTM for sequence prediction. The architecture and hyperparameters of each model were optimized, as detailed in Section II-D, to ensure that the models performed effectively in predicting ankle kinematics using the IMU data. This comparative analysis highlighted how each model adapted to the spatial-temporal nature of the IMU data, allowing us to identify the most suitable approach for ankle kinematics estimation.

D. Model Parameter Optimization

To find a proper setting for the window size and step size, a range of window sizes (N = 30, 50, 70, ..., 130, and 150) and step sizes (1, N/4, N/2, and N) were tested and compared for each ANN model. The optimal window size N and step size were found to be 70 and 1, respectively.

In accordance with section II.A, gait data was collected using four IMUs, each providing 6-axis acceleration and angular rate. This resulted in 42 different combinations, denoted as S1_j, S2_j, S3_j, S4_j, S1S2_j, S1S3_j, S1S4_j, S2S3_j, S2S4_j, S3S4_j, S1S2S3_j, S1S2S4_j, S1S3S4_j, S2S3S4_j, S1S2S3S4_j. These combinations were classified by the type of data they contained, including acceleration data from accelerometers (j=Acc), angular velocity data from gyroscopes (j=Gyro), and data from both accelerometers and gyroscopes (j=AccGyro). Although all of them could be used as inputs for the models, to simplify the process of model architecture and parameter optimization, we selected the acceleration and angular velocity data from S4, i.e. S4_AccGyro, as the representative input for the ANN models.

The structure of the ANN models was optimized through a trail-and-error approach, focusing on adjusting the number of layers, the number of units in each layer, and the inclusion or exclusion of dropout layers. When selecting the model architecture, we considered not only the prediction accuracy but also model size. This is because the intended use of the model is for real-time gait analysis on a microcontroller in the future, where the size and complexity of the model are crucial. The ANN models in this research are expected to be about 100 KB, which can be reduced by a factor of 10 (i.e. to approximately 10 KB) by applying pruning and post-training quantization techniques to the models [31]. This would keep the memory requirements below the limitations of the majority of the ARM Cortex-M series microcontrollers ($\geq$ 16 KB). The hyperparameters were tuned by using the random search technique in TensorFlow. The optimized model architecture and hyperparameters were listed in Table I. More detailed results of each ANN algorithm with different structures can be found in Supplementary A.

TABLE I
SUMMARY OF THE OPTIMIZED ANN ARCHITECTURES AND HYPERPARAMETERS.

	FC	CNN	LSTM	TX	CNN-LSTM
No. of hidden layers	2	4	2	1 TX blocks, 8 heads with size of 2, 4 feedforward units, 16 multi-layer perceptron	3
No. of units in the 1st hidden layer	16	16	8		8 (CNN)
No. of units in the 2nd hidden layer	16	16	32		8 (LSTM)
No. of units in the 3rd hidden layer	0	16	0		16 (LSTM)
No. of units in the 4th hidden layer	0	16	0		0
Learning rate	0.001	0.0001	0.001	0.001	0.001
Optimizer	Adam	Adam	Adam	Adam	Adam
Batch size	32	8	16	32	32
Maximum allowed No. of epochs	1000	1000	1000	1000	1000
Patience for early stopping	10	10	10	10	10
Model size (KB)	113	114	110	111	95

III. RESULTS

In this section, we presented the results from both the intra-subject and inter-subject tasks, comparing the performance of various models and inputs to identity the best model and the optimal sensing location/signal for each task.

A. Intra-subject Tasks

For the intra-subject tasks, all 42 input combinations were evaluated using the five optimized ANN models which were

outlined in section II-D. Each model was trained and tested five times to ensure reliability and stability. Performance metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (r2 score), were calculated using (1). The RMSE and MAE quantify the differences between predicted and actual ankle motion values, with lower values indicating more accurate predictions. The r2 score, ranging from -1.0 to 1.0, measures the goodness of fit, where a score of 1.0 represents a perfect fit and negative values indicate the model performs worse than a baseline horizontal line.

$$
\begin{aligned}
avg\_RMSE &= \frac{1}{5}\sum_{j=1}^{5}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i))^2} \\
avg\_MAE &= \frac{1}{5}\sum_{j=1}^{5}\sqrt{\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i|} \\
avg\_r2(y_i,\hat{y}_i) &= \frac{1}{5}\sum_{j=1}^{5}(1-\frac{\sum_{i=1}^{N}(y_i-\hat{y}_i))^2}{\sum_{i=1}^{N}(y_i-\bar{y}_i)^2})
\end{aligned}
\tag{1}
$$

where *avg_RMSE*, *avg_MAE*, and *avg_r2* represent the mean values of the RMSEs, MAEs, and r2 scores of the five repeated tests, respectively. The parameter N denotes the number of the samples. The parameters $\hat{y}$ and $\bar{y}$ represent the predicted value and the mean value of y, respectively.

1) Model Performance Comparison: To evaluate model performance, we averaged the RMSE, MAE, and r2 scores across all 42 input combinations for each model (i.e., the average of the 42 *avg_RMSE* values calculated using (1), along with the corresponding averages for MAE and r2 scores). The results were presented in Fig.4. This plot provided a visual comparison of the models' accuracy in predicting ankle kinematics using IMU data relative to the ground truth values from the MoCap system. The LSTM model clearly outperformed the others, followed by the CNN model.

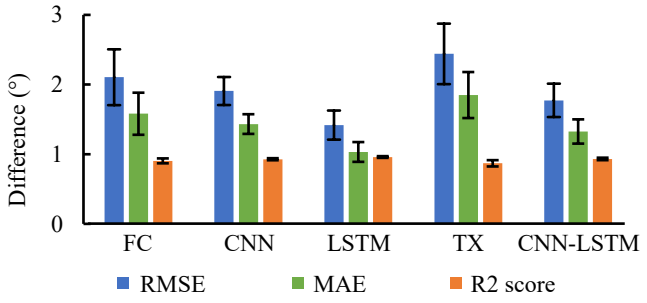

Fig. 4. Model performance across all 42 inputs in intra-subject tasks. The y-axis represents the difference between model predictions based on IMU data and the MoCap ground truth.

Additionally, we conducted a statistical analysis of the RMSEs and MAEs obtained from 42 different inputs. This analysis was essential to verify the observed trends in Fig.4 and to rigorously assess the significance of performance differences among the models. By quantifying these differences, we aimed to ensure that the most reliable and robust model was selected for intra-subject tasks. For conciseness, we focused primarily on discussing the RMSE results in this section, while the statistical results for MAEs were available in Supplementary B.1.

Based on the measurement and data features, we applied an ANOVA with repeated measures, incorporating the Greenhouse-Geisser correction, to analyze the RMSEs. The results revealed statistically significant differences among the RMSEs for the five developed models ($F(1.266, 51.916) = 400.019$, $p < 0.001$). To further identify the specific models which exhibited significant differences, we conducted the Bonferroni post hoc test, and the outcomes were presented in Table II.

TABLE II
PAIRWISE RMSE COMPARISONS OF PREDICTIONS FROM FIVE MODELS USING IMU DATA AGAINST MOCAP GROUND TRUTH IN THE INTRA-SUBJECT TASKS.

Model (I)	Model (J)	Mean difference (I-J)	std	p [b]	95% CI for difference [b]	
					Lower bound	Upper bound
FC	CNN	**0.198***	0.035	**<0.001**	0.094	0.302
	LSTM	**0.687***	0.034	**<0.001**	0.587	0.787
	TX	**-0.317***	0.014	**<0.001**	-0.360	-0.274
	CNN-LSTM	**0.332***	0.031	**<0.001**	0.240	0.424
CNN	FC	**-0.198***	0.035	**<0.001**	-0.302	-0.094
	LSTM	**0.489***	0.007	**<0.001**	0.470	0.509
	TX	**-0.515***	0.034	**<0.001**	-0.617	-0.413
	CNN-LSTM	**0.134***	0.009	**<0.001**	0.106	0.162
LSTM	FC	**-0.687***	0.034	**<0.001**	-0.755	-0.619
	CNN	**-0.489***	0.007	**<0.001**	-0.502	-0.476
	TX	**-1.004***	0.033	**<0.001**	-1.071	-0.937
	CNN-LSTM	**-0.355***	0.008	**<0.001**	-0.371	-0.339
TX	FC	**0.317***	0.014	**<0.001**	0.288	0.346
	CNN	**0.515***	0.034	**<0.001**	0.445	0.585
	LSTM	**1.004***	0.033	**<0.001**	0.937	1.071
	CNN-LSTM	**0.649***	0.031	**<0.001**	0.585	0.713
CNN-LSTM	FC	**-0.332***	0.031	**<0.001**	-0.395	-0.269
	CNN	**-0.134***	0.009	**<0.001**	-0.153	-0.115
	LSTM	**0.355***	0.008	**<0.001**	0.339	0.371
	TX	**-0.649***	0.031	**<0.001**	-0.713	-0.585

Based on estimated marginal means
*. The mean difference is significant at the 0.05 level.
b. Adjustment for multiple comparisons: Bonferroni.

The results shown in Table II revealed a statistically significant distinction in the RMSEs among all models ($p < 0.001$ for each pair comparison). A closer examination of the "Mean Difference" column highlighted that the RMSEs of the LSTM model were lower than those of the other models (negative Mean difference values), indicating its notably accurate predictions in the intra-subject tasks, consistent with the findings

from Fig.4. In contrast, the TX model showed considerably higher RMSEs (positive Mean difference values). This was possibly because the TX model architecture was constrained by the required model size, resulting in a small model that could not effectively capture the relationship. Moreover, the limited length of time-series data could have further impeded the TX model's ability to fully leverage its full potential advantages, particularly when compared to natural language data.

2) Input Performance Across Models: To identify the optimal input, the averaged RMSE, MAE, and r2 score, calculated using (1), along with their standard deviations, were plotted in Fig. 5.

The results revealed that, in cases of using data from the same IMUs, the combination of acceleration and gyroscope's readings (i.e., *j*=AccGyro) provided the highest prediction accuracy, followed by only using the gyroscope's signal (i.e. *j*=Gyro), with accelerations alone (i.e. *j*=Acc) providing the worst prediction performance. These findings were consistent across all five networks as shown in Fig. 5. Taking the FC model for an example (see Fig. 5(a)), the input S3_AccGyro yielded superior ankle motion estimation (RMSE: 2.31°±0.06°, MAE: 1.75°±0.05°, and r2 score: 0.89±0.01) in comparison with S3_Gyro (RMSE: 2.54°±0.02°, MAE: 1.89°±0.02°, and r2 score: 0.87±0.00), which provided better predictions than S3_Acc (RMSE: 2.88°±0.05°, MAE: 2.18°±0.04°, and r2 score: 0.83±0.01). Among all the 42 inputs, the estimation performance with S1S2S3S4_AccGyro outperformed the others.

Additionally, when tri-axis acceleration data from an individual IMU (i.e. S*i*_Acc, i = 1,2,3, and 4) were input to the models, S3 and S4 generally provided better estimations than S1 and S2. This could be attributed to the placement of S3 and S4 being closer to the ankle joint of interest compared to S1 and S2, resulting in more accurate and informative measurement data from S3 and S4. When accelerations from two IMUs were fed into the ANN models (i.e. S*i*S*j*_Acc, i, j $\in$ [1,2,3,4], $i \leq j$), in general S1S3 and S1S4 presented lower estimation errors than the other combinations. This could be because S3 and S4 were located close to each other, resulting in high coherence in their measurements that may lead to redundancy. However, since they were distant from S1, the measurements from S1 might provide a beneficial complementary effect for S3 and S4. For three IMUs (i.e. S1S2S3_Acc, S1S2S4_Acc, and S2S3S4_Acc), they demonstrated similar prediction accuracy in most cases. These findings also applied to the inputs that contained either Gyro or AccGyro data.

Compared to the best prediction achieved using the combination of all four IMU readings (S1S2S3S4_AccGyro), using S3_AccGyro as the model input led to an increase in RMSEs by 0.95° (70.2%), 0.32° (20.2%), 0.34° (32.1%), 0.65° (35.6%), and 0.47° (35.3%) for the FC, CNN, LSTM, TX, and CNN-LSTM models, respectively. Similarly, when using S4_AccGyro as the input, the RMSEs increased by 0.59° (43.5%) for FC, 0.31° (19.5%) for CNN, 0.31° (30.0%) for LSTM, 0.41° (22.5%) for TX, and 0.37° (27.9%) for CNN-LSTM.

B. Inter-subject tasks

For the inter-subject tasks, we carefully selected seven distinct combinations of AccGyro data as representative inputs for the models, namely S3_AccGyro, S4_AccGyro, S1S4_AccGyro, S3S4_AccGyro, S1S2S3_AccGyro, S1S2S4_AccGyro, and S1S2S3S4_AccGyro. These combinations were chosen based on their superior prediction accuracy in the intra-subject tasks, as illustrated in Fig. 5. These specific combinations represented either the top two accurate predictions or the single best prediction achieved using inputs derived from one, two, three, or four IMUs. By utilizing these specific combinations, we can effectively save time and computational resources while still expecting to achieve excellent prediction performance.

As illustrated in Fig. 2(b), the data splitting method for the inter-subject estimation differed from that of the intra-subject tasks, as the IMU data from each participant was used as the testing set in turn while the remaining data was used for model training and validation. Specifically, when the IMU data from the participant P*m* (m = 1,2,…,11) was chosen for model testing, the data from the other 10 participants were employed for model training and validation. The model training and testing were repeated five times for each case. The average RMSE, MAE, and r2 score were calculated using (2).

$$
\begin{aligned}
avg2\_RMSE &= \frac{1}{11}\sum_{m=1}^{11}\frac{1}{5}\sum_{j=1}^{5}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2)} \\
avg2\_MAE &= \frac{1}{11}\sum_{m=1}^{11}\frac{1}{5}\sum_{j=1}^{5}\sqrt{\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i|} \\
avg2\_r2(y_i,\hat{y}_i) &= \frac{1}{11}\sum_{m=1}^{11}\frac{1}{5}\sum_{j=1}^{5}(1-\frac{\sum_{i=1}^{N}(y_i-\hat{y}_i))^2}{\sum_{i=1}^{N}(y_i-\bar{y_i})^2})
\end{aligned}
\tag{2}
$$

where *avg2_RMSE*, *avg2_MAE*, and *avg2_r2* represent the mean values of the RMSEs, MAEs, and r2 scores of the 5 repeated tests with the data from each of 11 participants being the testing set in turn. The parameter N denotes the number of the samples. The parameters $\hat{y}$ and $\bar{y}$ represent the predicted value and the mean value of y.

1) Model Performance Comparison: For the inter-subject tasks, model performance was evaluated by averaging the RMSE, MAE, and r2 scores across the seven selected input combinations for each model (i.e., the average of the 7 *avg2_RMSE* values calculated using (2), along with the corresponding averages for MAE and r2 scores). The results were visualized in Fig. 6, which showed that the CNN model exhibited the lowest RMSEs and MAEs compared to the ground truth values, followed closely by the CNN-LSTM model.

Furthermore, we conducted an analysis of variance (ANOVA) with repeated measures to examine both the RMSEs and the MAEs. Since the results of both metrics were found to be similar, here we would focus on the RMSEs for the sake of simplicity and conciseness. The detailed statistical results of MAEs in the inter-subject tasks can be found in Supplementary B.2. The analysis revealed statistically significant differences

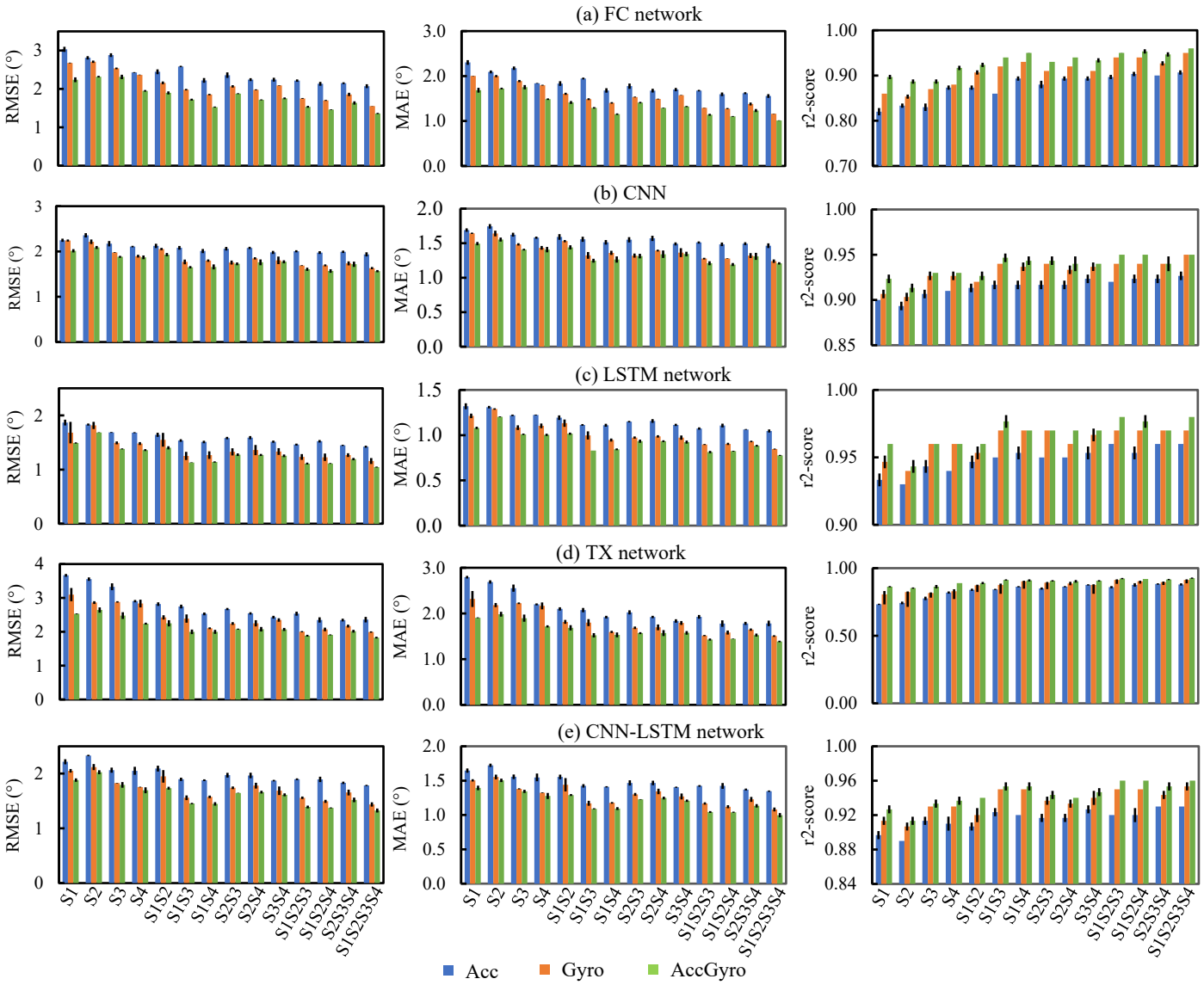

Fig. 5. The root mean square errors (RMSEs), mean absolute errors (MAEs), and coefficient of determination (r2 score) of different ANN models with different inputs in the intra-subject tasks (signal from a single or multiple IMUs. Acc: acceleration data from accelerometers; Gyro: angular velocity data from gyroscopes; and AccGyro: readings from both accelerometer and gyroscope).

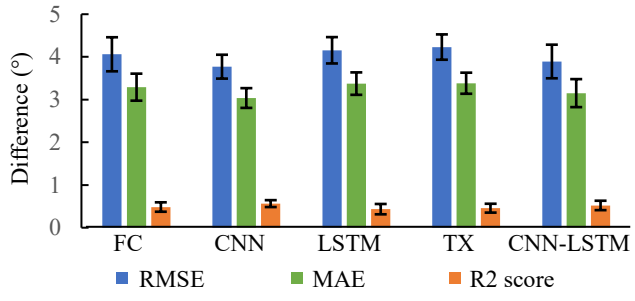

Fig. 6. Model performance across different inputs in inter-subject tasks (7 selected inputs). The y-axis represents the difference between model predictions based on IMU data and the MoCap ground truth.

in the performance of the five developed models based on the RMSEs ($F(4, 24) = 13.303$, $p < 0.001$). To gain further insights into the specific model that exhibited differentiation, a Bonferroni post hoc test was conducted and the corresponding results were presented in Table III. The results of the analysis indicated that the CNN model demonstrated statistically significant differences in RMSEs when compared to the FC model ($p = 0.049$), the LSTM model ($p = 0.011$), and the TX model ($p < 0.001$). However, no significant difference was observed when compared to the CNN-LSTM model ($p = 1.000$). Considering the "Mean Difference" column, it became apparent that both the CNN and CNN-LSTM models had lower RMSEs compared to the FC, LSTM, and TX models (negative Mean difference values). This finding, along with the results in Fig. 6, suggested that the CNN and CNN-LSTM models exhibited more accurate predictions in the inter-subject tasks. Furthermore, by examining the "Mean Difference" column, we observed that the mean RMSE of

the CNN model was slightly lower than that of the CNN-LSTM model by 0.122°, indicating a slightly better overall performance of the CNN model for the inter-subject tasks.

TABLE III
PAIRWISE RMSE COMPARISONS OF PREDICTIONS FROM FIVE MODELS USING IMU DATA AGAINST MOCAP GROUND TRUTH IN THE INTER-SUBJECT TASKS.

Model (I)	Model (J)	Mean difference (I-J)	std	p [b]	95% CI for difference [b]	
					Lower bound	Upper bound
FC	CNN	**0.290***	0.067	**0.049**	0.001	0.580
	LSTM	-0.094	0.093	1.000	-0.493	0.306
	TX	-0.169	0.087	0.989	-0.543	0.205
	CNN-LSTM	**0.169***	0.036	**0.036**	0.012	0.326
CNN	FC	**-0.290***	0.067	**0.049**	-0.580	-0.001
	LSTM	**-0.384***	0.066	**0.011**	-0.668	-0.100
	TX	**-0.459***	0.028	**<0.001**	-0.581	-0.338
	CNN-LSTM	-0.122	0.063	1.000	-0.393	0.150
LSTM	FC	0.094	0.093	1.000	-0.306	0.493
	CNN	**0.384***	0.066	**0.011**	0.100	0.668
	TX	-0.075	0.079	1.000	-0.415	0.264
	CNN-LSTM	0.262	0.101	0.404	-0.172	0.697
TX	FC	0.169	0.087	0.989	-0.205	0.543
	CNN	**0.459***	0.028	**<0.001**	0.338	0.581
	LSTM	0.075	0.079	1.000	-0.264	0.415
	CNN-LSTM	0.338	0.079	0.051	-0.002	0.678
CNN-LSTM	FC	**-0.169***	0.036	**0.036**	-0.326	-0.012
	CNN	0.122	0.063	1.000	-0.150	0.393
	LSTM	-0.262	0.101	0.404	-0.697	0.172
	CNN-LSTM	-0.338	0.079	0.051	-0.678	0.002

Based on estimated marginal means
*. The mean difference is significant at the 0.05 level.
b. Adjustment for multiple comparisons: Bonferroni.

2) Input Performance Across Models: To identify the optimal input for the inter-subject tasks, the average RMSE, MAE, and r2 score, calculated using (2), along with their corresponding standard deviations, were presented in Fig. 7.

It can be found that the RMSEs for most cases fell within the accepted mean error limit of 5°, which is considered reliable by the American Medical Association for evaluating movement impairments in a clinical context [32]. Moreover, in general, using multiple IMUs yielded slightly better results compared to using a single IMU in terms of prediction accuracy of ankle motion. Among the ANN models tested, the best estimation performance was achieved with the input S1S2S3S4_AccGyro for the FC, CNN, and LSTM models, and with S1S2S4_AccGyro for the TX and CNN-LSTM models. For each of the proposed models, the input S4_AccGyro exhibited similar prediction errors to S3_AccGyro, but with significantly higher standard deviations, particularly for the r2

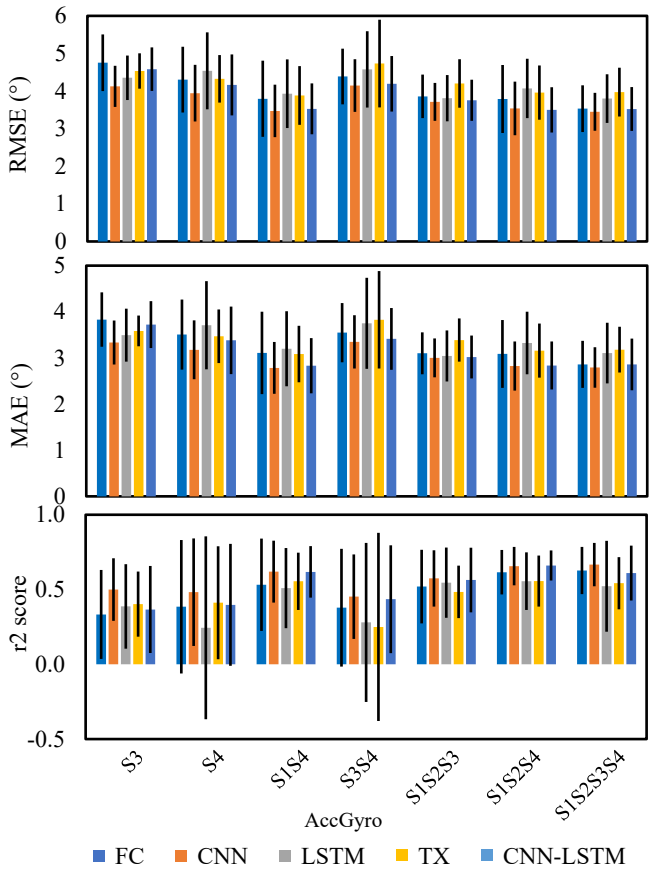

Fig. 7. The RMSE, MAE, and coefficient of determination (r2 score) of different ANN models for ankle plantar/dorsi-flexion and inversion/eversion estimations in the inter-subject tasks.

scores. This may be attributed to variations in the placement of S4 caused by the different shapes of participants' shoes. In comparison with the best predictions, using S3_AccGyro as the model input, the RMSE was increased by 1.23° (34.8%), 0.68° (19.7%), 0.56° (14.7%), 0.57° (14.4%), and 1.06° (30.1%) and the MAE went up by 0.97° (33.9%), 0.55° (19.8%), 0.45° (14.8%), 0.50° (16.2%), and 0.88° (31.0%) for the FC, CNN, LSTM, TX, and CNN-LSTM models, respectively.

IV. REAL-TIME TEST TRAIL

To evaluate real-time performance, the CNN model trained on the IMU data from all 11 participants was selected for deployment on a Teensy 3.6 microcontroller. This model was chosen based on its superior performance in both intra-subject and inter-subject tasks. To enhance inference efficiency, particularly in terms of memory utilization, the model was optimized using pruning and full integer quantization via the *TensorFlow Model Optimization Toolkit*. The optimized model (13 KB) was then converted into a *C* source file required for deployment on the microcontroller.

Three healthy participants (A, B, and C), who were not involved in model training, were recruited for this trial. An IMU sensor (S3, as shown in Fig. 1) was placed at the heel of each participant to capture ankle motion. In this trail,

the raw IMU data were processed on board in real time to estimate ankle kinematics, with the predictions stored on an SD card. While this trail utilized local storage, future iterations could incorporate wireless transmission via WiFi or Bluetooth for real-time visual visualization. Participants walked at their preferred speed along a 10-meter walkway in the Movement Laboratory. Simultaneously, the VICON MoCap system tracked reflective markers to provide ground truth data for comparison with the IMU-based estimations.

The real-time estimations of ankle kinematics were presented in Fig. 8, alongside the ground truth values obtained from the MoCap system. It can be found that the IMU-based estimations generally followed the trends of the ground truth values across all three participants. For Participant A, the system demonstrated an RMSE of 2.82°, MAE of 2.26°, and an r2 score of 0.73. For Participant B, the system showed an RMSE of 3.97°, MAE of 3.39°, and an r2 score of 0.66. Participant C had an RMSE of 3.29°, MAE of 2.56°, and an r2 score of 0.57. Across all three participants, the system achieved an overall RMSE of 3.34°±0.48°, MAE of 2.68°±0.46°, and an r2 score of 0.63±0.07.

V. DISCUSSION

This paper presented an innovative framework for real-time ankle kinematics predictions in multiple planes using time-series data from IMUs. Five ANN-based algorithms, namely the FC, LSTM, CNN, TX, and CNN-LSTM models, were developed for data analysis. Four IMUs were placed at different locations on the lower extremity to collect gait data, resulting in 42 combinations for model training. This enabled us to investigate the capability of different ANN models in processing the multimodal IMU data for ankle kinematics analysis, as well as to determine the minimal number of IMUs and their optimal location required for accurate predictions.

The results from the intra-subject and inter-subject tasks revealed notable differences in the models' generalization capabilities. In the intra-subject tasks, the LSTM network provided the highest prediction accuracy, followed by the CNN-LSTM network and then the CNN model, as shown in Fig. 4 and Table II. For instance, with the S3_AccGyro input, the RMSEs were 1.38°±0.01° (LSTM), 1.79°±0.05° (CNN-LSTM), and 1.88°±0.02° (CNN). This indicated the LSTM network's strong ability to generalize to new tasks within the same subject. In contrast, for the inter-subject tasks, the CNN and CNN-LSTM models demonstrated comparable performance based on the statistical analysis, but they exhibited statistically superior estimation accuracy compared to the other models when considering various inputs, as evidenced by the results presented in Fig. 6 and Table. III. Notably, when a single IMU was utilized as the model input, such as S3_AccGyro, the CNN model achieved the lowest RMSE of 4.13°±0.55° and MAE of 3.33°±0.48° among the proposed models, suggesting its capability to generalize well to new subjects. The differences in model performance between the intra-subject and inter-subject estimations could be attributed to the training conditions. Intra-subject tasks allowed the models to train and test on data from the same individuals, enabling them to learn adaptive and specific characteristics and patterns inherent to the data of each individual. This facilitated effective generalization within the same group, resembling a standard independent and identically distribution (IID) setting, which is beneficial for ensuring strong generalization performance. Conversely, the inter-subject tasks required the models to adapt to data from individuals not present during training. This scenario represents an out-of-distribution (OOD) setting, making generalization more challenging due to varying movement patterns and styles among different participants. The fact that the LSTM model performed better in the intra-subject task, was possibly due to its ability to capture individual-specific patterns, enhancing its accuracy for IID data. On the other hand, the architectures of the CNN and CNN-LSTM models inherently provide a form of regularization. Their ability to extract spatial features (CNN) and model temporal dependencies (LSTM) contributes to their enhanced performance in OOD settings, as they can better adapt to the variability in movement styles across different subjects. Furthermore, according to scaling laws [33], TX-based ANNs typically require extensive data for pretraining large models in an auto-regressive manner to fully demonstrate their superior learning capabilities. We hypothesize that, due to the limited dataset size and the smaller number of model parameters in our case, the TX-based model struggled to outperform the other lightweight models.

Using multiple IMUs generally improved prediction accuracy, as shown in Fig. 5 and Fig. 7. However, employing a single IMU still achieved clinically acceptable performance with a mean error no more than 5°. Specifically, when feeding the time-series data from a single IMU into the models, all the RMSEs were less than 3.67° in the intra-subject tasks and below 4.75° for all five models in the inter-subject tasks. Therefore, the use of a single IMU sensor in combination with the proposed ANN-based algorithms demonstrated considerable potential for clinical and daily applications, providing clinically meaningful estimations of ankle motion across multiple planes. This highlights its promise in meeting the needs of both healthcare practitioners and the wider market.

The placement of the IMU at the superior heel (S3) yielded more accurate and consistent predictions. In the intra-subject tasks, S3_AccGyro and S4_AccGyro outperformed the other single IMU inputs in most cases, with S3_AccGyro exhibiting slightly larger errors than S4_AccGyro, as shown in Fig. 5. In the inter-subject tasks, S4_AccGyro showed marginally better prediction accuracy than S3_AccGyro but had significantly higher standard deviations for each model (refer to Fig. 7). This might be related to the relatively large variations in the placement of S4. It is worth noting that the hyperparameters utilized in the model instances were optimized using S4_AccGyro as the input. Therefore, we hypothesize that by fine-tuning the model's hyperparameters, S3_AccGyro has the potential to yield even better estimation performance than currently observed.

Real-time estimations, as illustrated in Fig. 8, generally aligned well with the ground truth values, although accuracy varied across participants. Participant A showed the highest correlation (r2 score: 0.73), while Participants B and C

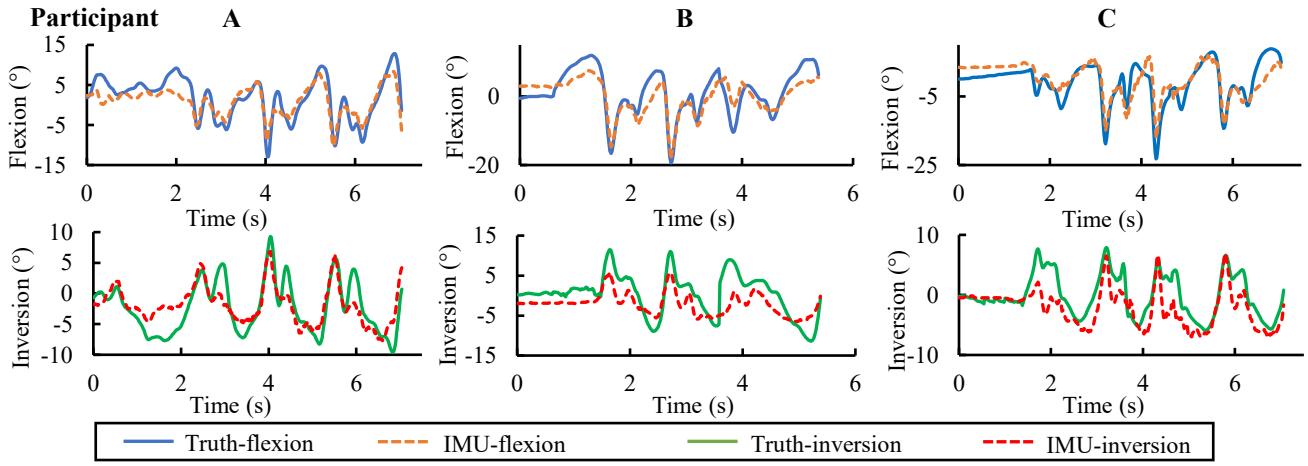

Fig. 8. Real-time estimations of ankle motion in the sagittal plane (plantar-/dorsi-flexion) and the frontal plane (inversion and eversion). Negative values for ankle flexion indicate dorsiflexion, while negative values for inversion indicate eversion.

exhibited slightly lower accuracy, likely due to differences in gait patterns or sensor positioning. The overall performance (RMSE: 3.34°±0.48°, MAE: 2.68°±0.46°, r2 score: 0.63±0.07) and the absence of noticeable time delays supported the feasibility of real-time estimation, though further refinement and dataset expansion may improve prediction accuracy.

Table IV summarizes the RMSEs reported in relevant research studies regarding the estimation of ankle kinematics using ANN-based methods, including our own KEEN system which employed S3_AccGyro input. The focus in this table is solely on ANN-based methods, as conventional algorithms such as the extended Kalman filter require precise sensor-to-segment calibration and detailed knowledge of human body segment parameters, including joint centre. ANNs, offer a more flexible and data-driven approach to estimate ankle kinematics. The KEEN system, based on ANN models, reduces the dependency on calibration and can generalize better in real-world scenarios. These qualities make ANNs particularly useful in dynamic, uncontrolled environments, where conventional methods may struggle with accuracy. Therefore, comparing ANNs with conventional methods is not meaningful due to the differences in underlying principles and assumptions. By limiting our comparisons to other ANN-based methods, we can better understand the strengths and limitations of our approach.

As shown in Table IV, all the models proposed in this study outperformed the previous ANN-based methods for the intra-subject tasks, with our CNN model achieving the most accurate predictions for the inter-subject tasks using the S3_AccGyro input. It is important to acknowledge that differences in datasets, the number of IMUs, their positions, and the trail conditions may affect direct comparisons with other studies. Nevertheless, the table provides a convenient overview of model performance and helps to identify the strengths and limitations of our current work. Additionally, the real-time capability of the KEEN system, demonstrated through the deployment of the CNN model on a microcontroller, offers a significant advantage over other research. This feature allows for immediate feedback on ankle motion, which is critical for clinical applications such as injury prevention and sports monitoring.

TABLE IV
SUMMARY AND COMPARISON WITH THE REPORTED RMSES FOR ANKLE MOTION ESTIMATIONS.

Network	No. of IMUs	Intra-subject tasks			Inter-subject tasks		
		PF/DF	INV/EV	mean	PF/DF	INV/EV	mean
GAN [18]	3	\	\	\	5.90°	3.20°	4.55°
LSTM [19]	1	3.06°	\	3.06°	4.35°	\	4.35°
FC [20]**	5	\	\	\	3.96°	6.25°	5.11°
DeepBBWAE-Net [23]	1	\	\	\	3.87°	\	3.87°
LSTM [28]	1	2.57°	\	2.57°	\	\	
FC*	1	2.54°	2.06°	2.30°	4.83°	4.45°	4.64°
CNN*	1	2.16°	1.55°	1.86°	**4.20°**	**3.91°**	**4.06°**
LSTM*	1	**1.57°**	**1.16°**	**1.37°**	4.51°	4.02°	4.27°
TX*	1	2.59°	2.36°	2.48°	4.69°	4.23°	4.46°
CNN-LSTM*	1	1.99°	1.57°	1.78°	4.51°	4.43°	4.47°

Note: PF/DF denotes plantar flexion and dorsiflexion, and INV/EV denotes ankle inversion and eversion. '\' means that the metrics is not available. '*' indicates that the results are from our work with the input of S3_AccGyro. '**' interpreted from plots.

While the KEEN system provides substantial advantages, particularly in elimination of sensor-to-segment calibration requirements, improved generalization, and achievement of real-time analysis, it also faces trade-offs that must be addressed through enhanced dataset, careful model design, strategic application, etc. First, the small sample size of 11 participants limits the generalizability of the findings. A larger and more diverse dataset, including participants of varying ages, body types, and gait patterns, is essential to improve the robustness and reliability of ankle kinematic predictions. In particular,

increasing the sample size would enhance the model's ability to generalize in OOD settings, where new subjects exhibit movement patterns that differ from those in the training data. Additionally, future work could explore the use of simulated IMU data to augment the existing dataset, which may improve the model's adaptability and capacity to generalize across different populations and scenarios [20], [27]. Second, the sensitivity of IMU placement is another key challenge. Although utilizing ANN models has demonstrated the system's resilience to slight misalignments, further research is needed to develop methods to mitigate sensor misalignment or to design adaptive models that account for this variability, which is essential for practical, real-world deployment. Third, in this research, all data were collected on a treadmill in a controlled lab environment. Previous studies [34], [35] have reported that walking on a treadmill might show different joint kinematics from ground walking. To improve the model's relevance for real-life applications, future studies should incorporate data from both treadmill and ground walking conditions. This will not only enhance the system's accuracy but also enables a more comprehensive validation of the model in dynamic and uncontrolled environments. Finally, this study focused on the walking conditions only; we plan to further investigate the ability of the system to distinguish between walking-related ankle movements and other random motions in future work.

Overall, this research highlights the potential for deploying the developed ankle kinematics estimation system in real-world applications. The successful deployment of the CNN model in Teensy3.6 microcontroller demonstrated its capability for real-time estimations on ankle motion, which can benefit areas such as injury prevention, rehabilitation, and wearable health monitoring. However, challenges remain, particularly in ensuring accuracy in uncontrolled environments. Collaboration with clinicians will be essential to refine the system for practical use and ensure its data can support patient assessments and treatment plans. Integrating the system with existing healthcare protocols and electronic health records can further enhance its utility, enabling more informed decision-making and personalized care. Addressing these challenges will be crucial for maximizing the system's practical application.

VI. Conclusion

This research investigated the feasibility of using IMU data and lightweight ANN models (FC, CNN, LSTM, TX, and CNN-LSTM networks) to predict ankle motion in both the sagittal and frontal planes. Through the evaluation of the five models, we identified the minimum number of IMUs and their optimal location required for accurate predictions. A system equipped with four IMUs was developed to measure accelerations and angular rates during treadmill walking, creating 42 different combinations which were used as the model inputs. The results obtained from both the intra-subject and the inter-subject tasks revealed that a single IMU located at the heel could provide clinically acceptable predications, while multiple IMUs demonstrated higher estimation accuracy. Among the models evaluated, the LSTM model exhibited the most accurate estimations for intra-subject tasks, whereas the CNN and the CNN-LSTM models showed comparable performance but achieved statistically significant improvement in prediction accuracy compared to the other models with various inputs for inter-subject tasks. Notably, when utilizing a single IMU (i.e., S3_AccGyro) as the input, the CNN model exhibited the highest prediction accuracy among the proposed models in the inter-subject tests. These findings provide valuable insights into the effectiveness of using a single IMU for ankle kinematics prediction in different tasks. Specifically, the proposed LSTM model proved to be efficient for new tasks within the same subject while the CNN model exhibited promising potential for adoption when dealing with new participants. The deployment of CNN into the Teensy 3.6 microcontroller indicated the real-time estmation capability of the KEEN system. Future work will focus on augmenting the dataset by simulating IMU data and recruiting more participants to perform activities in various indoor and outdoor environments, which is expected to enhance the generalizability of the model. Overall, this study contributes to the development of more accurate and efficient methods for predicting ankle motion, which has potential applications in various fields such as sports, rehabilitation, and robotics. The proposed framework can also be extended to estimate other human kinematics, including for knee, hip, and elbow.

Acknowledgment

We gratefully acknowledge the support of the UK Engineering and Physical Sciences Research Council (EPSRC grant EP/V004654/1 and EP/V025279/1), the computing resources from the Northern Ireland High Performance Computing (NI-HPC) service funded by EPSRC (EP/T022175).

References

[1] M. W. Cooke, S. Lamb, J. Marsh, and J. Dale, "A survey of current consultant practice of treatment of severe ankle sprains in emergency departments in the United Kingdom," *Emerg. Med. J.*, vol. 20, no. 6, pp. 505–507, Nov. 2003.

[2] B. R. Waterman, B. D. Owens, S. Davey, M. A. Zacchilli, and P. J. Belmont Jr, "The epidemiology of ankle sprains in the United States," *J. Bone Joint Surg. Am.*, vol. 92, no. 13, pp. 2279–2284, Oct. 2010.

[3] C. E. Hiller, E. J. Nightingale, J. Raymond, S. L. Kilbreath, J. Burns, D. A. Black, and K. M. Refshauge, "Prevalence and impact of chronic musculoskeletal ankle disorders in the community," *Arch. Phys. Med. Rehabil.*, vol. 93, no. 10, pp. 1801–1807, Oct. 2012.

[4] J. M. van Ochten, M. van Middelkoop, D. Meuffels, and S. M. Bierma-Zeinstra, "Chronic complaints after ankle sprains: a systematic review on effectiveness of treatments," *J. Sports Phys. Ther.*, vol. 44, no. 11, pp. 862–C23, Nov. 2014.

[5] T. Bestwick-Stevenson, L. A. Wyatt, D. Palmer, A. Ching, R. Kerslake, F. Coffey, M. E. Batt, and B. E. Scammell, "Incidence and risk factors for poor ankle functional recovery, and the development and progression of posttraumatic ankle osteoarthritis after significant ankle ligament injury (SALI): the SALI cohort study protocol," *BMC Musculoskelet. Disord.*, vol. 22, no. 362, pp. 1–11, Apr. 2021.

[6] L. J. Scott, T. Jones, M. R. Whitehouse, P. W. Robinson, and W. Hollingworth, "Exploring trends in admissions and treatment for ankle fractures: a longitudinal cohort study of routinely collected hospital data in England," *BMC Health Serv. Res.*, vol. 20, no. 811, pp. 1–10, Aug. 2020.

[7] J. D. Stull, S. B. Bhat, J. M. Kane, and S. M. Raikin, "Economic burden of inpatient admission of ankle fractures," *Foot Ankle Int.*, vol. 38, no. 9, pp. 997–1004, Sept. 2017.

[8] Z. Shuai, A. Dong, H. Liu, and Y. Cui, "Reliability and validity of an inertial measurement system to quantify lower extremity joint angle in functional movements," *Sensors*, vol. 22, no. 3, p. 863, Ian. 2022.

[9] H. Rouhani, J. Favre, X. Crevoisier, and K. Aminian, "Measurement of multi-segment foot joint angles during gait using a wearable system," *J. Biomech Eng.*, vol. 134, no. 6, p. 061006, Jun. 2012.

[10] S. Tadano, R. Takeda, and H. Miyagawa, "Three dimensional gait analysis using wearable acceleration and gyro sensors based on quaternion calculations," *Sensors*, vol. 13, no. 7, pp. 9321–9343, Jul. 2013.

[11] S. Choi, Y. B. Shin, S.-Y. Kim, and J. Kim, "A novel sensor-based assessment of lower limb spasticity in children with cerebral palsy," *J. Neuroeng. Rehabil.*, vol. 15, no. 1, pp. 1–16, Jun. 2018.

[12] S. Šlajpah, R. Kamnik, and M. Munih, "Kinematics based sensory fusion for wearable motion assessment in human walking," *Comput. Methods Programs Biomed.*, vol. 116, no. 2, pp. 131–144, Sept. 2014.

[13] G. Cooper, I. Sheret, L. McMillian, K. Siliverdis, N. Sha, D. Hodgins, L. Kenney, and D. Howard, "Inertial sensor-based knee flexion/extension angle estimation," *J. Biomech.*, vol. 42, no. 16, pp. 2678–2685, Sept. 2009.

[14] V. Joukov, V. Bonnet, M. Karg, G. Venture, and D. Kulić, "Rhythmic extended kalman filter for gait rehabilitation motion estimation and segmentation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 407–418, Feb. 2017.

[15] S. Majumder and M. J. Deen, "Wearable IMU-based system for real-time monitoring of lower-limb joints," *IEEE Sens. J.*, vol. 21, no. 6, pp. 8267–8275, Mar. 2020.

[16] J. Cockcroft, J. H. Muller, and C. Scheffer, "A novel complimentary filter for tracking hip angles during cycling using wireless inertial sensors and dynamic acceleration estimation," *IEEE Sens. J.*, vol. 14, no. 8, pp. 2864–2871, Aug. 2014.

[17] L. Meng, U. Martinez-Hernandez, C. Childs, A. A. Dehghani-Sanij, and A. Buis, "A practical gait feedback method based on wearable inertial sensors for a drop foot assistance device," *IEEE Sens. J.*, vol. 19, no. 24, pp. 12 235–12 243, Dec. 2019.

[18] D. Senanayake, S. Halgamuge, and D. C. Ackland, "Real-time conversion of inertial measurement unit data to ankle joint angles using deep neural networks," *J. Biomech.*, vol. 125, p. 110552, Aug. 2021.

[19] J. Sung, S. Han, H. Park, H.-M. Cho, S. Hwang, J. W. Park, and I. Youn, "Prediction of lower extremity multi-joint angles during overground walking by using a single IMU with a low frequency based on an LSTM recurrent neural network," *Sensors*, vol. 22, no. 1, p. 53, Jan. 2022.

[20] 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," *Front. Bioeng. Biotechnol.*, vol. 8, no. 41, pp. 1–16, Feb. 2020.

[21] 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, Jul. 2021.

[22] I. Weygers, M. Kok, M. Konings, H. Hallez, H. De Vroey, and K. Claeys, "Inertial sensor-based lower limb joint kinematics: A methodological systematic review," *Sensors*, vol. 20, no. 3, p. 673, Jan. 2020.

[23] 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 J. Biomed. Health Inform.*, vol. 26, no. 8, pp. 3906–3917, Aug. 2022.

[24] T. Seel, J. Raisch, and T. Schauer, "IMU-based joint angle measurement for gait analysis," *Sensors*, vol. 14, no. 4, pp. 6891–6909, Apr. 2014.

[25] 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, Aug. 2020.

[26] E. Rapp, S. Shin, W. Thomsen, R. Ferber, and E. Halilaj, "Estimation of kinematics from inertial measurement units using a combined deep learning and optimization framework," *J. Biomech.*, vol. 116, p. 110229, Feb. 2021.

[27] 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," *Front. Bioeng. Biotechnol.*, vol. 8, p. 604, Jun. 2020.

[28] J. Conte Alcaraz, S. Moghaddamnia, and J. Peissig, "Efficiency of deep neural networks for joint angle modeling in digital gait assessment," *EURASIP J. Adv. Signal Process.*, vol. 2021, no. 10, pp. 1–20, Feb. 2021.

[29] W. Niswander, W. Wang, and K. Kontson, "Optimization of imu sensor placement for the measurement of lower limb joint kinematics," *Sensors*, vol. 20, no. 21, p. 5993, Oct. 2020.

[30] E. M. Murtagh, J. L. Mair, E. Aguiar, C. Tudor-Locke, and M. H. Murphy, "Outdoor walking speeds of apparently healthy adults: A systematic review and meta-analysis," *Sports Med.*, vol. 51, no. 1, pp. 125–141, Jan. 2021.

[31] Google, *Pruning in Keras example*, Mountain View, CA, USA, (accessed on Apr. 16 2023). [Online]. Available: https://www.tensorflow.org/model_optimization/guide/pruning/pruning_with_keras

[32] H. Zheng, N. D. Black, and N. D. Harris, "Position-sensing technologies for movement analysis in stroke rehabilitation," *Med. Biol. Eng. Comput.*, vol. 43, no. 4, pp. 413–420, Jul. 2005.

[33] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, "Scaling laws for neural language models," *arXiv preprint arXiv:2001.08361*, 2020.

[34] M. Jung and S. Koo, "Physical factors that differentiate body kinematics between treadmill and overground walking," *Front. Bioeng. Biotechnol.*, no. 888691, pp. 1–13, Aug. 2022.

[35] F. Alton, L. Baldey, S. Caplan, and M. Morrissey, "A kinematic comparison of overground and treadmill walking," *Clin. Biomech.*, vol. 13, no. 6, pp. 434–440, Sept. 1998.