BiLSTM-Based Joint Torque Prediction From Mechanomyogram During Isometric Contractions: A Proof of Concept Study

Jongsang Son[ID], *Member, IEEE*, Fandi Shi[ID], and William Zev Rymer[ID], *Life Member, IEEE*

Abstract—**Quantifying muscle strength is an important measure in clinical settings; however, there is a lack of practical tools that can be deployed for routine assessment. The purpose of this study is to propose a deep learning model for ankle plantar flexion torque prediction from time-series mechanomyogram (MMG) signals recorded during isometric contractions (i.e., a similar form to manual muscle testing procedure in clinical practice) and to evaluate its performance. Four different deep learning models in terms of model architecture (based on a stacked bidirectional long short-term memory and dense layers) were designed with different combinations of the number of units (from 32 to 512) and dropout ratio (from 0.0 to 0.8), and then evaluated for prediction performance by conducting the leave-one-subject-out cross-validation method from the 10-subject dataset. As a result, the models explained more variance in the untrained test dataset as the error metrics (e.g., root-mean-square error) decreased and as the slope of the relationship between the measured and predicted joint torques became closer to 1.0. Although the slope estimates appear to be sensitive to an individual dataset, > 70% of the variance in nine out of 10 datasets was explained by the optimal model. These results demonstrated the feasibility of the proposed model as a potential tool to quantify average joint torque during a sustained isometric contraction.**

Index Terms—**Manual muscle testing, muscle/joint strength, machine/deep learning model, clinical assessment.**

I. INTRODUCTION

MUSCLE strength is considered an important measure in our clinics, as significantly reduced muscle strength is closely associated with severe motor impairments across a

wide range of clinical populations [1], [2], [3], [4]. In fact, previous studies have demonstrated a strong correlation between muscle strength and reduced motor function in a variety of activities such as balance, walking, reaching, and grasping [5]. This underscores the significance of assessing and quantifying muscle strength in routine clinical practice so that clinicians can not only monitor the progress of rehabilitation over time but also tailor rehabilitation strategies based on the individual's function.

Manual muscle testing (MMT), using a 6-point grading scale, is widely considered to be the gold standard to assess muscle strength in real clinical settings. However, the MMT scale is quite subjective and less reliable probably due to differences in experience and strength of examiners [6], [7]. To address these limitations, some investigators have attempted to use a dynamometer (e.g., an isokinetic device) or a hand-held device, which is available to better quantify muscle strength [8], [9]. Unfortunately, such approaches appear to be impractical for a routine assessment in real-life clinical settings. Thus, there is a need of an alternative, more practical approach for quantifying muscle strength-related measures.

It has been widely accepted that the electromyogram (EMG) is closely related to muscle force [10]. Indeed, several approaches have leveraged the EMG signals to estimate muscle force [11], [12]. However, the acquisition of EMG signals can be challenging due to technical constraints; for example, proper placement of the EMG electrodes, careful skin preparation and management of electrode-skin contact, and environmental noise (for review, see [13]). In addition to these technical challenges, other barriers can limit its routine application in real clinical settings [14], [15].

For the past several decades, the mechanomyogram (MMG) has been used to record small vibration signals from skeletal muscle. Skeletal muscle is known to generate low-frequency mechanical vibrations (< 100 Hz) during muscle contractions, which are presumably linked to changes in the shape of muscle fibers (e.g., cross-sectional area and length) in response to neural excitation of active motor units (for review, see [16]). Given that such mechanical signals propagate through soft tissues, MMG can provide some advantages over EMG, in terms of the sensor location, sensor-skin contact, and environmental noise [17], [18], [19].

Along with these advantages, a large number of studies have demonstrated significant relationships between several features (e.g., root-mean-square (RMS) MMG, median/mean frequency, or zero-crossing of the MMG signals) and mean muscle force/joint torque during a sustained isometric contraction [20], [21], [22], [23], [24]. Moreover, there have been attempts to estimate joint torque from some selected features of the MMG signals based on the machine learning models, demonstrating a reasonably good performance [25], [26], [27], [28]. However, in most of the previous studies, data collected from the same participant were used for both training and test datasets, and it is thus uncertain whether the models would perform well with untrained test data from a new individual. Given that there is no consistent relationship established, it is plausible that such features may not always contain key necessary kinetic information required for predicting muscle force/joint torque accurately. In addition, since a slope of the relationships between observed and predicted values has not been often reported, there is a lack of information on whether the previous models underpredict or overpredict the outcomes.

Recently, deep learning models have been extensively deployed in many applications, likely due to their ability to automatically extract features from raw sensor signals. For example, some studies demonstrated a deep learning model application with features from a few seconds of raw (or processed) MMG signals. However, most studies sought to address human movement recognition or muscle action recognition [29], [30].

Thus, the purpose of this proof-of-concept study is to propose a deep learning model for ankle plantar flexion torque prediction from MMG signals and to evaluate its performance. Considering that an isometric contraction protocol in research is a similar form to the MMT procedure in clinical practice, our deep learning model was designed to predict a single numeric value corresponding to an average ankle plantar flexion torque for a given time-series MMG signals during isometric plantar flexion contraction.

II. METHODS

A. Dataset

This study used the dataset collected from our previous study with 10 healthy young individuals (age: 25.8 ± 2.2 years; height: 170.0 ± 10.3 cm; body mass: 66.9 ± 12.5 kg; F/M: 4/6) without any relevant medical history of neuromuscular diseases in their lower limb [24]. Briefly, the participants performed three isometric plantar flexion contractions at each of different contraction intensities, i.e., 10%, 20%, 30%, 50%, 70%, and 100% of maximum voluntary isometric contraction (MVIC), and at each of different ankle joint angles, i.e., plantar flexion of 26°, plantar flexion of 10°, and dorsiflexion of 3° based on the neutral position (0° defined as perpendicular between the shank and sole). Given that the average maximum ankle plantar flexion torque was observed at plantar flexion of 10°, only the data collected at plantar flexion of 10° were used for this study. During the contractions, ankle torque signals (in N m) were recorded, using a 6-axis force measuring device (Omega160, ATI Industrial Automation, Apex, NC, USA), and MMG signals (in mm/s²)
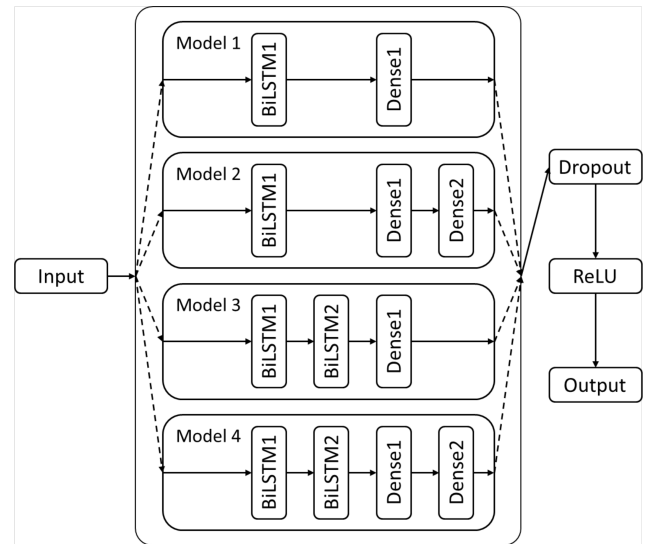

Fig. 1. Architecture of the proposed bidirectional long short-term memory (BiLSTM)-based models for a joint torque prediction from time-series 3-axis MMG signals during a sustained isometric contraction. Four different models are based on a stacked BiLSTM and dense layers. The hyperparameter values for the number of units of the BiLSTM and dense layers and the dropout ratio of the dropout layer are summarized in Table I.

were collected from the medial gastrocnemius (MG) muscle, using a 3-axis accelerometer with z-direction perpendicular to the skin surface over MG muscle (ADXL335, Analog Devices, Wilmington, MA, USA; sensitivity: 300 mV/g). These signals were collected at a sampling frequency of 2 kHz (NI USB-6259 BNC, National Instrument, Austin, TX, USA). All procedures described above were approved by the Northwestern University's Institutional Review Board, and the subjects gave informed consent forms before participating in this study.

B. Deep Learning Model Architecture

Since our application is to predict an average ankle plantar flexion torque value from time-series MMG signals during isometric plantar flexion contraction for typically 3-10 seconds (i.e., many-to-one regression problem), our deep learning model architecture was based on a stacked bidirectional long short-term memory (BiLSTM) and dense layers (Figure 1). Typically, input sequential data for BiLSTM are supposed to have the same data length (i.e., the same time step) across trials, and were thus zero-padded to the maximum length among all of the trials. A masking layer was added to use original sequential data for training a model. Afterward, as summarized in Table I, different combinations of BiLSTM, dense, and dropout layers were implemented to explore the effects of the model architecture and hyperparameters on the performance of the ankle plantar flexion torque prediction. A rectified linear unit activation layer and a dense layer with one unit were then added to output a single, positive plantar flexion torque value.

C. Data Preprocessing

The torque signals were lowpass filtered using a zero-phase 4th order Butterworth filter with a cutoff frequency

TABLE I
MODEL ARCHITECTURES

	Number of units				Ratio
	BiLSTM1 (n)	BiLSTM2	Dense1	Dense2	Dropout
Model 1	32, 64, 128, 256, or 512	N/A	n	N/A	0.0, 0.2, 0.5, or 0.8
Model 2		N/A		n / 2	
Model 3		n / 2		N/A	
Model 4		n / 2		n / 2	

of 6 Hz. For each trial, a 4-s segment in the middle of the sustained contraction was automatically selected with the minimum standard deviation of the segment, and the mean ankle plantar flexion torque value of the segment was then calculated. These mean values were normalized to body mass, and the normalized mean torque data (i.e., N m/kg) were used as output data of the deep learning models.

The 3-axis MMG signals were bandpass filtered using a zero-phase 4th order Butterworth filter with a passband of 4–40 Hz. In order to improve the performance of the trained models, data augmentation was applied to the filtered signals. Nine random durations ranging between 0 and 3 s with an interval of 0.1 s were chosen, and the data points corresponding to each of the durations from the beginning and end of the signals were removed (e.g., for a given 1-s duration, the first 1-s data and the last 1-s data were removed), leading to 10 different filtered signals (i.e., original plus nine random durations). A moving RMS filter with a 100-ms window size was applied to each axis of the filtered MMG signals. Principal component analysis was conducted to determine the scores of the moving RMS-filtered signals, and the three time-series scores were then downsampled 10 Hz to ensure that the maximum data length is less than 300 (i.e., up to 30-s long data). Finally, these processed signals were used as input data for the deep learning models. This preprocessing process was performed in MATLAB (R2022b, The MathWorks, Inc., Natick, MA, USA).

D. Model Training and Evaluation

The proposed models were implemented, trained, and tested in Python 3 using Keras (v2.10.0) [31] and TensorFlow (v2.10.0) [32]. Most of the layer arguments were not defined, so default arguments were used. Only the Dense1 and Dense2 layers in Figure 1 had an L2 kernel regularizer with a regularization factor of 0.01. The adaptive momentum (Adam) optimization algorithm was used to minimize the loss function (i.e., mean squared error) during the training of the proposed models. The batch size and maximum number of epochs were set to 32 and 500, respectively. During the training process, the validation loss was calculated at the end of each epoch, and the model with the best validation loss out of 500 epochs was saved as the final model.

Given that the purpose of this study is to develop a deep learning model that predicts an average ankle plantar flexion torque value for given time-series 3-axis MMG signals from a new individual, the prediction performance of each model was evaluated by conducting a leave-one-subject-out cross-validation (LOSOCV) method. The data from all subjects except one (i.e., nine subjects in this study) were randomly

separated into training set (80%) and validation set (20%) and were then used to train a model. The remaining data from the left-out subject were used only to test the model prediction performance on untrained test data. This procedure was applied to each subject, resulting in 10 different subject models for each combination of the hyperparameters.

The mean absolute error (MAE), RMS error (RMSE), and RMSE normalized to interquartile range (RMSE_{IQR}) between the measured and predicted normalized mean torque data were calculated as performance metrics to assess the prediction performance of each trained model. Moreover, linear regression analyses were conducted to determine the slope and coefficient of determination (R^2) of relationships between the measured and predicted normalized mean torque data. Briefly, as the predicted value is closer to the measured value, both slope and R^2 values are closer to 1.0. Considering that these performance metrics may vary according to different datasets randomly separated into training and validation, the LOSOCV procedure for each subject model was repeated 10 times (i.e., a repeated random sub-sampling method) and the average value of the performance metrics from the 10 models was then used as a representative prediction performance for each subject model. Finally, the prediction performance for each combination of the hyperparameters was assessed by the overall mean and standard deviation of the representative performance metrics across the 10 different subject models for each combination of the hyperparameters. These model evaluations were done in MATLAB (R2022b, The MathWorks, Inc., Natick, MA, USA).

Three-way analysis of variance (ANOVA) was conducted to evaluate the main effects of the model architecture, the number of units, and the dropout ratio as well as the three two-way interaction effects on each of the aforementioned performance metrics. Multiple comparisons were done with Tukey's HSD test. These statistical analyses were performed using JMP Pro (16, JMP Statistical Discovery LLC., Cary, NC, USA) with a significance level (α) of 0.05.

III. RESULTS

The average MAE, RMSE, and RMSE_{IQR} across the models were 0.021 ± 0.016, 0.028 ± 0.021, and 0.066 ± 0.050 for training dataset, 0.033 ± 0.016, 0.054 ± 0.019, and 0.129 ± 0.044 for validation dataset, and 0.219 ± 0.006, 0.281 ± 0.007, and 0.668 ± 0.016 for untrained test dataset, respectively. The linear regression analyses demonstrated that the average slope and R^2 were close to 1 for the training dataset (slope: 0.962 ± 0.064, R^2: 0.995 ± 0.005) and for the validation dataset (slope: 0.937 ± 0.063, R^2: 0.976 ± 0.013). Based on the average slope value of 0.830 ± 0.062 and R^2 value of 0.684 ± 0.032 for the untrained test dataset, it appears that the models predict the normalized torque data reasonably well, although the models slightly underpredicted the joint torques.

In general, the ANOVA results demonstrated a significant interaction between the model architecture and the number of units and between the number of units and the dropout ratio for all the performance metrics ($p < 0.05$). The interaction between the model architecture and the dropout ratio was significant for MAE ($p = 0.013$) and slope ($p = 0.016$) but
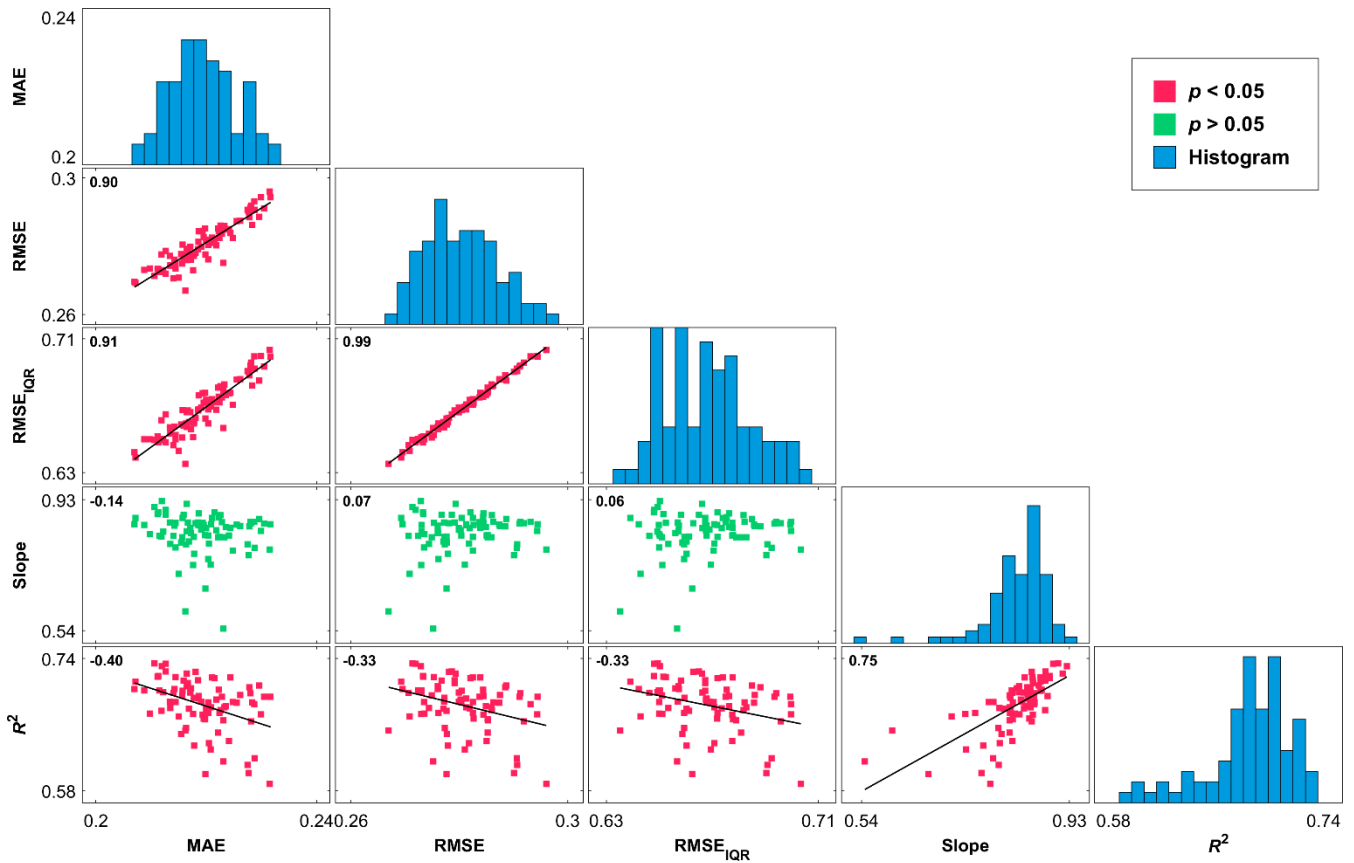
Fig. 2. Correlation matrix of performance metrics including mean average error (MAE), root-mean-square error (RMSE), RMSE normalized to interquartile range (RMSE$_{IQR}$), and slope (Slope) and coefficient of determination (R^2) of the regression between predicted and measured ankle plantar flexion torque. A significance level of the correlation is presented as a different marker color: red ($p < 0.05$) or green ($p > 0.05$). The diagonal elements present the histogram of the performance metrics.

TABLE II
PERFORMANCE METRICS BY MODEL ARCHITECTURE

	MAE	RMSE	RMSE$_{IQR}$	Slope	R^2
Model 1	0.219[b,c,d]	0.281[b]	0.670[b]	0.819[c,d]	0.662[c,d]
	(0.005)	(0.006)	(0.014)	(0.052)	(0.039)
Model 2	0.216[a,c,d]	0.276[a,c,d]	0.657[a,c,d]	0.810[c,d]	0.683
	(0.005)	(0.005)	(0.013)	(0.077)	(0.033)
Model 3	0.222[a,b]	0.283[b]	0.673[b]	0.849[a,b]	0.693[a]
	(0.005)	(0.006)	(0.014)	(0.037)	(0.017)
Model 4	0.221[a,b]	0.282[b]	0.672[b]	0.842[a,b]	0.696[a]
	(0.007)	(0.008)	(0.019)	(0.070)	(0.024)

Model 1: models with 1 BiLSTM and 1 Dense layers; Model 2: models with 1 BiLSTM and 2 Dense layers; Model 3: models with 2 BiLSTM and 1 Dense layers; Model 4: models with 2 BiLSTM and 2 Dense layers. a, b, c, and d indicate a significant difference with Model 1, 2, 3, and 4, respectively.

TABLE III
PERFORMANCE METRICS BY NUMBER OF UNITS

	MAE	RMSE	RMSE$_{IQR}$	Slope	R^2
NU 32	0.224[b,d,e]	0.284[b,d,e]	0.677[b,d,e]	0.772[b,c,d,e]	0.644[b,c,d,e]
	(0.004)	(0.008)	(0.019)	(0.098)	(0.035)
NU 64	0.220[a,e]	0.281[a,e]	0.668[a,e]	0.817[a,c,d,e]	0.680[a,e]
	(0.005)	(0.007)	(0.017)	(0.045)	(0.022)
NU 128	0.222[d,e]	0.282[e]	0.672[e]	0.836[a,b,e]	0.691[a]
	(0.006)	(0.006)	(0.013)	(0.027)	(0.018)
NU 256	0.219[a,c,e]	0.280[a,e]	0.667[a,e]	0.850[a,b,e]	0.696[a]
	(0.005)	(0.005)	(0.013)	(0.026)	(0.023)
NU 512	0.212[a,b,c,d]	0.276[a,b,c,d]	0.656[a,b,c,d]	0.876[a,b,c,d]	0.707[a,b]
	(0.004)	(0.004)	(0.010)	(0.030)	(0.018)

NU x: models with the number of units of x. a, b, c, d, and e indicate a significant difference with NU 32, 64, 128, 256, and 512, respectively.

not for RMSE ($p = 0.076$), RMSE$_{IQR}$ ($p = 0.087$), and R^2 ($p = 0.160$).

A. Effects of Model Architecture on Prediction Performance

Table II summarizes performance metrics for the model architectures. There was a significant difference in the average values of MAE, RMSE, RMSE$_{IQR}$, slope, and R^2 among four different model architectures ($p < 0.0001$). The models with 1 BiLSTM and 2 Dense layers (Model 2) yielded the smallest average values of MAE, RMSE, and RMSE$_{IQR}$ ($p < 0.05$). The models with 2 BiLSTM and 1 Dense layer (Model

3) yielded the best slope of 0.849 ± 0.037. Model 3 and the models with 2 BiLSTM and 2 Dense layers (Model 4) yielded a significantly better slope value than the models with 1 BiLSTM and 1 Dense layer (Model 1) and Model 2 ($p < 0.05$) as well as a significantly higher R^2 value than Model 1 ($p < 0.05$).

B. Effects of Number of Units on Prediction Performance

Table III summarizes performance metrics for the models with different numbers of units. There was a significant difference in the average values of MAE, RMSE, RMSE$_{IQR}$, slope, and R^2 among the models with five different numbers
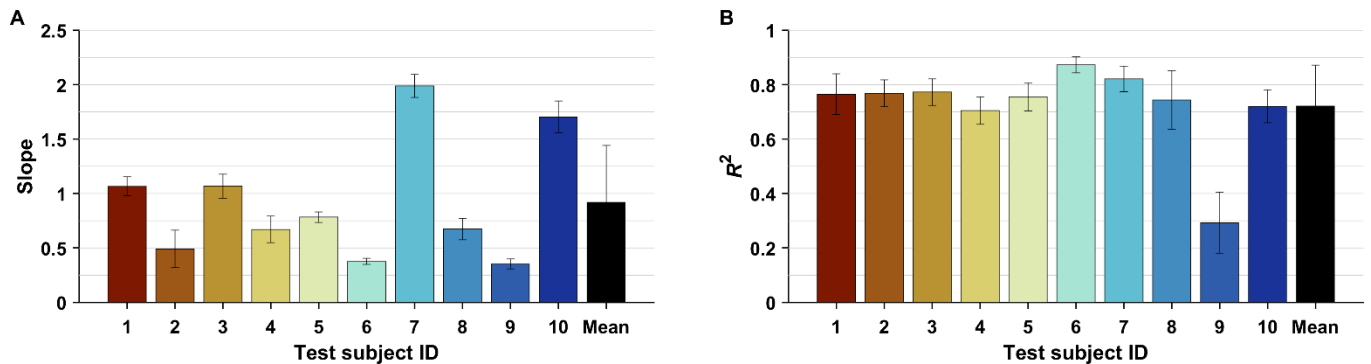
Fig. 3. Slope and coefficient of determination (R^2) of the regression between predicted and measured ankle plantar flexion torque by conducting a leave-one-subject-out cross-validation method.

TABLE IV
PERFORMANCE METRICS BY DROPOUT RATIO

	MAE	RMSE	$RMSE_{IQR}$	Slope	R^2
DR 0.0	0.219[b,c,d] (0.006)	0.280[b,c,d] (0.006)	0.666[b,c,d] (0.015)	0.857[d] (0.043)	0.701[b,c,d] (0.034)
DR 0.2	0.222[a,d] (0.006)	0.283[a,d] (0.007)	0.675[a,d] (0.015)	0.853[d] (0.033)	0.686[a,d] (0.032)
DR 0.5	0.221[a,d] (0.006)	0.283[a,d] (0.006)	0.673[a,d] (0.015)	0.841[d] (0.032)	0.680[a,d] (0.026)
DR 0.8	0.216[a,b,c] (0.005)	0.276[a,b,c] (0.005)	0.657[a,b,c] (0.012)	0.701[a,b,c] (0.083)	0.668[a,b,c] (0.028)

DR x: models with the dropout ratio of x. The superscripts a, b, c, and d indicate a significant difference with DR 0.0, 0.2, 0.5, and 0.8, respectively.

of units ($p < 0.0001$). There was a general trend of improved performance metrics with an increase in the number of units. On average, the models with the number of units of 512 yielded the smallest values of MAE, RMSE, and $RMSE_{IQR}$ ($p < 0.05$), demonstrating the best slope (0.876 ± 0.030) and R^2 (0.707 ± 0.018) values.

C. Effects of Dropout Ratio on Prediction Performance

Table IV summarizes performance metrics for the models with different dropout ratios. The models with each of the four different dropout ratios revealed a significant difference in the average values of MAE, RMSE, $RMSE_{IQR}$, slope, and R^2 ($p < 0.0001$). The models with the dropout ratio of 0.8 (DR 0.8) yielded the smallest average values of MAE, RMSE, and $RMSE_{IQR}$ compared to the others ($p < 0.05$), but the average values of the slope (0.771 ± 0.083) and R^2 (0.668 ± 0.028) were also significantly smaller ($p < 0.05$). On average, the models with the dropout ratio of 0.0 predicted the best slope (0.857 ± 0.043) and best explained the variance in the normalized mean torque data by ~70% ($p < 0.05$), yielding comparable average values of MAE, RMSE, and $RMSE_{IQR}$ to DR 0.8.

D. Summary of Prediction Performance Across Models

The correlation matrix revealed that across the models, there are strong correlations ($r > 0.9$; $p < 0.001$) between error metrics (i.e., MAE, RMSE, and $RMSE_{IQR}$) and moderate negative correlation ($r < -0.3$; $p < 0.01$) of R^2 with the error metrics (Figure 2). The slope-R^2 relationship was significant ($p < 0.001$), but the slope did not significantly correlate with

the error metrics ($p > 0.05$). These correlations indicate that R^2 increases likely as the error metrics decrease and as the slope becomes closer to 1.0. Interestingly, the R^2-slope plot demonstrates that most models predicted the mean torque value for the untrained test dataset with a reasonable slope (i.e., close to 1) and a moderate to strong effect size (i.e., $R^2 > 0.65$).

Based on the performance metrics, the model with 2 BiLSTM and 1 Dense layers with the number of units of 512 and the dropout ratio of 0.0 was selected as an optimal model for this study. Although the overall mean slope was close to 1.0, the slope for each subject dataset was not always close to 1.0, ranging between 0.35 and 1.99 (Figure 3A). However, the prediction from the selected optimal model had a strong effect size for the untrained test dataset (Figure 3B), explaining >70% of the variance in the normalized plantar flexion torque except one subject dataset (ID 9, $R^2 = 0.293$).

IV. DISCUSSION

The purpose of this proof-of-concept study was to propose a BiLSTM-based model that predicts an average ankle plantar flexion torque value for a given time-series 3-axis MMG signal recorded during a sustained isometric contraction. We designed the different deep learning models in terms of model architecture, the number of units, and dropout ratio, and then evaluated its prediction performance by conducting the LOSOCV method. Across the models, there was a general trend showing that the models tend to better account for the variance in untrained test datasets as the error metrics decrease and as the slope becomes closer to 1.0. From the optimal model prediction for the untrained test dataset, the slope estimates appear to be sensitive to an individual dataset, but >70% of the variance in nine out of 10 datasets was explained by the model. These results demonstrate the feasibility of the proposed model as a potential tool to quantify average joint torque during a sustained isometric contraction.

There have been attempts to estimate joint torque from some features of the MMG signals (e.g., RMS MMG, mean/median power frequency, zero-crossing, etc.) based on machine learning models, demonstrating a good performance assessed by the coefficient of determination (R^2). For example, Youn and Kim [26] introduced a multiple linear regression model that predicted the time-series elbow joint torque with an average R^2

of 0.84 for the same-subject validation and of 0.82 for cross-subject validation. Ibitoye, et al. [27] developed a support vector regression model that predicted an average knee extensor torque evoked by neuromuscular electrical stimulation with R^2 of 0.89 for stratified random sampling validation. Other studies also demonstrated promising results (R^2 range: 0.77–0.87) based on artificial neural network models [25], [26], [28]. However, in most of these studies, data collected from the same participant were used for both training and test datasets, likely yielding a better prediction performance with a higher R^2 value compared to the LOSOCV method that was used in this study. Thus, it is uncertain whether the previous models would perform well with untrained test data from a new individual. In addition, as these studies did not report a slope of the relationships between observed and predicted values, there is a lack of information on whether the previous models underpredict or overpredict the outcomes.

Given that our primary goal of the proposed model was to predict an average joint torque value for given time-series 3-axis MMG signals during a sustained isometric contraction from a new individual (i.e., a similar form to the MMT procedure in clinical practice), the model prediction performance was evaluated by conducting the LOSOCV method. In general, the overall mean slope was reasonably close to 1.0 regardless of the tested models, whereas the slope estimates for each untrained test subject ranged between 0.35 (i.e., underpredict) and 1.99 (i.e., overpredict). We confirmed that this wide range of the slope estimates was unlikely due to the proposed model issues, by conducting a repeated random sub-sampling method (i.e., split the whole dataset into training (80%), validation (10%), and test (10%) dataset randomly 10 times) with the optimal model (for the test dataset, slope: 0.975 ± 0.033 ranging between 0.889 and 1.008; R^2: 0.978 ± 0.020 ranging between 0.936 and 0.997). Given this superior performance of the proposed model over the previous models, it is plausible that the MMG signals contain subject-specific features that make muscle force/joint torque prediction from MMG signals more challenging.

Indeed, MMG signals are known to be affected by many factors. For example, the MMG amplitude tends to increase as the muscle contraction intensity increases [21], [24], [33], [34]. Regarding the mean/median frequency of the MMG signals during contraction at different intensities, there was mixed evidence demonstrating an increase (e.g., first dorsal interosseus [22], biceps brachii [20], [21]) or decrease (e.g., tibialis anterior [34]). These changes in the mechanical response appear to be associated with neuromuscular factors such as global motor unit behavior (e.g., recruitment and rate coding) [20], muscle architecture [34], muscle fiber type composition [35], muscle length [24], muscle stiffness [36], intramuscular pressure [37], subcutaneous thickness [38], skinfold thickness [39], and arm circumference [40]. Given these uncertainties, widely accepted variables that feature the amplitude and frequency components of MMG signals may not likely contain all the necessary information to predict kinetic variables such as joint torque. In this regard, we proposed a BiLSTM-based model that predicts an average joint torque from time-series 3-axis MMG signals during a sustained

isometric contraction, rather than features of the signals. The proposed model outperformed the previous machine learning models when evaluated with a comparable cross-validation method (an average R^2 of 0.97), promising that the proposed deep learning model may be more capable of learning subject-specific features, especially from a larger training dataset.

It is also worth considering the possible effects of other factors on the relationship between the MMG signals and ankle plantar flexion torques across participants. Since this study only used the MMG signals collected from MG (i.e., one of the plantar flexors), it is plausible that the contribution of the other plantar flexors to the net joint torque (i.e., load-sharing) may be different across trials and/or participants. Moreover, co-contraction of other muscles (e.g., dorsiflexors) could also contribute to the net joint torque slightly differently across trials and/or participants. In addition, these neighboring muscle contractions may likely influence the MMG signals collected from MG (i.e., crosstalk), as observed in the forearm muscles [41], [42], elbow flexors [43], and quadriceps femoris muscles [44]. Considering that multiple sensors may improve the prediction performance [45], [46], it would be interesting to test whether the proposed model prediction better performs when using MMG signals from more muscles (e.g., agonists and antagonists) and/or from multiple measurement sites.

The aforementioned factors are likely altered after disuse or neurologic diseases. For instance, compared to young neurologically intact individuals, older adults and stroke survivors demonstrated altered motor unit behavior (e.g., reduced mean firing rates, altered recruitment threshold) [47], [48], [49], [50], changes in muscle size, volume, and composition (e.g., reduced cross-sectional area, increased non-contractile tissues) [51], [52], [53], changes in the contractile properties (e.g., reduced force-generating capacity, altered twitch response, changes in fiber type composition, force-length curves) [49], [53], [54], [55], [56], changes in muscle architecture (e.g., fascicle length, pennation angle, muscle thickness) [53], [57], [58], [59], [60], [61], and changes in mechanical properties of muscle-tendon unit tissue (e.g., increased muscle stiffness, decreased tendon stiffness) [53], [58], [62], [63], [64]. These neuromuscular changes likely influence muscle vibratory behaviors and thus MMG signals. However, it appears that such alterations may not change the general shape of the relationship between the MMG features (e.g., RMS MMG, median power frequency) and strength measures [19], [23], [65]. Future studies are guaranteed to evaluate the prediction performance of the proposed model with data collected from these populations.

There are several other important considerations in this study. First, the MMG signals were collected from the distal part of the MG muscle rather than the belly part. Given the previous findings that the MMG signals may be different in amplitude and frequency according to the measurement site [17], [18], [34], [66], it will be interesting to test the sensitivity of the proposed model using data collected from different measurement sites. Second, the current study was tested only with a dataset collected from MG at the ankle joint. Future studies are required to test other muscles in terms of location, muscle architecture, muscle fiber type composition,

etc. Lastly, the RMS-filtered MMG signals were used as an input to the proposed model, which may lose the frequency information of the MMG signals to some extent. It would be interesting to determine whether adding frequency information can improve model prediction performance.

REFERENCES

[1] J. Chae, G. Yang, B. K. Park, and I. Labatia, "Muscle weakness and cocontraction in upper limb hemiparesis: Relationship to motor impairment and physical disability," *Neurorehabilitation Neural Repair*, vol. 16, no. 3, pp. 241–248, Sep. 2002, doi: 10.1177/154596830201600303.

[2] T. Cattagni et al., "Neural and muscular factors both contribute to plantar-flexor muscle weakness in older fallers," *Experim. Gerontology*, vol. 112, pp. 127–134, Oct. 2018, doi: 10.1016/j.exger.2018.09.011.

[3] R. I. Wadman et al., "Muscle strength and motor function throughout life in a cross-sectional cohort of 180 patients with spinal muscular atrophy types 1c–4," *Eur. J. Neurol.*, vol. 25, no. 3, pp. 512–518, Mar. 2018, doi: 10.1111/ene.13534.

[4] L. A. Ingram et al., "Quantifying upper limb motor impairment in chronic stroke: A physiological profiling approach," *J. Appl. Physiol.*, vol. 131, no. 3, pp. 949–965, Sep. 2021, doi: 10.1152/japplphysiol.00078.2021.

[5] R. W. Bohannon, "Selected determinants of ambulatory capacity in patients with hemiplegia," *Clin. Rehabil.*, vol. 3, no. 1, pp. 47–53, Feb. 1989, doi: 10.1177/026921558900300107.

[6] E. Frese, M. Brown, and B. J. Norton, "Clinical reliability of manual muscle testing: Middle trapezius and gluteus medius muscles," *Phys. Therapy*, vol. 67, no. 7, pp. 1072–1076, 1987, doi: 10.1093/ptj/67.7.1072.

[7] U. Aasa, S. Jaric, M. Barnekow-Bergkvist, and H. Johansson, "Muscle strength assessment from functional performance tests: Role of body size," *J. Strength Conditioning Res.*, vol. 17, no. 4, p. 664, 2003.

[8] R. C. Li et al., "The development, validity, and reliability of a manual muscle testing device with integrated limb position sensors," *Arch. Phys. Med. Rehabil.*, vol. 87, no. 3, pp. 411–417, Mar. 2006, doi: 10.1016/j.apmr.2005.11.011.

[9] C. T. Wadsworth, R. Krishnan, M. Sear, J. Harrold, and D. H. Nielsen, "Intrarater reliability of manual muscle testing and hand-held dynametric muscle testing," *Phys. Therapy*, vol. 67, no. 9, pp. 1342–1347, Sep. 1987, doi: 10.1093/ptj/67.9.1342.

[10] C. Disselhorst-Klug, T. Schmitz-Rode, and G. Rau, "Surface electromyography and muscle force: Limits in sEMG-force relationship and new approaches for applications," *Clin. Biomechanics*, vol. 24, no. 3, pp. 225–235, Mar. 2009, doi: 10.1016/j.clinbiomech.2008.08.003.

[11] J. Rosen, M. B. Fuchs, and M. Arcan, "Performances of hill-type and neural network muscle models—Toward a myosignal-based exoskeleton," *Comput. Biomed. Res.*, vol. 32, no. 5, pp. 415–439, Oct. 1999, doi: 10.1006/cbmr.1999.1524.

[12] J. Hashemi, E. Morin, P. Mousavi, K. Mountjoy, and K. Hashtrudi-Zaad, "EMG–force modeling using parallel cascade identification," *J. Electromyogr. Kinesiol.*, vol. 22, no. 3, pp. 469–477, Jun. 2012, doi: 10.1016/j.jelekin.2011.10.012.

[13] C. J. De Luca, "The use of surface electromyography in biomechanics," *J. Appl. Biomech.*, vol. 13, no. 2, pp. 135–163, 1997, doi: 10.1123/jab.13.2.135.

[14] R. Merletti, I. Campanini, W. Z. Rymer, and C. Disselhorst-Klug, "Editorial: Surface electromyography: Barriers limiting widespread use of sEMG in clinical assessment and neurorehabilitation," *Frontiers Neurol.*, vol. 12, Feb. 2021, Art. no. 642257, doi: 10.3389/fneur.2021.642257.

[15] H. A. Feldner, D. Howell, V. E. Kelly, S. W. McCoy, and K. M. Steele, "'Look, your muscles are firing!': A qualitative study of clinician perspectives on the use of surface electromyography in neurorehabilitation," *Arch. Phys. Med. Rehabil.*, vol. 100, no. 4, pp. 663–675, 2019, doi: 10.1016/j.apmr.2018.09.120.

[16] T. W. Beck et al., "Mechanomyographic amplitude and frequency responses during dynamic muscle actions: A comprehensive review," *Biomed. Eng. OnLine*, vol. 4, no. 1, p. 67, Dec. 2005, doi: 10.1186/1475-925x-4-67.

[17] C. Cescon, P. Madeleine, T. Graven-Nielsen, R. Merletti, and D. Farina, "Two-dimensional spatial distribution of surface mechanomyographical response to single motor unit activity," *J. Neurosci. Methods*, vol. 159, no. 1, pp. 19–25, Jan. 2007, doi: 10.1016/j.jneumeth.2006.06.011.

[18] C. Cescon, P. Madeleine, and D. Farina, "Longitudinal and transverse propagation of surface mechanomyographic waves generated by single motor unit activity," *Med. Biol. Eng. Comput.*, vol. 46, no. 9, pp. 871–877, Sep. 2008, doi: 10.1007/s11517-008-0357-4.

[19] C. Meagher et al., "New advances in mechanomyography sensor technology and signal processing: Validity and intrarater reliability of recordings from muscle," *J. Rehabil. Assistive Technol. Eng.*, vol. 7, Jan. 2020, Art. no. 205566832091611, doi: 10.1177/2055668320916116.

[20] C. Orizio, R. Perini, B. Diemont, M. Maranzana Figini, and A. Veicsteinas, "Spectral analysis of muscular sound during isometric contraction of biceps brachii," *J. Appl. Physiol.*, vol. 68, no. 2, pp. 508–512, Feb. 1990, doi: 10.1152/jappl.1990.68.2.508.

[21] K. Akataki, K. Mita, M. Watakabe, and K. Itoh, "Mechanomyogram and force relationship during voluntary isometric ramp contractions of the biceps brachii muscle," *Eur. J. Appl. Physiol.*, vol. 84, nos. 1–2, pp. 19–25, Feb. 2001, doi: 10.1007/s004210000321.

[22] K. Akataki, K. Mita, M. Watakabe, and K. Itoh, "Mechanomyographic responses during voluntary ramp contractions of the human first dorsal interosseous muscle," *Eur. J. Appl. Physiol.*, vol. 89, no. 6, pp. 520–525, Aug. 2003, doi: 10.1007/s00421-003-0835-1.

[23] X. L. Hu, K. Y. Tong, and L. Li, "The mechanomyography of persons after stroke during isometric voluntary contractions," *J. Electromyogr. Kinesiol.*, vol. 17, no. 4, pp. 473–483, Aug. 2007, doi: 10.1016/j.jelekin.2006.01.015.

[24] F. Shi, W. Z. Rymer, and J. Son, "Mechanomyogram amplitude vs. Isometric ankle plantarflexion torque of human medial gastrocnemius muscle at different ankle joint angles," *J. Electromyogr. Kinesiol.*, vol. 61, Dec. 2021, Art. no. 102609, doi: 10.1016/j.jelekin.2021.102609.

[25] W. Youn and J. Kim, "Estimation of elbow flexion force during isometric muscle contraction from mechanomyography and electromyography," *Med. Biol. Eng. Comput.*, vol. 48, no. 11, pp. 1149–1157, Nov. 2010, doi: 10.1007/s11517-010-0641-y.

[26] W. Youn and J. Kim, "Feasibility of using an artificial neural network model to estimate the elbow flexion force from mechanomyography," *J. Neurosci. Methods*, vol. 194, no. 2, pp. 386–393, Jan. 2011, doi: 10.1016/j.jneumeth.2010.11.003.

[27] M. Ibitoye, N. Hamzaid, A. A. Wahab, N. Hasnan, S. Olatunji, and G. Davis, "Estimation of electrically-evoked knee torque from mechanomyography using support vector regression," *Sensors*, vol. 16, no. 7, p. 1115, Jul. 2016, doi: 10.3390/s16071115.

[28] M. A. Dzulkifli, N. A. Hamzaid, G. M. Davis, and N. Hasnan, "Neural network-based muscle torque estimation using mechanomyography during electrically-evoked knee extension and standing in spinal cord injury," *Frontiers Neurorobotics*, vol. 12, p. 50, Aug. 2018, doi: 10.3389/fnbot.2018.00050.

[29] H. Wu, Q. Huang, D. Wang, and L. Gao, "A CNN-SVM combined model for pattern recognition of knee motion using mechanomyography signals," *J. Electromyogr. Kinesiol.*, vol. 42, pp. 136–142, Oct. 2018, doi: 10.1016/j.jelekin.2018.07.005.

[30] C. S. M. Castillo, S. Wilson, R. Vaidyanathan, and S. F. Atashzar, "Wearable MMG-Plus-One armband: Evaluation of normal force on mechanomyography (MMG) to enhance human–machine interfacing," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 196–205, 2021, doi: 10.1109/TNSRE.2020.3043368.

[31] F. Chollet et al., "Keras," Tech. Rep., 2015. [Online]. Available: https://keras.io

[32] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, *arXiv:1603.04467*.

[33] K. T. Ebersole, T. J. Housh, G. O. Johnson, T. K. Evetovich, D. B. Smith, and S. R. Perry, "The effect of leg flexion angle on the mechanomyographic responses to isometric muscle actions," *Eur. J. Appl. Physiol.*, vol. 78, no. 3, pp. 264–269, Jul. 1998, doi: 10.1007/s004210050418.

[34] C. Cescon, D. Farina, M. Gobbo, R. Merletti, and C. Orizio, "Effect of accelerometer location on mechanomyogram variables during voluntary, constant-force contractions in three human muscles," *Med. Biol. Eng. Comput.*, vol. 42, no. 1, pp. 121–127, Jan. 2004, doi: 10.1007/bf02351021.

[35] T. J. Herda et al., "A noninvasive, log-transform method for fiber type discrimination using mechanomyography," *J. Electromyogr. Kinesiol.*, vol. 20, no. 5, pp. 787–794, Oct. 2010, doi: 10.1016/j.jelekin.2010.01.004.

[36] S. Longo, E. Cè, S. Rampichini, M. Devoto, E. Limonta, and F. Esposito, "Mechanomyogram amplitude correlates with human gastrocnemius medialis muscle and tendon stiffness both before and after acute passive stretching," *Experim. Physiol.*, vol. 99, no. 10, pp. 1359–1369, Oct. 2014, doi: 10.1113/expphysiol.2014.080366.

[37] K. Søgaard, C. Orizio, and G. Sjøgaard, "Surface mechanomyogram amplitude is not attenuated by intramuscular pressure," *Eur. J. Appl. Physiol.*, vol. 96, no. 2, pp. 178–184, Jan. 2006, doi: 10.1007/s00421-004-1211-5.

[38] E. M. Scheeren, L. Mineiro, E. B. Neves, E. Krueger, G. N. N. Neto, and P. Nohama, "Influence of subcutaneous fat on mechanomyographic signals at three levels of voluntary effort," *Res. Biomed. Eng.*, vol. 32, no. 4, pp. 307–317, Jan. 2017, doi: 10.1590/2446-4740.03615.

[39] A. Jaskólska, W. Brzenczek, K. Kisiel-Sajewicz, A. Kawczyński, J. Marusiak, and A. Jaskólski, "The effect of skinfold on frequency of human muscle mechanomyogram," *J. Electromyogr. Kinesiol.*, vol. 14, no. 2, pp. 217–225, Apr. 2004, doi: 10.1016/j.jelekin.2003.08.001.

[40] I. Talib, K. Sundaraj, and C. K. Lam, "Association of anthropometric parameters with amplitude and crosstalk of mechanomyographic signals during forearm flexion, pronation and supination torque tasks," *Sci. Rep.*, vol. 9, no. 1, p. 16166, Nov. 2019, doi: 10.1038/s41598-019-52536-4.

[41] A. Islam, K. Sundaraj, R. B. Ahmad, S. Sundaraj, N. U. Ahamed, and M. A. Ali, "Analysis of crosstalk in the mechanomyographic signals generated by forearm muscles during different wrist postures," *Muscle Nerve*, vol. 51, no. 6, pp. 899–906, Jun. 2015, doi: 10.1002/mus.24454.

[42] M. A. Islam, K. Sundaraj, R. B. Ahmad, S. Sundaraj, N. U. Ahamed, and M. A. Ali, "Cross-talk in mechanomyographic signals from the forearm muscles during sub-maximal to maximal isometric grip force," *PLoS ONE*, vol. 9, no. 5, May 2014, Art. no. e96628, doi: 10.1371/journal.pone.0096628.

[43] I. Talib, K. Sundaraj, and C. K. Lam, "Analysis of the crosstalk in mechanomyographic signals along the longitudinal, lateral and transverse axes of elbow flexor muscles during sustained isometric forearm flexion, supination and pronation exercises," *J. Musculoskeletal Neuronal Interact.*, vol. 20, no. 2, pp. 194–205, 2020.

[44] T. W. Beck, J. M. DeFreitas, and M. S. Stock, "An examination of cross-talk among surface mechanomyographic signals from the superficial quadriceps femoris muscles during isometric muscle actions," *Human Movement Sci.*, vol. 29, no. 2, pp. 165–171, Apr. 2010, doi: 10.1016/j.humov.2009.11.005.

[45] Y. Zhang and C. Xia, "A preliminary study of classification of upper limb motions and forces based on mechanomyography," *Med. Eng. Phys.*, vol. 81, pp. 97–104, Jul. 2020, doi: 10.1016/j.medengphy.2020.05.009.

[46] N. Hondo and T. Tsuji, "Torque estimation of knee flexion and extension movements from a mechanomyogram of the femoral muscle," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1120–1126, 2022, doi: 10.1109/TNSRE.2022.3169225.

[47] G. Kamen and C. J. De Luca, "Unusual motor unit firing behavior in older adults," *Brain Res.*, vol. 482, no. 1, pp. 136–140, Mar. 1989, doi: 10.1016/0006-8993(89)90550-7.

[48] B. L. Tracy, K. S. Maluf, J. L. Stephenson, S. K. Hunter, and R. M. Enoka, "Variability of motor unit discharge and force fluctuations across a range of muscle forces in older adults," *Muscle Nerve*, vol. 32, no. 4, pp. 533–540, Oct. 2005, doi: 10.1002/mus.20392.

[49] B. H. Dalton, J. M. Jakobi, B. L. Allman, and C. L. Rice, "Differential age-related changes in motor unit properties between elbow flexors and extensors," *Acta Physiologica*, vol. 200, pp. 45–55, Sep. 2010, doi: 10.1111/j.1748-1716.2010.02100.x.

[50] J. Kallio, K. Søgaard, J. Avela, P. Komi, H. Selänne, and V. Linnamo, "Age-related decreases in motor unit discharge rate and force control during isometric plantar flexion," *J. Electromyogr. Kinesiol.*, vol. 22, no. 6, pp. 983–989, Dec. 2012, doi: 10.1016/j.jelekin.2012.05.009.

[51] C. J. Hasson, J. A. Kent-Braun, and G. E. Caldwell, "Contractile and non-contractile tissue volume and distribution in ankle muscles of young and older adults," *J. Biomechanics*, vol. 44, no. 12, pp. 2299–2306, Aug. 2011, doi: 10.1016/j.jbiomech.2011.05.031.

[52] J. W. Ramsay, P. J. Barrance, T. S. Buchanan, and J. S. Higginson, "Paretic muscle atrophy and non-contractile tissue content in individual muscles of the post-stroke lower extremity," *J. Biomechanics*, vol. 44, no. 16, pp. 2741–2746, Nov. 2011, doi: 10.1016/j.jbiomech.2011.09.001.

[53] L. Stenroth, J. Peltonen, N. J. Cronin, S. Sipila, and T. Finni, "Age-related differences in Achilles tendon properties and triceps surae muscle architecture in vivo," *J. Appl. Physiol.*, vol. 113, no. 10, pp. 1537–1544, Nov. 2012, doi: 10.1152/japplphysiol.00782.2012.

[54] F. Gao and L.-Q. Zhang, "Altered contractile properties of the gastrocnemius muscle poststroke," *J. Appl. Physiol.*, vol. 105, no. 6, pp. 1802–1808, Dec. 2008, doi: 10.1152/japplphysiol.90930.2008.

[55] D. M. Callahan et al., "Reduced rate of knee extensor torque development in older adults with knee osteoarthritis is associated with intrinsic muscle contractile deficits," *Experim. Gerontology*, vol. 72, pp. 16–21, Dec. 2015, doi: 10.1016/j.exger.2015.08.016.

[56] K. Severinsen et al., "Skeletal muscle fiber characteristics and oxidative capacity in hemiparetic stroke survivors," *Muscle Nerve*, vol. 53, no. 5, pp. 748–754, May 2016, doi: 10.1002/mus.24907.

[57] L. Li, K. Y. Tong, and X. Hu, "The effect of poststroke impairments on brachialis muscle architecture as measured by ultrasound," *Arch. Phys. Med. Rehabil.*, vol. 88, no. 2, pp. 243–250, Feb. 2007, doi: 10.1016/j.apmr.2006.11.013.

[58] H. Zhao, Y. Ren, E. J. Roth, R. L. Harvey, and L.-Q. Zhang, "Concurrent deficits of soleus and gastrocnemius muscle fascicles and Achilles tendon post stroke," *J. Appl. Physiol.*, vol. 118, no. 7, pp. 863–871, Apr. 2015, doi: 10.1152/japplphysiol.00226.2014.

[59] F. Berenpas, A.-M. Martens, V. Weerdesteyn, A. C. Geurts, and N. van Alfen, "Bilateral changes in muscle architecture of physically active people with chronic stroke: A quantitative muscle ultrasound study," *Clin. Neurophysiol.*, vol. 128, no. 1, pp. 115–122, Jan. 2017, doi: 10.1016/j.clinph.2016.10.096.

[60] J. Son, W. Z. Rymer, and S. S. M. Lee, "Limited fascicle shortening and fascicle rotation may be associated with impaired voluntary force-generating capacity in pennate muscles of chronic stroke survivors," *Clin. Biomechanics*, vol. 75, May 2020, Art. no. 105007, doi: 10.1016/j.clinbiomech.2020.105007.

[61] M. Narici et al., "Age-related alterations in muscle architecture are a signature of sarcopenia: The ultrasound sarcopenia index," *J. Cachexia, Sarcopenia Muscle*, vol. 12, no. 4, pp. 973–982, Aug. 2021, doi: 10.1002/jcsm.12720.

[62] U. Svantesson, H. Takahashi, U. Carlsson, A. Danielsson, and K. S. Sunnerhagen, "Muscle and tendon stiffness in patients with upper motor neuron lesion following a stroke," *Eur. J. Appl. Physiol.*, vol. 82, no. 4, pp. 275–279, Jul. 2000, doi: 10.1007/s004210000216.

[63] S. S. M. Lee, S. Spear, and W. Z. Rymer, "Quantifying changes in material properties of stroke-impaired muscle," *Clin. Biomechanics*, vol. 30, no. 3, pp. 269–275, Mar. 2015, doi: 10.1016/j.clinbiomech.2015.01.004.

[64] K. L. Jakubowski, A. Terman, R. V. C. Santana, and S. S. M. Lee, "Passive material properties of stroke-impaired plantarflexor and dorsiflexor muscles," *Clin. Biomechanics*, vol. 49, pp. 48–55, Nov. 2017, doi: 10.1016/j.clinbiomech.2017.08.009.

[65] K. Akataki, K. Mita, M. Watakabe, and K. Ito, "Age-related change in motor unit activation strategy in force production: A mechanomyographic investigation," *Muscle Nerve*, vol. 25, no. 4, pp. 505–512, Apr. 2002, doi: 10.1002/mus.10076.

[66] P. Kaczmarek, J. Celichowski, H. Drzymała-Celichowska, and A. Kasiński, "The image of motor units architecture in the mechanomyographic signal during the single motor unit contraction: In vivo and simulation study," *J. Electromyogr. Kinesiol.*, vol. 19, no. 4, pp. 553–563, Aug. 2009, doi: 10.1016/j.jelekin.2008.03.007.