IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 13, 2021, accepted May 26, 2021, date of publication June 3, 2021, date of current version June 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3085660

Deep Learning Techniques in Estimating Ankle Joint Power Using Wearable IMUs

ARNAB BARUA[1], UMME ZAKIA[2], (Student Member, IEEE), CARLO MENON[2,3], (Senior Member, IEEE), AND XIANTA JIANG[1], (Member, IEEE)

[1]Department of Computer Science, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada
[2]School of Mechatronics Systems Engineering, Simon Fraser University, Vancouver, BC V6B 5K3, Canada
[3]Department of Health Sciences and Technology, ETH Zürich, 8092 Zürich, Switzerland

Corresponding author: Arnab Barua (abarua@mun.ca)

This work was supported by the Seed, Bridge, and Interdisciplinary, Memorial University.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Office of Research Ethics at Simon Fraser University under Application No. 2018s0355, and performed in line with the TCPS.

ABSTRACT Estimating ankle joint power can be used to identify gait abnormalities, which is usually achieved by employing a complicated biomechanical model using heavy equipment settings. This paper demonstrates deep learning approaches to estimate ankle joint power from two Inertial Measurement Unit (IMU) sensors attached at foot and shank. The purpose of this study was to investigate deep learning models in estimating ankle joint power in practical scenarios, in terms of variance in walking speeds, reduced number of extracted features and inter-subject model adaption. IMU data was collected from nine healthy participants during five walking trials at different speeds on a force-plate-instrumented treadmill while an optical motion tracker was used as ground truth. Three state-of-the-art deep neural architectures, namely Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN) and, fusion of CNN and LSTM (CNN-LSTM), were developed, trained, and evaluated in predicting ankle joint power by extracting few simple, meaningful features. The proposed architectures were found efficient and promising with higher estimation accuracies (correlation coefficient, $R > 0.92$ and adjusted R-squared value $> 83\%$) and lower errors (mean squared error < 0.06, and mean absolute error < 0.13) in inter-participant evaluations. Performance evaluations among the three deep regressors showed that LSTM performed comparatively better. Lower standard deviations in mean squared error (0.029) and adjusted R-squared value (5.5%) proved the proposed model's efficiency for all participants.

INDEX TERMS Ankle joint power, Inertial Measurement Units, deep neural regressor, LSTM, CNN, feature extraction.

I. INTRODUCTION

Human locomotion during walking or running requires full-body musculoskeletal coordination with the nervous system. Muscles and tendons associated with different joints such as the hip, knee, and ankle are considered to produce mechanical power for propulsion [1], [2]. Especially, ankle power generated at the ankle joint complex to control the lower limb for mobility is a major contributor to human gait. Reduced joint power happens in abnormal gait or impairment, an indication of muscle weakness, deteriorating health, and disease. Therefore, gait analysis has been studied since the 1960s in the clinical diagnosis of the physiological disorder in the elderly population, neurological disorder in cerebral palsy (CP), Parkinson's disease (PD), rehabilitation training for stroke people, and limb prosthesis [3]–[6]. It is also studied in sports for athletes to observe performance and prevent injuries [7]. So, monitoring ankle joint power can be a good indicator for gait abnormality analysis and helps to recover normal, active life. Usually, in physical modeling, the foot is represented as a rigid body and requires inverse kinematics to estimate joint power [8], [9]. This becomes difficult and time-consuming to generate and tune the parameters of the biomechanical model outside of a clinical environment. Modern machine learning algorithms are proven data-driven approaches and more acceptable for

The associate editor coordinating the review of this manuscript and approving it for publication was Razi Iqbal.

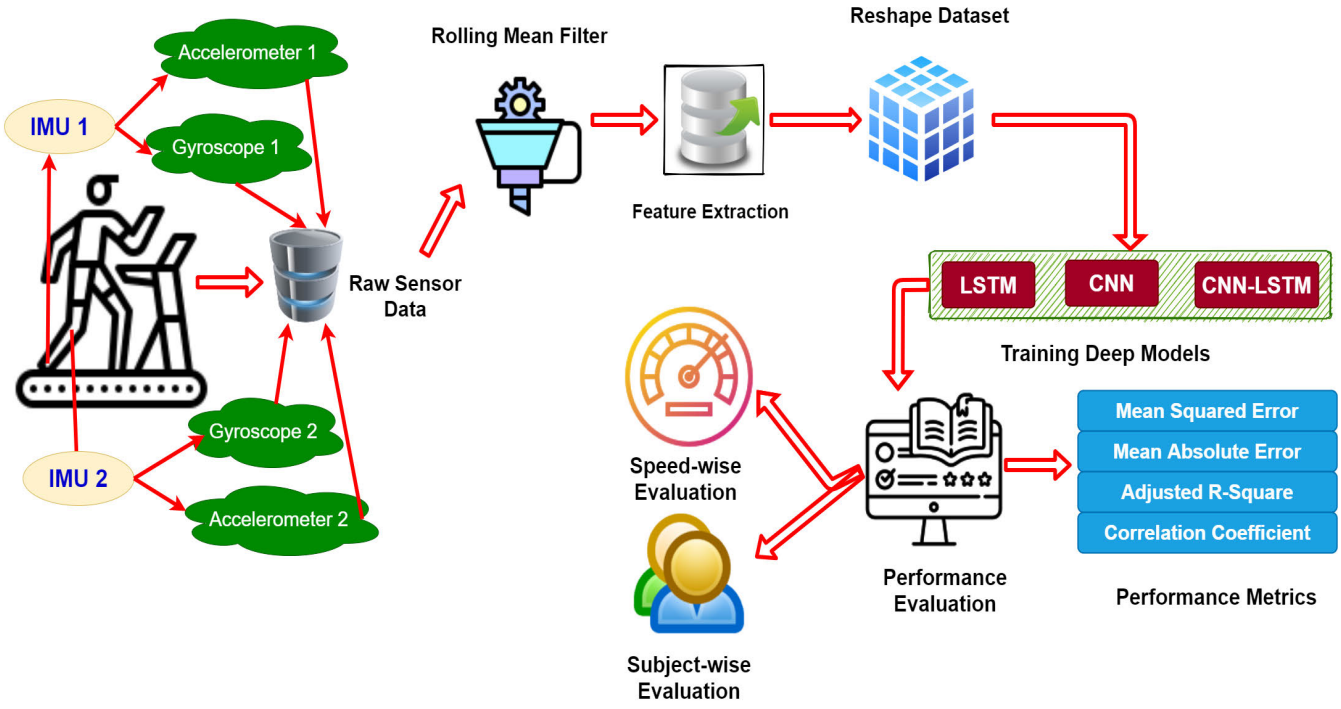


FIGURE 1. Pictorial view of the overall process followed in this study.

real-world applications. Monitoring gait using a variety of sensory [surface electromyography (sEMG), force myography (FMG), inertia measurement unit (IMU)], motion tracking, radar, Wi-Fi, or visionary systems have become easier and doable [10]–[16]. Most of these approaches rely on machine learning algorithms to estimate gait cycles or joint power. Furthermore, using wearable devices would be advantageous in detecting gait events [17]–[19] than the traditional approach of a controlled environment with instrumented walkways, cameras, and treadmills [20], [21]. Such a wearable device using inertia and measurement using (IMUs) to analyze the gait cycle has gained growing interest [22]–[24].

Recent advances in deep learning have intriguingly shifted researchers from conventional machine learning approaches towards deep learning techniques. Most recently, gait analysis is conducted using wearable IMUs, magnetometer, 3D marker, skeleton data, sEMG with convolutional neural network (CNN), long-short-term memory (LSTM), recurrent neural network (RNN) [25]–[31]. As a state-of-the-art technology, our focus is to investigate deep learning algorithms to analyze gait and estimate ankle joint power using wearable IMUs. For sports or rehabilitation, a wearable ankle joint power monitoring system in outdoor with no clinical setup would be effective for real-time assessment. Using conventional ML algorithms, this is a bit constraint because of extensive feature extraction requirements and longer data preprocessing. Using deep learning would reduce this process and could perform better and faster, which is essential in real-time estimations. In our previous study, measuring ankle joint power using two IMUs was found effective using a random forest algorithm (RF) [32]. But RF required huge feature extraction (256 features) and could not perform well in peak power.

In this paper, we propose IMU-based deep learning (DL) techniques with the added novelty of reduced features (8 features) extraction process and estimating ankle joint power even in peak power strikes during stance. In this study, two wearable IMUs on the shank and foot of 9 participants are used for ankle joint power monitoring. We investigate three deep learning techniques: LSTM, CNN, and a hybrid-CNN-LSTM model for estimating ankle joint power, and their performances are promising with higher accuracies (correlation coefficient, $R > 0.92$). Hence, a wearable IMU monitoring system based on a deep learning algorithm would reduce the gap in faster and better ankle power estimations during practical gait cycles in real-world scenarios.

II. MATERIALS AND METHODS

A pictorial view of the overall procedure we followed in conducting this study is portrayed in Fig. 1. Detailed explanation about the steps is provided in following sections.

A. DATA ACCUMULATION

In this study, data was collected from nine healthy participants wearing two IMUs during walking on a treadmill, as mentioned in [32]. All young, healthy adult males [avg. height: 176±7 cm, avg. weight: 72±9 kg, avg. age: 27±8] participated voluntarily with their written consents approved

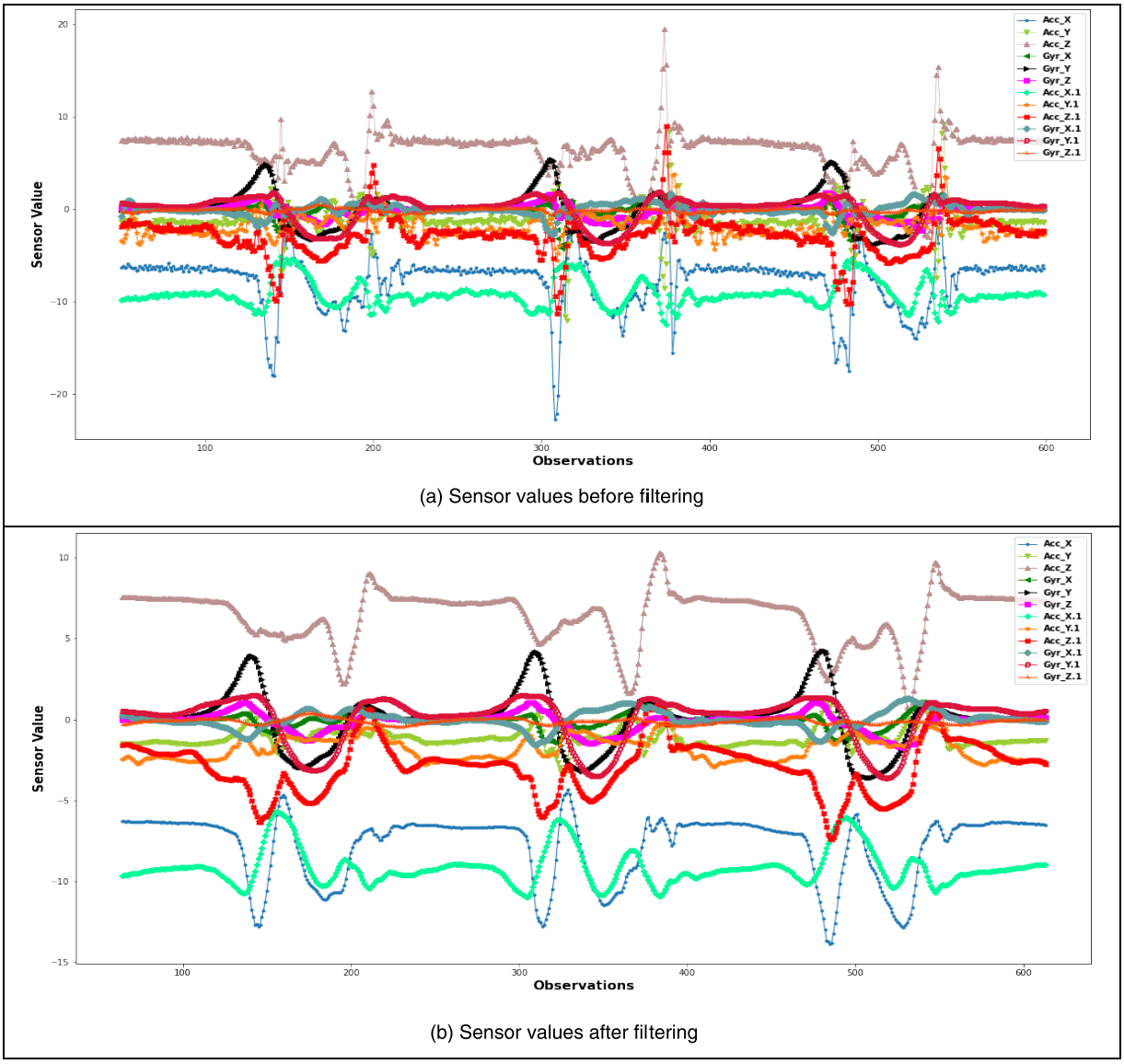


FIGURE 2. Smoothness of values of all 12 sensor axes (a) before, and (b) after median filtering.

by the Office of Research Ethics at Simon Fraser University. No female participants voluntarily participated in this study at the time of data collection before global pandemic (COVID-19), and data accumulation from any more participants was hardly possible afterward till the time of paper submission. Each participant performed 5-walking trials with divergent speeds of 0.4m/s, 0.7m/s, 1.0m/s,1.3 m/s and 1.6m/s. During the experiment, two wireless IMU units (MTw Awinda, Xsens, Enschede, The Netherlands) and seven motion tracker markers were mounted on the shank and foot of the left limb of a participant. Acceleration and angular velocity data were accumulated from these IMUs, and reference values were collected from the force-plate-instrumented treadmill and a Vicon motion capture system (Vicon, Oxford, UK). The reference ankle joint power was determined by channeling data from these two different set-ups through a biomechanical model developed for the individual participants. All data were collected synchronously at a sampling rate of 100Hz. Detailed participants' demographic data, experimental setup and protocol followed for data accumulation can be further read in [32].

B. DATA DISCIPLINING

Six-axis data (tri-axial accelerometer data and tri-axial gyroscope data) from each IMUs for estimating ankle joint power were utilized in this study. A total of 12 independent variables were generated from the two IMUs, as shown in Fig. 2. Several data preprocessing steps were taken to expand the ability of the deep learning model to learn more efficiently, as described below.

i **Displaced Rolling Mean Filter:** We commenced by directing the unrefined values of 12 independent features into

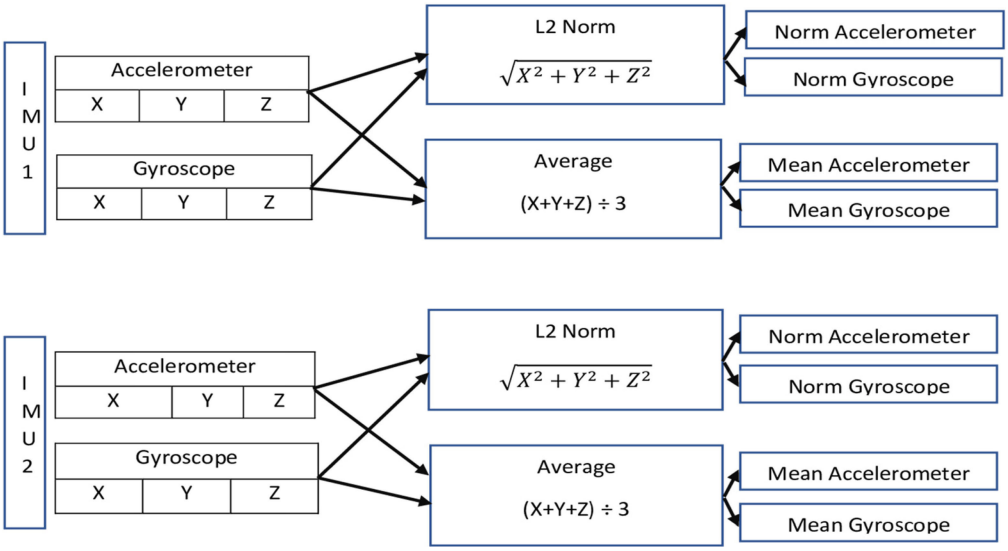


FIGURE 3. Procedure of extracting 8 features from 12 raw features.

a displaced (shifted to left by 15 instances) rolling mean filter with a window size of 15 with 14 overlapping [Fig. 2]. It has been observed that in terms of time series data, rolling mean can be beneficial to eliminate momentary fluctuations and bring out continual long-term patterns. In general, for a conventional rolling mean filter or trailing mean filter with a window size of 15, the average values of the predictors calculated from each window would correspond to the ankle joint power that corresponded to the last row (15^{th} row) of the window. However, the implemented displaced rolling mean filter shifted left by 15 instances did the opposite. It meant that the row generated by averaging the 15 rows in a window corresponding to the ankle joint power, which actually corresponded to the first row of each window. This was done to replicate the trend of the original raw data more accurately.

ii **Feature Extraction:** As mentioned earlier, we conducted this study with data collected using two IMUs. Each IMU contained two sensors, a tri-axial accelerometer, and a tri-axial gyroscope. So, for each IMU, we got 6 raw features (Accelerometer X, Accelerometer Y, Accelerometer Z, Gyroscope X, Gyroscope Y, and Gyroscope Z). In total, we attained 12 raw features from the two IMUs. For practical reasons, the wearable IMUs and other functional components were put off and worn again by a participant [32]. Ensuring similar positioning of the device could hardly be achievable. Consequently, in [32], it became difficult for the machine learning models to generalize inter-participant testing due to differing acquired data from position shifts. To resolve this issue, in this paper, we used 8 extracted features from the 12 raw features as a favorable alternative. The feature extraction was performed in a row-wise manner. Each row contained values for 12 raw features. From these 12 values of a row, we extracted 8 values that corresponded to 8 extracted features. For feature engineering, we considered the L2 norm and average of three axes of each accelerometer and gyroscope sensor. Two features (L2 norm and average) were extracted from each accelerometer and gyroscope sensor (three axes combinedly), obtaining 4 features from one IMU. Consequently, for 4 sensors (2 accelerometers and 2 gyroscopes) we computed 8 features from the two IMUs. Fig. 3. demonstrates the feature extraction procedure with a pictorial view.

iii **Principal Component Analysis (PCA):** After extraction of these 8 significant features, values of these 8 extracted features were further transformed using Principal Component Analysis [33] to reduce association and expand variance among the dimensions. This helped each dimension to be more influential for the learning models. By applying PCA on the 8 extracted features, we acquired 8 principal components, and we considered all 8 principal components as ultimate features in model training.

iv **Min-Max Normalization:** In the next step, predictors were normalized to a range between 0 and 1 to eliminate domination of any dimension when training the deep learning models. In addition, we also narrowed the scale of the target variable to a range between -1 and 1. It has been observed that, when using deep neural networks, lowering the scale of the target variable, assist to shrink the size of gradient descent, hence helping the network to update weights and result in a steadier model. However, lowering the scale of the target variable diminishes its resemblance to a real-world scenario and affects the metrics such as Mean Squared Error (MSE) to assess the performance of a regressor. To overcome this issue, an inverse transformation scheme was applied that converted the target variable to its original form after any type of prediction and before computing any evaluation metrics.

TABLE 1. Structural arrangement of all three deep regressors.

LSTM			CNN			CNN-LSTM		
Layers	Units	Activation function	Layers	Units	Activation function	Layers	Units	Activation function
LSTM	1024	tanh	Convolution	1024	tanh	Convolution	1024	tanh
Dense	256	relu	Dropout	N/A	N/A	Dropout	N/A	N/A
Dense	128	relu	Convolution	512	relu	Convolution	512	relu
Dense	64	relu	Convolution	256	relu	Pooling	N/A	N/A
Output	1	linear	Pooling	N/A	N/A	LSTM	512	relu
			Flatten	N/A	N/A	Dense	256	relu
			Dense	100	relu	Dense	128	relu
			Output	1	linear	Output	1	linear

C. PROPOSED ARCHITECTURES OF DEEP NEURAL REGRESSORS

Due to the time-series nature of the Gait cycles dataset, we preferred Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and a fusion of CNN and LSTM (CNN-LSTM Fusion) among commonly used regressors. LSTM and CNN algorithms are particularly designed for working with data containing time-series properties and patterns in data, respectively. The presence of gait cycles preserved both properties which influenced to consider these regressors. Long Short-Term Memory (LSTM) is considered as an upgraded variant of Recurrent Neural Network (RNN) which can remember long sequences of time series data. Using LSTM as a supervised learning approach requires it to have both features and a dependent variable. Convolutional Neural Network (CNN) is genuinely designed to aid image analysis, however, the emergence of 1D-CNN allowed it to work with time-series data. CNN-LSTM Fusion enables a model to utilize the ability to extract features and preserve long-term dependencies.

A brief description of these models' architectures is described below and provided in Table 1. We used 8 features derived from the IMUs [Section 2.B] as predictors and the dependent variable was ankle joint power. It should be mentioned that, for each training iteration, we used an early-stopping technique with a maximum epoch of 30. Moreover, for all three models, we used 'Adam' optimizer with a learning rate of 0.01 and 'Mean Squared Error (MSE)' as the loss function.

LSTM: We initiated our proposed LSTM architecture by adding an LSTM layer containing 1024 LSTM cells followed by three dense layers containing 256,128 and 64 neurons respectively. Finally, we added an output layer containing 1 neuron with a linear activation function to predict the ankle joint power. For the LSTM layer, we used tanh as the activation function and ReLu for all three dense layers.

CNN: In our proposed CNN architecture, firstly a convolution layer with 1024 filters and kernel of size 7 was introduced which was followed by a dropout layer with a dropout rate of 0.3 to penalize the previous layer to avoid overfitting. Two convolution layers were added after the first one with 512 filters and 256 filters, respectively. The size of kernels for these two convolutions layers was 5 and 3 respectively. Following, we added a pooling layer that performed average pooling operation and then flattened the pooled feature map into a single column to feed it into a dense layer with 100 neurons. Finally, an output layer with the same configuration as the proposed LSTM architecture was used to generate the forecasted version of ankle joint power. For the first convolution layer tanh activation was used, and all other layers used ReLu activation function except for the output layer.

CNN-LSTM: In the proposed CNN-LSTM architecture, we used two convolution layers having 1024 and 512 filters respectively and 1 pooling layer to perform the average pooling operation in the CNN part. For the LSTM portion, 1 LSTM layer with 512 cells followed by two dense layers containing 256 and 128 neurons respectively was employed. The final layer was configured in the same manner as LSTM and CNN. The sizes of the kernel for the first and second convolution layer were 7 and 5 respectively. A dropout layer was stacked on the first convolution layer with a dropout rate of 0.3. The first convolution layer used tanh as its activation function whereas, for all other layers, the activation function was ReLu.

D. PERFORMANCE METRICS

To evaluate the models' performance, we used conventional metrics and graphical representation which can demonstrate a model's power of prediction distinctly. As the evaluation metrics, we considered the four listed metrics given below,

Mean Squared Error (MSE): MSE is a negatively oriented metric defined by the average of the squared difference between true values and predicted values. For a negatively oriented value like MSE, a smaller value is always preferable for any regressor. MSE can be calculated using the following formula,

$$MSE = \frac{1}{n}\sum (true\ response - predicted\ response)^2$$

Mean Absolute Error (MAE): MAE is another negatively oriented metric that considers the average of absolute residuals of predictions. The following formula is used to compute MAE,

$$MAE = \frac{1}{n}\sum |true\ response - predicted\ response|$$

Adjusted R-Squared: R-squared value explains the power of a model's independent variables to explain the variations

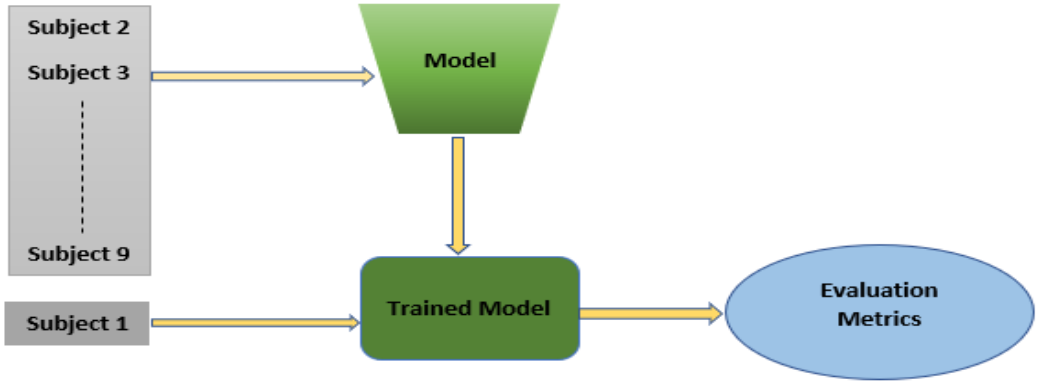


FIGURE 4. 1 of 27 train-test iterations (Leave-one-out cross validation technique).

that exist into the dependent variable, however with the increment in the number of input variables R-square remains the same or expands even if the increased number of variables have not significant association with the target variable. At this point, an upgraded version of the R-squared value which is known as the Adjusted R-squared value works better. It can penalize the addition of insignificant input variables. We usually expect a high value of adjusted R-square from a model.

Correlation Coefficient (R-value): The correlation coefficient also known as R-value is a statistical measurement that explains the measurement of closeness between the true response and estimated response. A desired model's responses tend to generate a greater R-value. It usually ranges between -1 and $+1$.

E. MODELS TRAINING AND TESTING

Feeding data to LSTM, CNN, and CNN-LSTM required the dataset to be shaped in multiple overlapped windows. In general, the dataset was in a 2-dimensional shape (number of rows$*$ number of predictor columns). For the deep learning architectures, the dataset was reshaped into a 3-dimensional structure (number of windows $\times$ window length $\times$ number of features). These windows contained past instances from which the models tried to remember (LSTM) or identify patterns (CNN) or both (CNN-LSTM). For our study, a window size of 7 with 6 overlapping was used for all models where each window contained 56 points (7 rows $\times$ 8 features). The label data (ankle joint power) was also arranged according to the selected window size. Each window was utilized to learn one value of ankle joint power, which belonged to the last instance of that window. After reshaping the dataset, we adapted the leave one out cross-validation (LOOCV) procedure for the inter-participant test [Fig. 4]. We had to train and test each type of model nine times (total participants $=$ 9), wherein each iteration data from one participant of all speeds were used to test, and data from all other participants were used to train. Thus, the final input dimension of training sets and test sets became (**412000**$\times$ **7** $\times$ **8**) and (**51500**$\times$ **7** $\times$ **8**), respectively. The varying number of windows was experienced due to the variations of participants in the training set and test set. We followed this iterative process and observed the regressors' performances for divergent participants to evaluate the models for generality and practical usability. Performance matrices generated from all LOOCV iterations were averaged to obtain an overall performance of each deep neural regressor.

In addition, every time a deep model was trained and evaluated using corresponding test dataset, we plotted a learning curve exhibiting mean squared error for training and test set for each epoch. Such a learning curve could show whether the model suffered from overfitting; we observed that there was no overfitting of these trained models. Cross-validation itself ensured that the model did not overfit while fine-tuning the parameters and adding dropout layers in the model's architecture actually helped reducing overfitting too. Fig. 5. shows an example of learning curve plotted while data from participant 8 was used as the test set.

Implementation of the all-pre-processing steps as well as deep neural architectures was performed using Python 3.7.10 in Google Colabotory environment [34]. We used modules from libraries named Keras 2.4.3 and TensorFlow 2.4.1 to implement the deep neural regressors.

III. RESULTS

All the evaluation metrics were computed by averaging the value of those metrics generated by each type of regressor for all participants. Table 2 exhibits a cumulative evaluation for all participants. Reported results in Table 2 reveals that LSTM exhibited superior performance among the three models [MSE:0.059, MAE:0.131, Adjusted R-Squared: 84.75%, and Correlation Coefficient (R): 92.69%] compared to CNN (R $=$ 92.27%) and CNN-LSTM (R $=$ 92.07%). Although, for all four metrics, the difference between the outcome of LSTM and CNN was not that significant with a standard deviation of 0.13% in MSE, 0.38% in MAE, 0.6% in adjusted R-squared value, and 0.42% in R-value [Table 3]. In the case of CNN-LSTM, it also performed positively but could not outperform CNN and LSTM. Although LSTM exhibited relatively better scores in the other three metrics, it had a

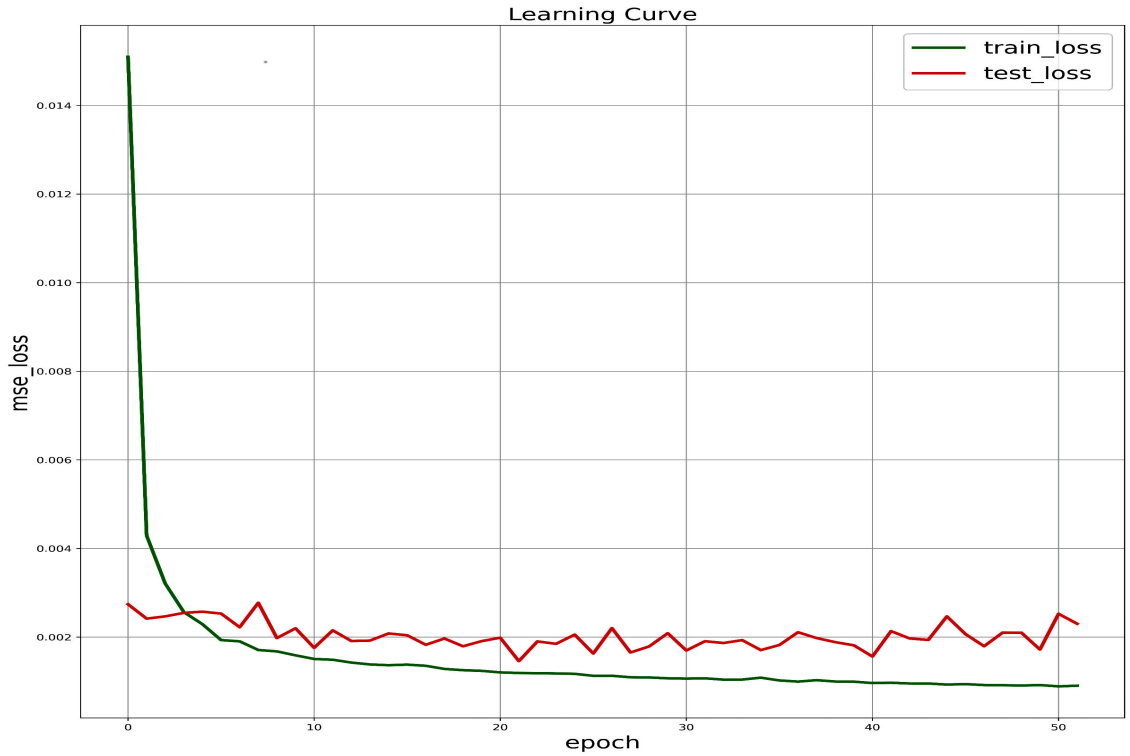


FIGURE 5. Learning curve representing training MSE and test MSE for each epoch.

TABLE 2. The averaged value of evaluation metrics for each deep neural framework and random forest.

Types of Regressor	MSE	MAE	Adjusted R-Squared (%)	Correlation Coefficient (R) (%)
LSTM	0.059	0.131	84.75	92.69
CNN	0.061	0.127	84.15	92.27
CNN-LSTM	0.065	0.129	83.10	92.07
Random Forest	0.138	0.184	64.08	81.75

higher MAE value than others. In contrary to MSE, MAE is not that sensitive to outliers in data which makes MSE a more significant metric in comparison with MAE.

In this study, we emphasized elaborating deep regressors' performances in estimating ankle joint power with a lower number of features. For a fair comparison, we included the performance of a canonical machine learning algorithm, namely Random Forest (RF), as shown in Table 2. Scikit-Learn 0.22.2.post1 library was used to implement RF. The hyper-parameter (number of trees) was tuned to 100 to get the best performance. All other hyper-parameters had their default values set by Scikit-Learn. Using RF, we achieved an MSE value of 0.138, which was ∼7% higher than MSE values achieved by deep regressors. RF exhibited an MAE value of 0.184, which also indicated low performance by exhibiting ∼6% higher MAE values than the deep regressors. Consequently, similar trends were seen regarding adjusted R-squared (64.08%) and R-value (81.75%), which were lower by ∼19% and ∼10%, respectively, than the deep regressors. As the overall performance of the RF was not satisfactory, we did not consider evaluating its performance for subject-wise and speed-wise evaluation.

i **Subject-wise Evaluation:** A model's effectiveness for an individual participant was retrieved by considering boxplots of MSE, MAE, and adjusted R-squared value, as shown in Fig. 6 (Median values). From the right-skewed boxplot of LSTM (MSE), it was deduced that 50% of total participants had MSE values lower than 0.05 whereas for the other two models the value was a little bit higher than 0.06. In terms of MAE, for all participants, the distribution seemed normal and identical for all the models although LSTM showed a little positive skewness. For adjusted R-squared values, almost all boxplots were positively skewed with a median value of 82.5% which indicated that 50% of participants rendered adjusted R-squared values less than the median value. However, the adjusted R-squared values for CNN and CNN-LSTM models had lower values than 75% and 70% respectively, which explained that these two models performed poorly for one or two participants.

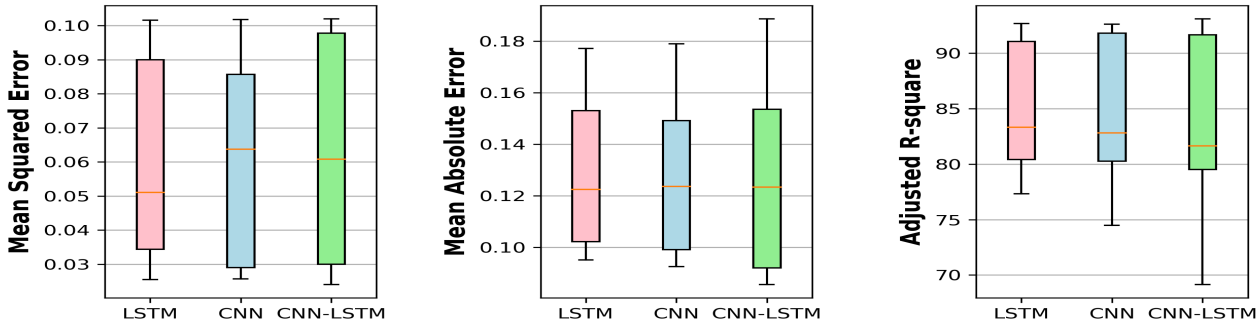


FIGURE 6. Bar-plots to represent the spread of metrics over 9 participants for each model.

TABLE 3. Standard deviation to represent the spread of evaluation metrics over all participants.

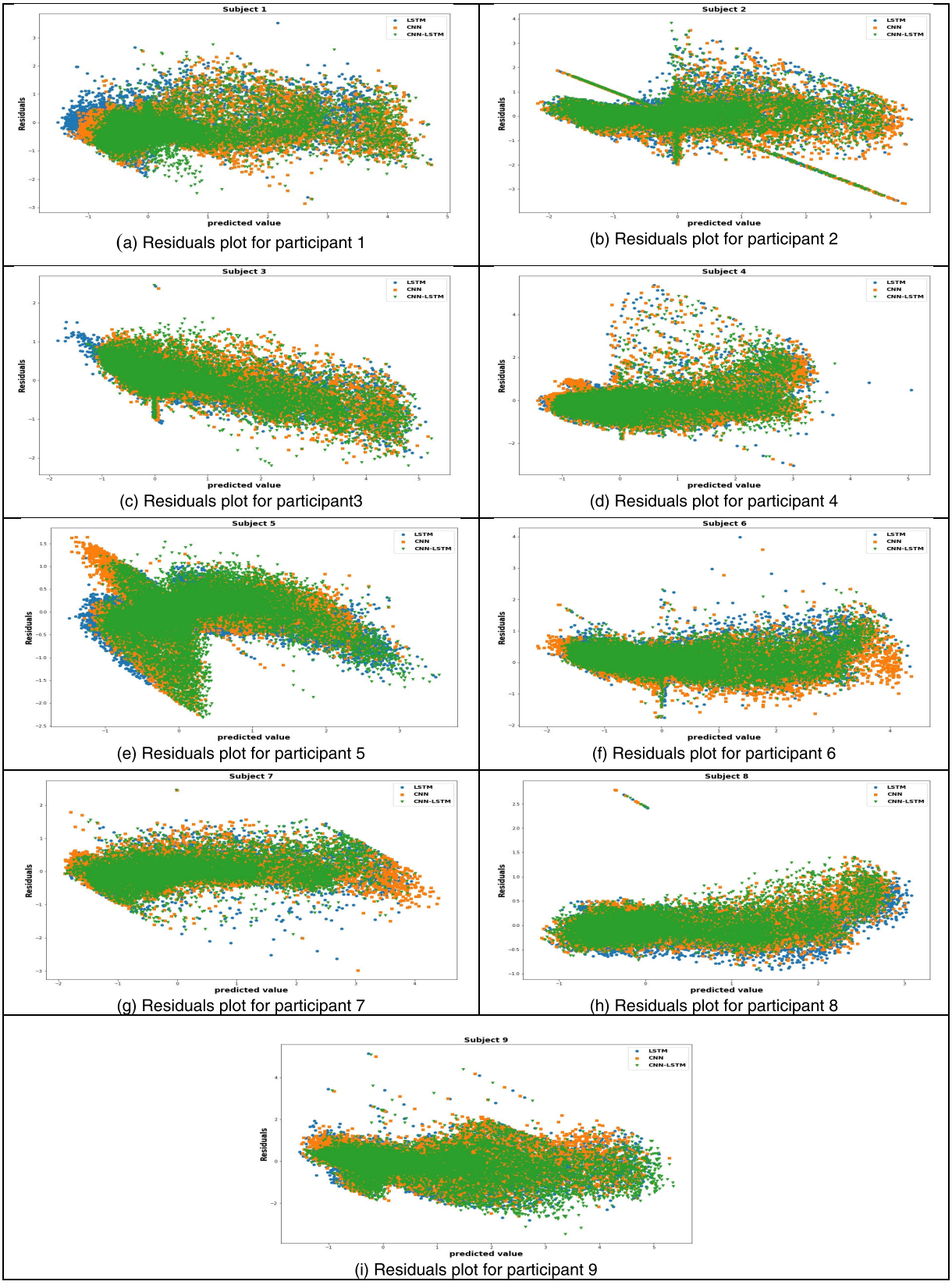
Model type\Metrics	MSE	MAE	Adjusted R-Square	R
	Standard Deviation of metrics values over all participants			
LSTM	0.029	0.029	5.5%	3.1%
CNN	0.029	0.030	6.6%	3.7%
CNN-LSTM	0.032	0.035	7.8%	4.4%

TABLE 4. Averaged value of evaluation metrics for five different speed trials over all participants.

	LSTM				CNN				CNN-LSTM			
Speed Trials	MSE	MAE	Adjusted R-Squared (%)	R (%)	MSE	MAE	Adjusted R-Squared (%)	R (%)	MSE	MAE	Adjusted R-Squared (%)	R (%)
0.4	0.011	0.072	65.06	81.25	0.010	0.063	70.02	83.09	0.010	0.065	68.80	83.19
0.7	0.021	0.093	77.75	89.72	0.021	0.088	79.37	90.53	0.018	0.086	80.30	90.84
1.0	0.043	0.120	84.82	93.54	0.038	0.111	86.52	94.22	0.041	0.111	85.04	93.80
1.3	0.087	0.159	83.85	92.17	0.091	0.159	82.81	91.88	0.097	0.162	81.52	91.64
1.6	0.140	0.212	84.67	93.45	0.151	0.215	82.67	91.75	0.163	0.221	81.14	92.06

For a better understanding of models' performance on an individual participant, we emphasized the residual plots (Y-axis) against predicted ankle joint power (X-axis), as portrayed in Fig. 7. For participants 3, 6, 7 and 8, the residuals seemed less spread and closer to 0, which explained that all three models obtained comparatively better performance for these 4 participants. For participants 2, 4 and 5, residuals plots were more scattered and few higher residuals indicated the existence of some noisy values as well as poor performance of the models. For all other participants, all three regressors generated average performance comparatively with the presence of some points exhibiting higher errors in predictions.

ii **Speed-wise Evaluation:** Comparative analysis of the proposed architectures was further investigated for different speed trials. As mentioned in section 2.A, each participant performed 5 walking trials with divergent speeds for each trial. Table 4 provides the averaged value of all evaluation metrics for each regressor for all five speed-trials. It was observed that all three regressors responded inadequately for trials with lower speeds (0.4 m/s and 0.7 m/s), but they performed better with increasing speeds (1.0 m/s,1.3 m/s, and 1.6 m/s). The reason behind this could be that the peak ankle joint power for lower speeds was not as large as it was for higher speeds. In predicting ankle joint power in lower speeds, LSTM performed poorer [R = 81.25%, 89.72%, and 93.54% for 0.4 m/s, 0.7 m/s and 1.0 m/s, respectively] in comparison with CNN [R = 83.09%, 90.53% and 94.22% for 0.4 m/s, 0.7 m/s and 1.0 m/s, respectively] and CNN-LSTM [R = 83.19%, 90.84% and R = 93.80% for 0.4 m/s, 0.7 m/s and 1.0 m/s, respectively]. Similar trends were observed for MSE and MAE at lower speeds. However, for higher speed-trials LSTM [R = 92.17% and 93.45% for 1.3 m/s and 1.6 m/s, respectively] outperformed CNN [R = 91.88%, and 91.75% for 1.3 m/s and 1.6 m/s, respectively] and CNN-LSTM [R = 91.64% and 92.06% for 1.3 m/s and 1.6 m/s, respectively]. Both MSE and MAE were noticed to have higher values when computed for dimensions with higher variances. We did not consider the Adjusted R-squared value in speed-trial comparisons as it could have a negative value for non-linear regression since there was no pre-defined intercept term. Bar plots of Fig.8a and Fig.8b showed the presence of outliers in adjusted R-squared value and R-value



(a) Residuals plot for participant 1

(b) Residuals plot for participant 2

(c) Residuals plot for participant3

(d) Residuals plot for participant 4

(e) Residuals plot for participant 5

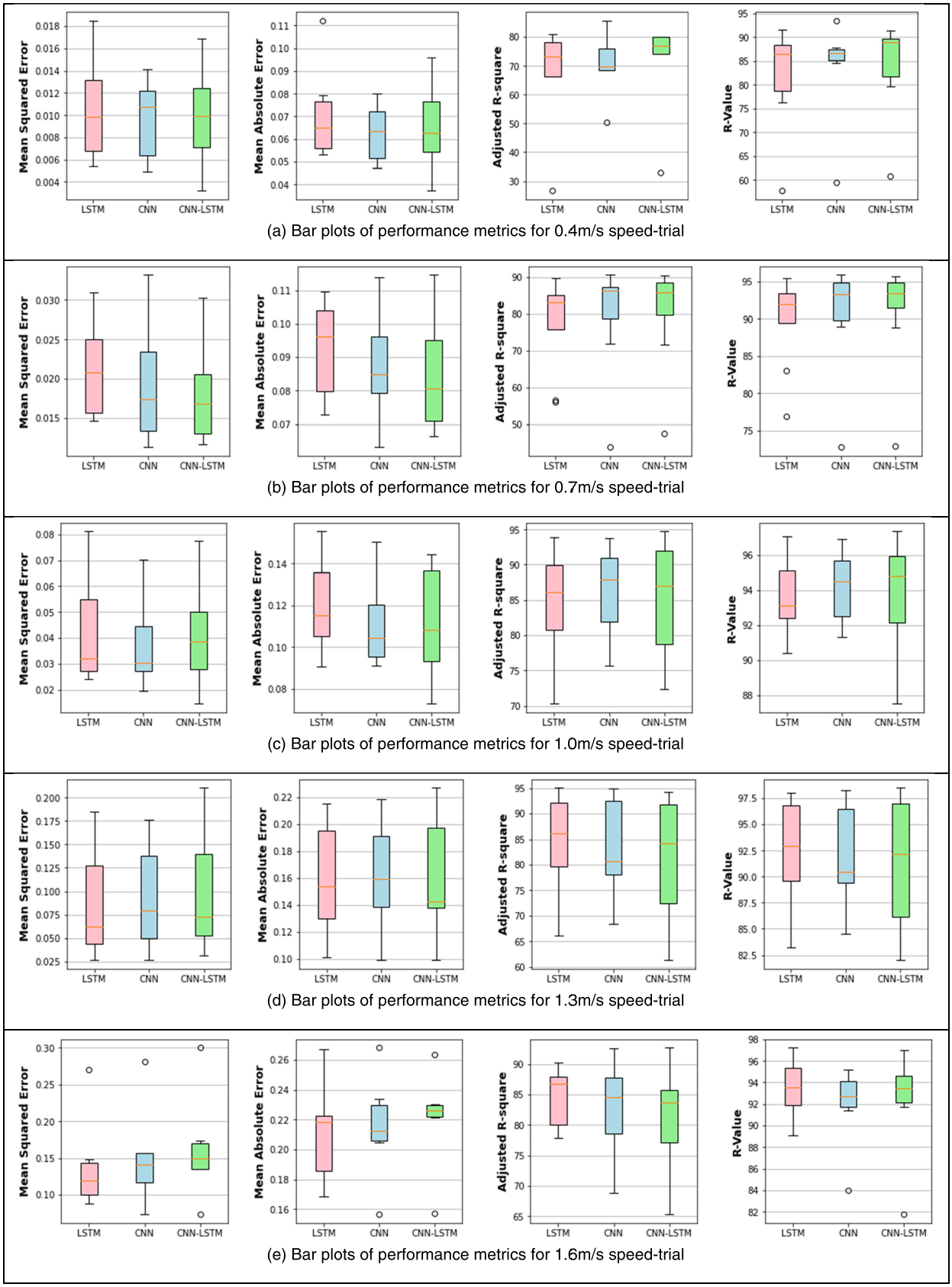
(f) Residuals plot for participant 6

(g) Residuals plot for participant 7

(h) Residuals plot for participant 8

(i) Residuals plot for participant 9

FIGURE 7. **Residual plots to exhibit widening of estimated residuals for all participant by each model.**



(a) Bar plots of performance metrics for 0.4m/s speed-trial

(b) Bar plots of performance metrics for 0.7m/s speed-trial

(c) Bar plots of performance metrics for 1.0m/s speed-trial

(d) Bar plots of performance metrics for 1.3m/s speed-trial

(e) Bar plots of performance metrics for 1.6m/s speed-trial

FIGURE 8. Bar plot of evaluation metrics for divergent speed-trials.

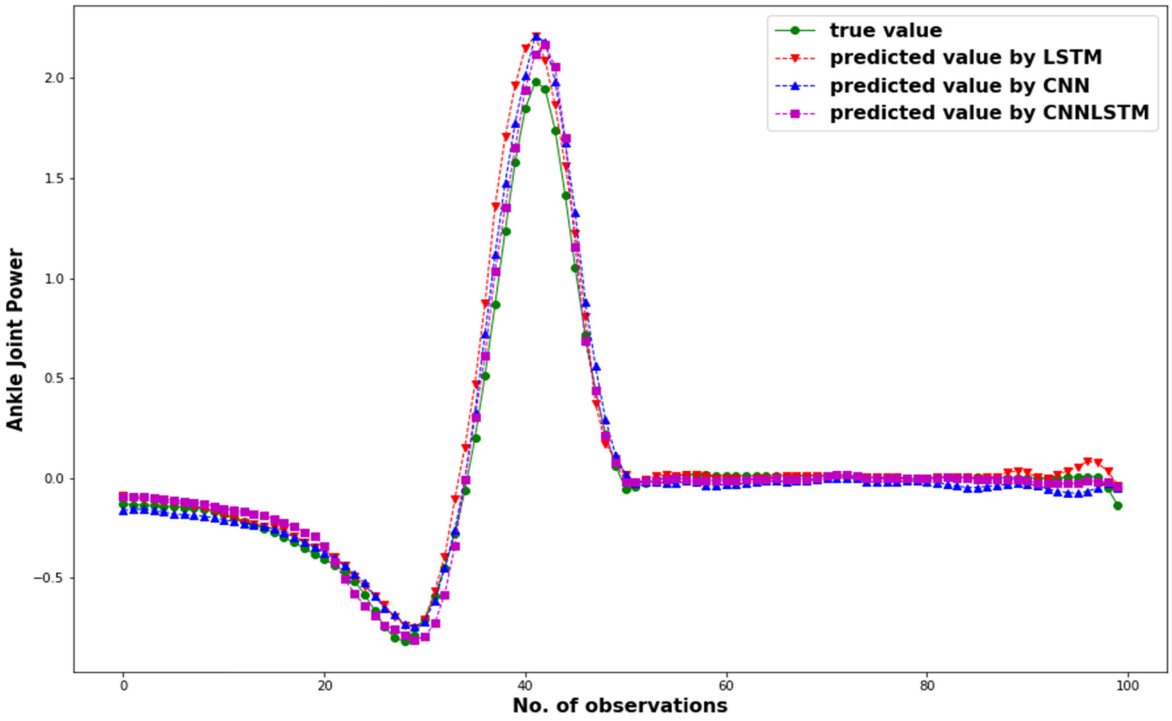


FIGURE 9. Closeness of ankle joint power predicted by the models to reference ankle joint power.

which caused the regressors to perform badly for few participants. Fig.8 [(a)-(e)] shows the speed trial responses of the three models. Besides, Fig. 9 provides a pictorial view of the closeness of predicted ankle joint power with reference values achieved by the proposed models.

IV. DISCUSSION

Deep learning architecture has been advanced so rapidly because of its' information productive nature which basic learning algorithms cannot offer. Prediction of ankle joint power using two IMUs were performed using Random Forest Regressor where an almost similar result was acquired ($R = 92\%$) [32]. However, to achieve this outcome, it was required to extract a bunch of features and required to make use of the Quantile Regression method to compensate for the error generated when predicting peak powers.

In this study, few extracted features from the raw IMU signals improved the proposed models' performances with higher ankle power estimation accuracies. During the study, the wearable IMUs were worn on and off which resembled practical scenarios. Such position shift of the IMUs introduced variations in the raw signals which could restrict model performances in inter-participant evaluations. Also, raw signals could result in poor prediction for the LSTM model. However, the CNN and the LSTM-CNN could predict well with raw signals ($R < 88\%$) but their performances improved significantly with extracted features like the LSTM ($R > 92\%$). Furthermore, the feature extraction process was simple and required minimal processing time (< 810 ms) which was negligible but worthy of implementation. Also, we could drastically reduce the required 256 extracted features mentioned in [32] to only 8 extracted features to implement the deep learning models. These features were simple and effective that resulted in higher estimation accuracies.

The speed-wise analysis showed that proposed models performed well for both lower and higher speeds. Although LSTM clearly outperformed CNN and CNN-LSTM in estimating ankle joint power for higher speed trials, it performed slightly inferior in lower speeds. So, speed-wise assessment demonstrated that prediction of ankle joint power was much facilitated if we took account of patterns (CNN) for lower speed trials and time-series nature (LSTM) for higher speed trials. In [32], the RF model could not predict better peak power during stance. Introducing such deep learning models could improve estimation accuracies even when participants walked in either lower or higher walking speeds with variations in peak power.

In our study extraction of lots of features was not required, rather we reduced the dimensions, and all three proposed frameworks performed effectively even in the prediction of peak powers without involving any further compensating algorithm. Table 3 shows that LSTM rendered better outcomes even for individual participants. Three out of the four metrics and positive skewness of the boxplots also demonstrated that LSTM exhibited better performances. Although CNN could not outperform LSTM, it had performed comparably to LSTM. In terms of standard deviation, MSE for both LSTM and CNN showed a comparatively lower spread

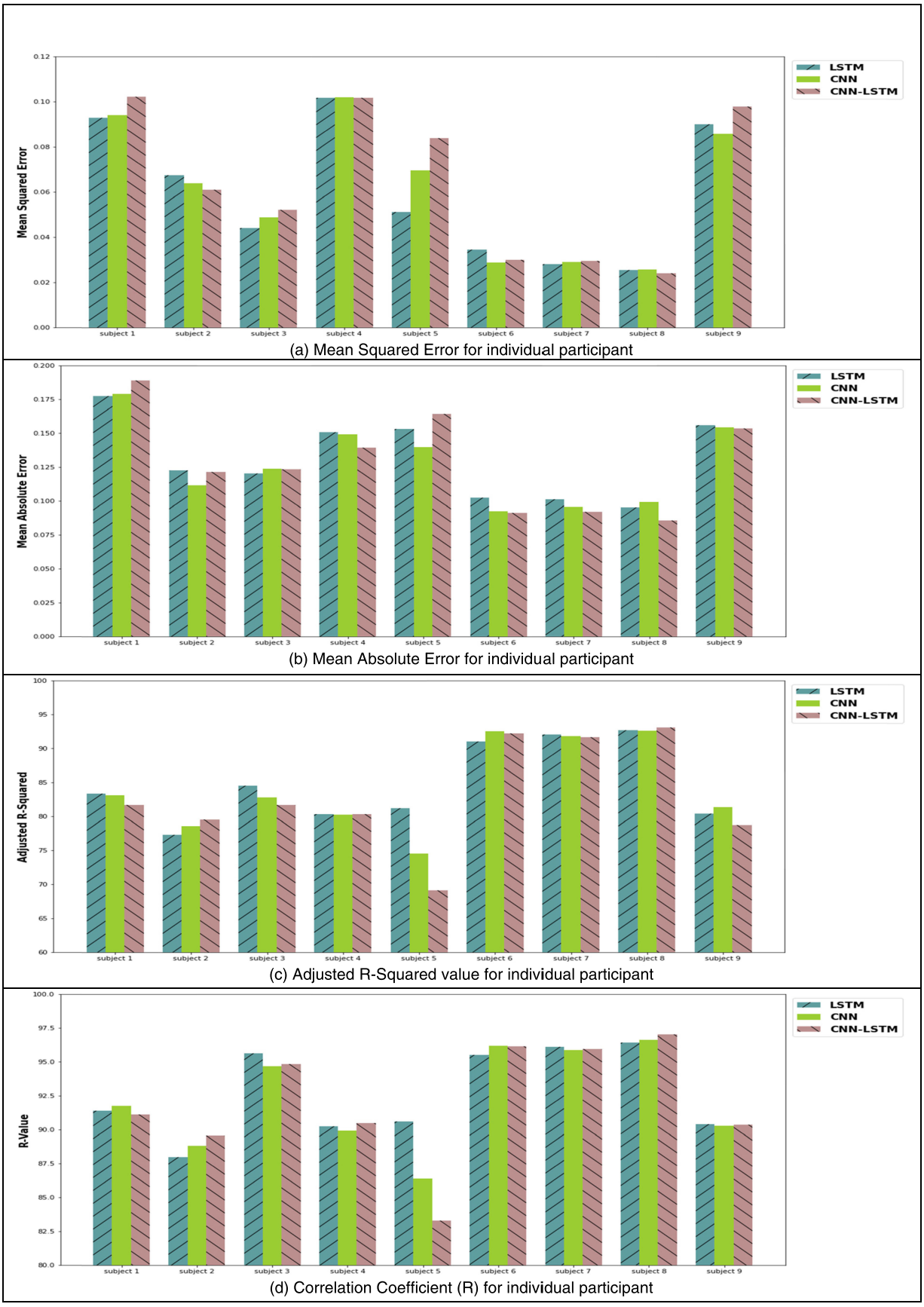


FIGURE 10. Comparative view of metrics values generated from each model type for individual participants.

0.029, which was almost similar for MAE too. Also, LSTM outperformed other models by exhibiting a lower standard deviation of 5.5% and 3.1% for Adjusted R-squared value and R-value, respectively. However, the hybrid model of CNN and LSTM could not render similar effectiveness. This lower potency could arise from the combination of both time-series and patterns in IMU signals, made the prediction process comparatively complex and lowered its performance for model generalization. In Fig. 10. although all three models performed poorly for participants 1, 2, 4, 5, and 9 [bar diagram of R-values], still LSTM seemed to be more effective for most of these participants, but further study is needed to support this leading nature of LSTM in the ankle joint power prediction.

V. LIMITATION AND FUTURE SCOPES

The employment of three divergent deep neural architectures to predict ankle joint power using two IMUs can help other researchers to assess their performance. Our study demonstrated that the prediction of ankle joint power can be assisted using both time-stamp and successive trends present in IMUs signals. The study data was accumulated in a confined environment where participants walked at some pre-defined speeds. Future studies can be conducted in an open environment with natural walking speed to investigate the models' performances in real-world scenarios. Moreover, the performance of each model on individual speed trials can also be assessed in the future to provide more comparative analysis among these proposed frameworks.

VI. CONCLUSION

Advancement in the prediction of ankle joint power with the instrumental set-up of lesser complexity can help in the development of applications associated with lower extremity monitoring appliances. Since ankle joint power can be a good indicator of players' movement in sports, estimating it with simple arrangement may lead to a facile method to observe players' activities associated with lower limbs. Besides, we evaluated our models by using the data amassed from a group of young people so the models can be employed to estimate the ankle joint power of players with fewer target training data using transfer learning. In our study, we exerted effort to accomplish such advancement by exhibiting the effectiveness of three deep neural architectures (LSTM, CNN, and CNN-LSTM). It was observed that deep neural frameworks could forecast ankle joint power with higher accuracy (maximum R = 92%) and lower error (minimum MSE = 0.059) employing only 8 simple features. In terms of inter-participant test, considering sensors positioning and shifting, it can be considered as significant progress and may pave other researchers to initiate further analysis.

REFERENCES

[1] D. A. Winter, "Kinetics: Force and movements of force," in *Biomechanics and Motor Control of Human Movement*. Hoboken, NJ, USA: Wiley, 1990.

[2] M. W. Whittle, "Gait analysis," *The Soft Tissues*. Amsterdam, The Netherlands: Elsevier, 1993, pp. 187–199.

[3] T. A. L. Wren, G. E. Gorton, S. Õunpuu, and C. A. Tucker, "Efficacy of clinical gait analysis: A systematic review," *Gait Posture*, vol. 34, no. 2, pp. 149–153, Jun. 2011.

[4] M. Plotnik, N. Giladi, and J. M. Hausdorff, "A new measure for quantifying the bilateral coordination of human gait: Effects of aging and Parkinson's disease," *Exp. Brain Res.*, vol. 181, no. 4, pp. 561–570, Jul. 2007.

[5] D. H. Sutherland and J. R. Davids, "Common gait abnormalities of the knee in cerebral palsy," *Clin. Orthopaedics Rel. Res.*, vol. 288, no. 288, pp. 139–147, Mar. 1993.

[6] J.-M. Belda-Lois, S. Mena-del Horno, I. Bermejo-Bosch, J. C. Moreno, J. L. Pons, D. Farina, M. Iosa, M. Molinari, F. Tamburella, A. Ramos, A. Caria, T. Solis-Escalante, C. Brunner, and M. Rea, "Rehabilitation of gait after stroke: A review towards a top-down approach," *J. Neuroeng. Rehabil.*, vol. 8, no. 1, p. 66, 2011.

[7] D. Gouwanda and S. Senanayake, "Emerging trends of body-mounted sensors in sports and human gait analysis," in *Proc. 4th Kuala Lumpur Int. Conf. Biomed. Eng.*, 2008, pp. 715–718.

[8] P. Zell and B. Rosenhahn, "A physics-based statistical model for human gait analysis," in *Proc. German Conf. Pattern Recognit.*, 2015, pp. 169–180.

[9] M. Perc, "The dynamics of human gait," *Eur. J. Phys.*, vol. 26, no. 3, p. 525, 2005.

[10] R. A. Bogey, A. J. Gitter, and L. A. Barnes, "Determination of ankle muscle power in normal gait using an EMG-to-force processing approach," *J. Electromyogr. Kinesiol.*, vol. 20, no. 1, pp. 46–54, Feb. 2010.

[11] X. Jiang, K. Chu, M. Khoshnam, and C. Menon, "A wearable gait phase detection system based on force myography techniques," *Sensors*, vol. 18, no. 4, p. 1279, Apr. 2018.

[12] A. K. Godiyal, H. K. Verma, N. Khanna, and D. Joshi, "A force myography-based system for gait event detection in overground and ramp walking," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 10, pp. 2314–2323, Oct. 2018.

[13] S. Chakraborty, A. Nandy, T. Yamaguchi, V. Bonnet, and G. Venture, "Accuracy of image data stream of a markerless motion capture system in determining the local dynamic stability and joint kinematics of human gait," *J. Biomech.*, vol. 104, May 2020, Art. no. 109718.

[14] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster, "Full body gait analysis with Kinect," in *Proc. Annu. Int. Conf.IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 1964–1967.

[15] C. Prakash, R. Kumar, and N. Mittal, "Recent developments in human gait research: Parameters, approaches, applications, machine learning techniques, datasets and challenges," *Artif. Intell. Rev.*, vol. 49, no. 1, pp. 1–40, Jan. 2018.

[16] R. Luo, S. Sun, X. Zhang, Z. Tang, and W. Wang, "A low-cost end-to-end sEMG-based gait sub-phase recognition system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 267–276, Jan. 2020.

[17] A. Muro-De-La-Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla, "Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications," *Sensors*, vol. 14, no. 2, pp. 3362–3394, Feb. 2014.

[18] W. Tao, T. Liu, R. Zheng, and H. Feng, "Gait analysis using wearable sensors," *Sensors*, vol. 12, no. 2, pp. 2255–2283, Feb. 2012.

[19] P. B. Shull, W. Jirattigalachote, M. A. Hunt, M. R. Cutkosky, and S. L. Delp, "Quantified self and human movement: A review on the clinical impact of wearable sensing and feedback for gait analysis and intervention," *Gait Posture*, vol. 40, no. 1, pp. 11–19, May 2014.

[20] A. Kalron, Z. Dvir, L. Frid, and A. Achiron, "Quantifying gait impairment using an instrumented treadmill in people with multiple sclerosis," *ISRN Neurol.*, vol. 2013, pp. 1–6, Jun. 2013.

[21] S. C. Wearing, L. F. Reed, and S. R. Urry, "Agreement between temporal and spatial gait parameters from an instrumented walkway and treadmill system at matched walking speed," *Gait Posture*, vol. 38, no. 3, pp. 380–384, Jul. 2013.

[22] L. Vargas-Valencia, A. Elias, E. Rocon, T. Bastos-Filho, and A. Frizera, "An IMU-to-body alignment method applied to human gait analysis," *Sensors*, vol. 16, no. 12, p. 2090, Dec. 2016.

[23] A. R. Anwary, H. Yu, and M. Vassallo, "Optimal foot location for placing wearable IMU sensors and automatic feature extraction for gait analysis," *IEEE Sensors J.*, vol. 18, no. 6, pp. 2555–2567, Mar. 2018.

[24] N. Abhayasinghe and I. Murray, "Human gait phase recognition based on thigh movement computed using IMUs," in *Proc. IEEE 9th Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process. (ISSNIP)*, Apr. 2014, pp. 1–4.

[25] M. Nagashima, S.-G. Cho, M. Ding, G. A. G. Ricardez, J. Takamatsu, and T. Ogasawara, "Prediction of plantar forces during gait using wearable sensors and deep neural networks," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 3629–3632.
[26] F. Young, G. Coulby, I. Watson, C. Downs, S. Stuart, and A. Godfrey, "Just find it: The Mymo approach to recommend running shoes," *IEEE Access*, vol. 8, pp. 109791–109800, 2020.
[27] M. Lempereur, F. Rousseau, O. Rémy-Néris, C. Pons, L. Houx, G. Quellec, and S. Brochard, "A new deep learning-based method for the detection of gait events in children with gait disorders: Proof-of-concept and concurrent validity," *J. Biomech.*, vol. 98, Jan. 2020, Art. no. 109490.
[28] M. D. Nguyen, K.-R. Mun, D. Jung, J. Han, M. Park, J. Kim, and J. Kim, "IMU-based spectrogram approach with deep convolutional neural networks for gait classification," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2020, pp. 1–6.
[29] A. Vaith, B. Taetz, and G. Bleser, "Uncertainty based active learning with deep neural networks for inertial gait analysis," in *Proc. IEEE 23rd Int. Conf. Inf. Fusion (FUSION)*, Jul. 2020, pp. 1–8.
[30] O. Dehzangi, M. Taherisadr, and R. ChangalVala, "IMU-based gait recognition using convolutional neural networks and multi-sensor fusion," *Sensors*, vol. 17, no. 12, p. 2735, Nov. 2017.
[31] S. Davarzani, D. Saucier, P. Peranich, W. Carroll, A. Turner, E. Parker, C. Middleton, P. Nguyen, P. Robertson, B. Smith, J. Ball, R. Burch, H. Chander, A. Knight, R. Prabhu, and T. Luczak, "Closing the wearable gap—Part VI: Human gait recognition using deep learning methodologies," *Electronics*, vol. 9, no. 5, p. 796, May 2020.
[32] X. Jiang, M. Gholami, M. Khoshnam, J. J. Eng, and C. Menon, "Estimation of ankle joint power during walking using two inertial sensors," *Sensors*, vol. 19, no. 12, p. 2796, Jun. 2019.
[33] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 374, no. 2065, Apr. 2016, Art. no. 20150202.
[34] E. Bisong, "Google colaboratory," in *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. New York, NY, USA: Springer, 2019, pp. 59–64.

ARNAB BARUA was born in Chittagong, Bangladesh, in 1995. He received the B.Sc. degree in computer science from International Islamic University Chittagong, in 2018. He is currently pursuing the M.Sc. degree in computer science with the Memorial University of Newfoundland.

Since 2020, he has been working as a Research Assistant with the Xianta Jiang's Laboratory. He has contributed to researches associated with human activity and the author of multiple publications related to human activity. His research interests include improvement of human activity recognition and gait analysis using divergent applications of machine learning.

UMME ZAKIA (Student Member, IEEE) received the B.Sc. degree in electronics and computer science from Jahangirnagar University, Bangladesh, in 2001, and the M.Sc. degree in computer science and engineering from North South University, Bangladesh, in 2007. She is currently pursuing the Ph.D. degree in engineering science with Simon Fraser University, BC, Canada. She is also working as a Research Assistant with the Menrva Research Group. She has published several articles in peer-reviewed journals and conferences. Her research interests include machine learning, robotics, HRI, communication networks, and VLSI.

CARLO MENON (Senior Member, IEEE) received the Laurea degree in mechanical engineering from the University of Padua, Italy, in 2001, and the Ph.D. degree from the University of Padua, in 2005. From 2005 to 2006, he was a Research Fellow with the European Space Agency, The Netherlands. In 2007, he joined Simon Fraser University (SFU), BC, Canada, and founded the Menrva Research Group. He is currently a Professor with ETH Zurich and leads the Biomedical and Mobile Health Technology Laboratory. He is also an Adjunct Professor with the Schools of Mechatronic Systems and Engineering Science, SFU. He has published over 300 articles, including both journal articles and conference papers. He received several research and career awards, including the International IAF Luigi G. Napolitano Award, the International BIONIS Award, the Career Investigator Award from the Michael Smith Foundation for Health Research (MSFHR), the New Investigator Award from the Canadian Institutes of Health Research (CIHR), and the Tier 1 Canada Research Chair (CRC).

XIANTA JIANG (Member, IEEE) received the dual Ph.D. degree in computer science from Simon Fraser University and Zhejiang University, in 2015, where his thesis research focused on mental workload assessment using eye-tracking technology. He completed a postdoctoral fellowship at the MENVA Research Group, Engineering Science, Simon Fraser University, working on the area of detecting human activity by processing muscle activities data with a machine learning algorithm. He latterly worked as a Senior Research Associate with the Surgical Simulation Research Laboratory (SSRL), Department of Surgery, University of Alberta, studying team cognition using augmented reality (AR) and eye-tracking technologies. Since June 2019, he holds an assistant professor position with the Department of Computer Science, MUN. His research interests include intelligent human–machine interaction, ubiquitous computing in healthcare, human behavior recognition, bio-signal processing, eye-tracking, pupil diameter, and human factors.