

## Article

# Treatment Outcome Prediction Using Multi-Task Learning: Application to Botulinum Toxin in Gait Rehabilitation

Adil Khan <sup>1,2,\*</sup> , Antoine Hazart <sup>1</sup>, Omar Galarraga <sup>3</sup> , Sonia Garcia-Salicetti <sup>4</sup>  and Vincent Vigneron <sup>1,\*</sup> <sup>1</sup> Informatique, Bio-Informatique et Systèmes Complexes (IBISC) EA 4526, Univ Evry, Université Paris-Saclay, 91020 Evry, France<sup>2</sup> Department of Computer Science, Sukkur IBA University, Sukkur 65200, Sindh, Pakistan<sup>3</sup> UGECAM Ile-de-France, Movement Analysis Laboratory, 77170 Coubert, France<sup>4</sup> SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, 91764 Palaiseau, France\* Correspondence: [adil.khan@universite-paris-saclay.fr](mailto:adil.khan@universite-paris-saclay.fr) or [adilkhan@iba-suk.edu.pk](mailto:adilkhan@iba-suk.edu.pk) (A.K.); [vincent.vigneron@univ-evry.fr](mailto:vincent.vigneron@univ-evry.fr) (V.V.)

**Abstract:** We propose a framework for optimizing personalized treatment outcomes for patients with neurological diseases. A typical consequence of such diseases is gait disorders, partially explained by command and muscle tone problems associated with spasticity. Intramuscular injection of botulinum toxin type A is a common treatment for spasticity. According to the patient's profile, offering the optimal treatment combined with the highest possible benefit-risk ratio is important. For the prediction of knee and ankle kinematics after botulinum toxin type A (BTX-A) treatment, we propose: (1) a regression strategy based on a multi-task architecture composed of LSTM models; (2) to introduce medical treatment data (MTD) for context modeling; and (3) a gating mechanism to model treatment interaction more efficiently. The proposed models were compared with and without metadata describing treatments and with serial models. Multi-task learning (MTL) achieved the lowest root-mean-squared error (RMSE) (5.60°) for traumatic brain injury (TBI) patients on knee trajectories and the lowest RMSE (3.77°) for cerebral palsy (CP) patients on ankle trajectories, with only a difference of 5.60° between actual and predicted. Overall, the best RMSE ranged from 5.24° to 6.24° for the MTL models. To the best of our knowledge, this is the first time that MTL has been used for post-treatment gait trajectory prediction. The MTL models outperformed the serial models, particularly when introducing treatment metadata. The gating mechanism is efficient in modeling treatment interaction and improving trajectory prediction.

**Keywords:** multi-task learning; clinical gait analysis; gait rehabilitation; deep learning; long short-term memory; botulinum toxin



**Citation:** Khan, A.; Hazart, A.; Galarraga, O.; Garcia-Salicetti, S.; Vigneron, V. Treatment Outcome Prediction Using Multi-Task Learning: Application to Botulinum Toxin in Gait Rehabilitation. *Sensors* **2022**, *22*, 8452. <https://doi.org/10.3390/s22218452>

Academic Editors: Toshihisa Tanaka, Jordi Solé-Casals, Zhe Sun, Pere Marti-Puig and César F. Caiafa

Received: 5 October 2022

Accepted: 28 October 2022

Published: 3 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

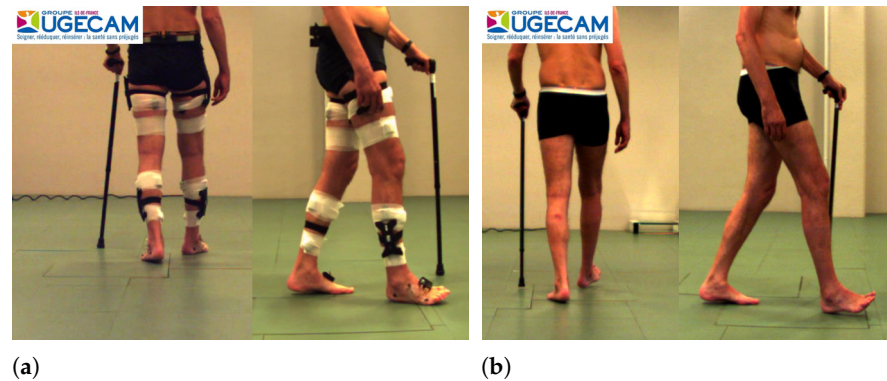


**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

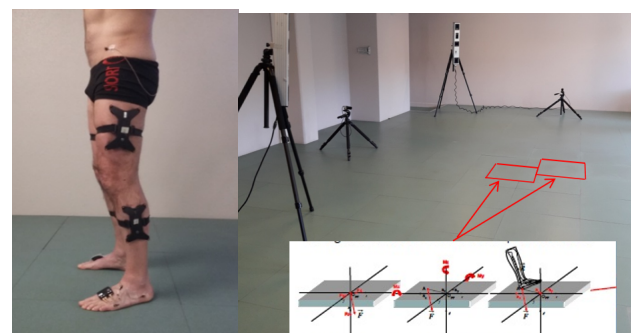
Fatigue, weakness, sensory loss, ataxia, and spasticity are among the usual causes of motor impairments due to neurological diseases such as multiple sclerosis (MS) [1], TBI, spinal cord injury (SCI), and CP, among others. For this reason, physicians often advise people with such impairments to be treated in rehabilitation as a supplement to their background pharmacologic treatment. Spasticity is a motor disorder characterized by a velocity-dependent increase in tonic stretch reflexes (muscle tone) with exaggerated tendon jerks resulting from the hyper-excitability of the stretch reflexes as one component of upper motor neuron syndrome [2]. Intramuscular injection of BTX-A is a standard treatment for spasticity. It has been shown that BTX-A produces improvements in lower and upper limb function [3], thereby improving movement such as walking [4] (see Figure 1) or fine motor skills. The minimum and maximum dose of BTX-A may vary depending on the muscle that is considered [5]. Furthermore, the total dose of BTX-A (sum of doses for all treated muscles) should not exceed the recommended amount according to the patient and the considered muscles (i.e., upper limbs and lower limbs).

BTX-A is a relatively expensive pharmaceutical product, and its consumption has increased in recent years [6,7]. Although its effect on muscle function is considered reversible, BTX-A treatment presents risks (i.e., undesirable effects), and injection sessions should be spaced by at least 3 months apart. For all these reasons, optimizing BTX-A treatment by choosing the right muscles to be treated and the dose distribution is a complex task of great relevance and requires careful study of the patient's condition.



**Figure 1.** Example of the outcome of BTX-A treatment on gait (a) before treatment (b) after BTX-A treatment.

In practice, decision-making is based on a patient's medical history, physical examination, and clinical movement analysis (CMA). CMA consists of studying movement troubles and identifying their plausible causes, based on bio-mechanical interpretation of instrumental measures [8] (Figure 2). If certain quality criteria are fulfilled, CMA data are sufficiently reliable for clinical interpretation [9]. CMA techniques can be used to analyze lower limb movement (e.g., walking, climbing stairs, running, etc.) or fine motor skills. Numerous scientific studies have shown that CMA, especially clinical gait analysis (CGA), provides considerable aid in the assessment and treatment decision for various neurological diseases such as CP [10], post-stroke hemiparesis [4], and MS [11], among others. Camardella et al. [12] support the idea of using machine learning to also predict clinical scores after robot-assisted rehabilitation as a decision-support tool for clinicians.



**Figure 2.** Clinical gait analysis. Different types of sensors are used to conduct kinematic and kinetic analyses of locomotion in gait labs. These may include optoelectronic motion capture, force platforms, electromyography, and IMU sensors, among others.

Artificial intelligence (AI) and machine learning (ML) techniques have become almost ubiquitous in our daily lives by guiding our decisions and providing recommendations. Therefore, it is not surprising that ML approaches are becoming increasingly popular in precision medicine and fulfill an increasing demand for new healthcare solutions, in particular a better understanding of pathological processes. Among AI and ML methods, the deep neural network (DNN) [13] has already shown spectacular results in aiding clinical decision-making [14]. The DNN requires a significant amount of data to be properly trained. However, available experimental databases are often limited in size, which makes them

impractical to construct DNNs for prediction models. Medical data are often heterogeneous, complex, incomplete, uncertain, multi-modal, and multilevel, drastically decreasing the amount of exploitable data and questioning the development of prediction models [15]. Machine learning (ML) models must be able to manage data of a different nature describing the patient (images, time series, discrete clinical data, etc.) and link them to data from treatment in nominal, categorical (type of treatment) [16], and/or discrete (doses) forms. This requires that the model be taught a regression task between the data after and before BTX-A treatment. Since these treatments are often a combination of several factors (e.g., several drug injections), it is necessary to be able to model their interactions. Therefore, we propose a strategy to create multi-task DNN. Indeed, MTL can cope with sparse data problems and build a more robust model by sharing knowledge among different tasks [17]. MTL has been widely applied in ML and in the biomedical field to address the diversity of the data [17]. In the CGA literature, several works exploited deep learning (DL) for predicting gait trajectories, most of them on healthy gait. Su et al. [18] predicted gait trajectories and the five gait phases (loading response, mid-stance, terminal stance, pre-swing, and swing) with a long short-term memory (LSTM) to help in the design of exoskeletons. They employed either 10 or 30 time steps as the input for predicting the next five or ten steps. Twelve people were enrolled in their experiment, and the data were collected using attached inertial measurement units (IMUs) on their body parts. Zhu et al. [19] used an attention-based convolutional neural network (CNN)-LSTM to forecast the joint trajectories of the knee and ankle, based on lower and upper limb data, for the next 60 milliseconds. Zaroug et al. [20] constructed an LSTM auto-encoder to forecast linear acceleration and angular velocity trajectories. To predict five or ten steps into the future, they considered several lengths of input time steps (five to 40 steps) of kinematic data of six male participants. Hernandez et al. [21] proposed a hybrid network combining an LSTM with a CNN (DeepConvLSTM) to estimate kinematic trajectories, reaching an average mean absolute error (MAE) of 3.6°. Jia et al. [22] constructed a DNN for trajectory prediction using LSTM units and a feature fusion layer. This layer uses EMG and joint angle data. Liu et al. [23] built a deep spatio-temporal model composed of LSTM units to forecast two-time steps into the future, using the kinematic data of 35 subjects. More recently, Kolaghassi et al. [24] worked on the pathological gait trajectories of children with neurological disorders. They used two deep learning models, an LSTM and a CNN, to forecast hip, knee, and ankle trajectories. Note that all these studies tackled the prediction of the same gait cycle. The issue we face in this study is much more complex since it is centered on the impact of several treatments (BTX-A) on gait trajectories.

Our contribution consists of proposing a new solution to predict the BTX-A post-treatment gait trajectory of the patient, and possibly the interaction between different treatments. This solution is an MTL architecture, which alleviates the drawbacks previously mentioned: dataset size (number of patients), sample size (number of features), and feature diversity. To the best of our knowledge, this is the first time that MTL has been used for post-treatment gait trajectory prediction. This architecture comprises a collection of LSTM-shaped sub-models, arranged in parallel or series. Each sub-model is used for one treatment, and each treatment corresponds to an injected muscle. These muscles are attached to the left and right knees and ankles. This MTL model learns to map pre-treatment gait sequences to post-treatment sequences. A gating mechanism is proposed with different architectures to control the treatments' influence on the final prediction.

Section 2 presents the data collection and their characteristics. Section 2.3 describes, more specifically, the different deep architectures used. The most prominent results are presented in Section 3. The paper ends with a conclusion and a short discussion.

## 2. Materials and Methods

### 2.1. Dataset Acquisition

Joint kinematics are typically acquired by optoelectronic systems [8] or inertial measurement unit (IMU) systems [25]. In this work, data were collected at the Movement

Analysis Laboratory of Rehabilitation Center of UGECAM Coubert (France), using an opto-electronic Codamotion system consisting of four CX1 cameras at 100 Hz. All the patients in this laboratory were adults with different types of gait issues. This database consists of patients with central neural system disorders, e.g., CP, SCI, or TBI, and all patients had undergone spasticity treatment with BTX-A injections.

In this retrospective study, all the considered data were obtained from patients who participated in clinical activities. The database is composed of  $N_{pat} = 38$  patients that underwent CGA before and after spasticity treatment with botulinum toxin. The usage of these data was approved by the institution's research ethics committee. The patients were informed about the research and did not oppose the utilization of their data.  $N_{uni} = 15$  patients (39.47%) were unilaterally affected (the right lower limb was affected in 6 of them and the left lower limb for the other 9), and  $N_{bil} = 23$  patients (60.53%) were bilaterally affected, which means that, in total,  $N_{limbs} = 61$  lower limbs had been modified. The data contain the CGA of patients before treatment, medical treatment details, and the CGA after treatment. The average age of the patients at the time of pre-treatment CGA, the time of injection, and the time of post-treatment CGA was 46.67 years old (yo), 46.76 yo, and 46.93 yo respectively. The range of age in the dataset is from 21 to 75 yo. There was approximately a 3-month gap between pre-treatment CGA and post-treatment CGA. The details of the patients are listed in Table 1. In this work, we considered injections into four muscles: soleus, gastrocnemius (medialis and lateralis), semitendinosus, and rectus femoris. We also defined a fifth category called “other muscles”, which groups all the other muscles that were treated (see Table 2). There were 28 different combinations of BTX-A injections of these four muscles. A treatment binary code vector:

$$s^j = (s_1^j, \dots, s_c^j)^T, s_i^j \in \{0, 1\}, i = 1 \dots c (c = 5 \text{ as shown in Table 1})$$

was attributed to each lower limb  $i$ , with  $s_i^j = 1$  if muscle  $i$  was injected in limb  $j$ , 0 otherwise, and  $d^j = (d_1^j, \dots, d_5^j)^T, d_i^j \in \{0, 1\}$  is a binary vector for the disease of patient's limb  $j$ . There are five diseases: CP, MS, TBI, SCI, and stroke.  $T$  is the transpose operator.

**Table 1.** Patient database description.

<b>Total Patients</b>	38
<b>Age (Mean <math>\pm</math> SD)</b>	46.76 $\pm$ 13.43
<b>Males/Females</b>	24/14
<b>Unilaterally/Bilaterally affected</b>	15/23
<b>Cerebral Palsy</b>	3
<b>Stroke</b>	9
<b>Multiple Sclerosis</b>	12
<b>Traumatic Brain Injury</b>	3
<b>Spinal Cord Injury</b>	11

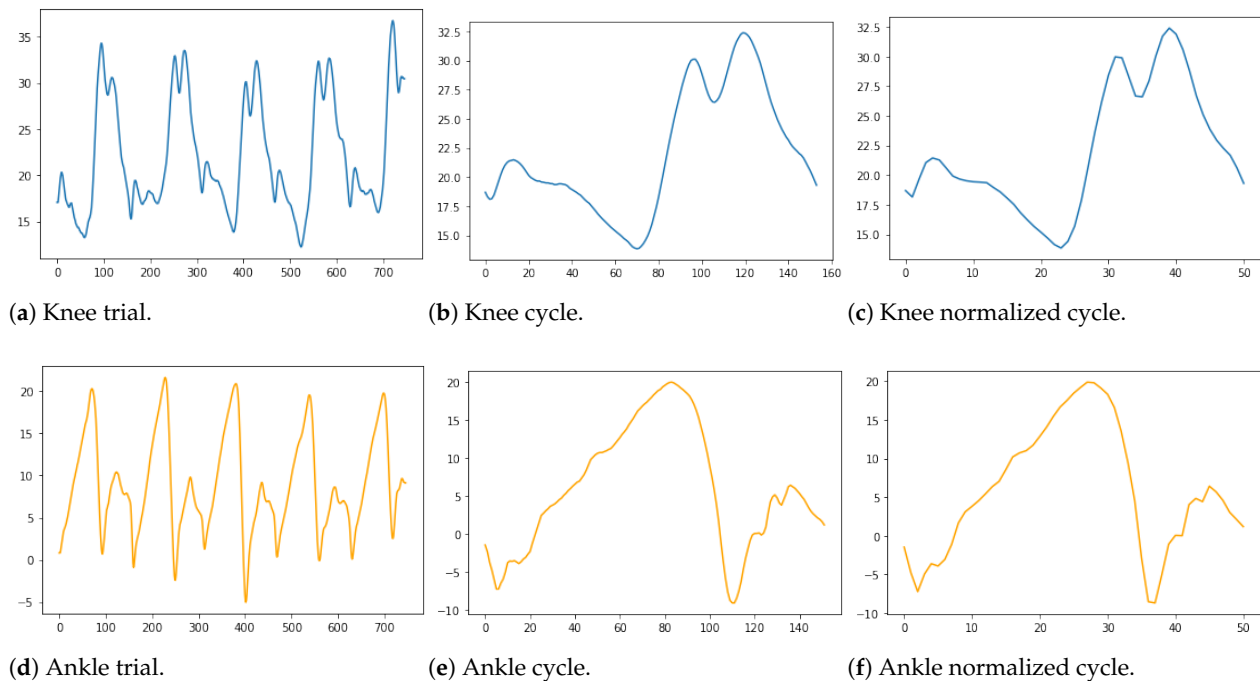
**Table 2.** Considered injected muscles and their frequencies in the database.

Muscle Number	Muscle/Category	Injections in Patient	
		Number	Proportion
1	Soleus	49	29.7%
2	Gastrocnemius (Medialis and/or Lateralis)	37	28.5 %
3	Rectus Femoris	18	10.8%
4	Semitendinosus	12	7.2%
5	Other Muscle	40	24.2 %

## 2.2. Data Preparation

Kinematic data were automatically segmented into gait cycles from initial contact (IC) to terminal swing (TS), utilizing the high-pass algorithm (HPA) [26]. Then, gait cycles were re-sampled and normalized to 51 points (2% of the gait cycle) as proposed by CGA [27],

so that the DL models were trained with fixed-length sequences, as illustrated in Figure 3. Mean gait cycles were computed for each limb. Combining both the pre- and post-treatment cycles of each patient led to a total of  $n = 1622$  gait strides. For any patient's limb  $j$ , the input vector is an angular time series  $x^j = (x_1^j, \dots, x_m^j)^T \in [-180, +180]^m$ , and the target vector is  $t^j = (t_1^j, \dots, t_m^j)^T$ , with  $m = 51 \times 2 = 102$ . Let  $\mathcal{D} = \{x^j, t^j, d^j, s^j\}_{j=1}^n$  be the input–target training set.



**Figure 3.** Process of converting one trial to one normalized cycle.

The patient's data consist of multiple gait cycles at the time of pre-treatment CGA and post-treatment CGA. Different trials were recorded for each patient. In one trial, there were multiple cycles of pre-treatment CGA. We extracted all the cycles of all patients and stored them. We separated a person's right and left cycles since we considered them as different samples in the data. We performed the same procedure for post-treatment CGA data. Each pre-treatment cycle was associated with a target post-treatment cycle.

Note that the number of cycles per patient varies from one patient to another.

There is a total of 5 joints (pelvis, hip, knee, ankle, and foot) and three signals per joint in our dataset, leading to 15 signals. These three signals represent the projections of the trajectory of each joint, respectively, on the sagittal, frontal, and transverse planes. In this study, we only considered

the knee and ankle on the sagittal plane, because most treatments were performed around these joints. Figure 3a,d show the sagittal plane signal (flexion/extension) of the ankle and knee for a patient's complete trial containing multiple cycles.

Figure 3b,e show a cycle extracted from the full knee and ankle trials, respectively. Figure 3c,f show the normalized cycle in 51 points. In the end, our dataset contains 1622 samples and 210 features; the first feature represents the ID (patient name); the second to 103rd are the features of the pre-treatment CGA; then,  $c = 5$  features describe the presence or absence of botulinum toxin injection according to the muscle categories; finally, the last 102 features concern the post-treatment CGA of a patient.

An input matrix  $X$  and a target output matrix  $Y$  were constructed using the parameters of  $n$  training samples,  $f$  features (the sagittal plane of the ankle and knee),  $l_{in}$  input size, and  $l_{out}$  output size. Pre- and post-treatment data were centered and reduced by the standard deviation. The goal was to construct a model with  $g()$  that maps  $\hat{Y} = g(X)$ , where  $\hat{Y}$  is a value that is very close to the actual value of  $Y$ .



### 2.3. Description of the Models

#### 2.3.1. Long Short-Term Memory

When training, early recurrent networks had difficulty remembering information for longer periods, such as several thousand time steps. Hochreiter et al. [28] introduced a particular memory cell capable of retaining information for long periods of time. The LSTM can read and write to its memory. More importantly, this memory never goes through an activation function. This effectively combats the [29] trailing gradient problem and makes the formation of this pattern very stable.

The original LSTM works with a series of input signals  $x_t$ . It has a so-called hidden state  $h_t$  and cell state  $c_t$  of the same size as  $x_t$ . The cell state  $c_t$  is the model's memory. The hidden state  $h_t$  is the model's prediction of  $x_t$ .

The LSTM equations are defined by the following set of matrix equations:

$$A = h_t \parallel x_t \quad (1)$$

$$f_t = \sigma(W_f A + b_f) \quad (2)$$

$$i_t = \sigma(W_i A + b_i) \quad (3)$$

$$o_t = \sigma(W_o A + b_o) \quad (4)$$

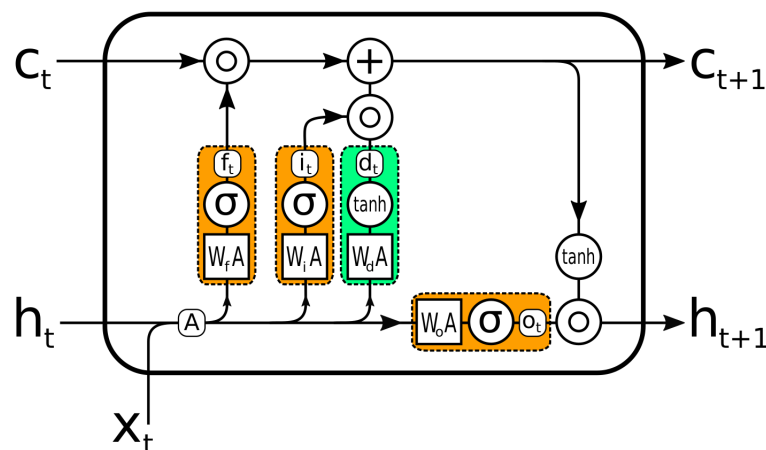
$$d_t = \tanh(W_d A + b_d) \quad (5)$$

$$c_{t+1} = f_t \circ c_t + i_t \circ d_t \quad (6)$$

$$h_{t+1} = o_t \circ \tanh(c_{t+1}) \quad (7)$$

where  $\parallel^1$  is the concatenation operator,  $\circ$  is the Hadamard product,  $\sigma$  is the logistic function,  $W$  are weight matrices, and  $b$  are biases. The basic idea is that the model takes the input  $x_t$  and the previous prediction of the current input  $h_t$ , updates its internal memory  $c_t$  to  $c_{t+1}$ , and then makes a new prediction  $h_{t+1}$  based on  $c_{t+1}$ ,  $h_t$ , and  $x_t$ .

The original LSTM could have multiple parallel memory cells  $c_t$ , but in practice, mostly, only one memory cell is used; the description of the LSTM was limited to one  $c_t$ . All the gate functions (Equations (2)–(4)) are fully connected layers,  $y = f(Wx + b)$  with a sigmoid activation function. The data flow in the LSTM is illustrated in Figure 4).



**Figure 4.** LSTM unit. The gates, which decide which part of the information to pass on, are orange. Green is the update to the memory cell.

Furthermore, the role of  $h_t$  is not strictly fixed to be a prediction of  $x_t$ . It can be any series of predictions that is connected to the input series  $x_t$ . For example, if  $x_t$  was the number of people who entered (or left) a building in the last hour, then  $h_t$  could be the current number of people inside the building (with appropriate scaling, so it fits the output range  $[-1, 1]$ ).

For this study, we used several variants of LSTM. The five categories of treatments are reported in Table 2: BTX-A injection of the first four muscles and the fifth category of injections in all other muscles. Each treatment is represented by an LSTM layer. Hidden states represent, according to the DL architecture used, the presence or absence of treatments by BTX-A in the five muscles.

While the LSTM is well suited to prediction tasks on time series, sometimes, knowledge about future events is necessary for the correct prediction. Therefore, the term “future” is relative to  $t$  and means the following data points. Of course, the next/future data points must already be known to be included in the prediction. Reference [30] identified two strategies to integrate the knowledge of future events into an LSTM model: bi-directional recurrent neural network (RNN) [31] and delayed input, the second approach consisting of delaying the signal by a delay  $\tau$ :

**Model 1** LSTM was used with pre-treatment CGA data and post-treatment CGA data. Treatments were not considered in this experiment. The model was implemented using five layers of LSTM units, with 51 units per layer, one unit for each point of a cycle. Note that each unit receives a pair of inputs for the knee and ankle, respectively. The final layer is fed into a dense layer of 102 neurons ( $2 \times 51$  values), which is then reshaped to obtain the desired output, as shown in Figure 5a. In this model, we initialized the values of the cell state and hidden state to 0.

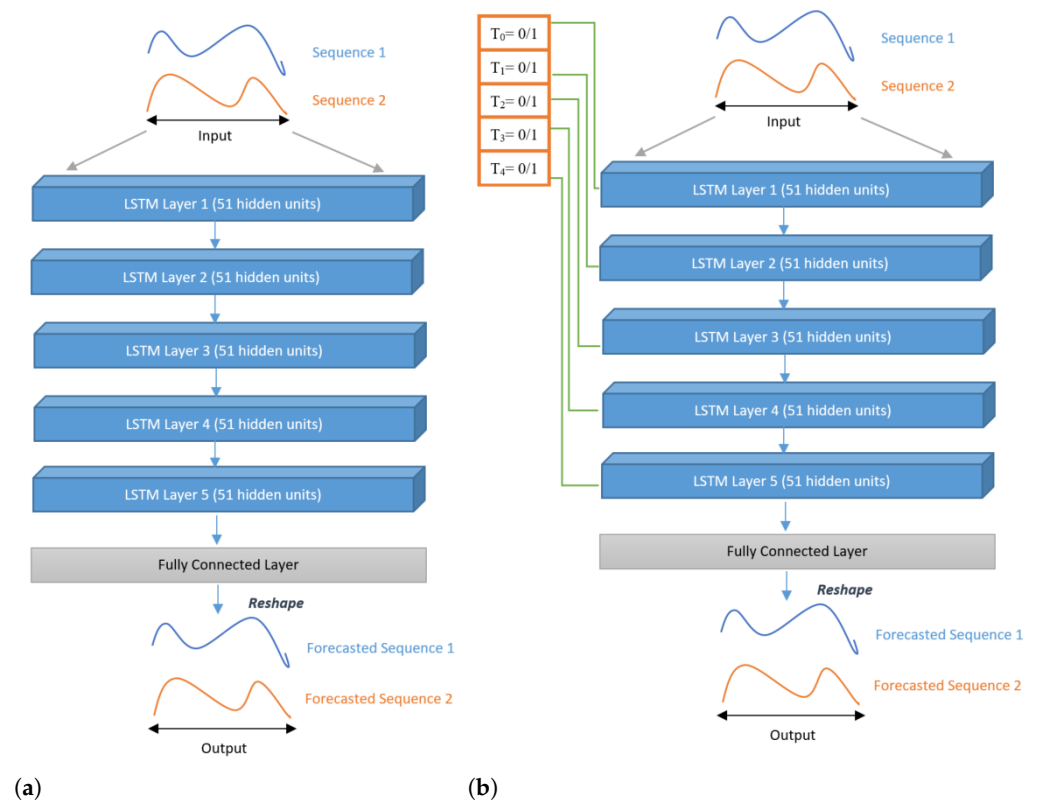
**Model 2** A total of five treatments, together with pre- and post-treatment CGA data, were included in this model, displayed in Figure 5b. In this architecture, the values were initialized according to the medical treatment. If one patient had muscle 1 and muscle 3 injected (Table 2), then each layer’s components of the hidden states vector in LSTM layer 1 and LSTM layer 3 are initialized to 1, and the other layers’ hidden states vector is initialized as 0. In this model, we also initialized the cell state as 0.

### 2.3.2. Bi-Directional LSTM

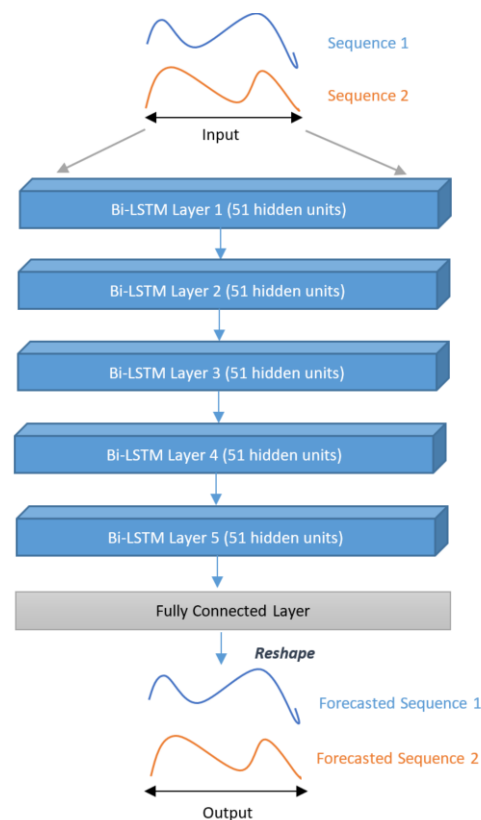
The entire signal must be known for this approach. Two LSTM models were trained in parallel, one on the input series (forward) and the other on the reverse input series (backward), starting with the last input and then the next-to-last, and so on. Thus, for each  $t$ , there were two hidden states:  $h_{1,t}$  and  $h_{2,t}$  among the two available models.  $h_{1,t}$  only contains information about the past, and  $h_{2,t}$  only contains information about the future. Together, they have information about the whole signal, and the final prediction  $f(h_{1,t}, h_{2,t})$  was made using the two hidden states. This method has the disadvantage that two models must be trained; therefore, the number of parameters and the training time are doubled.

We studied the bi-directional LSTM (Bi-LSTM) architecture and considered two experiments, namely with and without MTD, as previously presented for LSTM:

**Model 3** A Bi-LSTM, as depicted in Figure 6. As shown in Figure 6, the model has mainly the same structure as the previous Model 1 (same number of layers and units in each layer). The final layer’s hidden state is fed into a fully connected layer. As in Model 1, we initialized the values of the cell state and hidden state of each layer to 0.



**Figure 5.** LSTM architectures (Model 1 and Model 2) proposed in this work: (a) without MTD; (b) with MTD.

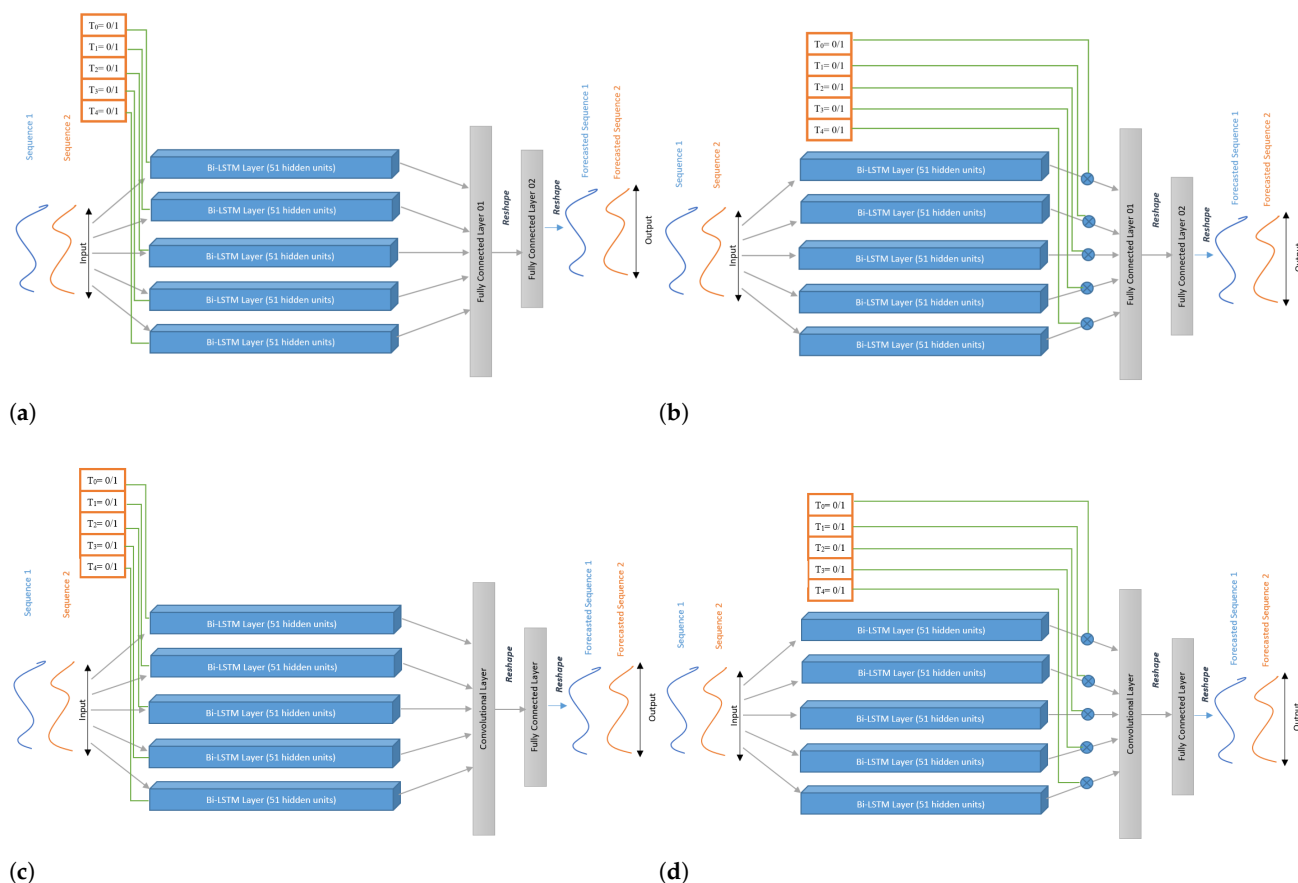


**Figure 6.** First Bi-LSTM architecture (Model 3) proposed in this work without considering MTD.



**Model 4** This model takes into account MTD in a multi-task architecture of Bi-LSTM models. Indeed, five Bi-LSTM models work in parallel while incorporating MTD as in Model 2. Each Bi-LSTM has 51 units, each receiving as input a pair for the knee and ankle, respectively. Input  $X$  is fed to the five Bi-LSTM sub-models, and the cell state of all such sub-models was initialized to 0. Furthermore, the hidden states of all sub-models were initialized according to the presence or absence of MTD (as discussed in Model 2). This architecture has two fully connected layers: the first layer concatenates the outputs of all the sub-models; the second maps the output of the first layer to 102 neurons as per the desired output, as shown in Figure 7a.

**Model 5** This model is also a multi-task architecture of Bi-LSTM sub-models, as in Model 4. However, in this case, MTD is considered differently, with a gating mechanism. Instead of passing MTD as a hidden state of each Bi-LSTM sub-model, we incorporated them at the end of such sub-models by multiplying each sub-model's output by its corresponding binary value of MTD. In other words, if there is any treatment, it will be used further in the model; otherwise, it will be discarded (multiplying with 0), as illustrated in Figure 7b. By performing this experiment, we wanted to assess the impact of this gating mechanism compared to MTD internal processing by each sub-model, as done in Model 4.



**Figure 7.** Multi-task learning architectures with Bi-LSTM sub-models; (a) Model 4: processing MTD internally in each sub-model; (b) Model 5: incorporating MTD through a gating mechanism; (c) Model 6: processing MTD internally in each sub-model using the Conv layer; (d) Model 7: incorporating MTD through a gating mechanism using the Conv layer.

**Models 6 and 7** In both models, we replaced the first fully connected layer (FC Layer 01) (see Figure 7a,b) with a convolutional layer (see Figure 7c,d), with kernel size (5,2) and stride (3,2). As there are five Bi-LSTM sub-models and each has an output of size  $2 \times 102$ , we concatenated such outputs and reshaped them into a matrix of size  $(10 \times 102)$ , then given as the input to the convolutional layer. Finally, the convolutional layer's output is fed into a fully connected layer of size 102. Model 6 incorporates MTD as in Model 4, through the internal states of the sub-models. Model 7 uses the gating mechanism as in Model 5.

### 2.3.3. Experimental Setup

CGA data consist of 1622 combination pre-treatment and post-treatment gait cycles of 38 patients. Leave-one-out cross validation was used to assess the models' performance. For each iteration, we used 37 patients for training the model and one for testing. In the end, we took the RMSE of all tested patients for each model. The mini-batches were used throughout the training process, and the size of each batch was 16. We chose the RMSE as the loss function for optimizing the deep learning models and used the ADAM optimizer for learning. We tried different learning rates and selected the best-possible values. We report in Table 3 all details concerning the models' hyper-parameters.

**Table 3.** Hyper-parameter selection for LSTM, Bi-LSTM, and other variants of the architecture. The MTD column is used for medical treatment data (included/not included).

Model No. and Fig. Reference	Model Type	MTD	LSTM Layers (Units)	Conv. Layer	FC Layers (Units)	Learning Rate
Model 1 (Figure 5a)	LSTM (serial)	No	5 layers (51)	None	1(102)	0.005
Model 2 (Figure 5b)	LSTM (serial)	Yes	5 layers (51)	None	1(102)	0.005
Model 3 (Figure 6)	Bi-LSTM (serial)	No	5 layers (51)	None	1(102)	0.005
Model 4 (Figure 7a)	MTL, 5 Bi-LSTMs	Yes	1 layer per sub-model (51)	None	2 (1020 & 102)	0.005
Model 5 (Figure 7b)	MTL, 5 gated Bi-LSTMs	Yes	1 layer per sub-model (51)	None	2 (1020 & 102)	0.005
Model 6 (Figure 7c)	MTL, 5 Bi-LSTMs + Conv Layer	Yes	1 layer per sub-model (51)	Kernel (5,2), stride (3,2)	1(102)	0.005
Model 7 (Figure 7d)	MTL, 5 gated Bi-LSTM + Conv Layer	Yes	1 layer per sub-model (51)	Kernel (5,2), stride (3,2)	1(102)	0.001

We calculated the RMSE to see how closely the predicted trajectories of the knee and ankle,  $\hat{Y}$ , matched the actual trajectories of the knee and ankle,  $Y$ . The following equation for the RMSE can be derived if we assume that  $n$  represents the number of testing samples,  $f$  represents the number of features, and  $l_{out}$  represents the output size.

$$RMSE = \sqrt{\frac{1}{nfl_{out}} \sum_{i=1}^n \sum_{j=1}^f \sum_{k=1}^{l_{out}} (y_{i,j,k} - \hat{y}_{i,j,k})^2} \quad (8)$$

We also calculated the standard error (SE) to measure the variation of the RMSE with respect to each disease and the  $R^2$  score to check how well the data fit the regression model. SE is calculated using

$$SE = \frac{s}{\sqrt{n}}. \quad (9)$$

where  $s$  is the standard deviation of prediction with respect to a particular disease and  $n$  is the total number of patients having a particular disease. The coefficient of determination ( $R^2$ ) as follows

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2}, \quad (10)$$

can be interpreted as the proportion of the variance in the dependent variable that is predictable from the independent variables (worst value  $-\infty$ , best value  $= +1$ ), opposite the MSE, which magnifies the error if the model outputs a very bad prediction (worst value  $+\infty$ , best value  $= 0$ ).

We compared and evaluated the performance of the models with the use of these measures.

### 3. Results

We evaluated Models 1 to 7 on our dataset with the above-mentioned metrics and display the results in Table 4. The lowest average RMSE values and the highest  $R^2$  scores are displayed in bold; they correspond to the best prediction model according to the diseases reported in Table 4.

From Table 4, we noticed that Model 4 outperformed other models in the prediction of post-treatment gait trajectories for patients having MS and TBI. Furthermore, Model 6 performed better for SCI patients than all other architectures. Model 7 outperformed other models of patients having stroke and CP. We noticed that, in all cases, the MTL architectures achieved better performance globally, on both knee and ankle signals.

The following two tables (Tables 5 and 6) report the performance scores of the prediction of gait trajectories for knee and ankle, respectively. In Table 5, the best prediction for the knee angle was obtained for TBI patients by Model 5 with an average RMSE of  $5.60^\circ$  and  $R^2 = 0.72$ . Furthermore, for all diseases, the MTL architectures outperformed the others. Model 6 gave the best prediction for MS and SCI in terms of RMSE and Model 4 in terms of the  $R^2$  score for the same pathologies. On the other hand, for stroke patients, Model 7 had the best average RMSE, and Model 6 had the best coefficient of determination. In Table 6, we notice that the best RMSE for the ankle was  $3.77^\circ$ , which is lower than that obtained for the knee ( $5.60^\circ$ ). However, even though the RMSE was usually lower (thus better) for the ankle, the  $R^2$  scores were usually lower (thus worse) as well. In particular, for stroke, all the  $R^2$  of the ankle angle were negative.

**Table 4.** Performance of different models in prediction of post-treatment gait trajectories with respect to different diseases.

Model No. and Fig. Reference	Model Type	Spinal Cord Injury (SCI)	Multiple Sclerosis (MS)	Stroke	Cerebral Palsy (CP)	Traumatic Brain Injury (TBI)
		11	12	No. of Patients 9	3	3
		474	530	No. of Cycles 322	148	148
		RMSE Mean $\pm$ Standard Error $R^2$ Score				
Model 1 (Figure 5a)	LSTM (serial)	6.82 $\pm$ 0.09 0.65	6.89 $\pm$ 0.10 0.69	8.11 $\pm$ 0.19 0.58	7.66 $\pm$ 0.14 0.71	5.87 $\pm$ 0.11 0.67
Model 2 (Figure 5b)	LSTM (serial)	6.71 $\pm$ 0.08 0.61	6.77 $\pm$ 0.08 0.65	8.03 $\pm$ 0.19 0.61	7.23 $\pm$ 0.11 0.70	7.63 $\pm$ 0.32 0.67
Model 3 (Figure 6)	Bi-LSTM (serial)	6.9 $\pm$ 0.10 0.72	6.38 $\pm$ 0.10 0.78	7.06 $\pm$ 0.18 0.71	7.2 $\pm$ 0.10 0.78	7.78 $\pm$ 0.22 0.59
Model 4 (Figure 7a)	MTL, 5 Bi-LSTMs	6.26 $\pm$ 0.08 <b>0.78</b>	<b>5.8 <math>\pm</math> 0.11</b> <b>0.79</b>	6.99 $\pm$ 0.019 0.72	6.57 $\pm$ 0.12 0.76	<b>5.24 <math>\pm</math> 0.13</b> <b>0.73</b>
Model 5 (Figure 7b)	MTL, 5 gated Bi-LSTMs	6.67 $\pm$ 0.08 0.75	6.11 $\pm$ 0.09 0.71	7.73 $\pm$ 0.29 <b>0.74</b>	6.22 $\pm$ 0.14 0.78	6.07 $\pm$ 0.21 0.63
Model 6 (Figure 7c)	MTL, 5 Bi-LSTMs + Conv Layer	<b>5.75 <math>\pm</math> 0.08</b> 0.73	6.08 $\pm$ 0.12 0.74	7.16 $\pm$ 0.24 0.71	6.2 $\pm$ 0.12 0.79	6.58 $\pm$ 0.14 0.64
Model 7 (Figure 7d)	MTL, 5 gated Bi-LSTMs + Conv Layer	6.31 $\pm$ 0.12 0.66	7.59 $\pm$ 0.13 0.70	<b>6.24 <math>\pm</math> 0.14</b> 0.66	<b>6.00 <math>\pm</math> 0.14</b> <b>0.80</b>	7.02 $\pm$ 0.07 0.46

Bold entries denote the lowest average RMSE and maximum  $R^2$  over all limbs having a given disease.

From a different perspective, the following graphs (in Figures 8 and 9) illustrate the trajectories (pre-treatment, real post-treatment, predicted post-treatment of the patient, and standard course of an adult) of two patients. The Y-axis represents the ankle dorsiflexion or knee flexion, and the X-axis represents the gait cycle of a patient.

Figure 8 compares the prediction of different models on the knee and ankle joints in a patient diagnosed with CP. These figures differentiate the prediction between the MTL models and others. Figure 8a–c illustrate the predictions on the knee angles made by Model 1, Model 2, and Model 3, which are not MTL models. Figure 8d shows the corresponding prediction of Model 7, which is an MTL model. The predictions of post-treatment gait from Model 7 were better than others. In other words, it was closer than that patient's expected post-treatment gait trajectory (average of all his/her target gait cycles in the training set). On the other hand, Figure 8e–h compare the prediction of the ankle joint of the same patient. Figure 8g,h illustrate the prediction of Model 1 and Model 3, respectively. Figure 8g,h show the predictions of Model 4 and Model 7, respectively, which are MTL models.

We noticed that the predicted post-treatment trajectory in Figure 8g was better than the first two models, which were serial, and we see in Figure 8h the significant improvement of the prediction at the end of the gait cycle, between 80% and 100%, compared to Figure 8g. On this patient, the MTL models also performed better on the ankle joint.

**Table 5.** Performance of different models in the prediction of post-treatment knee gait with respect to different diseases.

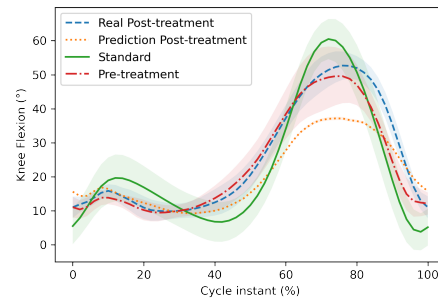
Model No. and Fig. Reference	Model Type	SCI	MS	Stroke	CP	TBI
		11	12	No. of Patients 9	3	3
		474	530	No. of Cycles 322	148	148
		RMSE Mean $\pm$ Standard Error R <sup>2</sup> Score				
Model 1	LSTM (serial)	7.73 $\pm$ 0.09 0.67	8.05 $\pm$ 0.11 0.65	8.62 $\pm$ 0.21 −0.08	10.16 $\pm$ 0.13 0.70	6.66 $\pm$ 0.09 0.59
Model 2	LSTM (serial)	7.58 $\pm$ 0.08 0.72	8.26 $\pm$ 0.09 0.62	7.85 $\pm$ 0.17 0.04	8.56 $\pm$ 0.12 0.70	8.05 $\pm$ 0.27 0.55
Model 3	Bi-LSTM (serial)	8.11 $\pm$ 0.13 0.67	7.41 $\pm$ 0.12 0.73	7.77 $\pm$ 0.21 0.16	7.42 $\pm$ 0.11 0.78	7.89 $\pm$ 0.28 0.41
Model 4	MTL, 5 Bi-LSTMs	7.51 $\pm$ 0.08 <b>0.76</b>	7.23 $\pm$ 0.14 <b>0.77</b>	7.14 $\pm$ 0.018 0.26	<b>6.75 <math>\pm</math> 0.11</b> <b>0.80</b>	5.81 $\pm$ 0.13 0.61
Model 5	MTL, 5 gated Bi-LSTMs	7.62 $\pm$ 0.10 0.71	7.23 $\pm$ 0.11 0.64	8.02 $\pm$ 0.25 0.45	7.00 $\pm$ 13 0.77	<b>5.60 <math>\pm</math> 0.06</b> <b>0.72</b>
Model 6	MTL, 5 Bi-LSTMs + Conv Layer	<b>6.94 <math>\pm</math> 0.09</b> 0.69	<b>6.78 <math>\pm</math> 0.14</b> 0.70	7.19 $\pm$ 0.25 <b>0.48</b>	8.63 $\pm$ 0.18 0.79	8.24 $\pm$ 0.17 0.62
Model 7	MTL, 5 gated Bi-LSTMs + Conv Layer	8.14 $\pm$ 0.12 0.59	8.52 $\pm$ 0.13 0.66	<b>6.21 <math>\pm</math> 0.14</b> 0.08	7.82 $\pm$ 0.14 0.78	5.94 $\pm$ 0.07 0.34

**Table 6.** Performance of different models in prediction of post-treatment ankle gait with respect to different diseases.

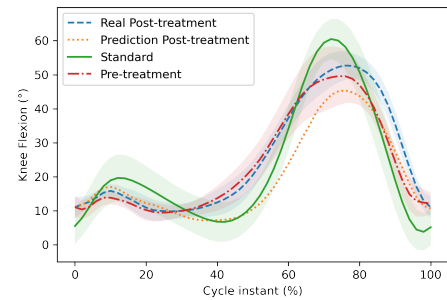
Model No. and Fig. Reference	Model Type	SCI	MS	Stroke	CP	TBI
		11	12	No. of Patients 9	3	3
		474	530	No. of Cycles 322	148	148
		RMSE Mean $\pm$ Standard Error R <sup>2</sup> Score				
Model 1	LSTM (serial)	5.91 $\pm$ 0.08 0.52	5.73 $\pm$ 0.08 0.50	7.61 $\pm$ 0.16 −3.75	5.16 $\pm$ 0.14 0.49	5.09 $\pm$ 0.13 0.08
Model 2	LSTM (serial)	5.85 $\pm$ 0.07 0.19	5.29 $\pm$ 0.07 0.37	8.21 $\pm$ 0.21 −4.48	5.89 $\pm$ 0.10 0.37	7.22 $\pm$ 0.36 0.15
Model 3	Bi-LSTM (serial)	5.69 $\pm$ 0.007 0.40	5.35 $\pm$ 0.07 0.68	6.34 $\pm$ 0.15 −3.94	6.99 $\pm$ 0.09 0.44	5.66 $\pm$ 0.15 0.34
Model 4	MTL, 5 Bi-LSTMs	5.01 $\pm$ 0.08 <b>0.54</b>	<b>4.38 <math>\pm</math> 0.08</b> <b>0.69</b>	6.85 $\pm$ 0.19 −4.13	6.4 $\pm$ 0.12 0.37	<b>4.68 <math>\pm</math> 0.13</b> <b>0.54</b>
Model 5	MTL, 5 gated Bi-LSTMs	4.56 $\pm$ 0.06 0.44	5.39 $\pm$ 0.10 0.47	7.14 $\pm$ 0.22 −4.20	<b>3.77 <math>\pm</math> 0.05</b> <b>0.50</b>	4.93 $\pm$ 0.10 0.37
Model 6	MTL, 5 Bi-LSTMs + Conv Layer	5.72 $\pm$ 0.06 0.46	5.00 $\pm$ 0.06 0.63	7.44 $\pm$ 0.32 −5.14	5.45 $\pm$ 0.14 0.47	6.54 $\pm$ 0.36 −0.03
Model 7	MTL, 5 gated Bi-LSTMs + Conv Layer	<b>4.49 <math>\pm</math> 0.06</b> 0.25	6.66 $\pm$ 0.19 0.46	<b>6.26 <math>\pm</math> 0.26</b> −4.85	4.17 $\pm$ 0.09 0.23	10.63 $\pm$ 0.26 −0.83

Figure 9 compares the trajectories of the knee and ankle joints of another patient diagnosed with MS. Figures 9a,b, represent the predictions of the knee angles made by Model 1 and Model 2, which are not MTL models. Figure 9c,d represent the prediction of the knee angles made by Model 4 and Model 6, respectively, which are MTL models. We can see that MTL models had better predictions than the first two. The predicted

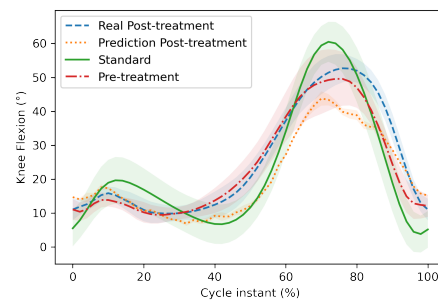
post-treatment trajectories were closer to the real post-treatment trajectories. Last four Figure 9e–h compare the trajectories of the ankle joint. Figure 9e–g represent the prediction of Model 1, Model 2, and Model 3. Although Model 3 is not an MTL model, its predictions were much better than the first two serial models. However, the prediction of the MTL model (Model 5) in Figure 9h was better than all other models for this particular patient. In general, as proven by Tables 4–6, for almost every patient, MTL performed better.



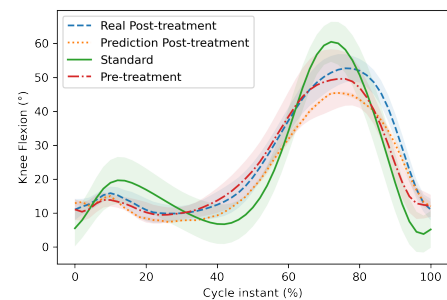
(a) Model 1: LSTM (serial)



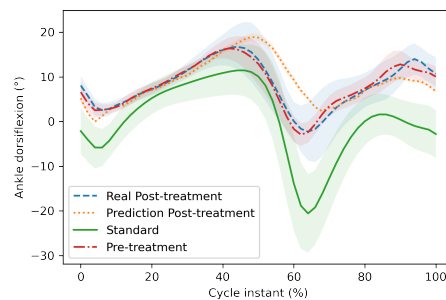
(b) Model 2: LSTM with MTD (serial)



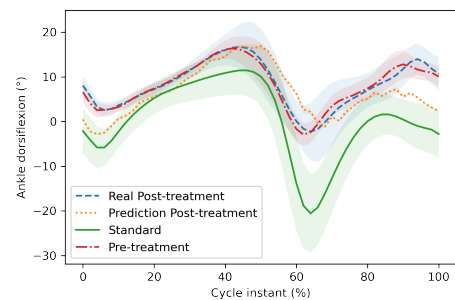
(c) Model 3: Bi-LSTM (serial)



(d) Model 7: MTL, 5 gated Bi-LSTMs + Conv Layer

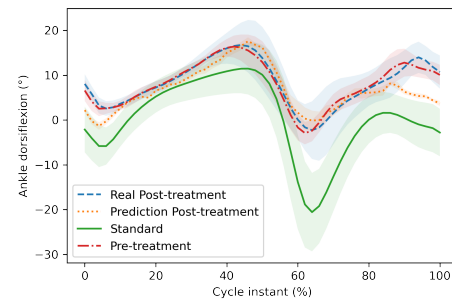


(e) Model 1: LSTM (serial)

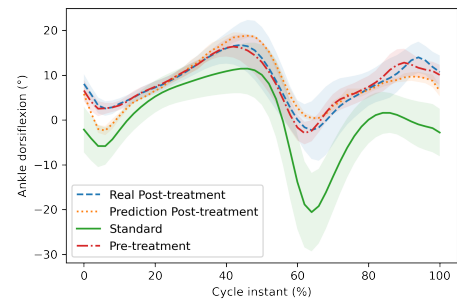


(f) Model 2: Bi-LSTM (serial)

Figure 8. Cont.

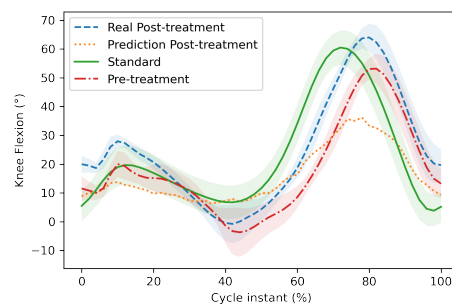


(g) Model 4: MTL, 5 Bi-LSTMs

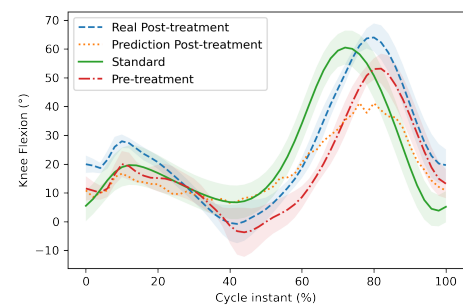


(h) Model 7: MTL, 5 gated Bi-LSTMs + Conv Layer

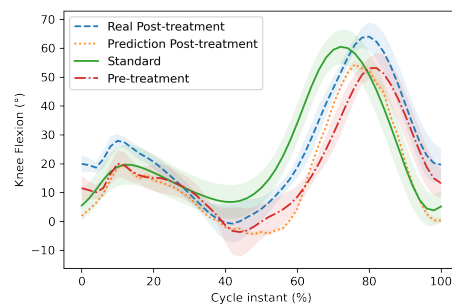
**Figure 8.** Comparison of the post-treatment gait trajectory of the knee and ankle joints in a patient diagnosed with CP. The first three models (a–c) are serial (Model 1, Model 2, and Model 3), and the fourth model (d) (Model 7) is an MTL model, which represents the prediction of the knee joint. The sixth and seventh models (e,f) are serial (Model 1 and Model 3), and the last two models (g,h) (Model 4 and Model 7) are MTL models, which represent the prediction of the ankle joint.



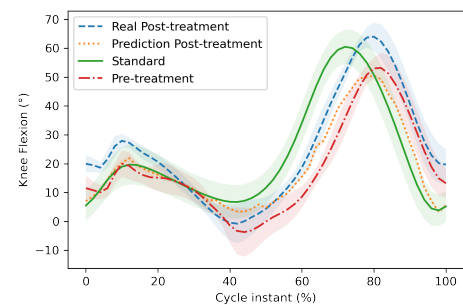
(a) Model 1: LSTM (Serial)



(b) Model 2: LSTM with MTD (Serial)



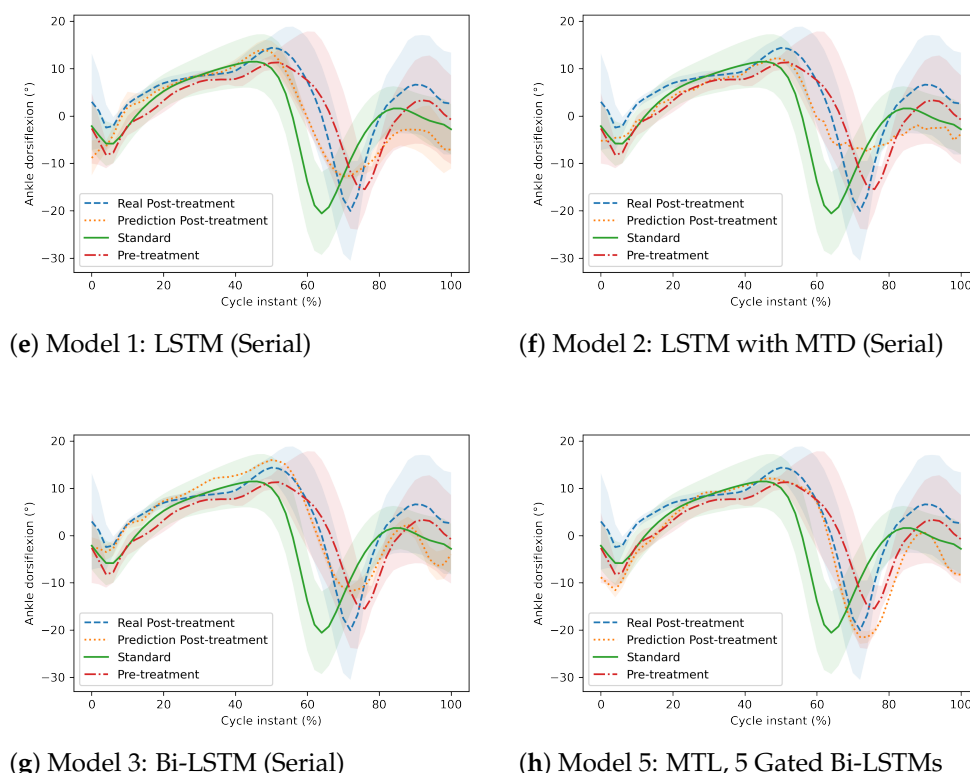
(c) Model 4: MTL, 5 Bi-LSTMs



(d) Model 6: MTL, 5 Bi-LSTMs + Conv Layer

**Figure 9.** Cont.





**Figure 9.** Comparison of the post-treatment gait trajectory of the knee and ankle joint in a patient diagnosed with MS. The first two models (a,b) are serial (Model 1 and Model 2), and the third and fourth models (c,d) are MTL models (Model 4 and Model 6), which represents the prediction of the knee joint. The fifth, sixth and seventh models (e–g) are serial (Model 1, Model 2, and Model 3), and the last model (h) is an MTL model (Model 4), which represents the prediction of the ankle joint.

#### 4. Discussion and Conclusions

In this study, we used MTL to design an LSTM model and its variants to predict the post-treatment trajectory of adults with an abnormal gait. To the best of our knowledge, this specific prediction task, which exhibits greater inter- and intra-subject variability compared to the courses of normal adults, has not been addressed before in the literature using MTL.

To forecast the trajectories of the knee and the ankle in the sagittal plane, we used LSTM-based models. LSTM was chosen because it has been successfully applied to sequential data, and it can capture long-term dependencies through its learning [32]. To better evaluate the performance of MTL on a given problem, we also implemented serial models using LSTM. The RMSE was used to compare the results of both sorts of models. The RMSE of the MTL models was lower for all types of patients (different pathologies). The MTL models also gave the highest  $R^2$ , better explaining the total variance of the target than the serial models. The MTL models performed better than the serial models in our problem of multiple treatment combinations. MTL architectures allow introducing the medical treatment metadata into the model. Instead of performing a simple post–pre regression task, our results imply that introducing the treatment information (i.e., muscles treated by BTX-A) improves performance.

Overall, the best prediction was obtained for TBI using the Bi-LSTM with MTL (Models 4) architecture. The results in Table 4 show that there was only a  $5.24^\circ$  average difference in actual and predicted trajectories and  $R^2 = 0.73$ . The best maximum average RMSE error between actual and predicted trajectories was  $6.24^\circ$  for stroke patients, using the MTL architecture with gated Bi-LSTM and a convolutional layer (Model 7). For the knee and ankle, the best results were  $6.75^\circ$  ( $R^2 = 0.80$ ) and  $3.77^\circ$  ( $R^2 = 0.5$ ), respectively, for CP patients. The RMSE was usually higher for the knee than for the ankle despite having higher coefficients of determination. This suggests that the models were able to explain the

variance of the knee angle better, but the amplitudes in the knee were higher than in the ankle. Moreover, no proposed model was able to adequately explain the variance of the ankle angle for patients with a stroke (only negative  $R^2$  scores).

#### 4.1. Comparison to Previous Works

Since this is the first time that the whole kinematic signals for knee and ankle on the sagittal plane were predicted for botulinum toxin treatment, it is difficult to compare our performance to other works. Nevertheless, we can compare our methods for the predictions of peak knee and ankle on sagittal planes reported by [4] for rectus femoris botulinum toxin injection of patients with stroke (Table 7). In this case, the  $R^2$  score of the proposed method for stroke was better for peak knee flexion, but worse for peak ankle dorsiflexion. Since the compared models were not trained and tested with the same databases, this comparison must be taken with caution.

We also compared our performances to the predictions of the whole postoperative kinematic curves for patients with CP. Even though the proposed methods were not tested on the same databases, these performances were better than the postoperative predictions for CP reported by Galarraga et al. [16], Niiler et al. [33], and Niiler [34], as shown in Table 7.

**Table 7.** Performance comparison of the prediction methods. LinReg and MLinReg correspond to linear regression and multiple LinReg, respectively, in [4,16]. PCA stands for principal component analysis. NN99 and NN01 correspond to feedforward neural networks, respectively, in [33,34].

Model	Knee Flexion	Ankle DorsiFlexion
$R^2$ score		
<b>Model 5 Stroke</b> <sup>†</sup>	0.62	−4.69
<b>Model 6 Stroke</b> <sup>†</sup>	0.49	−4.24
<b>LinReg Stroke</b> <sup>†</sup> [4]	0.24	0.43
Mean RMSE (°)		
<b>Model 4 CP</b>	6.8	6.4
<b>Model 5 CP</b>	7.0	3.8
<b>PCA+MLinReg CP</b> [16]	9.0	7.5
<b>NN99 CP</b> [33]	9.7	6.7
<b>NN01 CP</b> [34]	9.2	not reported

<sup>†</sup> Only peak flexion and not over the whole time series.

#### 4.2. Limitations

Besides the lack of external validation, the main limitation of the proposed models is the relatively small size of the database. DL models usually need large amounts of data to be properly trained. Unfortunately, this is rarely the case in biomedical databases. Another limitation of the model is that it does not consider other aspects of the patient, such as psychological factors, age, stress, and social environment, among others, which play a major role in the rehabilitation and, thus, in the treatment outcome.

#### 4.3. Conclusions

It was concluded from the results that the number of patients and type of disease did not directly affect the model's performance. More precisely, we can say that inter- and intra-subject variability affected the model's performance more than the number of patients (samples) and type of disease. Table 1 gives a detailed description of the number of patients with each disease, and Tables 4–6 report the number of training samples. The minimum number of patients was 3 with CP and TBI diseases, while the maximum number of patients was 12 with MS disease. We noticed that the RMSE of CP patients and TBI patients were 6.00° and 5.24°, respectively. On the other hand, the RMSE of MS patients

was 5.8°. This showed that having four-times more patients for a given disease than others did not significantly affect the RMSE value.

Finally, Bi-LSTM combined with MTL was highly effective at increasing the total quantity of information accessible to the model, enhancing the context provided to the algorithm. Future work will focus on MTL models with Bi-LSTM networks to exploit more precise information about treatments, such as the dose information, to further enhance the context given to the model.

**Author Contributions:** Conceptualization, A.K., O.G., S.G.-S. and V.V.; methodology, A.K. and A.H.; software, A.K. and A.H.; validation, A.K. and A.H.; formal analysis, A.K., O.G., S.G.-S. and V.V.; investigation, A.K., O.G., S.G.-S. and V.V.; resources, O.G., S.G.-S. and V.V.; data curation, A.K. and O.G.; writing—original draft preparation, A.K.; writing—review and editing, A.K., O.G., S.G.-S. and V.V.; visualization, A.K.; supervision, S.G.-S. and V.V.; project administration, S.G.-S. and V.V.; funding acquisition, O.G. and V.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was part of a Master’s thesis funded by FéDeV and a Ph.D. project funded by HEC Pakistan.

**Institutional Review Board Statement:** This retrospective study was approved by the local research ethics committee.

**Informed Consent Statement:** Since this is a retrospective study, no informed consent was needed. Patients were informed about the research and did not oppose to the utilization of their data.

**Data Availability Statement:** Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data are not available.

**Acknowledgments:** The authors would like to thank FéDev and HEC Pakistan for funding part of this research. We thank the staff of the Movement Analysis Laboratory of UGECAM Coubert, who acquired all the data used in this work. We would also like to acknowledge the anonymous Reviewers for their valuable observations and suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
ML	Machine learning
DL	Deep learning
MTL	Multi-task learning
CP	Cerebral palsy
CGA	Clinical gait analysis
LSTM	Long short-term memory
CNN	Convolutional neural network
MAE	Mean absolute error
RMSE	Root-mean-squared error
SE	Standard error
BTX-A	Botulinum toxin type A
TBI	Traumatic brain injury
SCI	Spinal cord injury
MS	Multiple sclerosis
CMA	Clinical movement analysis
DNN	Deep neural network
RNN	Recurrent neural network
IMU	Inertial measurement unit
HPA	High-pass algorithm
Bi-LSTM	Bi-directional LSTM
MTD	Medical treatment data

## References

- McLoughlin, J.; Barr, C.; Crotty, M.; Lord, S.; Sturnieks, D. Association of postural sway with disability status and cerebellar dysfunction in people with multiple sclerosis: a preliminary study. *Int. J. Care* **2015**, *17*, 146–151. [\[CrossRef\]](#) [\[PubMed\]](#)
- Blumhardt, L. *Multiple Sclerosis Dictionary*; Taylor & Francis: Abingdon, UK, 2004.
- Sun, L.C.; Chen, R.; Fu, C.; Chen, Y.; Wu, Q.; Chen, R.; Lin, X.; Luo, S. Efficacy and Safety of Botulinum Toxin Type A for Limb Spasticity after Stroke: A Meta-Analysis of Randomized Controlled Trials. *BioMed Res. Int.* **2019**, *2019*, 1–17. [\[CrossRef\]](#) [\[PubMed\]](#)
- Roche, N.; Boudarham, J.; Hardy, A.; Bonnyaud, C.; Bensmail, D. Use of gait parameters to predict the effectiveness of botulinum toxin injection in the spastic rectus femoris muscle of stroke patients with stiff knee gait. *Eur. J. Phys. Rehabil. Med.* **2015**, *51*, 361–370. [\[PubMed\]](#)
- Notice Patient—DYSPORT 500 UNITES SPEYWOOD, Poudre Pour Solution Injectable—Base de Données Publique des médicaments. Available online: <https://base-donnees-publique.medicaments.gouv.fr/affichageDoc.php?specid=60242321&typedoc=N> (accessed on 29 April 2021).
- Tribouillard, H. Elaboration d’une Methodologie de Validation des Indications hors Autorisation de mise sur le Marche de la Toxine Botulique de Type A au Centre Hospitalo-Universitaire de Lille: Exemple du Bavage Chez l’adulte. Ph.D. Thesis, Faculte de Pharmacie, Universite de Lille, Lille, France, 2018.
- Battagli, D. Utilisation des Toxines Botuliques aux Hospices Civils de Lyon: Bilan 2015 des Indications. Ph.D. Thesis, Faculte de Pharmacie, Universite Claude Bernard—Lyon 1, Villeurbanne, France, 2017.
- Baker, R.W. *Measuring Walking: A Handbook of Clinical Gait Analysis*, 1st ed.; MacKeith Press: London, UK, 2013.
- McGinley, J.L.; Baker, R.; Wolfe, R.; Morris, M.E. The reliability of three-dimensional kinematic gait measurements: A systematic review. *Gait Posture* **2009**, *29*, 360–369. [\[CrossRef\]](#)
- Moon, D.; Esquenazi, A. Instrumented Gait Analysis: A Tool in the Treatment of Spastic Gait Dysfunction. *JBJS Rev.* **2016**, *4*, 1. [\[CrossRef\]](#)
- Zörner, B.; Filli, L.; Reuter, K.; Kapitza, S.; Lörincz, L.; Sutter, T.; Weller, D.; Farkas, M.; Easthope, C.S.; Czaplinski, A.; et al. Prolonged-release fampridine in multiple sclerosis: Improved ambulation effected by changes in walking pattern. *Mult. Scler.* **2016**, *22*, 1463–1475. [\[CrossRef\]](#)
- Camardella, C.; Cappiello, G.; Curto, Z.; Germanotta, M.; Aprile, I.; Mazzoleni, S.; Scoglio, A.; Frisoli, A. A Random Tree Forest decision support system to personalize upper extremity robot-assisted rehabilitation in stroke: a pilot study. *IEEE Int. Conf. Rehabil. Robot.* **2022**, *2022*, 1–6. [\[CrossRef\]](#)
- LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)
- Topol, E.J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. [\[CrossRef\]](#)
- Barnes, S.; Saria, S.; Levin, S. An Evolutionary Computation Approach for Optimizing Multilevel Data to Predict Patient Outcomes. *J. Healthc. Eng.* **2018**, *2018*. [\[CrossRef\]](#) [\[PubMed\]](#)
- Galarraga, C.O.A.; Vigneron, V.; Dorizzi, B.; Khouri, N.; Desailly, E. Predicting postoperative gait in cerebral palsy. *Gait Posture* **2017**, *52*, 45–51. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhang, Y.; Yang, Q. A survey on multi-task learning. *IEEE Trans. Knowl. Data Eng.* **2021**. [\[CrossRef\]](#)
- Su, B.; Gutierrez-Farewik, E.M. Gait trajectory and gait phase prediction based on an LSTM network. *Sensors* **2020**, *20*, 7127. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhu, C.; Liu, Q.; Meng, W.; Ai, Q.; Xie, S.Q. An Attention-Based CNN-LSTM Model with Limb Synergy for Joint Angles Prediction. In Proceedings of the 2021 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Sapporo, Japan, 11–15 July 2021; pp. 747–752.
- Zaroug, A.; Lai, D.T.; Mudie, K.; Begg, R. Lower limb kinematics trajectory prediction using long short-term memory neural networks. *Front. Bioeng. Biotechnol.* **2020**, *8*, 362. [\[CrossRef\]](#)
- Hernandez, V.; Dadkhah, D.; Babakeshizadeh, V.; Kulić, D. Lower body kinematics estimation from wearable sensors for walking and running: A deep learning approach. *Gait Posture* **2021**, *83*, 185–193. [\[CrossRef\]](#)
- Jia, L.; Ai, Q.; Meng, W.; Liu, Q.; Xie, S.Q. Individualized Gait Trajectory Prediction Based on Fusion LSTM Networks for Robotic Rehabilitation Training. In Proceedings of the 2021 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Sapporo, Japan, 11–15 July 2021; pp. 988–993.
- Liu, D.X.; Wu, X.; Wang, C.; Chen, C. Gait trajectory prediction for lower-limb exoskeleton based on Deep Spatial-Temporal Model (DSTM). In Proceedings of the 2017 2nd International Conference on Advanced Robotics and Mechatronics (ICARM), Kunming, China, 10–20 May 2017; pp. 564–569.
- Kolaghassi, R.; Al-Hares, M.K.; Marcelli, G.; Sirlantzis, K. Performance of Deep Learning Models in Forecasting Gait Trajectories of Children with Neurological Disorders. *Sensors* **2022**, *22*, 2969. [\[CrossRef\]](#)
- Cardarelli, S.; Mengarelli, A.; Tigrini, A.; Strazza, A.; Nardo, F.D.; Fioretti, S.; Verdini, F. Single IMU Displacement and Orientation Estimation of Human Center of Mass: A Magnetometer-Free Approach. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 5629–5639. [\[CrossRef\]](#)
- Desailly, E.; Daniel, Y.; Sardain, P.; Lacouture, P. Foot contact event detection using kinematic data in cerebral palsy children and normal adults gait. *Gait Posture* **2009**, *29*, 76–80. [\[CrossRef\]](#)

27. Schwartz, M.H.; Rozumalski, A. The Gait Deviation Index: A new comprehensive index of gait pathology. *Gait Posture* **2008**, *28*, 351–357. [[CrossRef](#)]
28. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
29. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
30. Salehinejad, H.; Baarbe, J.; Sankar, S.; Barfett, J.; Colak, E.; Valaee, S. Recent Advances in Recurrent Neural Networks. *arXiv* **2018**, arXiv:1801.01078.
31. Schuster, M.; Paliwal, K. Bidirectional recurrent neural networks. *Signal Process. IEEE Trans.* **1997**, *45*, 2673–2681. [[CrossRef](#)]
32. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
33. Nailer, T.A.; Richards, J.G.; Miller, F.; Sun, J.Q.; Castagno, P. Reliability of predictions of post-operative gait in rectus transfer patients using FFT neural networks. *Gait Posture* **1999**, *9*.
34. Nailer, T.A. Efficacy of Predictions of Post-Operative Gait in Rectus Transfer Patients Using Neural Networks. Ph.D. Thesis, University of Delaware, Newark, Delaware, 2001.