https://doi.org/10.31436/iiumej.v26i1.3379

BILATERAL GROUND REACTION FORCE PREDICTION WITH DEEP LEARNING MODELS AND CUSTOM FORCE PLATE

YING HENG YEO, MUHAMMAD FAUZINIZAM RAZALI*, ZAIDI MOHD RIPIN, NUR-AKASYAH J., MOHAMAD IKHWAN ZAINI RIDZWAN, ALEXANDER WAI TENG TAN, JIA YI TAY

Neurorehabilitation Engineering and Assistance Systems Research, School of Mechanical Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia.

**Corresponding author: mefauzinizam@usm.my*

(Received: 24 July 2024; Accepted: 5 December 2024; Published online: 10 January 2025)

ABSTRACT: Several low-cost force plates have been proposed as alternatives for laboratory-grade force plates. Nevertheless, the inability to quantify bilateral ground reaction force (GRF) prevents these inexpensive force plates from being used for biomechanical analysis and certain clinical metric acquisition. This study developed deep-learning models, such as autoencoder and U-net, to predict bilateral GRF from vertical GRF measured using a low-cost custom force plate during sit-to-stand, gait initialization, and gait. Results indicated that the U-net model, which utilized STFT vertical GRF as input, performed the best. In addition to predicting the mediolateral GRF measured during sit-to-stand, the model accurately predicted the anterior-posterior and mediolateral GRF for sit-to-stand, gait initialization, and gait in the test dataset, achieving high Pearson's correlation coefficient, coefficient of determination, and intraclass correlation coefficient values of over 0.90, 0.79, and 0.89, respectively. The model demonstrated a higher Pearson's correlation coefficient compared to three related previous studies that utilized different methods to predict anterior-posterior GRF and six studies in inferring mediolateral GRF. The results demonstrated the potential of TFU and custom force plate as a GRF measurement tool to perform bio-mechanical analysis.

ABSTRAK: Beberapa plat daya kos rendah telah dicadangkan sebagai alternatif kepada plat daya berkualiti makmal. Walau bagaimanapun, ketidakmampuan untuk mengukur daya reaksi tanah (GRF) secara bilateral menghalang plat daya yang murah ini daripada digunakan untuk analisis biomekanik dan pengambilan metrik klinikal tertentu. Kajian ini membangunkan model pembelajaran mendalam, seperti autoencoder dan U-net, untuk meramalkan GRF bilateral daripada GRF menegak yang diukur menggunakan plat daya khas kos rendah semasa pergerakan duduk-ke-berdiri, permulaan berjalan, dan berjalan. Hasil menunjukkan bahawa model U-net, yang menggunakan GRF menegak STFT sebagai input, memberikan prestasi terbaik. Selain meramalkan GRF mediolateral yang diukur semasa duduk-ke-berdiri, model ini juga meramalkan dengan tepat GRF anterior-posterior dan mediolateral untuk duduk-ke-berdiri, permulaan berjalan, dan berjalan dalam set data ujian, mencapai nilai koefisien korelasi Pearson, koefisien penentuan, dan koefisien korelasi intrakelas yang tinggi melebihi 0.90, 0.79, dan 0.89, masing-masing. Model ini menunjukkan koefisien korelasi Pearson yang lebih tinggi berbanding tiga kajian terdahulu yang berkaitan yang menggunakan kaedah berbeza untuk meramalkan GRF anterior-posterior dan enam kajian dalam menyimpulkan GRF mediolateral. Hasil kajian menunjukkan potensi TFU dan plat daya khas sebagai alat pengukuran GRF untuk melakukan analisis biomekanik.

KEYWORDS: *Force plate; deep learning model; ground reaction force; prediction*

https://doi.org/10.31436/iiumej.v26i1.3379

1. INTRODUCTION

The force plate is the gold standard for measuring the ground reaction force (GRF) and the center of pressure (CoP) exerted by a person's movement, where the obtained data can be utilized for biomechanical analysis. The GRF represents external forces acting on the human body from the ground, while the CoP represents the position upon which the GRF works. A laboratory-grade force plate usually consists of a six-axis load cell enclosed in a flat rectangular case, which measures the three-dimensional GRF and moments. The CoP is then computed based on the recorded GRF and moments. Both GRF and CoP are necessary for determining the forces and torques withstood by various human body components in biomechanical analysis. In addition, the force plate is essential in acquiring various GRF- or CoP-derived metrics that assess the elderly or patients' health conditions in clinical settings. These applications demonstrate the importance of force plates in biomechanical and clinical domains.

In laboratory-grade force plates, the high production cost remains one of the concerns. For example, AMTI, Bertec, and Kistler force plates cost between USD 10,000 and 20,000. These prices restrict the availability of force plates in clinical or biomechanical fields, mainly when the budget is tight. The costing issue frequently occurs in clinical facilities worldwide, where the spending is directed towards other instruments with a higher return on investment than the force plate. In third-world countries, the affordability problem is even more severe, where less capital is invested in research. As a result, the clinical and biomechanical research that employs force plates is hampered in developing countries. On the other hand, the scalability of a force plate is infeasible due to the high cost. An example that demands force plate scalability involves a task under study, which must be performed over a long distance, requiring many force plates to provide sufficient coverage.

Several techniques that avoid the use of force plates have been introduced to address the concerns. This usage often involves inverse dynamic optimization and machine learning models. Accelerometers are also used instead of a complete inertial motion capture system to simplify the GRF prediction [6] further. However, when external pressures other than GRF are applied to the participants, the accuracy of these methods may be compromised [7]. Therefore, pressure-sensing insoles offer another alternative that can consider external forces [8-10]. However, these commercialized insoles are usually costly [8; 10]. Moreover, when fewer pressure sensors are used during movements (such as calf raises), the GRF prediction model applied with a low-cost customized insole may be unreliable [9].

Meanwhile, previous studies suggested low-cost force plates as alternatives to laboratory-grade force plates [11; 12]. A low-cost force plate typically consists of four single-axis load cells positioned vertically to measure the vertical GRF. However, no load cell faces the lateral direction to measure the bilateral GRF. When an individual performs tasks that generate perceivable bilateral GRF, the bilateral GRF will not be identified and analyzed.

The reaction of a single-axis load cell to off-axis loading may provide insight into addressing the issue of low-cost force plates. The load cell strain gauge generates values that depend on the force-induced strain. Generally, strain is highest in the direction of the primary applied force, while the off-axis loading creates minor strain. This minor strain is due to the Poisson effect, which influences the signals. The changes in readings can then determine the extent of off-axis loading. In other words, the vertical GRF readings from the four independent load cells on the inexpensive force plate estimate the off-axis bilateral GRF. To the authors' knowledge, no previous studies have investigated the algorithm for computing the bilateral GRF from the vertical GRF recorded using low-cost force plate load cells.

https://doi.org/10.31436/iiumej.v26i1.3379

This study aimed to develop a low-cost custom force plate to acquire bilateral GRF, in which deep learning models were trained to predict bilateral GRF from vertical GRF. Considering the necessity of bilateral GRF in biomechanical analysis and clinical settings, the prediction of bilateral GRF for sit-to-stand, gait initialization, and gait was investigated. With the proposed method, the low-cost force plates, which previously were unable to measure bilateral GRF, could now be applied to predict bilateral GRF. The improved low-cost force plates, which were more affordable and easily accessible to clinicians and researchers with tight budgets, could potentially be used for biomechanical analysis with enhanced accuracy and acquisition of bilateral GRF-derived clinical assessment metrics. The scalability issue in biomechanical analysis could be overcome by acquiring multiple low-cost force plates at a reasonable price as compared to laboratory-grade force plates.

2. MATERIALS AND METHOD

2.1. Participants

As the intraclass correlation coefficient (ICC) was applied to evaluate the model accuracy in this study, as mentioned in Section "Statistical Analysis," the sample size to achieve an ICC precision of ±0.02 was calculated [13]. The expected ICC was set at 0.98 based on a study on low-cost force plates in measuring GRF during sit-to-stand [11]. The confidence interval was set as 95%, with two raters (laboratory-grade and custom force plates). The following criteria were used for participant selection:

- Being 18 years of age or older;
- Being able to stand for more than one hour;
- Being able to engage in physical activity without restriction;
- Not having a musculoskeletal or neurological disorder;
- Not taking medication that affects body balance;
- Not being pregnant;
- Being able to provide informed consent.

This study, with the protocol code USM/JEPeM/PP/23020164, was approved by the Ethics Committee of Universiti Sains Malaysia. Table 1 summarizes an overview of the participants' information.

Table 1. Summary of the information regarding the participants in this study

Information	Value
Gender (male/female)	10 / 10
Age (year, mean ± standard deviation)	30.65 ± 8.48
Height (cm, mean ± standard deviation)	165.05 ± 7.99
Weight (kg, mean ± standard deviation)	63.59 ± 10.68

2.2. Apparatus

The custom force plate with dimensions of 676 mm × 400 mm × 49 mm comprised a rectangular board, an ESP32, four HX711, and four half-bridge strain gauge single-axis load cells, which were used in typical digital bathroom scales. The cost of the custom force plate was less than USD$50, at least 200 times cheaper than a laboratory-grade force plate. The custom force plate has been granted an exemption from the registration of medical devices by the Medical Device Authority (MDA), Ministry of Health Malaysia, under protocol number

https://doi.org/10.31436/iiumej.v26i1.3379

CIU-20230226-11. Figure 1 exhibits the top and bottom views of a custom force plate. A 4060-05 Bertec (laboratory-grade force plate) was used as the gold standard to evaluate the accuracy of the model.

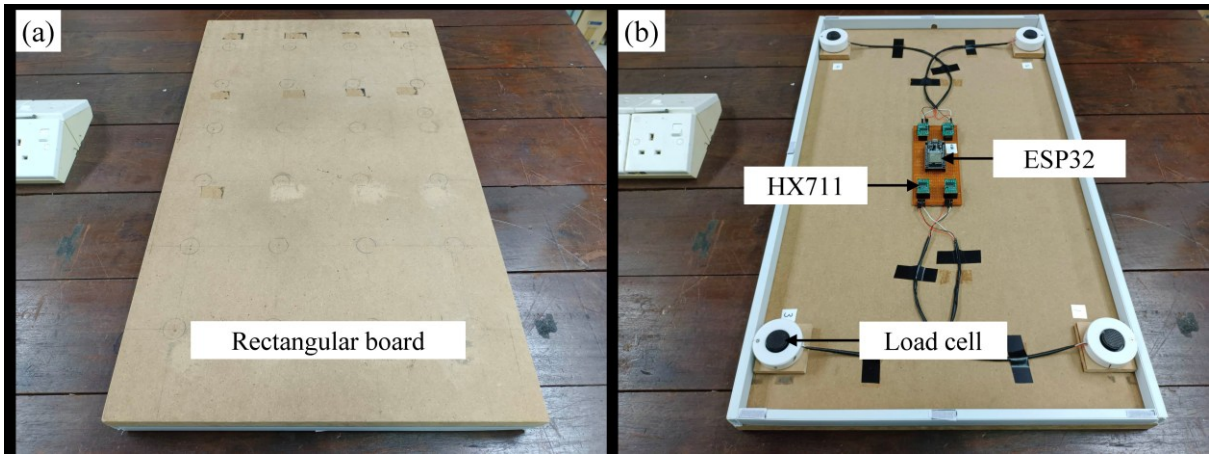

Figure 1. (a) The top view of the custom force plate. (b) The bottom view of the custom force plate.

2.3. Experimental Procedures

Before the experiment began, the custom force plate was placed on the Bertec force plate. The participants were asked to perform three tasks on the stacked force plates based on the procedures described in previous works: sit-to-stand [12], gait initialization [14], and gait [15]. When the participants conducted the tasks, forces were transmitted through the contact surfaces of the feet to the stacked force plates. Subsequently, vertical and bilateral GRF, which acted on the participants's feet, were generated according to the types of tasks. The three-dimensional GRF readings from the Bertec force plate and vertical GRF measurements from the custom force plate were then recorded at 500 Hz and 80 Hz sampling rates, respectively, and each task was repeated 10 times.

2.4. Data Preprocessing

A median filter with a window size of five was applied to custom force plate readings to remove outliers. Bertec force plate values were low-pass filtered using a zero-phase fourth-order low-pass Butterworth filter. The critical frequency was adjusted to half of the sampling rate of the custom force plate. The values measured with the Bertec force plate were then interpolated to match the sampling rate of the custom force plate.

2.5. Dataset Preparation

The complete dataset of 20 participants was randomly divided into training [fourteen participants (70 %)], validation [three participants (15 %)], and test datasets [three participants (15 %)], as shown in Figure 2. The dataset splitting ratio was applied according to previous research on gait cycle prediction models [16]. Each dataset included patches from the Bertec force plate's anterior-posterior and mediolateral GRF (model output) and the custom force plate's individual load cell vertical GRF (model input). To expand the dataset size, the readings were randomly cropped into smaller patches with a size of ntt. A total of 40 patches were cropped from each sit-to-stand trial recording, whereas 20 patches were generated from a recording of the gait initialization or a gait trial, as the sit-to-stand trial took longer. The training, validation, and test datasets comprised 16800, 3600, and 3600 patches, respectively.

https://doi.org/10.31436/iiumej.v26i1.3379

Following the rule of 10 events per variable [17], the minimum amount of data required was 10 times the amount of input required for the deep learning models. Considering the maximum length of model input, $ntt = 112$, as indicated in Table 2, and channel number 4 corresponding to the four load cell readings, the minimum number of training dataset readings was 4480 (10 × 112 × 4). Therefore, the number of training data utilized in this study, 16800, was deemed appropriate. The mean and standard deviation obtained from the training dataset were used to standardize the patches from all datasets.

The Short-Time Fourier Transform (STFT) was used to convert the patches into the time-frequency domain input and output to explore if representing the input in time-frequency form could enhance model accuracy. A Fourier transform window of 62 data points was employed to produce real and imaginary values, with 32 points on a frequency axis. The number of strides of the window sliding was set as 16, significantly less than the window size of 62, to generate a smooth output. An additional 62 subsequent data points from full trial readings were added to the patches to ensure enough data points were accommodated in the STFT window.

2.6. Model Architecture

Autoencoder [18] and U-net [19] were used in this study to predict bilateral GRF. The autoencoder was a neural network trained to compress input into a latent representation with a smaller dimension through an encoder. Later, a decoder was employed to reconstruct the output from the latent representation. Throughout the training process, the autoencoder acquired efficient representations, allowing for dimensionality reduction, feature extraction, anomaly detection, and even data generation through manipulation of the latent space. U-Net was similar in that it consisted of an encoder and a decoder. The difference was that the encoder's hidden layers were concatenated to the decoder's hidden layers with similar dimensions. This reduced the information loss, which may happen in an autoencoder [20], and allowed precise localization in generating the output [19].

https://doi.org/10.31436/iiumej.v26i1.3379

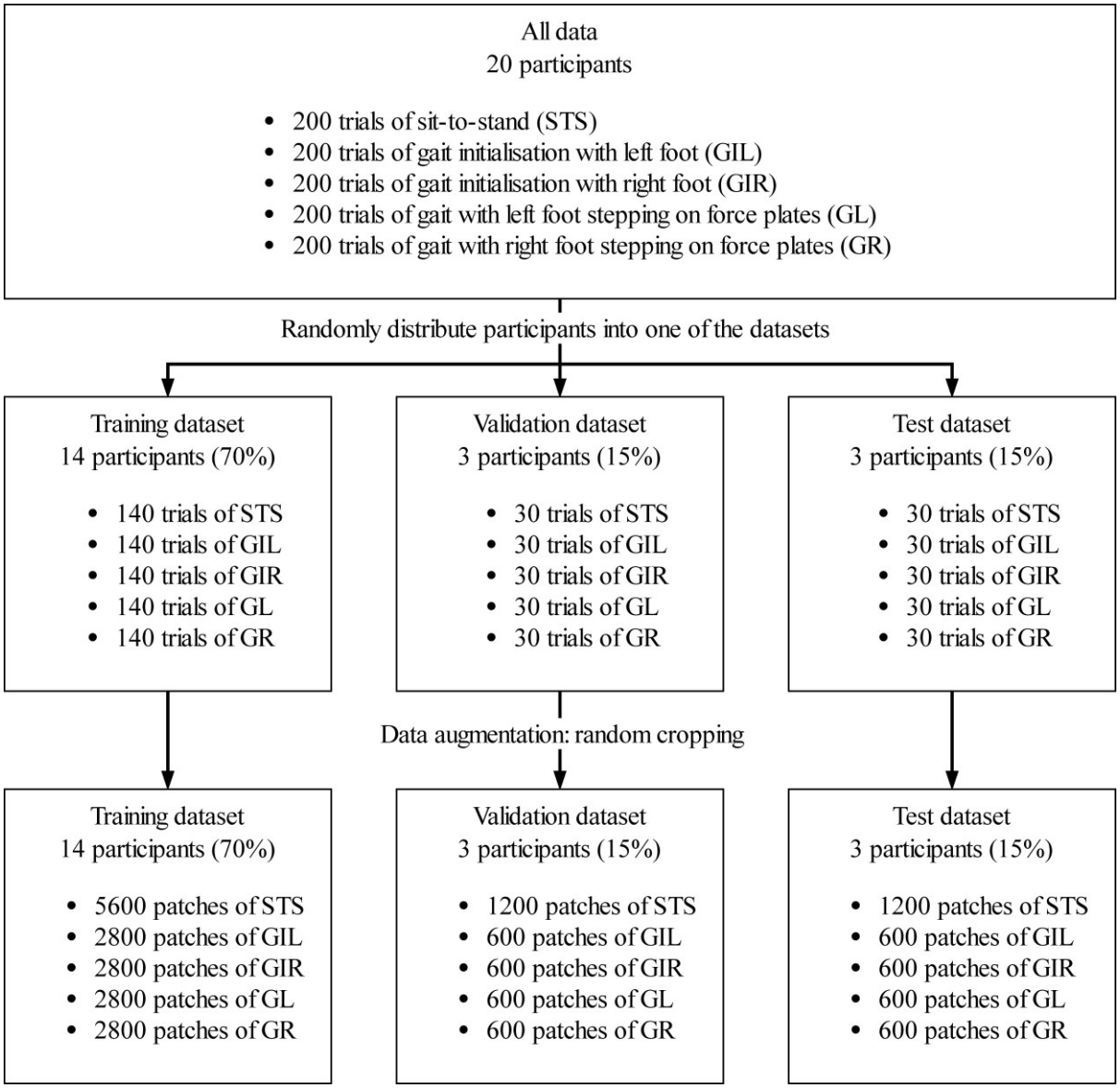

Figure 2. A flow chart shows the dataset preparation for model training.

Different model architectures were developed to receive input and produce output in the time and time-frequency domains (see Section "Data preparation"). Subsequently, four different model architectures, time domain autoencoder (TA), time domain U-Net (TU), time-frequency domain autoencoder (TFA), and time-frequency domain U-Net (TFU), were constructed. Several hyperparameters related to the model architectures were then modified in the hyperparameter optimization to acquire the optimal model architectures (see Section "Hyperparameter optimization"). Figures 3 and 4 depict the optimal architectures for the TA, TU, TFA, and TFU models determined through hyperparameter optimization. The following discussion of the model architecture and hyperparameter optimization was conducted using the figures.

https://doi.org/10.31436/iiumej.v26i1.3379

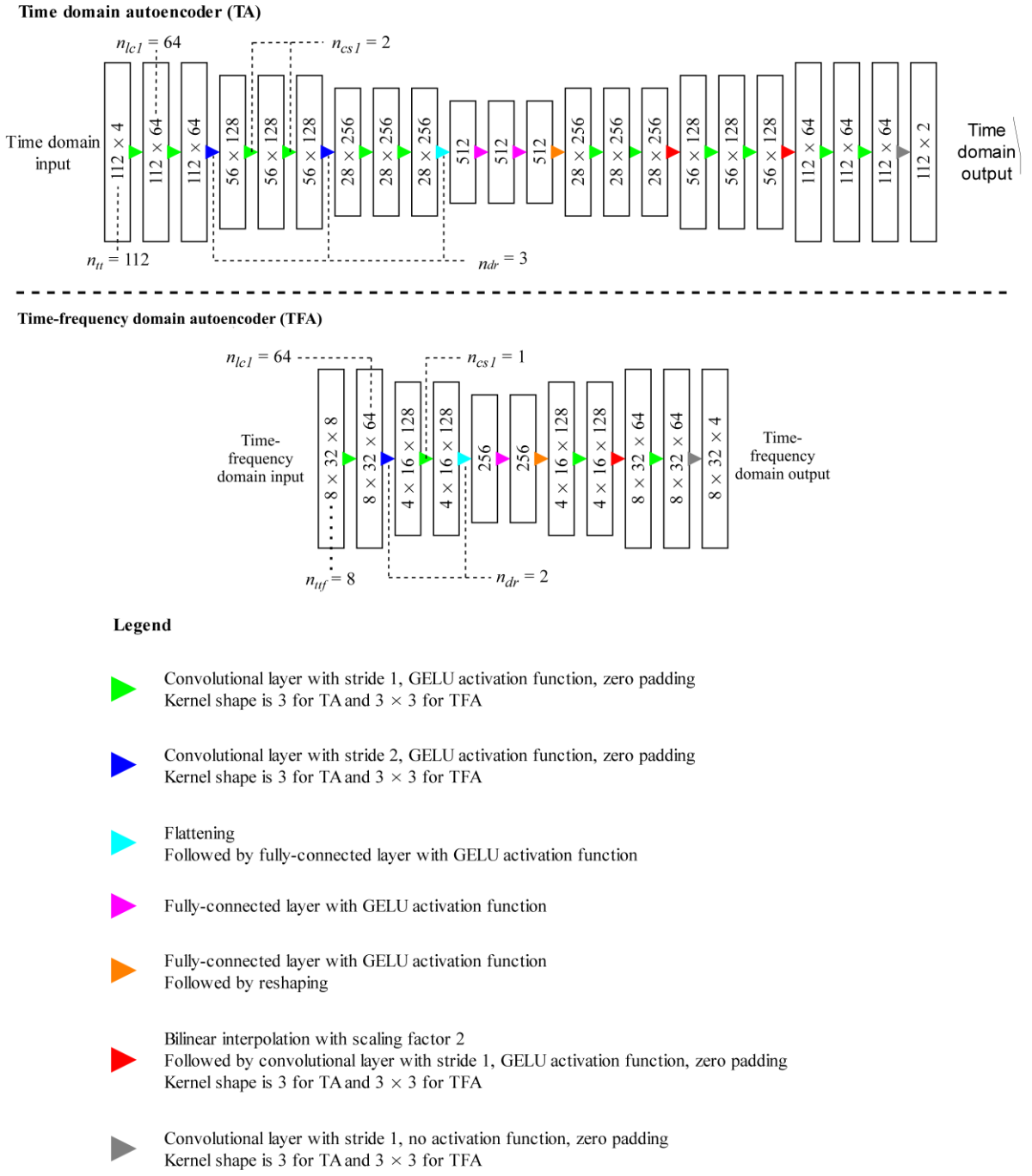

Figure 3. The optimal autoencoder model architectures are determined from hyperparameter optimization. The dotted lines are not part of the model architectures.

https://doi.org/10.31436/iiumej.v26i1.3379

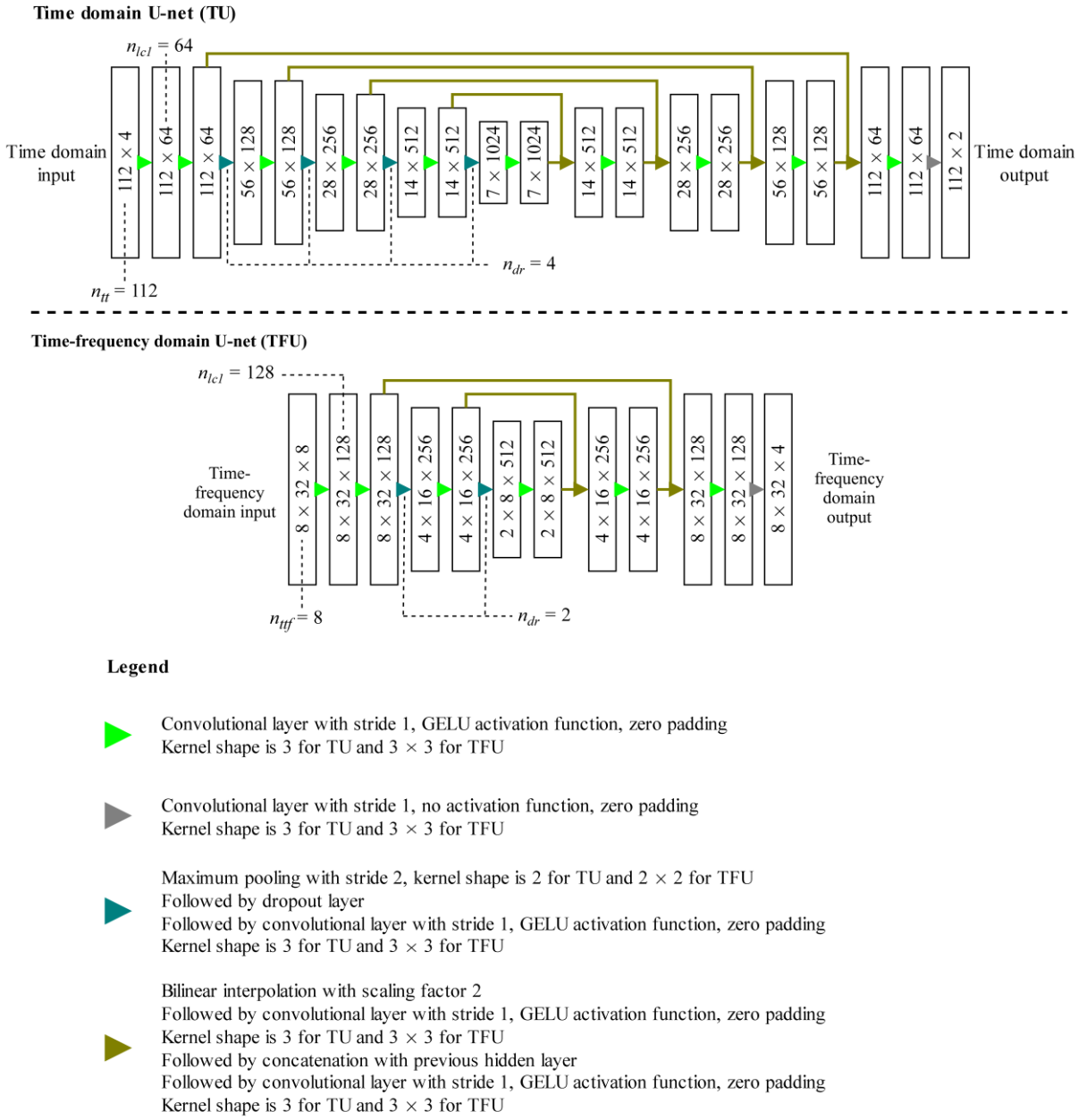

Figure 4. The optimal U-net model architectures are determined from the hyperparameter optimization. The dotted lines are not part of the model architectures.

The input for the time domain models (TA and TU) consisted of a two-dimensional patch with size $ntt \times 4$. Based on the random cropping operation described in Section "Data preparation," the first dimension represented the time axis with size ntt. The vertical GRF data points from the four custom force plate load cells were represented by the last dimension (channel having a size of four). The output was a two-dimensional patch with size $ntt \times 2$, where the anterior-posterior and mediolateral GRF of the models was predicted using the channel with size two. A three-dimensional $nttf \times 32 \times 8$ patch functioned as the input for the time-frequency domain models (TFA and TFU). The time and frequency axes were represented by the first (which had a size nttf) and the second dimensions (which had a size 32), respectively. Additionally, the real and imaginary values generated with STFT from the four custom force plate load cell measurements were stored in the eight channels. The output

was a three-dimensional patch with size $nttf \times 32 \times 4$. The four channels also represented the real and imaginary values of the model-predicted STFT-transformed anterior-posterior and mediolateral GRF. Different ntt and nttf were applied in hyperparameter optimization to investigate the optimal input receptive field.

In a typical convolutional layer, l, the value of the output feature, $y^{l}_{i,j,k}$ located at the *i*th row and jth column in the kth feature map was calculated with Eq. (1) [21],

$$y^{l}_{i,j,k} = \boldsymbol{w}^{l}_{k}{}^{T} \boldsymbol{x}^{l}_{i,j} + b^{l}_{k} \tag{1}$$

where $\boldsymbol{w}^{l}_{k}$ is the weight, b^{l}_{k} is the bias of the k^{th} filter, and $\boldsymbol{x}^{l}_{i,j}$ is the input value at the i^{th} row and j^{th} column of the l^{th} layer. Gaussian error linear unit (GELU) activation function is shown in Eq. (2) [22], where x is the function input while y is the function output.

$$y = 0.5x\left(1 + \tanh\left(\sqrt{\frac{2}{\pi}}(x + 0.044715x^{3})\right)\right) \tag{2}$$

A fully-connected layer is the matrix multiplication between a one-dimensional input vector, x, and a two-dimensional weight matrix, W, as stated in Eq. (3) [23],

$$y = Wx + b \tag{3}$$

where b is the scalar bias, and $\boldsymbol{y}$ is the one-dimensional output vector. Bilinear interpolation is described by Eq. (4) and (5) [24],

$$\omega(t) = \begin{cases} 1 - |t|, 0 \le |t| \le 1 \\ 0, others \end{cases} \tag{4}$$

$$y = \begin{bmatrix} \omega(\Delta i) & \omega(1 - \Delta i) \end{bmatrix} \begin{bmatrix} x_{00} & x_{01} \\ x_{10} & x_{11} \end{bmatrix} \begin{bmatrix} \omega(\Delta j) \\ \omega(1 - \Delta j) \end{bmatrix} \tag{5}$$

where y is the interpolated value, x represents the values at the four adjacent corners of a square with coordinates (0, 0), (0, 1), (1, 0), (1, 1), Δi and Δj denote the positions of y in relative to x with coordinate (0, 0). The maximum pooling layer is defined by Eq. (6) [21],

$$y^{l}_{i,j,k} = \text{pool}\left(x^{l}_{m,n,k}\right), \forall (m, n) \in \mathcal{R}_{i,j} \tag{6}$$

where $y^{l}_{i,j,k}$ is the maximum value located at the ith row and jth column in the kth feature map of the lth layer obtained from $\mathcal{R}_{i,j}$, which is a group of adjacent values around the location (i, j).

In Figures 3 and 4, nlc1 represents the number of the first hidden layer channels. Following a dimensional reduction of the time or frequency axis, the number of channels in the subsequent encoder hidden layers was consistently doubled. In autoencoder models like TA and TFA, $ncs1$ signifies the number of convolutional layers with stride as one applies after each dimensionality reduction. The hyperparameters $nlc1$, ndr, and $ncs1$ are optimized in the hyperparameter optimization to achieve adequate model capacity and complexity [25]. The hyperparameter optimization was conducted on U-net models (TU and TFU) to identify the optimal dropout rate (rd) for reducing overfitting in the training dataset [19].

Autoencoder and U-net were selected as they performed effectively in the signal analysis. [26-28]. Even though there were several previous studies that applied machine learning models to predict GRF [3; 6; 8; 9; 29], these models were not used as they only accepted data points from a single time frame. It was doubted that a data point of vertical GRF recorded with the four load cells on a custom force plate could provide enough information to carry out bilateral

https://doi.org/10.31436/iiumej.v26i1.3379

GRF prediction. On the other hand, as compared to recurrent neural network (RNN)-based models, autoencoder and U-net models were relatively less computationally expensive to train, considering the backpropagation in the time dimension, especially when the RNN sequence was long [30].

2.7. Model Training

Google Jax and the NVIDIA RTX 2060 Super GPU were used to implement the model training. During each training iteration, a batch of inputs with a size of 64 was fed into the model to generate the expected outputs. A batch size of 64 was chosen for its positive impact on model accuracy and its compatibility with a higher learning rate, facilitating accelerated model training [31]. The loss was calculated using the mean square error. The weight decay rate (rwd) was implemented across all models to apply L2 norm regularization to the weights, reducing overfitting to the training dataset.

In the case of the autoencoder models (TA and TFA), L1 norm regularisation of hidden unit activations was conducted with a sparsity constant (ks) to explore whether a sparse autoencoder could generalize more effectively than one without sparse representation. Only the bottleneck hidden unit activations were considered in the L1 norm regularisation to prevent excessive regularization of the models. The values of rwd and ks were optimized using hyperparameter optimization. The losses of the autoencoder (lossA) and U-net (lossU) are obtained using Eq. (7) and (8), respectively, as follows:

$$\mathrm{loss_A} = \left\|\widehat{\mathrm{Y}} - \mathrm{Y}\right\|_2 + \mathrm{r_{wd}} \cdot \|\mathrm{W}\|_2 + \mathrm{k_s} \cdot \|\mathrm{H}\|_1 \tag{7}$$

$$\mathrm{loss_U} = \left\|\widehat{\mathrm{Y}} - \mathrm{Y}\right\|_2 + \mathrm{r_{wd}} \cdot \|\mathrm{W}\|_2 \tag{8}$$

where $\widehat{\mathrm{Y}}$ is the predicted output; Y is the actual output, rwd is the weight decay rate; W refers to all layer weights; ks is the sparsity constant; H refers to the bottleneck hidden unit activations.

The weights and biases were tuned using the Adam optimizer. The learning rate (rl) was optimized using hyperparameter optimization. Two phases constituted the model training, during which the models were trained for 100 epochs in the initial phase. Early stopping was performed using the epoch number determined in the first phase to reduce the overfitting of the training dataset.

2.8. Hyperparameter Optimization

Hyperparameter optimization was necessary to tune hyperparameters, which must be fixed before model training and could not be estimated through data learning [32]. The optimal hyperparameters for training models that produced good outcomes were obtained after employing hyperparameter optimization. Table 2 tabulates the list of hyperparameters that are optimized in this study. Remarkably, the range of rl was determined based on the learning rate used in the previous work that analyzed signals with autoencoder, 10-4, 10-3, and 10-2 [33]. Apart from these values, smaller numbers such as 10-6 and 10-5 were used for rl to avoid exploding gradient as the number of layers applied in the models could be more than that in the previous work [33]. The limited time and computational resources required to train the models led to the implementation of Bayesian optimization [34]. The training of 50 models was conducted for each architecture, including TA, TFA, TU, and TFU. The model with the lowest minimal validation loss score was then selected as the most effective model for the corresponding architecture. Minimal validation loss was used as the model selection metric because it accounted for overfitting to the training dataset compared to training loss [35]. Even though test loss exhibited a similar characteristic to minimal validation loss, test loss was not

applied in model selection to avoid train-test contamination in evaluating the performance of the selected models [36].

Table 2. Summary list of the hyperparameters to be optimized, description, targeted models, and the values.

Hyperparameter	Description	Targeted Models	Values
n_{tt}	Dimension of time axis for time-domain input and output	TA, TU	16, 48, 112
n_{ttf}	Dimension of time axis for time-frequency domain input and output	TFA, TFU	2, 4, 8 (corresponded to 16, 48, and 112 values of n_{tt} after STFT)
n_{lc1}	Number of channels of first hidden layer	TA, TFA, TU, TFU	16, 32, 64, 128
n_{dr}	Number of dimensional reductions in time and frequency axis of encoder	TA, TFA, TU, TFU	2, 3, 4, 5
n_{cs1}	Number of convolutional layers with stride as one after each dimensional reduction	TA, TFA	1, 2, 3
r_d	Dropout rate	TU, TFU	0, 0.25, 0.5, 0.75
r_{wd}	Weight decay rate	TA, TFA, TU, TFU	10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}
k_s	Sparsity constant	TA, TFA	0, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}
r_l	Learning rate	TA, TFA, TU, TFU	10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}

2.9. Model Deployment

The best model of each architecture was employed to predict the bilateral GRF readings of complete sit-to-stand, gait initialization, and gait trials in the training dataset (inclusive of the validation dataset) and test dataset. Initially, if a time-frequency domain model (such as TFA and TFU) was used, STFT was applied to the complete vertical GRF readings of the individual load cell. The trial readings were then cropped into patches with ntt or $nttf$ input size and fed into the model. For the cropping operation, sliding windows with a stride of 50% of the input size were applied. Subsequently, the average of the overlapping outputs for the model was calculated. If the time-frequency domain model was implemented, inverse STFT was applied to the averaged output. Lastly, the expected time series bilateral GRF represented the model's output.

2.10. Statistical Analysis

Prior to the analysis, the bilateral GRF was normalized using the participant's weight to obtain a value expressed as a percentage of body weight (% BW). To determine the extent of the difference between predicted readings and their actual counterparts, the Root Mean Square Error (RMSE) and relative RMSE (rRMSE) between the actual and predicted bilateral GRF readings were computed [37]. The Pearson correlation coefficient (ρ) was also calculated to evaluate the linearity between predicted and actual readings. The value of ρ was important for assessing the accuracy of predicted readings. If predicted readings showed a strong linear correlation with actual readings, the model could infer the trend of actual readings with minimal residual noise. The equation of ρ is shown in Eq. (9) [38],

$$\rho = \frac{\sum(x_p - \bar{x}_p)(x_a - \bar{x}_a)}{\sqrt{(\sum(x_p - \bar{x}_p)^2 \sum(x_a - \bar{x}_a)^2)}} \quad (9)$$

https://doi.org/10.31436/iiumej.v26i1.3379

where x_p is predicted reading, $\bar{x}_p$ is the mean of predicted readings, x_a is actual reading, and $\bar{x}_a$ is the mean of actual readings. Based on the ranges of ρ, the strength of the correlation is considered [39] as follows:

- 0.9 is "very strong"
- 0.7–0.9 is "strong"
- 0.4–0.7 is "moderate"
- 0.2–0.4 is "weak"
- < 0.2 is "very weak"

The coefficient of determination (R2) was calculated to assess the fit of the predicted readings to the actual readings. A good fit between predicted and actual readings was suggested, with R2 exceeding 0.8 [40]. The agreement between predicted and actual readings was assessed using the intraclass correlation coefficient [ICC (2, 1)] through the application of the two-way mixed-effects model (single rater type and defining the definition as absolute agreement) [41]. The definition is relevant to measuring the model performance, where the actual and predicted readings acquired from the same tasks must be equal to achieve absolute agreement. ICC (2, 1) is calculated according to Eq. (10),

$$\mathrm{ICC}\,(2,1) = \frac{\mathrm{MS_R} - \mathrm{MS_E}}{\mathrm{MS_R} + (\mathrm{k_r}-1)\mathrm{MS_E} + \frac{\mathrm{kr}}{\mathrm{n}}(\mathrm{MS_C} - \mathrm{MS_E})} \tag{10}$$

where MSR is the mean square for trial, MSE is the mean square for error, MSC is the mean square for actual and predicted readings, kr is the number of raters (which is one since only actual readings are used as the gold standard), and n is the number of readings. The agreement is rated based on the ICC (2, 1) ranges [41] as follows:

- 0.9 is "excellent"
- 0.75–0.9 is "good"
- 0.5–0.75 is "moderate"
- < 0.5 is "poor"

3. RESULTS

The optimal hyperparameter sets were obtained via hyperparameter optimization, as tabulated in Table 3. Model training with these hyperparameters yielded TA, TFA, TU, and TFU models with the lowest minimal validation loss.

Table 3. Summary of the optimal hyperparameter sets for TA, TFA, TU, and TFU models

Hyperparameter	Autoencoder		U-net	
	TA	TFA	TU	TFU
n_{tt}	112	-	112	-
n_{ttf}	-	8	-	8
n_{lc1}	64	64	64	128
n_{dr}	3	2	4	2
n_{cs1}	2	1	-	-
r_d	-	-	0	0.5
r_{wd}	10^{-6}	10^{-7}	10^{-5}	10^{-5}
k_s	10^{-5}	10^{-5}	-	-
r_l	10^{-4}	10^{-4}	10^{-4}	10^{-3}

https://doi.org/10.31436/iiumej.v26i1.3379

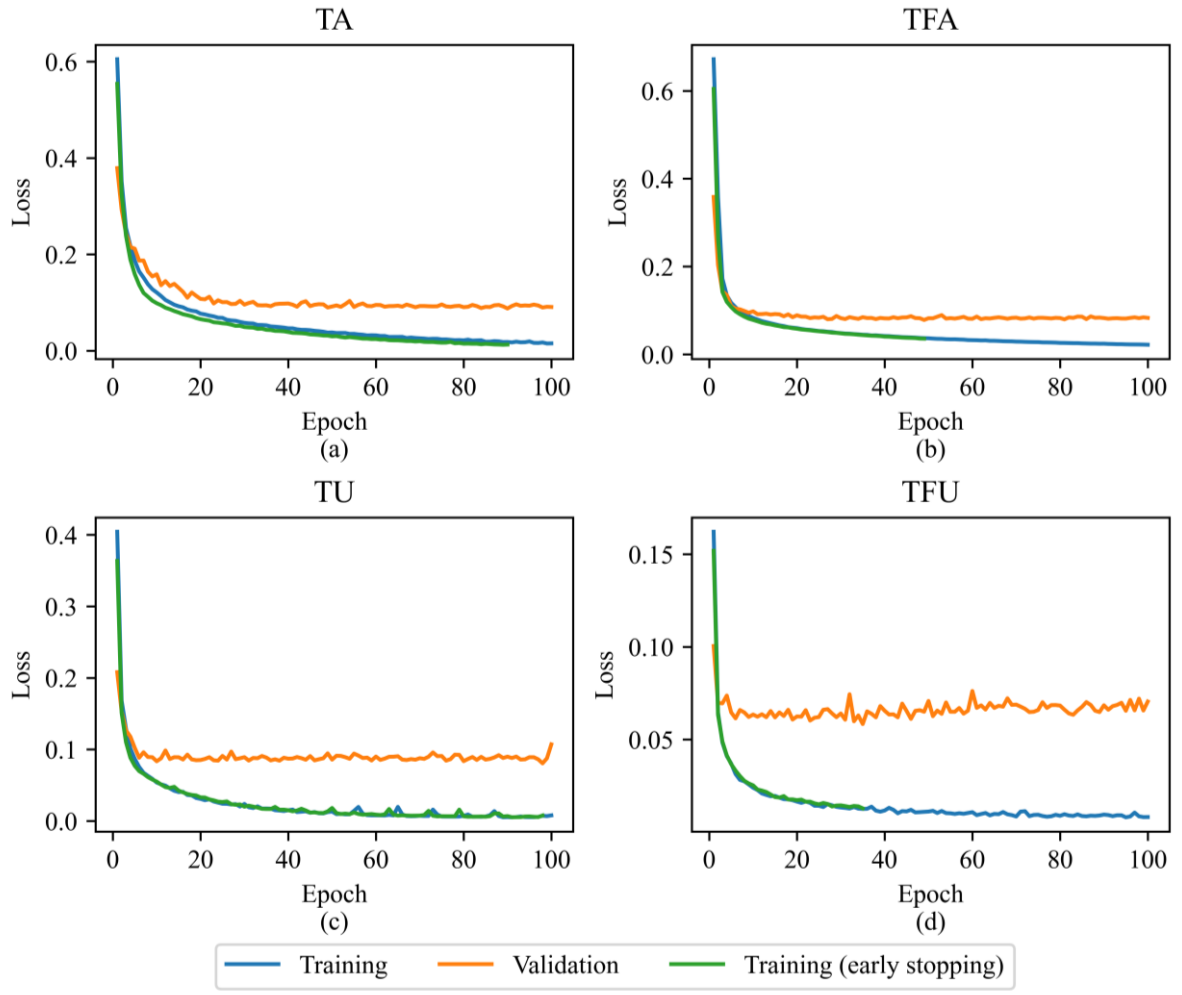

Figure 5. Graphs of loss versus epoch for training of different models.

The graphs of loss versus epoch are plotted in Figure 5 for the training of each model listed in Table 3. The training and validation losses during model training without and with early stopping, as highlighted in Section "Model training," are shown. For all models, the losses decreased and converged towards a horizontal level over training epochs, indicating that the configurations listed in Table 3 were suitable for conducting stable model training. Validation losses did not exhibit a clear U-shaped curve, indicating that overfitting to the training dataset did not occur to a significant extent. By implementing early stopping at the epoch when minimum validation loss was recorded, overfitting to the training dataset could be further reduced.

https://doi.org/10.31436/iiumej.v26i1.3379

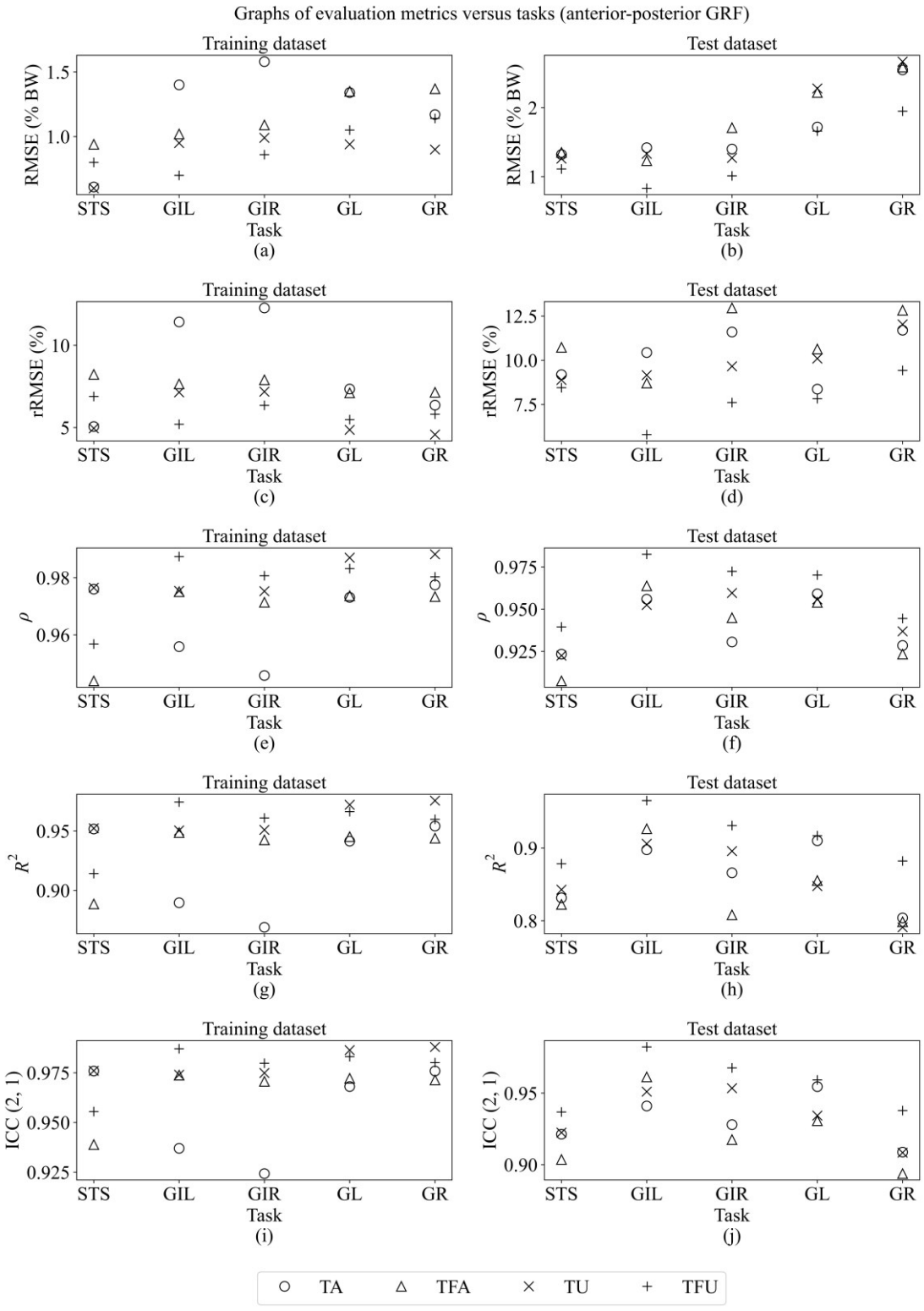

Figure 6. Graphs of RMSE, rRMSE, ρ, R^2, and ICC (2, 1) versus task in measuring anterior-posterior GRF. The tasks include sit-to-stand (STS), gait initialization with the left foot (GIL), gait initialization with the right foot (GIR), gait with the left foot stepping on a force plate (GL), and gait with the right foot stepping on a force plate (GR). The validation dataset was included in the training dataset.

https://doi.org/10.31436/iiumej.v26i1.3379

Figure 6 portrays the RMSE, rRMSE, ρ, R2, and ICC (2, 1) between the predicted and actual anterior-posterior GRF readings. According to the figure, the TFU model yielded the best results for the test dataset. The TFU model surpassed the TA, TFA, and TU models, achieving lower RMSE and rRMSE for anterior-posterior GRF in all tasks. The TFU model also attained the highest test dataset ρ, R2, and ICC (2, 1) for all tasks. According to the test dataset, the TFU model demonstrated excellent agreement (ICC > 0.9), good fitting (R2 > 0.8), and very strong linear correlation (ρ > 0.9) for all tasks in predicting anterior-posterior GRF.

According to Figure 7, the TFU model records the lowest RMSE and rRMSE in inferring the test dataset mediolateral GRF for all tasks, except the rRMSE obtained for sit-to-stand, where the TA model achieves the lowest value. Moreover, the TFU model obtained the highest values of ρ, R2, and ICC (2, 1) in the test dataset for all tasks. The TFU model demonstrated excellent agreement (ICC > 0.9), good fitting (R2 > 0.8), and very strong linear correlation (ρ > 0.9) in predicting the mediolateral GRF for gait initialization and gait with the right foot stepping on a force plate. A strong linear correlation ($0.7 < \rho \leq 0.9$) and moderate agreement ($0.5 < ICC \leq 0.75$) were also observed for the mediolateral GRF during sit-to-stand. In addition, the TFU model successfully predicted the mediolateral GRF during gait with the left foot stepping on a force plate, showing a strong linear correlation ($0.7 < \rho \leq 0.9$) and good agreement ($0.75 < ICC \leq 0.9$). The values of RMSE, rRMSE, ρ, R2, and ICC (2, 1) displayed in Figures 6 and 7 are listed in Supplementary File 1 in the form of tables.

Figure 8 presents examples of anterior-posterior GRF versus time graphs for sit-to-stand, gait initialization, and gait. The trial data for the graphs was randomly chosen from the test dataset. Each predicted curve fitting to the actual measurements was demonstrated with the R^2 values. The TFU model generated the highest R^2 values throughout all examined trials. Compared to other models, the TFU model fit more closely to the actual readings. The predicted measurements of the TFA model were obviously smoother than the actual equivalents. On the contrary, the prediction results of time domain models (TA and TU) suggested undesirable abrupt changes [see Figure 8(b) to 8(e)].

https://doi.org/10.31436/iiumej.v26i1.3379

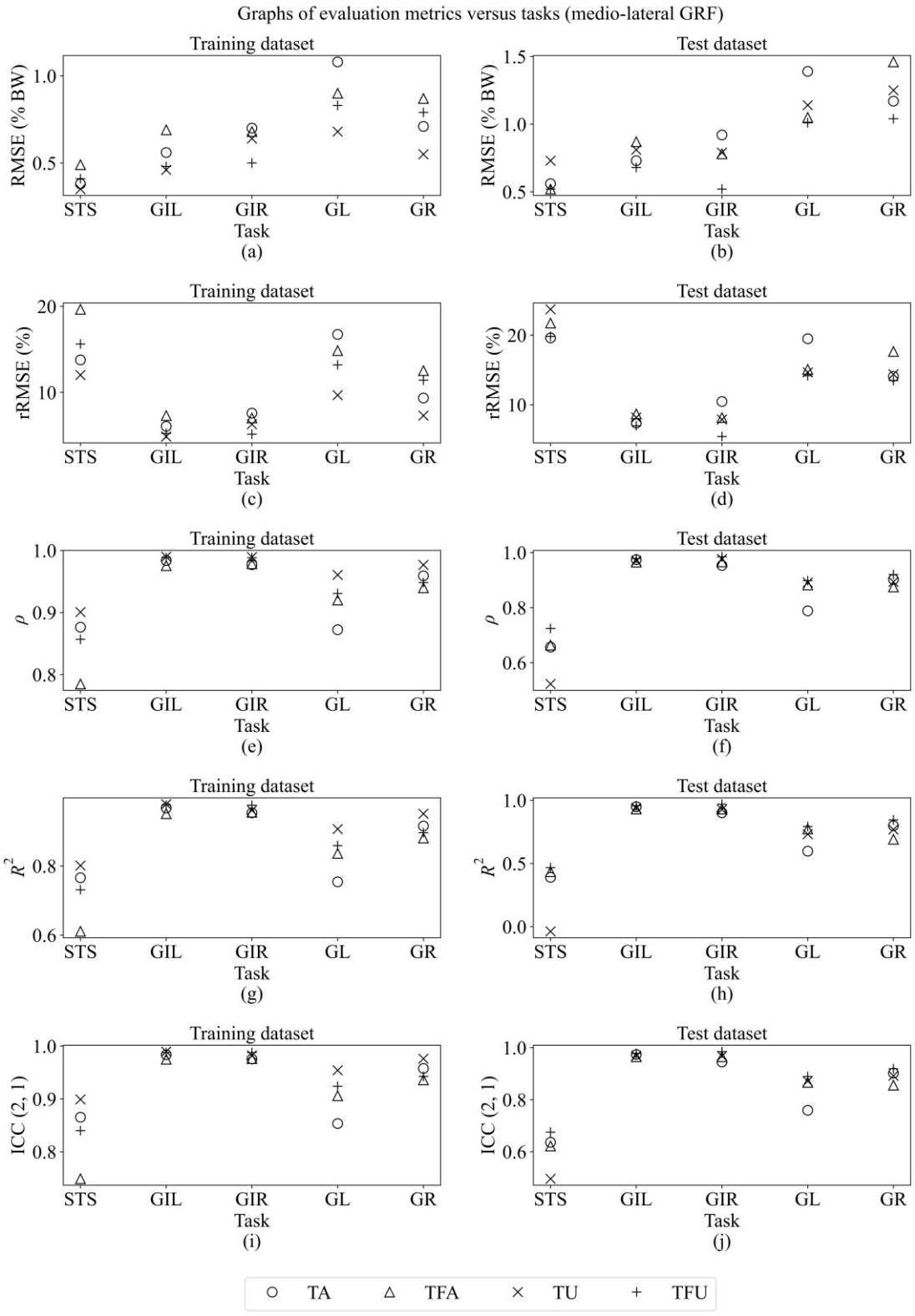

Figure 7. Graphs of RMSE, rRMSE, ρ, R^2, and ICC (2, 1) versus task in measuring mediolateral GRF. The tasks include sit-to-stand (STS), gait initialization with the left foot (GIL), gait initialization with the right foot (GIR), gait with the left foot stepping on a force plate (GL), and gait with the right foot stepping on a force plate (GR). The validation dataset was included in the training dataset.

https://doi.org/10.31436/iiumej.v26i1.3379

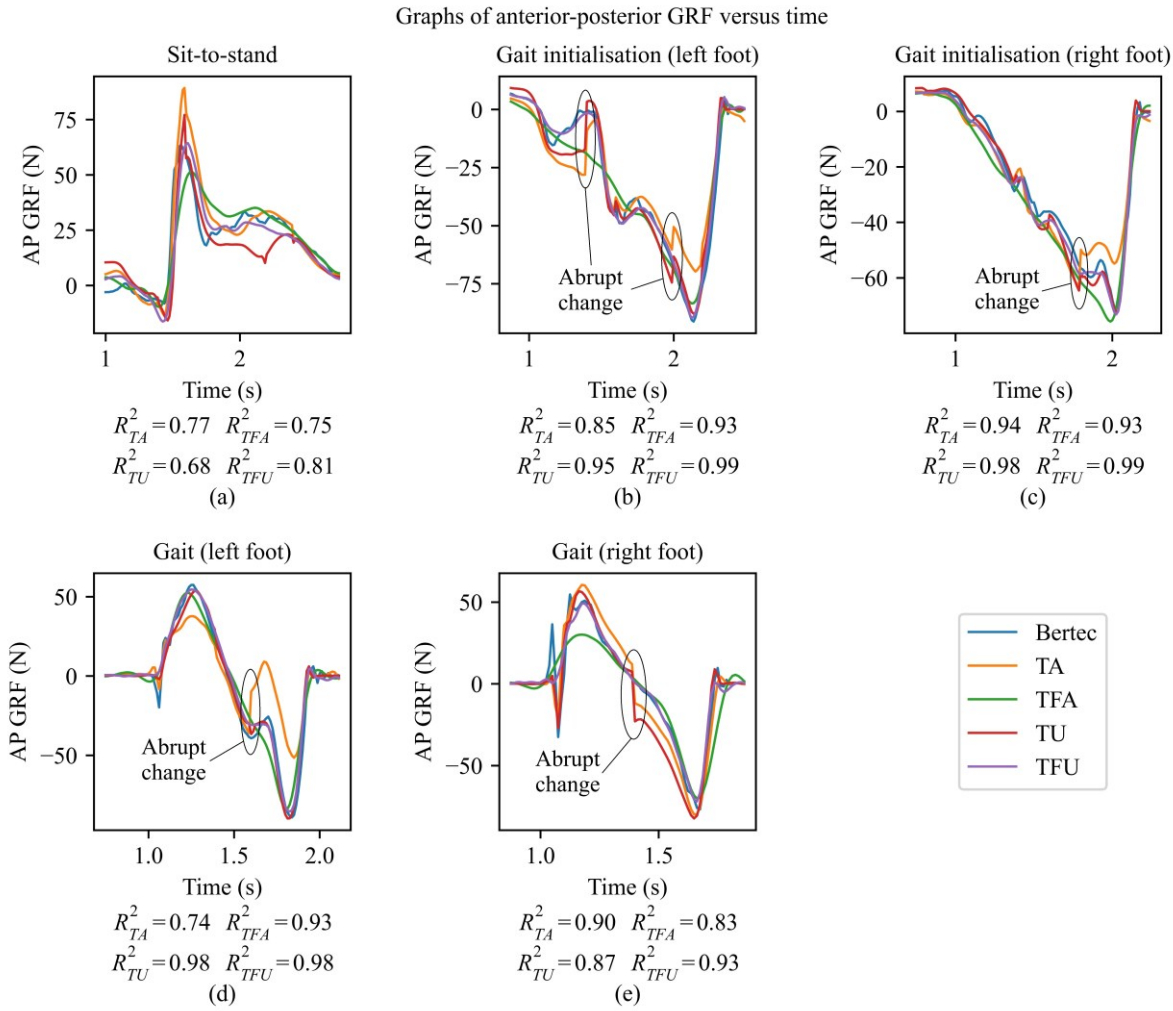

Figure 8. Graphs of anterior-posterior GRF versus time for various tasks. AP represents anterior-posterior.

The graphs of mediolateral GRF versus time plotted with randomly chosen trial data are shown in Figure 9. The TFU model achieved the highest R2 values for most of the selected trials, except in predicting the mediolateral GRF for gait where the right foot stepped on the force plate. Again, the TFA model-generated curves are the smoothest, while abrupt changes in TA and TU model readings can be observed based on Figures 9(b) and 9(d). The mediolateral GRF measured during sit-to-stand can barely be predicted by all the models based on Figure 9(a) and the corresponding R2 values.

https://doi.org/10.31436/iiumej.v26i1.3379

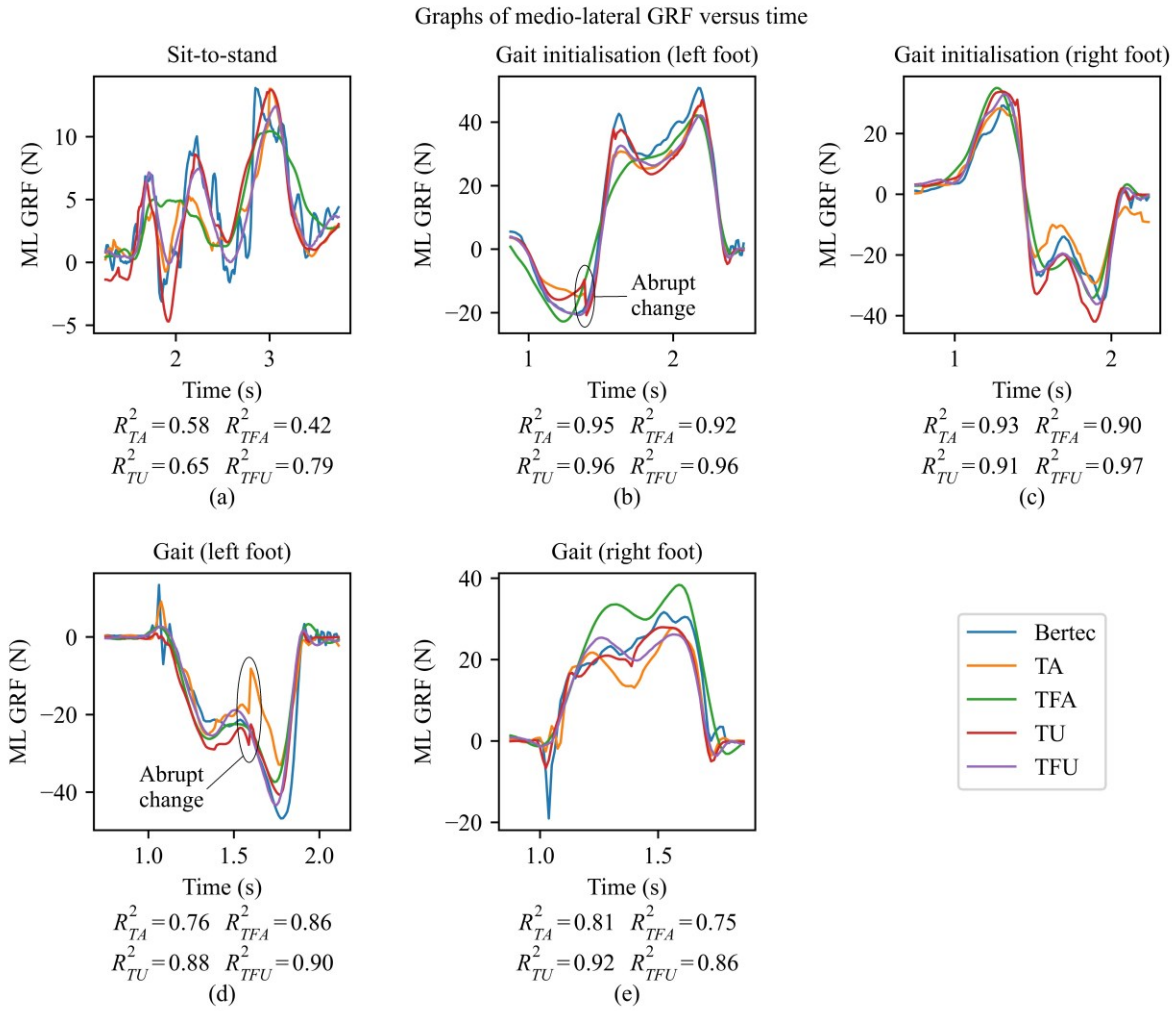

Figure 9. Graphs of mediolateral GRF versus time for various tasks. ML represents mediolateral.

4. DISCUSSION

Four deep learning models (TA, TFA, TU, and TFU) were developed in this study to predict the bilateral GRF from the vertical GRF measured using a cost-effective custom force plate. In Table 3, it was noted that inputs with the largest dimensions (ntt and $nttf$) were selected to train the best models, suggesting that these models needed data from the majority of vertical GRF readings to predict the bilateral GRF accurately. This shows that the autoencoder bottleneck's sparsity helped extract features from the input to improve the output inference.

Figs. 6 and 7 highlight that the TFU model is the most accurate and demonstrates the most generalisability. Figs. 8 and 9 further indicated that, compared to other models, the prediction of the TFU model matched the actual readings to a closer extent. The absence of overshoot in TFU-predicted readings provided insight into the model architecture and hyperparameters applied, which helped mitigate overfitting through appropriate model complexity.

Compared to other tasks, the actual pattern of mediolateral GRF recorded during sit-to-stand was more amorphous. Thus, this observation explained the challenging prediction task for sit-to-stand movement. The satisfactory results indicated that the information extracted from four vertical GRF readings recorded using custom force plate load cells was insufficient

https://doi.org/10.31436/iiumej.v26i1.3379

to accurately predict the mediolateral GRF during sit-to-stand. A subject may produce multiple sit-to-stand vertical readings that are similar but exhibit inconsistent mediolateral GRF patterns due to minor variations in sit-to-stand movements. To enhance accuracy, prediction models may require additional sources of information, such as human body posture and kinematics.

The values predicted by the TFA were significantly smoother than the actual measurements, suggesting that the model could not capture the high-frequency components of the bilateral GRF. Conversely, the swift changes in outputs generated by the time-domain models (TA and TU) may have been due to distinct values at the boundaries of overlapping output that could not be smoothed out through the averaging process (refer to Section "Model deployment"). The absence of sudden changes in the time-frequency domain model, TFA, and TFU outputs may have been due to the extensive STFT sliding window overlap. The STFT sliding window size was 62, significantly larger than the number of strides of 16, resulting in a substantial overlap between consecutive sliding windows. This overlap prevented a decrease in reading magnitude at the window function's boundary, leading to smoother readings [42]. Reducing the undesirable quick changes in predicted readings allowed for a better fit between predicted and actual readings. This explained the accuracy improvement when STFT was applied.

Two explanations were proposed to justify the accurate prediction. The initial explanation (refer to Section 1) suggested that the model could have extracted features from the strain gauge elongation caused by off-axis bilateral GRF exertion to make the prediction. Secondly, prediction could possibly be conducted due to the consistent patterns of model inputs and outputs. The vertical GRF readings obtained by individual load cells of the custom force plate during various tasks followed several recognizable and consistent trends (see Figure 10). For example, the vertical GRF readings for sit-to-stand obtained by load cells began with a decrease, progressed through a peak and a local minimum, and finally ended with decreasing fluctuation. In contrast, the vertical GRF readings recorded during gait initialization started with fluctuation, followed by a peak, and a drop to zero. Clear patterns were observed in gait vertical GRF readings, with values increasing from zero to a peak and then decreasing back to zero at varying rates.

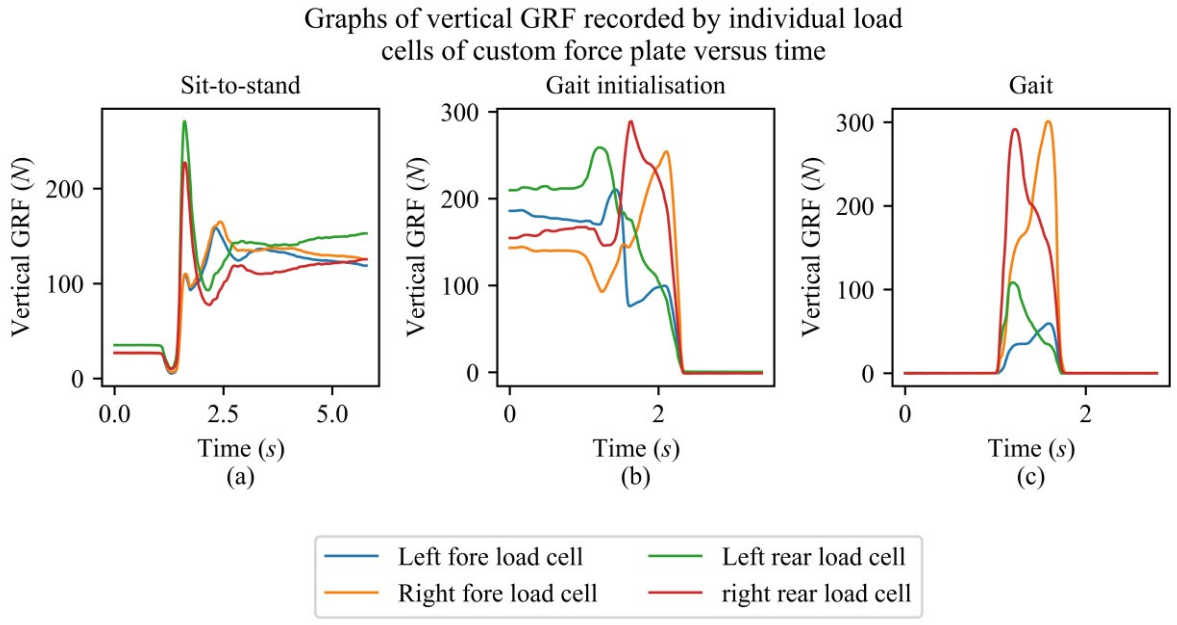

Figure 10. Graphs of vertical GRF recorded by individual load cells of custom force plate versus time for various tasks. The readings recorded exhibit perceivable patterns.

https://doi.org/10.31436/iiumej.v26i1.3379

On the other hand, the bilateral GRF recorded for most tasks also illustrated predictable patterns (see Figures 8 and 9). For instance, the anterior-posterior GRF measurements during gait initialization indicated a deep valley, showing the propulsive force exerted by the participant's feet as they moved forward. The anterior-posterior GRF during gait included a peak representing the peak braking force and a valley indicating the peak propulsive force. Distinct trends were noticeable in mediolateral GRF, particularly during gait initialization, where the readings started with a peak or valley, followed by an 'M' or 'W' pattern, depending on the foot used to start the gait. Thus, the model could have linked the input pattern to the corresponding output trend.

The findings of this study were compared with those of earlier studies. In previous studies, RMSE stated in the unit "N/kg" was transformed into % BW. Due to the limited information in previously published studies, the sit-to-stand findings were not compared. Only one previous study investigated the GRF prediction during gait initialization [1]. The RMSE of the anterior-posterior GRF (0.83–1.01 % BW) and mediolateral GRF (0.52–0.68 % BW) recorded in this study for gait initialization using the TFU model were lower than the previous study (1.63–2.96 % BW for anterior-posterior GRF and 1.73–1.94 % BW for mediolateral GRF) [1]. Consequently, the ρ values of gait initialization anterior-posterior GRF (0.97–0.98 versus 0.94–0.97) and mediolateral GRF (0.98 versus 0.84–0.90) values obtained in this study were greater than the earlier study [1]. Hence, the bilateral GRF prediction during gait initialization was improved in this work as compared to before.

Figure 11 reveals the comparison of the result for gait between this work and previous studies, in which the TFU model prediction results are concluded as follows when anterior-posterior GRF during gait is taken into consideration:

- Lower RMSE range than seven studies [1; 2; 4; 5; 9; 10; 29]
- Lower rRMSE range than three studies [5; 8; 29]
- Higher ρ range than the three studies [4; 5; 29]

https://doi.org/10.31436/iiumej.v26i1.3379

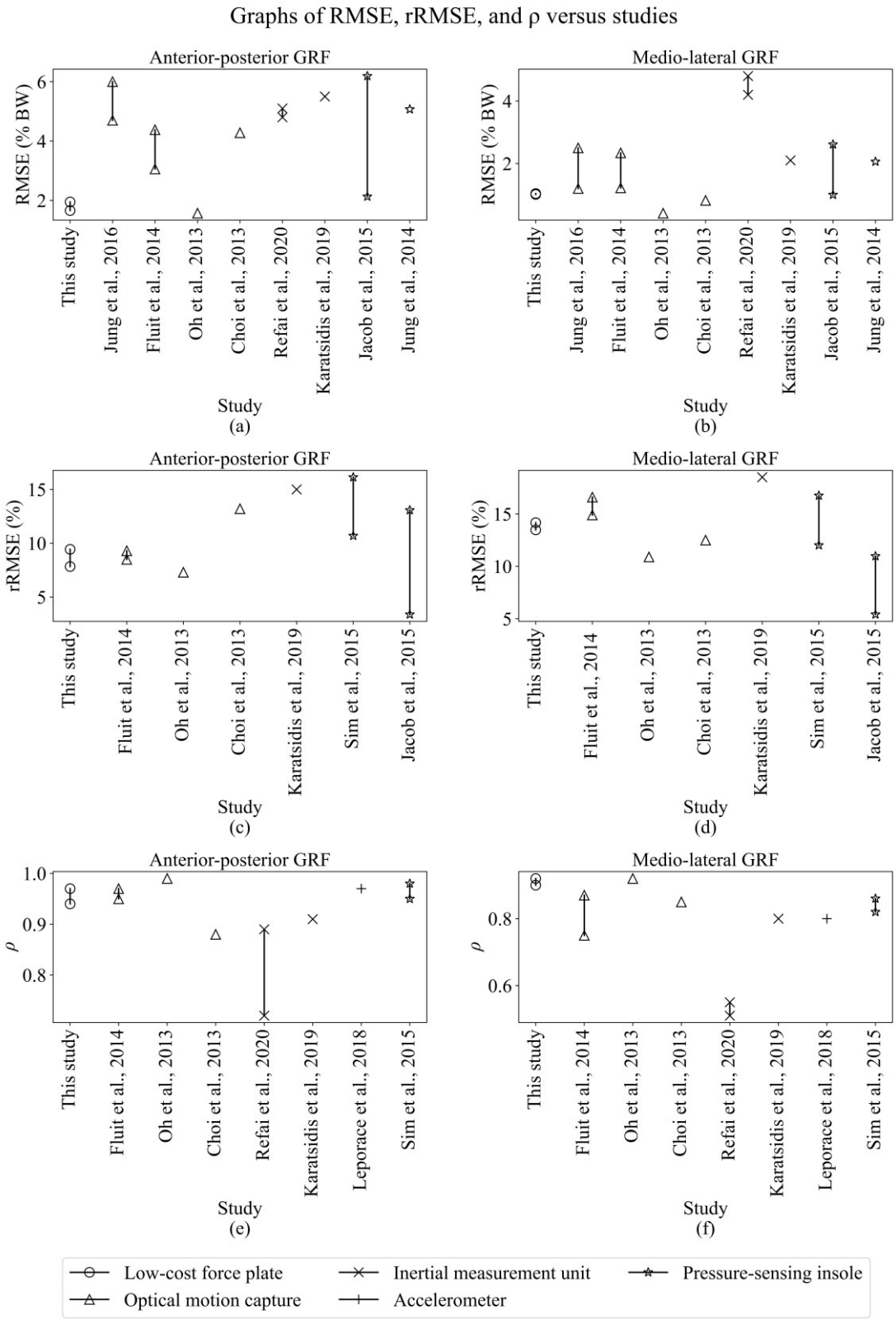

Figure 11. The RMSE, rRMSE, and ρ in this study and previous studies. The data investigates bilateral GRF prediction in gait. The results from this study are derived from the test dataset.

Furthermore, the mediolateral gait GRF results achieved in this study are observed to exhibit the following:

- Lower RMSE range than five studies [1; 2; 4; 5; 10]
- Lower rRMSE range than two studies [1; 5]
- Higher ρ range than six studies [1; 4-6; 8; 29]

Therefore, the comparison showed that the method in this study generated a relatively good prediction.

A larger dataset could be collected to improve the model performance and generalization in future work. By collecting more data from different individuals, the models could learn the various patterns of bilateral GRF that do not exist in the current dataset. This could potentially improve the prediction accuracy and generalization of data from different subjects, especially for the outputs that were challenging to infer, such as mediolateral GRF during sit-to-stand tasks. Furthermore, the larger dataset could include GRF recorded during other tasks, such as running and jumping. This enhanced the model's ability to generalize across a wider range of tasks, thereby improving its applicability in various research settings.

In dealing with noisy input, high-frequency noises, and outliers were suppressed with low-pass and medium filters, as highlighted in the Section "Data preprocessing." However, the application of external force on the subjects who performed tasks on the force plates could cause perturbation in GRF readings, which could not be reduced with low-pass and medium filters. In future studies, sensitivity analysis will be conducted to identify the effect of the noises due to the application of external forces. The model would be trained to generate the correct bilateral GRF readings under this condition.

5. CONCLUSION

The bilateral GRF predictions in this study were successfully implemented for sit-to-stand, gait initialization, and gait. The best model was determined to be the TFU model. Based on the test dataset readings for all tasks, apart from the mediolateral GRF readings for sit-to-stand, the model could generalize to a good extent with a high ICC (2, 1) of more than 0.88. The accuracy of the low-cost custom force plate and TFU model in bilateral GRF prediction was demonstrated to be comparable with and even higher than other previous alternatives. The results could encourage biomechanical studies even in budget-limited conditions by using a low-cost bilateral GRF sensing force plate. The proposed method could enhance the accuracy of force and torque acting on musculoskeletal models calculated during biomechanical analysis compared to any other low-cost force plate that cannot quantify bilateral GRF. Moreover, the bilateral GRF-based metrics used to assess health conditions may soon be obtained using a low-cost force plate.

ACKNOWLEDGEMENT

This work was supported by Collaborative Research in Engineering, Science & Technology Center (CREST), School of Mechanical Engineering, Universiti Sains Malaysia, and SAS Institute Sdn. Bhd. under Grant 304.PMEKANIK.6050419.C121.

REFERENCES

[1] Fluit, R., Andersen, M.S., Kolk, S., Verdonschot, N., Koopman, H.F. (2014). Prediction of ground reaction forces and moments during various activities of daily living. J Biomech, 47(10):2321-2329.10.1016/j.jbiomech.2014.04.030.

[2] Jung, Y., Jung, M., Ryu, J., Yoon, S., Park, S.K., Koo, S. (2016). Dynamically adjustable foot-ground contact model to estimate ground reaction force during walking and running. Gait Posture, 45:62-68.10.1016/j.gaitpost.2016.01.005.

[3] Oh, S.E., Choi, A., Mun, J.H. (2013). Prediction of ground reaction forces during gait based on kinematics and a neural network model. J Biomech, 46(14):2372-2380.10.1016/j.jbiomech.2013.07.036.

[4] Refai, M.I.M., Beijnum, B.J.F.v., Buurke, J.H., Veltink, P.H. (2020). Portable Gait Lab: Estimating 3D GRF Using a Pelvis IMU in a Foot IMU Defined Frame. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(6):1308-1316.10.1109/TNSRE.2020.2984809.

[5] Karatsidis, A., Jung, M., Schepers, H.M., Bellusci, G., de Zee, M., Veltink, P.H., et al. (2019). Musculoskeletal model-based inverse dynamic analysis under ambulatory conditions using inertial motion capture. Med Eng Phys, 65:68-77.https://doi.org/10.1016/j.medengphy.2018.12.021.

[6] Leporace, G., Batista, L., Nadal, J. (2018). Prediction of 3D ground reaction forces during gait based on accelerometer data. Research on Biomedical Engineering, 34.10.1590/2446-4740.06817.

[7] Faber, G.S., Chang, C.C., Kingma, I., Dennerlein, J.T., van Dieën, J.H. (2016). Estimating 3D L5/S1 moments and ground reaction forces during trunk bending using a full-body ambulatory inertial motion capture system. J Biomech, 49(6):904-912.10.1016/j.jbiomech.2015.11.042.

[8] Sim, T., Kwon, H., Oh, S.E., Joo, S.B., Choi, A., Heo, H.M., et al. (2015). Predicting Complete Ground Reaction Forces and Moments During Gait With Insole Plantar Pressure Information Using a Wavelet Neural Network. J Biomech Eng, 137(9).10.1115/1.4030892.

[9] Jacobs, D.A., Ferris, D.P. (2015). Estimation of ground reaction forces and ankle moment with multiple, low-cost sensors. J Neuroeng Rehabil, 12:90.10.1186/s12984-015-0081-x.

[10] Jung, Y., Jung, M., Lee, K., Koo, S. (2014). Ground reaction force estimation using an insole-type pressure mat and joint kinematics during walking. J Biomech, 47(11):2693-2699.10.1016/j.jbiomech.2014.05.007.

[11] Abujaber, S., Gillispie, G., Marmon, A., Zeni, J., Jr. (2015). Validity of the Nintendo Wii Balance Board to assess weight bearing asymmetry during sit-to-stand and return-to-sit task. Gait Posture, 41(2):676-682.10.1016/j.gaitpost.2015.01.023.

[12] Yamako, G., Chosa, E., Totoribe, K., Fukao, Y., Deng, G. (2017). Quantification of the sit-to-stand movement for monitoring age-related motor deterioration using the Nintendo Wii Balance Board. PLoS One, 12(11):e0188165.10.1371/journal.pone.0188165.

[13] Arifin, W.N. (2018). A Web-based Sample Size Calculator for Reliability Studies. 10:67-76.10.21315/eimj2018.10.3.8.

[14] Onuma, R., Masuda, T., Hoshi, F., Matsuda, T., Sakai, T., Okawa, A., et al. (2021). Measurements of the centre of pressure of individual legs reveal new characteristics of reduced anticipatory postural adjustments during gait initiation in patients with post-stroke hemiplegia. J Rehabil Med, 53(7):jrm00211.10.2340/16501977-2856.

[15] Eltoukhy, M., Kuenze, C., Andersen, M.S., Oh, J., Signorile, J. (2017). Prediction of ground reaction forces for Parkinson's disease patients using a kinect-driven musculoskeletal gait analysis model. Med Eng Phys, 50:75-82.10.1016/j.medengphy.2017.10.004.

[16] Izhar, C., Zakaria, H., Maruzuki, M., Sulaiman, M., Rahim, A. (2021). Gait cycle prediction model based on gait kinematic using machine learning technique for assistive rehabilitation device. IAES International Journal of Artificial Intelligence (IJ-AI), 10:752.10.11591/ijai.v10.i3.pp752-763.

[17] Vittinghoff, E., McCulloch, C.E. (2007). Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol, 165(6):710-718.10.1093/aje/kwk052.

[18] Bank, D., Koenigstein, N., Giryes, R. (2020). Autoencoders. arXiv preprint arXiv:2003.05991.

[19] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 5-9 October 2015;Cham; pp 234-241.

[20] Yu, S., Príncipe, J.C. (2019). Understanding autoencoders with information theoretic concepts. Neural Networks, 117:104-123.https://doi.org/10.1016/j.neunet.2019.05.003.

[21] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent advances in convolutional neural networks. Pattern Recognition, 77:354-377.https://doi.org/10.1016/j.patcog.2017.10.013.

[22] Lee, M. (2023). Mathematical Analysis and Performance Evaluation of the GELU Activation Function in Deep Learning. Journal of Mathematics, 2023(1):4229924.https://doi.org/10.1155/2023/4229924.

[23] Mu, R., Zeng, X. (2019). A Review of Deep Learning Research. KSII Trans. Internet Inf. Syst., 13:1738-1764.

[24] Sun, N., Li, H. (2019). Super Resolution Reconstruction of Images Based on Interpolation and Full Convolutional Neural Network and Application in Medical Fields. IEEE Access, 7:186470-186479.10.1109/ACCESS.2019.2960828.

[25] Goodfellow, I., Yoshua, B., Courville, A. (2016). Capacity, overfitting and underfitting. In Deep Learning. MIT press; pp. 108 - 114.

[26] Zeghidour, N., Luebs, A., Omran, A., Skoglund, J., Tagliasacchi, M. (2021). Soundstream: An end-to-end neural audio codec. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30:495-507.

[27] Kristoffersen, M., Møller, M., Martinez Nuevo, P., Ostergaard, J. (2021). Deep Sound Field Reconstruction in Real Rooms: Introducing the ISOBEL Sound Field Dataset. ArXiv, abs/2102.06455.https://arxiv.org/abs/2102.06455.

[28] Li, Y., Tagliasacchi, M., Rybakov, O., Ungureanu, V., Roblek, D. (2021). Real-time speech frequency bandwidth extension. In Proceedings of ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP): 6-11 June 2021; pp 691-695.

[29] Choi, A., Lee, J.-M., Mun, J.H. (2013). Ground reaction forces predicted by using artificial neural network during asymmetric movements. International Journal of Precision Engineering and Manufacturing, 14(3):475-483.10.1007/s12541-013-0064-4.

[30] Orvieto, A., Smith, S.L., Gu, A., Fernando, A., Gulcehre, C., Pascanu, R., et al., (2023). Resurrecting Recurrent Neural Networks for Long Sequences, in: Andreas, K., Emma, B., Kyunghyun, C., Barbara, E., Sivan, S., Jonathan, S. (Eds.), Proceedings of the 40th International Conference on Machine Learning. PMLR, Proceedings of Machine Learning Research, pp. 26670--26698.

[31] Kandel, I., Castelli, M. (2020). The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. ICT Express, 6(4):312-315.https://doi.org/10.1016/j.icte.2020.04.010.

[32] Yang, L., Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing, 415:295-316.https://doi.org/10.1016/j.neucom.2020.07.061.

[33] Nogales, A., Caracuel-Cayuela, J., García-Tejedor, Á.J., (2024). Analyzing the Influence of Diverse Background Noises on Voice Transmission: A Deep Learning Approach to Noise Suppression, Applied Sciences.

[34] Snoek, J., Larochelle, H., Adams, R.P. (2012). Practical bayesian optimization of machine learning algorithms. Advances in neural information processing systems, 25.

[35] Goodfellow, I., Yoshua, B., Courville, A. (2016). Early stopping. In Deep Learning. MIT press; pp. 241 - 249.

[36] Magar, I., Schwartz, R. (2022). Data Contamination: From Memorization to Exploitation. ArXiv, abs/2203.08242.

[37] Ren, L., Jones, R.K., Howard, D. (2008). Whole body inverse dynamics over a complete gait cycle based only on measured kinematics. J Biomech, 41(12):2750-2759.https://doi.org/10.1016/j.jbiomech.2008.06.001.

[38] Šverko, Z., Vrankić, M., Vlahinić, S., Rogelj, P., (2022). Complex Pearson Correlation Coefficient for EEG Connectivity Analysis, Sensors.

[39] Learner, J., Goodman, N.W. (1996). Descriptions of correlation. Lancet, 348(9021):199-200.10.1016/s0140-6736(05)66149-1.

[40] Chicco, D., Warrens, M.J., Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci, 7:e623.10.7717/peerj-cs.623.

[41] Koo, T.K., Li, M.Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med, 15(2):155-163.10.1016/j.jcm.2016.02.012.

[42] Heinzel, G., Rüdiger, A.O., Schilling, R. (2002). Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new at-top windows.