A Transferable Deep Learning Prognosis Model for Predicting Stroke Patients' Recovery in Different Rehabilitation Trainings

Ping-Ju Lin, Xiaoxue Zhai, Wei Li, Tianyi Li, *Graduate Student Member, IEEE*, Dandan Cheng, Chong Li, Yu Pan, and Linhong Ji

***Abstract*—Since the underlying mechanisms of neurorehabilitation are not fully understood, the prognosis of stroke recovery faces significant difficulties. Recovery outcomes can vary when undergoing different treatments; however, few models have been developed to predict patient outcomes toward multiple treatments. In this study, we aimed to investigate the potential of predicting a treatment's outcome using a deep learning prognosis model developed for another treatment. A total of 15 stroke survivors were recruited in this study, and their clinical and physiological data were measured before and after the treatment (clinical measurement, biomechanical measurement, and electroencephalography (EEG) measurement). Multiple biomarkers and clinical scale scores of patients who had completed manual stretching rehabilitation training were analyzed. Data were used to train deep learning prognosis models, yielding an 87.50% prognosis accuracy. Pre-trained prognosis models were then applied to patients who completed robotic-assisted stretching training, yielding a prognosis accuracy of 91.84%. Interpretation of the deep learning models revealed several key factors influencing patients' recoveries, including the plantar-flexor active range of movement ($r = 0.930$, $P = 0.02$), dorsiflexor strength ($r = 0.932$, $P = 0.002$), plantar-flexor strength ($r = 0.930$, $P = 0.002$), EEG power spectrum density and EEG functional connectivities in the occipital, central parietal, and parietal areas. Our results suggest (i) that deep learning can be a promising method for accurate prediction of the recovery potential of stroke patients in clinical scenarios and (ii) that it can be successfully applied to different rehabilitation trainings with explainable factors.**

Manuscript received 11 June 2022; revised 9 August 2022; accepted 4 September 2022. Date of publication 9 September 2022; date of current version 6 December 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 51805288, in part by Young Elite Scientists Sponsorship Program by CAST under Grant 2019QNRC001, in part by Tsinghua University Precision Medicine Research Program under Grants 10001020124 and 10001020129, and in part by the Capital's Funds for Health Improvement and Research Program under Grant 2022-2Z-2242. *(Corresponding authors: Chong Li; Yu Pan.)*

Ping-Ju Lin, Wei Li, Tianyi Li, Chong Li, and Linhong Ji are with the Division of Intelligent and Bio-mimetic Machinery, The State Key Laboratory of Tribology, Tsinghua University, Beijing 100084, China (e-mail: lbr20@mails.tsinghua.edu.cn; lw_tsinghua15@126.com; kittyshusky@gmail.com; chongli@tsinghua.edu.cn; jilh@tsinghua.edu.cn).

Xiaoxue Zhai, Dandan Cheng, and Yu Pan are with the Department of Physical Medicine and Rehabilitation, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing 100084, China (e-mail: zxxa02445@btch.edu.cn; fdda02901@btch.edu.cn; panyu@btch.edu.cn).

Digital Object Identifier 10.1109/JBHI.2022.3205436

***Index Terms*—Stroke prognosis, deep learning, electroencephalography, robot-assisted training, manual stretching.**

I. Introduction

STROKE is the leading cause of death and long-term functional impairment worldwide [1]. Stroke survivors commonly suffer from motor impairments that negatively impact their daily lives [2]. Fortunately, several specialized therapies have emerged in recent years to improve the motor abilities of stroke survivors [3]. Furthermore, new bioinformation can improve the efficiency of stroke rehabilitation. Clinical scales, demographic information, and neurophysiological data are valuable to predict recovery outcomes for stroke patients. For instance, a clinical scale of stroke patients has been used to predict the recovery of upper extremity function in the following 3 months [4], [5]. Demographic information and clinical scales have been used to create a new biomarker, CoRisk, to identify patients who were less likely to exhibit favorable outcomes within a 3-month period [6]. Liu combined all three types of biometrics to predict whether a stroke patient can achieve effective recovery [7], [8]. In addition, psychological biomarker was also applied in certain model to prognosticate stroke recovery [9]. The above prognostic models were developed based on the correlations between recovery efficacy and biometrics.

Currently, technological rehabilitation treatments, including robot-assisted training, functional electrical stimulation, and transcranial magnetic stimulation (TMS), have been validated in clinical settings and have shown competence in stroke patients' motor recovery [3], [10], [11]. However, since stroke patients are a heterogeneous group, recovery outcomes vary among patients who receive the same treatment [12], indicating an even greater dissimilarity between patients from different treatments. The diversity among patients makes it difficult to determine whether a specific type of treatment would benefit one specific patient [12]. Additionally, having an accurate prediction model can help clinicians avoid challenges in stroke recovery due to uncertainties and delays of treatments [13], [14]. Therefore, the prognosis of individual stroke survivors' recovery outcomes under a specific rehabilitation method is a significant problem.

Artificial intelligence has become a promising method for predicting stroke survivors' recovery outcomes [15], [16]. The

gradient boosting decision tree was created to cope with various complicated inputs in order to forecast motor progress in stroke survivors [8]. A combination of initial clinical scale scores, demographic information, and structural magnetic resonance imaging (MRI) data has been selected to predict the outcome after repetitive TMS intervention [12]. The study applied five machine learning models to predict individual motor impairment using MRI results and other information [12]. A neurophysiological biomarker signal, electroencephalography (EEG), was used to predict stroke patients' outcomes after brain–computer interface (BCI) upper limb rehabilitation training [17], [18]. Another machine learning method, partial least square, was applied to predict the functional recovery based on EEG data [19]. Another study used a deep neural network to predict mortality based on stroke patients' EEG data [20]. Since deep learning remains a "black box" solution [11], another focus of this research was to understand the mechanism behind the prognosis model and the factors influencing neurorehabilitation. Interpretation methods, including the Shapley additive explanations (SHAP) value [21], attention maps, and gradient-weighted class activation mapping, have been used in the visualization of medical biomarkers [22], computed tomography scans [23], and EEG signals [24], respectively. The above methods might have the potential to find out the recovery mechanisms from deep learning–supported prognosis.

In the present study, we aimed to investigate the potential of using a deep learning-based prognosis model developed for a specific treatment to predict the recovery outcome of another treatment. Therefore, a deep learning model was created to prognosticate the efficacy of motor recovery after 2 weeks of manual stretching training by learning a combination of clinical scale scores, demographic information, and resting-state EEG data before training. Then, we applied the pre-trained model to learn the corresponding data collected from robot-assisted stretching training to prognosticate the efficiency of motor recovery. In addition, the interpretation method was also employed to identify the critical factors that influenced the prognosis result. Section II introduces the architecture of the proposed model, and Section III presents the prediction results for stroke patients after 2-week rehabilitation training. The paper ends with a discussion on the merits of the pre-trained model and the factors influencing outcomes.

II. Materials and Methods

A. Participants

The Institutional Ethical Committee of Tsinghua Changgung Hospital approved this study. This study was conducted according to the principles expressed in the Declaration of Helsinki (18172-0-01) and registered as a clinical trial (ChiCTR2000030108).

Between May 2019 and November 2020, 15 individuals were recruited based on our inclusion criteria. Each participant gave written informed consent prior to their eligibility assessment. No adverse events related to the study occurred. All participants were recruited at the Rehabilitation Department of Beijing Tsinghua Changgung Hospital. The main inclusion criteria were: (i) participants should have no prior history of recurrent stroke; (ii) rehabilitation within 6 months from stroke onset; (iii) age ranges from 18 to 75 years; (iv) participants should understand the physician's orders and must not have cognitive deficits that could prevent them from undertaking the rehabilitation tasks; (v) the participants need to have the ability to remain in a stable state without spasticity during the EEG measurements and training sessions. Table I reveals the patients' information.

TABLE I
Patients' Information

	Manual	Robot-assisted	*P*
Subject	8	7	
Demographics			
Gender (woman)	25%	14%	1
Age	59.88(±10.10)	63.71(±12.19)	0.35
Clinical Measurements			
Aphasia	13%	14%	1
Cognitive Impairment	0%	14%	0.47
Ipsilateral hemiplegia	13%	57%	0.12
Hemorrhage	88%	100%	1
Course of Disease	100.13 (±48.22)	64.29 (±29.36)	0.13
Berg Balance Test	42.63 (±9.82)	38.29 (±13.59)	0.64
Postural Assessment scale	30.13 (±4.29)	28.29 (±6.70)	0.72
Activate Daily Living	71.88 (±28.65)	57.86 (±24.13)	0.27
Fugl-Meyer (Lower limb)	24.88 (±5.41)	19.43(±8.73)	0.15
Biomechanical Measurements			
Dorsiflexor AROM	5.25 (±5.75)	4.29 (±5.47)	0.71
Dorsiflexor PROM	17.50 (±3.81)	15.00 (±2.45)	0.12
Plantar-flexion AROM	18.00 (±17.89)	12.43 (±10.55)	0.55
Plantar-flexion PROM	39.50 (±5.43)	36.71 (±4.39)	0.15
Dorsiflexor strength	55.50 (±64.10)	62.29 (±65.47)	0.91
Plantar flexor strength	65.88 (±64.52)	93.29 (±74.68)	0.35

AROM, active range of motion; PROM, passive range of motion
*Statistical result of two groups in different treatments. No significant difference was observed (in demographic characteristics, clinical measurement, and biomechanical measurement).

Each participant was evaluated by clinical, biomechanical, and EEG measurements. All measured data were applied in our deep learning model to prognosticate patients' recovery outcome.

B. Clinical Measurements

All participants completed the following functional assessments during the clinical evaluations. The Fugl–Meyer Assessment of Lower Extremity (FMA-LE) is the evaluation standard for measuring lower limb sensorimotor recovery (from 0 to 34 points, hemiplegia to healthy) [25]. The Berg Balance Scale (BBS) is used for assessing the balancing function of participants (from 0 to 56 points, imbalance to balance) [26]. The Postural Assessment Scale for Stroke (PASS) is a 12-item performance-based scale used for assessing and monitoring postural control following stroke (from 0 to 36, worst to normal). The Activities of Daily Living (ADL) scale is designed to describe fundamental skills required to care for oneself independently or collectively (from 0 to 100, worst to normal) [27]. Clinical evaluations were performed immediately before and after the 2-week intervention (see Table I).

C. Biomechanical Measurements

In the biomechanical evaluations, measurements of muscle strength, range of motion (ROM), and ankle stiffness (Nm/°)

were conducted. Ankle stiffness was measured by an ankle robot [28]. The dorsiflexor (DF) and plantar-flexor (PF) muscle strength were measured on each patient's affected side. Ankle stiffness includes the DF stiffness and PF stiffness, which were calculated as the ratio of the reaction moment from ground to the angular deflection of the specific joint as T/θ [29], where T is the passive torque increment during the amount of ankle movement. When θ is an infinitely small angle, the quasi-static stiffness can be approximated by the slope of a tangent line of the torque–angle curve at an ankle position. The peak stretching velocity was set at 5 degrees/second during the passive range of motion (PROM) to avoid reflex responses induced by our experimental criteria.

Additionally, the PROM stiffness of DF at 10° and of PF at 30° was measured during the PROM. The PROMs of DF and PF (PF PROM/PF PROM) and active ROMs (AROMs) of DF and PF (DF AROM/PF AROM) were measured by a HogganMicroFET3 portable device (Hoggan Health Industries, Inc. Salt Lake City, USA). All biomechanical evaluations were performed immediately before and after the 2-week intervention (see Table I).

D. Neurophysiological Measurements

Before treatment intervention, all participants' EEG signals were recorded during neurophysiological measurement sessions. The goal of these sessions is to look for patterns in the participants' EEG power spectrum and functional connectivities while they are at rest. A NeuSen W system (Neuracle Tech., Changzhou, China) was used to capture signals from 32 EEG channels covering the entire scalp at a sample rate of 1000 Hz. Raw EEG data were pre-processed following the method described in our previous study; they were first averaged by an average reference and then filtered to a frequency range of 0.5–45 Hz using a finite impulse response band-pass filter [31]. EEG waves were subjected to independent component analysis to eliminate human interference. The processed EEG data were then separated into 2-s intervals [14]. To ensure data consistency between our prognosis model and displayed results, EEG data collected from ipsilateral hemiplegia patients with lesions in the right hemisphere have been flipped to corresponding channels in the left hemisphere [30]. Thus, the two hemisphere areas correspond to the information from electrodes FP2, AF4, F4, F8, FC2, FC6, C4, T8, CP2, CP6, P4, P8, PO4, and O2 (affected hemisphere) and FP1, AF3, F3, F7, FC1, FC5, C3, T7, CP1, CP5, P3, P7, PO3, and O1 (unaffected hemisphere).

E. EEG Power Spectrum Density and Functional Connectivities

MATLAB's built-in function pwelch was used to compute each segment's EEG power spectrum density from pre-processed data (R2018b, MathWorks, Natick, MA, USA). The Welch technique was modified in this work using the following parameters: 0.5 Hz frequency resolution, 1-s Hanning window overlap, and 0 phase shift. Each trial was averaged into a single matrix to eliminate biased prediction due to experimental error. Consequently, a two-dimensional matrix was created, with the two dimensions representing EEG signal channels and EEG data frequency (0.5–45 Hz, interval of 0.5 Hz). For EEG power data, the matrix size was (32, 90).

This study focused on six recognized EEG signal bands discovered in earlier studies as significant indicators for EEG functional connectivities, namely Delta (1–3 Hz), Theta (4–7 Hz), Alpha (8–13 Hz), Beta Low (13–18 Hz), Beta Medium (18–21 Hz), and Beta High (21–30 Hz) [31], [32]. EEG functional connectivities was calculated using a previously described method [14]. The columns represent the 496 two-electrode combinations (for 32 electrodes, $1 + 2 + \ldots + 31 = 496$). The rows represent the six bands used to illustrate connectedness (6).

The EEG power spectrum density and functional connectivities results were normalized based on the z-score, which boosted the deep learning model's prediction accuracy [33].

F. Treatments

This study was an assessor-blinded, randomized controlled experiment. Each participant was randomly assigned to one of the two different rehabilitation treatments, namely robot-assisted ankle stretching or manual stretching training, for 10 sessions in 2 weeks (five times a week, 20 min/session).

A physiotherapist with at least 5 years of experience in clinical rehabilitation gave manual stretching training. Each participant's PROM ankle score was required to be measured to ensure safety before the manual stretching training. During the manual stretching training, the physiotherapist gently stretched the participant's affected side ankle from PF to DF in the sagittal plane and then held it at the extreme DF position for 5 s. In all stretching training sessions, the participants remained in a relaxed position (Fig. 2(a)).

The ankle stretching robot can assist a stroke patient with ankle stretching; this robot was described in our previous study [28]. During treatment, the participants were required to look at the display monitor; when the monitor displayed the phrase "ankle joint," the robotic device assisted the patients in conducting ankle stretching training, moving from the DF position to the PF position (Fig. 2(b)).

G. Labeling of the Development Datasets

Previous studies have shown that recoveries from stroke usually plateau at ∼70% from initial performance to complete recovery, regardless of patient age or sex, stroke type, and therapy dose [34], [35], [36], [37], [38]. Therefore, proportional recovery was used as a standard in clinical rehabilitation to determine whether a patient can recover to expectation under specific rehabilitation training. We used the increase between FMA-LE scores before and after intervention from clinical evaluation to estimate motor recovery performance. The maximum value was 34. The following function was used to calculate the FMA score if the expected 70% improvement from the initial assessment was achieved:

$$\Delta FMA_{predict} = 0.7 \times (34 - FMA_{initial}) \quad (1)$$

The residual was defined as the difference between predicted and observed improvement from the initial assessment to the assessment after intervention [7], [8], [34]. In the present study, the median value of differences between predicted and observed

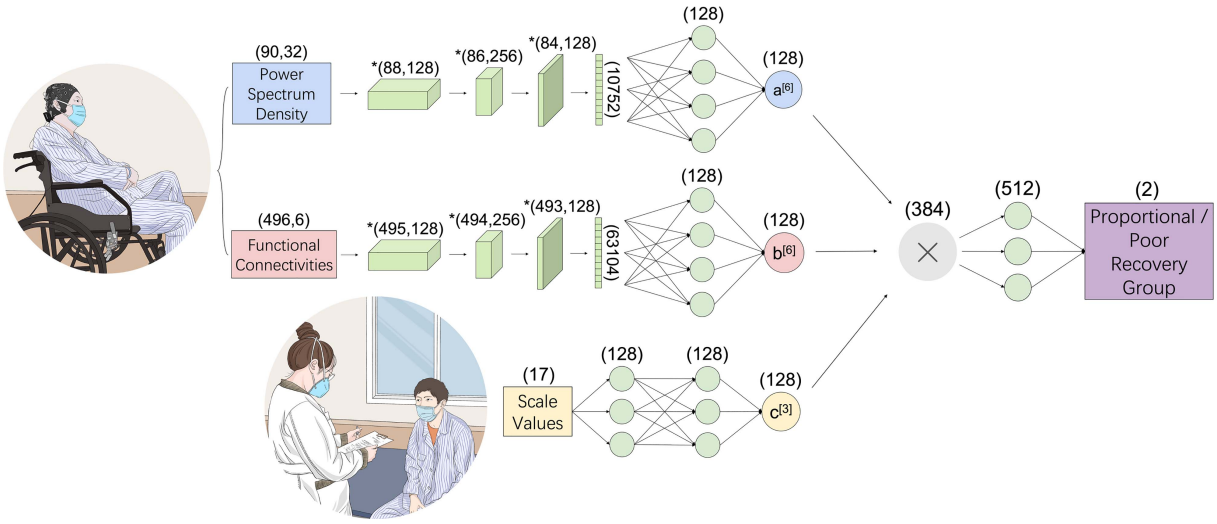

Fig. 1. Framework of the deep learning prognosis model for stroke survivors. The following inputs were applied to predict rehabilitation outcome: (i) Scaled values including clinical scales and biomechanical measurements, (ii) EEG functional connectivities, and (iii) EEG power spectra. *Matrices have a Conv-BatchNorm-ReLU structure.

(a)

(b)

Fig. 2. (a) The treatment of Manual stretching (Physiotherapist is assisting stroke survivor ankle movement). (b) The treatment of Robot-assisted stretching (The Robotic is assisting stroke survivor ankle movement).

values was 6. Therefore, patients were assigned to the proportional recovery group if their residuals were smaller than 6 and to the poor recovery group if their residuals were greater than 6.

H. Deep Learning Model

A supervised deep learning architecture was designed for the prognosis model for predicting recovery outcomes. Three inputs were selected: binary and scaled values obtained from clinical assessments and biomechanical measurements (size: [17]) (Table I), functional connectivities based on EEG data (size: [496, 6]), and the power spectrum density based on EEG data (size: [90, 32]).

Two 128-node fully connected layers in the prognosis model were used to learn features from clinical and biomechanical assessment. In addition, the convolution layers were used to learn features from the EEG power spectrum and functional connectivities. This prognosis model utilized Nesterov-accelerated Adaptive Moment Estimation (Nadam) gradient descent as the stochastic optimizer for our prognosis model. The Nadam optimizer combines the advantages of Nesterov adaptive, momentum optimizer, and RMS-Prop optimizer and is thus capable of working with large datasets [39]. With the Nadam optimizer, our deep learning prognosis model can automatically learn multiple features from a large training dataset with adaptive rates. Moreover, this prognosis model utilizes the categorical cross-entropy as a loss function. For forwarding and back-propagation, 200 epochs with a batch size of 32 were used.

EEG inputs were first applied on the convolution layer, then on the batch normalization layer, and then activated by the Rectified Linear Unit (ReLU) activation function that is used in all layers (Conv-BatchNorm-ReLU). The complete sequence was performed three times to extract features. Moreover, the kernel size was different in the EEG power spectrum density (3) and EEG functional connectivities (2). After three repeating

sequences, the matrix would be flattened. Next, a fully connected layer (dense) with 128 nodes was added to the resulting matrix after the feature extraction layer.

Finally, the concatenate layer collected matrices from those three inputs, namely, scale information, the EEG power spectrum density, and EEG functional connectivities. These matrices were placed in a 512-node dense layer with all important characteristics preserved. The output was a prediction of each patient's recovery category, namely, poor recovery or proportional recovery.

To validate the prognosis model's reliability, leave-one-out cross-validation was applied. When verifying deep learning models with a small sample set, leave-one-out cross-validation is frequently utilized [40]. We applied this method to ensure the reliability of the prediction for each model; for each training session of the leave-one-out method, a single stroke patient's data were "left out" to serve as the validation set, while the remaining data from the training set comprised the other sets.

Our proposed architecture was implemented in CUDA and run on an RTX 3070Ti GPU with 8 GB memory (NVIDIA). The standard implementations run on an Intel i7-11700k CPU at 3.60 GHz. Keras was the platform used for building our deep learning models' architecture. The average time for activating the model was 68.16 s.

I. SHAP Interaction Values

SHAP values were used to explain the machine learning model. This method is a game-theoretic approach to explain the output of the machine learning model. However, this study used the deep learning method to build a prognosis model. Therefore, to explain the output of the deep learning method, the function GradientExplainer was applied in this study. GradientExplainer is a technique that combines ideas from Integrated Gradients, SHAP, and SmoothGrad.

In the present study, we validated the reliability of the deep learning model with the leave-one-out method. Therefore, the number of parameter sets of the prognosis model was the same as the number of datasets from each patient. Each prognosis model has its SHAP interaction values because it uses a different training set. Pearson's correlation was applied to find features that were simultaneously important across all models generated from leave-one-out validation.

J. Statistical Analysis

The Shapiro–Wilk test and the homogeneity of variance test were applied to evaluate the differences between proportional recovery group and poor recovery group datasets. Fisher's exact test was used to compare categorical variables, and the Wilcoxon rank-sum test was used to compare continuous variables in baseline characteristics.

By applying the Shapiro–Wilk test and the homogeneity of variance test to obtained EEG signals, we could verify if the power spectrum density and functional connectivities were normally distributed [41]. The two-sample t-test or the Wilcoxon rank-sum test was applied to examine differences between the proportional and poor recovery groups. The Pearson or Spearman correlation coefficient was applied to find the correlation.

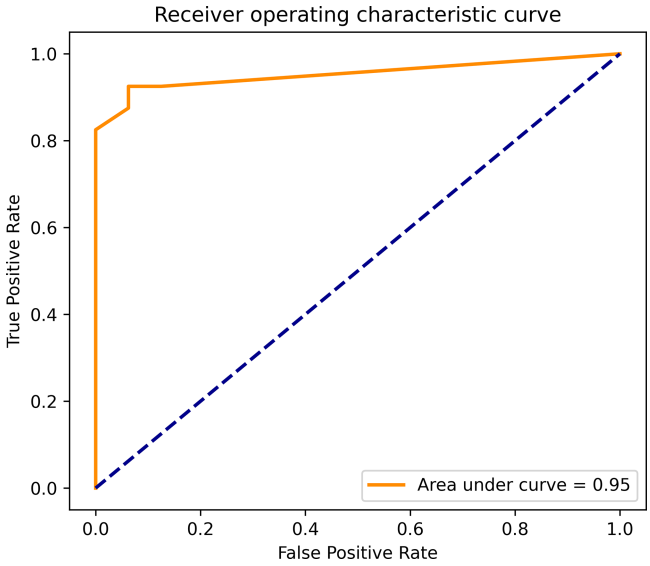

Fig. 3. Receiver operating characteristic curve of manual stretching prognosis models was generated by the leave-one-out method (orange curve). The area under the characteristic curve was 0.95. An area under the curve of 1 indicates absolute prognosis accuracy.

The false discovery rate (FDR) was used to avoid Type I errors, and a T-distribution was applied to obtain the 95% confidence interval (95% CI). All statistical analyses were conducted using MATLAB and SPSS version 26.0 (IBM Corporation, Armonk, NY, USA).

III. Results

A. Performance of the Prognosis Model for Predicting the Outcome of Manual Stretching Training

Clinical data, biomechanical data, EEG functional connectivities, and EEG power spectra were applied to our prognosis model as inputs (as shown in Fig. 1) to predict rehabilitation outcomes.

In the leave-one-out cross-validation method, the number of parameter sets of the prognosis model generated by this method is the same as the number of available training sets, which is eight. The prognosis model was used as a classification model to predict whether the patient can reach proportional recovery. The prediction results of the prognosis model generated by the leave-one-out method. The prediction accuracy is 87.50%. Fig. 3 shows the receiver operating characteristic (ROC) curve area. The area under the ROC curve (AUC) was 0.95. The confusion matrix of the manual stretching prognosis model shows that the number of true positives is 3 and that the number of true negatives is 4 (Fig. 4). The sensitivity was 1.00, and the specificity was 0.80.

B. Performance of Pre-Trained Prognosis Model for Predicting the Outcome of Robot-Assisted Training

The prognosis model generated from manual stretching datasets was used to predict the rehabilitation outcomes of

TABLE II
PRE-TRAINING MANUAL STRETCHING PROGNOSIS MODEL APPLIED ON ROBOT-ASSISTED TRAINING

Participants	Model_1	*Model_2	Model_3	Model_4	Model_5	Model_6	Model_7	Model_8	Avg	Avg_nE
Accuracy	100%	85.72%	85.72%	85.72%	71.43%	100%	100%	100%	91.07%	91.84%
Sensitivity	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00	0.941	0.933
Specificity	1.00	0.80	0.80	0.80	0.80	1.00	1.00	1.00	0.897	0.912

Avg, the average from a total of eight sets of parameters of the prognosis model;
Avg_nE, the average accuracy from a total of seven sets of parameters of the prognosis model, excluding the erroneous model in manual stretching rehabilitation training.
*Model 2 is the erroneous model in manual stretching rehabilitation training.

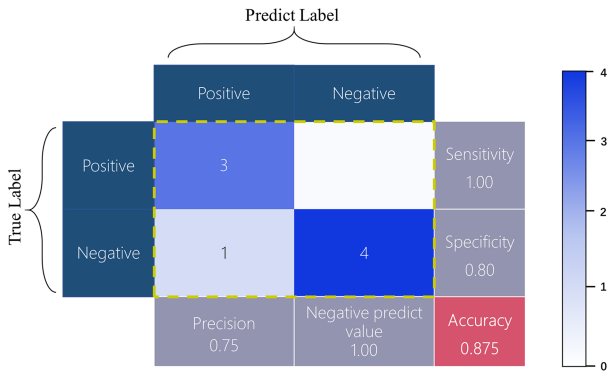

Fig. 4. Confusion matrix of manual stretching prognosis models was generated by the leave-one-out method.

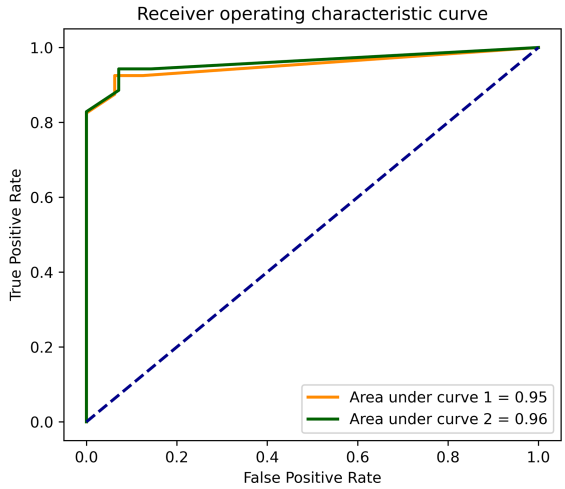

Fig. 5. Orange receiver operating characteristic curve represents average prognosis accuracy for all eight patients who received robot-assisted treatment (curve 1). The green receiver operating characteristic curve represents the average prognosis accuracy for seven patients, excluding one out-of-bounds model (curve 2). The areas under characteristic curves can be found in the legend (curve 2).

robotic-assisted stretching. The inputs of the pre-trained prognosis model were the same as for the prognosis model used above.

The prognosis results from the pre-trained prognosis model are presented in Table II. The prediction accuracy showed competitive performance in classifying between the proportional recovery group and the poor recovery group after robot-assisted training. The average accuracy was 91.07% (sensitivity is 0.941, specificity is 0.897), and the average accuracy, excluding the out-of-bound parameter set, in manual stretching rehabilitation training was 91.84% (sensitivity is 0.933; specificity is 0.912). The operation was based on a previous study, which can avoid the influence from an erroneous model [42]. The lowest prediction accuracy for one pre-trained prognosis model was 71%. Four among the rest seven parameter sets presented 100% accuracy in predicting robotic-assisted stretching training outcomes. Fig. 5 shows the receiver operating characteristic (ROC) curve area of average and excluding the out-of-bound parameter sets. The area under the ROC curve (AUC) of the average was 0.95 and the area under the ROC curve (AUC) of excluding the out-of-bound parameter was 0.96.

C. Key Factors of Prognosis Model

Results from GradientExplainer reveal key factors from all different parameter sets' prognosis models, including clinical data, biomechanical data, the EEG power spectrum density, and EEG functional connectivities. Key factors determining whether a patient can make a proportional recovery are thus analyzed using a statistical method.

Characteristics of biomechanical measurements were revealed to show significant differences in PF AROM, DF strength, and PF strength. After Pearson correlation coefficient analysis, the correlation and credibility of biomechanical measurements were as follows: PF AROM ($r = 0.930$, $P = 0.02$), DF strength ($r = 0.932$, $P = 0.002$), and PF strength ($r = 0.930$, $P = 0.002$). The above analyses all showed high r values, which indicates high positive correlation between these biomarkers and patient recovery.

The results obtained from the EEG power spectrum density through the Pearson correlation coefficient analysis showed a positive correlation between the power density in the occipital lobe (O2) and patient recovery; the range of this positive correlation included the Alpha band to the Beta High band (Alpha band, $r = 0.956$, $P = 0.001$; Beta Low band, $r = 0.782$, $P = 0.038$; Beta Medium band, $r = 0.897$, $P = 0.006$; Beta High band, $r = 0.782$, $P = 0.038$). Power density in the central parietal lobe (CP2) showed a positive correlation with patient recovery in the Theta band ($r = 0.812$, $P = 0.026$). Power density in the right parietal lobe (P8) showed a positive correlation with patient recovery in the Alpha band ($r = 0.94$, $P = 0.002$). Power density in the left parietal lobe (P3) showed a positive correlation with patient recovery in the Beta Medium band ($r = 0.868$, $P = 0.011$) (Fig. 6(a)).

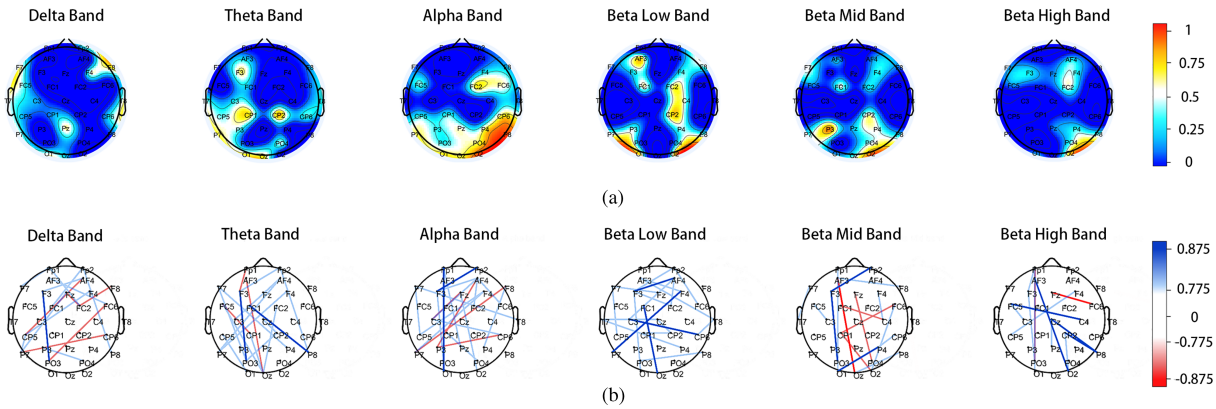

Fig. 6. (a) SHAP interaction values obtained from power spectrum density (warmer color tones represent higher positive correlation). (b) SHAP interaction values obtained from functional connectivities (bright colors mark areas of high correlation, both positive and negative).

The Pearson correlation coefficient analysis results for EEG functional connectivities showed that the following interhemispheric functional connectivities are positively correlated with patient recovery: frontal cortex and central cortex (FC5–C4); frontal cortex and parietal cortex (F3–PO3, F4—P7); frontal cortex and occipital cortex (FP1–O1, FC1–Oz, AF3–O2); parietal cortex and central cortex (P8–CZ, P8–CP2); occipital cortex and parietal cortex (O1–P4). The following functional connectivities within the frontal areas are positively correlated with patient recovery: F3–FP2 and F3–AF4. The values of functional connectivities between the following areas are negatively correlated with patient recovery: frontal cortex and occipital cortex (AF3–Oz) (Fig. 6(b)).

D. Patient Characteristics

In the manual stretching training, significant differences in clinical information between the proportional recovery group and the poor recovery group are found in PASS ($P = 0.039$) and FMA-LE ($P < 0.01$). Significant differences in biomedical information between the two groups were found in DF AROM ($P = 0.030$), PF AROM ($P = 0.010$), and PF strength ($P = 0.014$). Patients from the proportional recovery group were observed to have a consistently higher value for these biomarkers than patients from the poor recovery group. On the contrary, EEG power spectrum density and functional connectivities showed no significant difference between the proportional recovery group and the poor recovery group when evaluated by the Wilcoxon rank-sum test.

In the robotic-assisted stretching training, significant differences in clinical information between the proportional and poor recovery groups were found in BBS ($P = 0.047$) and FMA-LE ($P = 0.004$). Biomedical information between the two groups showed a significant difference in PF PROM ($P = 0.042$), DF strength ($P = 0.012$), and PF strength ($P = 0.002$), in which the proportional recovery group had higher values for all biomarkers than the poor recovery group. Similar to manual stretching training, the EEG power spectrum density and functional connectivities between the two groups showed no significant difference when evaluated by the Wilcoxon rank-sum test.

IV. Discussion

The main objective of this study was to explore the feasibility of creating a transferable deep learning prognosis model compatible with various rehabilitation treatments. Accurate prognosis results obtained in this study proved the possibility of creating a prognosis model that can be transferred between similar rehabilitation treatments.

Our deep learning prognosis model based on clinical data, biomechanical information, and neurophysiological signals has been developed to predict lower limb recovery with competitive performance. The resulting models were validated by leave-one-out cross-validation. A total of eight models are presented based on leave-one-out cross-validation. Only one of these models was inaccurate, leading to an average prognosis accuracy of 87.50%. The prediction results agree with our hypothesis that using all biomarkers in conjunction to predict stroke survivor outcomes under a specific rehabilitation training can result in prediction results with high enough accuracy for use in real-life settings. The prognosis models for manual stretching training accurately classify the proportional and poor recovery groups. Different treatments generally have different feature domains, posing significant difficulties for transfer learning. Regardless, we can conclude from our result that the prognosis models can have high accuracy by evaluating all biomarkers in conjunction with each other and limiting the evaluation to a specific treatment.

Applying the pre-trained prognosis model from manual stretching treatment on stroke survivors with robot-assisted training (a concept from transfer learning) yielded a prediction accuracy of 91.84%. This high accuracy may be due to the advantages of transfer learning [43]. The first advantage is the ability of the pre-trained model to avoid data bias. Since each patient's recovery varies with his/her initial capability, it would be difficult to have an unbiased comparison between the two comparison groups [12]. Furthermore, in clinical settings, patient data are hard to acquire because each patient could face issues such as stability issues, consciousness issues, and other problems that could occur during training [44]. A review study showed that the sample sizes for EEG studies using deep learning are generally around 10 participants per dataset [45]. With such a small sample size, a few-shot learning study shows that the biased data would cause a considerable impact (a 17% drop in

accuracy) [46]. Therefore, a pre-trained model generated from an unbiased dataset could achieve a more accurate result. Our study's pre-trained application result shows an average accuracy of 91.84%. The second advantage of adapting transfer learning in our methodology is applying one prognosis model to datasets from a different training method [43], [47] without having to train a new model for every rehabilitation method. Relational knowledge transfer, one of the transfer learning methods, can deal with the classification or regression problems only if the source and target domains have similarities [43]. A previous study has identified similarities between robot-assisted training and manual stretching rehabilitation training from a clinical aspect [48]. It is thus reasonable to suggest that the prognosis efficacies of these two treatments have similar domains of influential factors, allowing us to apply transfer learning to a broader range of rehabilitation methods.

GradientExplainer can be applied to prognosis models to visualize the key features of stroke survivors who met the criteria of a proportional recovery group after rehabilitation. Analyzing features from biomechanical measurements showed a high correlation coefficient between patient recovery and biomarkers, including PF AROM, PF strength, and DF strength. A previous study showed that ankle DF and PF strength correlate with walking distances [49], supporting our finding that DF and PF strength are essential factors for stroke survivors to have a good recovery. Another study found that AROM is an influential factor in predicting upper extremity function, which also supports our findings to some extent [4], [50]. SHAP interaction values based on the power spectrum density show that the power spectrum density in the Alpha and Beta Low bands correlates with stroke survivors' recovery. A long-term study showed a difference in EEG signals at different post-stroke times, supporting our findings [55]. In our research, we further found that key features in the occipital lobe, central parietal lobe, and parietal lobe are positively correlated with stroke survivors' recovery. SHAP interaction values based on functional connectivities showed the correlation between interhemispheric and intrahemispheric connectivities and patient recovery.

Interhemispheric EEG connectivities between the frontal cortex and central cortex, the frontal cortex and parietal cortex, the frontal cortex and occipital cortex, the occipital cortex and central cortex, and the occipital cortex and parietal cortex showed high correlation coefficients with stroke patients' recovery. Similarly, intrahemispheric EEG connectivities within the frontal, central, and parietal cortex showed high correlation coefficients with stroke patients' recovery. Several functional MRI (fMRI) studies support our SHAP findings with respect to the EEG power spectrum density and functional connectivities [52], [53], [54], [55]. An fMRI study analyzed the longitudinal changes of resting-state functional connectivities in which the stroke survivors have lower connectivities in the contralesional central and occipital cortex than healthy patients, supporting our results [52]. The present study found that high signal activities in the central and occipital cortex correlate with a high probability of proportional recovery. Another study investigating functional recovery through analyzing functional connectivities matched our findings with respect to interhemispheric connections [53]. One study showed that the frontal-parietal cortex is correlated with motor recovery [54]. The results of a brain structural analysis in a robot-assisted rehabilitation training study also match our findings [55]. It is thus reasonable to suggest that by applying deep learning interpretation methods to a sufficiently accurate model, researchers can identify hidden mechanisms and factors influencing patient recovery with relative ease and accuracy.

Differences between the proportional recovery group and the poor recovery group in manual and robot-assisted stretching training showed significant resemblance. The differences between the proportional recovery group and the poor recovery group are virtually the same for FMA-LE scores, PF strength, and DF strength ($P = 0.053$) in both rehabilitation methods. These findings are the same as the GradientExplainer results for biomechanical data. These results indicate that the recovery mechanism of manual stretching training is similar to that of robotic-assisted training, which may contribute to transfer of the proposed prognosis model.

Aside from helping doctors to make better judgments, accurate prognosis results may help boost the confidence of patients. For patients in the proportional recovery group, an accurate forecast of recovery has the potential to boost confidence and stimulate a positive attitude, which was shown to boost motor recovery [56]. However, to date, there is no research that categorically states a strong correlation between prognosis accuracy and patient mood. Further research on this topic can be conducted based on our prognosis model.

V. Conclusion

In summary, the multi-input design (including demographic, clinical, biomechanical, and neurophysiological data), along with the flexible nature of deep learning, allows our prognosis model to be transferable between different scenarios with high accuracy. Successful application of transfer between manual and robot-assisted rehabilitation shows potential in reducing the amount of model required to generate a prognosis for different rehabilitation treatments.

References

[1] C. W. Tsao et al., "Heart disease and stroke statistics-2022 update: A report from the American Heart Association," *Circulation*, vol. 145, no. 8, pp. e153–e639, Feb. 2022.
[2] B. H. Dobkin, "Rehabilitation after stroke," *New England J. Med.*, vol. 352, no. 16, pp. 1677–1684, Jul. 2005.
[3] M. W. O'Dell, C. C. Lin, and V. Harrison, "Stroke rehabilitation: Strategies to enhance motor recovery," *Annu. Rev. Med.*, vol. 60, pp. 55–68, 2009.
[4] J. A. Beebe and C. E. Lang, "Active range of motion predicts upper extremity function 3 months after stroke," *Stroke*, vol. 40, no. 5, pp. 1772–1779, May 2009.
[5] C. Counsell, M. Dennis, M. McDowall, and C. Warlow, "Predicting outcome after acute and subacute stroke development and validation of new prognostic models," *Stroke*, vol. 33, no. 4, pp. 1041–1047, Apr. 2002.
[6] G. M. De Marchis et al., "A novel biomarker-based prognostic score in acute ischemic stroke: The CoRisk score," *Neurology*, vol. 92, no. 13, pp. e1517–e1525, Mar. 2019.
[7] G. Liu et al., "Motor recovery prediction with clinical assessment and local diffusion homogeneity after acute subcortical infarction," *Stroke*, vol. 48, no. 8, pp. 2121–2128, Aug. 2017.
[8] G. Liu et al., "Machine learning for predicting motor improvement after acute subcortical infarction using baseline whole brain volumes," *Neuro Rehabil. Neural Repair*, vol. 36, no. 1, pp. 38–48, Jan. 2022.

[9] R. Squitti et al., "Prognostic value of serum copper for post-stroke clinical recovery: A pilot study," *Front. Neurol.*, vol. 9, May 2018, Art. no. 333.

[10] X. Zhai et al., "Effects of robot-aided rehabilitation on the ankle joint properties and balance function in stroke survivors: A randomized controlled trial," *Front. Neurol.*, vol. 12, Oct. 2021, Art. no. 719305.

[11] Y. L. Kuo et al., "Transcranial magnetic stimulation to assess motor neurophysiology after acute stroke in the United States: Feasibility, lessons learned, and values for future research," *Brain Stimulation*, vol. 15, no. 1, pp. 179–181, Jan./Feb. 2022.

[12] C. Tozlu et al., "Machine learning methods predict individual upper-limb motor impairment following therapy in chronic stroke," *Neuro Rehabil. Neural Repair*, vol. 34, no. 5, pp. 428–439, May 2020.

[13] P. Langhorne, J. Bernhardt, and G. Kwakkel, "Stroke rehabilitation," *Lancet*, vol. 377, pp. 1693–1702, May 2011.

[14] C. M. Stinear, "Prediction of motor recovery after stroke: Advances in biomarkers," *Lancet Neurol.*, vol. 16, no. 10, pp. 826–836, Sep. 2017.

[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[16] A. Bivard, L. Churilov, and M. Parsons, "Artificial intelligence for decision support in acute stroke—current roles and potential," *Nature Rev. Neurol.*, vol. 16, no. 10, pp. 575–585, Oct. 2020.

[17] R. Mane et al., "Prognostic and monitory EEG-biomarkers for BCI upper-limb stroke rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 8, pp. 1654–1664, Aug. 2019.

[18] P. J. Lin et al., "CNN-based prognosis of BCI rehabilitation using EEG from first session BCI training," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1936–1943, 2021.

[19] A. M. Chiarelli et al., "Electroencephalography-derived prognosis of functional recovery in acute stroke through machine learning approaches," *Int. J. Neural Syst.*, vol. 30, no. 12, Dec. 2020, Art. no. 2050067.

[20] S. Zhang et al., "Interpretability analysis of one-year mortality prediction for stroke patients based on deep neural network," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 4, pp. 1903–1910, Apr. 2022.

[21] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 4768–4777.

[22] S. M. Lundberg et al., "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery," *Nature Biomed. Eng.*, vol. 2, no. 10, pp. 749–760, Oct. 2018.

[23] H. Lee et al., "An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets," *Nature Biomed. Eng.*, vol. 3, no. 3, pp. 173–182, Mar. 2019.

[24] S. Jonas et al., "EEG-based outcome prediction after cardiac arrest with convolutional neural networks: Performance and visualization of discriminative features," *Hum. Brain Mapping*, vol. 40, no. 16, pp. 4606–4617, Nov. 2019.

[25] A. R. Fugl-Meyer, L. Jääskö, I. Leyman, S. Olsson, and S. Steglind, "The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance," *Scand. J. Rehabil. Med.*, vol. 7, no. 1, pp. 13–31, 1975.

[26] K. Berg, S. W. Dauphinee, J. I. Williams, and D. Gayton, "Measuring balance in the elderly: Preliminary development of an instrument," *Physiotherapy Canada*, vol. 41, no. 6, pp. 304–311, Jul./Aug. 1989.

[27] C. V. Granger, L. S. Dewis, N. C. Peters, C. C. Sherwood, and J. E. Barrett, "Stroke rehabilitation: Analysis of repeated Barthel index measures," *Arch. Phys. Med. Rehabil.*, vol. 60, no. 1, pp. 14–17, Jan. 1979.

[28] L. Q. Zhang et al., "Intelligent stretching of ankle joints with contracture/spasticity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 10, no. 3, pp. 149–157, Sep. 2002.

[29] F. Gao, Y. Ren, E. J. Roth, R. Harvey, and L. Q. Zhang, "Effects of repeated ankle stretching on calf muscle–tendon and ankle biomechanical properties in stroke survivors," *Clin. Biomech.*, vol. 26, no. 5, pp. 516–522, Jun. 2011.

[30] W. Li, C. Li, Q. Xu, X. Guan, and L. Ji, "Influence of focal vibration over Achilles tendon on the activation of sensorimotor cortex in healthy subjects and subacute stroke patients," *Neuro Rep.*, vol. 30, no. 16, pp. 1081–1086, Nov. 2019.

[31] W. Li et al., "Study of the activation in sensorimotor cortex and topological properties of functional brain network following focal vibration on healthy subjects and subacute stroke patients: An EEG study," *Brain Res.*, vol. 1722, Nov. 2019, Art. no. 146338.

[32] J. M. Cassidy et al., "Low-frequency oscillations are a biomarker of injury and recovery after stroke," *Stroke*, vol. 51, no. 5, pp. 1442–1450, May 2020.

[33] N. Fei, Y. Gao, Z. Lu, and T. Xiang, "Z-score normalization, hubness, and few-shot learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 142–151.

[34] E. R. Buch et al., "Predicting motor improvement after stroke with clinical assessment and diffusion tensor imaging," *Neurology*, vol. 86, no. 20, pp. 1924–1925, Apr. 2016.

[35] C. Winters, E. E. van Wegen, A. Daffertshofer, and G. Kwakkel, "Generalizability of the proportional recovery model for the upper extremity after an ischemic stroke," *Neuro Rehabil. Neural Repair*, vol. 29, no. 7, pp. 614–622, Aug. 2015.

[36] M. C. Smith, W. D. Byblow, P. A. Barber, and C. M. Stinear, "Proportional recovery from lower limb motor impairment after stroke," *Stroke*, vol. 48, no. 5, pp. 1400–1403, May 2017.

[37] W. D. Byblow, C. M. Stinear, P. A. Barber, M. A. Petoe, and S. J. Ackerley, "Proportional recovery after stroke depends on corticomotor integrity," *Ann. Neurol.*, vol. 78, no. 6, pp. 848–859, Dec. 2015.

[38] J. M. Veerbeek, G. Kwakkel, E. E. H. van Wegan, J. C. F. Ket, and M. W. Heymans, "Early prediction of outcome of activities of daily living after stroke: A systematic review," *Stroke*, vol. 42, no. 5, pp. 1482–1488, May 2011.

[39] T. Dozat, "Incorporating nesterov momentum into adam," in *Proc. Workshop Track (ICLR)*, 2016, pp. 1–4.

[40] J. Lever, M. Krzywinski, and N. Altman, "Model selection and overfitting," *Nature Methods*, vol. 13, no. 9, pp. 703–704, Sep. 2016.

[41] M. A. Salehinejad et al., "Cognitive functions and underlying parameters of human brain physiology are associated with chronotype," *Nature Commun.*, vol. 12, no. 1, pp. 1–19, Aug. 2021.

[42] B. C. Ahn et al., "Clinical decision support algorithm based on machine learning to assess the clinical response to anti–programmed death-1 therapy in patients with non–small-cell lung cancer," *Eur. J. Cancer*, vol. 153, pp. 179–189, Aug. 2021.

[43] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[44] A. Biasiucci et al., "Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke," *Nature Commun.*, vol. 9, no. 1, pp. 1–13, Jun. 2018.

[45] Y. Roy et al., "Deep learning-based electroencephalography analysis: A systematic review," *J. Neural Eng.*, vol. 16, no. 5, Aug. 2019, Art. no. 051001.

[46] M. Ochal, M. Patacchiola, A. Storkey, J. Vazquez, and S. Wang, "Few-shot learning with class imbalance," 2021, *arXiv:2101.02523*.

[47] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," *Adv. Neural Inf. Process. Syst.*, vol. 27, pp. 3320–3328, 2014.

[48] H. Rodgers et al., "Robot assisted training for the upper limb after stroke (RATULS): A multicentre randomised controlled trial," *Lancet*, vol. 394, no. 10192, pp. 51–62, Jul. 2019.

[49] S. S. Ng and C. W. Hui-Chan, "Contribution of ankle dorsiflexor strength to walking endurance in people with spastic hemiplegia after stroke," *Arch. Phys. Med. Rehabil.*, vol. 93, no. 6, pp. 1046–1051, Jun. 2012.

[50] C. M. Stinear et al., "PREP2: A biomarker-based algorithm for predicting upper limb function after stroke," *Ann. Clin. Transl. Neurol.*, vol. 4, no. 11, pp. 811–820, Nov. 2017.

[51] B. Carla et al., "Quantitative EEG and functional outcome following acute ischemic stroke," *Clin. Neurophysiol.*, vol. 129, no. 8, pp. 1680–1687, Jan. 2018.

[52] C. H. Park et al., "Longitudinal changes of resting-state functional connectivities during motor recovery after stroke," *Stroke*, vol. 42, no. 5, pp. 1357–1362, May 2011.

[53] C. Grefkes and G. R. Fink, "Connectivity-based approaches in stroke and recovery of function," *Lancet Neurol.*, vol. 13, no. 2, pp. 206–216, Feb. 2014.

[54] A. Baldassarre et al., "Dissociated functional connectivities profiles for motor and attention deficits in acute right-hemisphere stroke," *Brain*, vol. 139, no. 7, pp. 2024–2038, Jul. 2016.

[55] E. Formaggio, S. Masiero, A. Bosco, F. Izzi, F. Piccione, and A. Del Felice, "Quantitative EEG evaluation during robot-assisted foot movement," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 9, pp. 1633–1640, Sep. 2017.

[56] G. S. Seale, I.-M. Berges, K. J. Ottenbacher, and V. G. Ostir, "Change in positive emotion and recovery of functional status following stroke," *Rehabil. Psychol.*, vol. 55, no. 1, pp. 33–39, Feb. 2010.