Enhanced Deep Transfer Learning Model based on Spatial-Temporal driven Scalograms for Precise Decoding of Motor Intent in Stroke Survivors

Oluwarotimi Williams Samuel*, *IEEE Senior Member*, Mojisola G. Asogbon, Frank Kulwa, Alireza Rezaie Zangene, Tolulope T. Oyemakinde, Tobore Igbe, Alistair A. McEwan, Yongcheng Li, Guanglin Li*, *IEEE Senior Member*

***Abstract*— Motor function loss greatly impacts post-stroke survivors while performing activities of daily living. In the recent years, intelligent rehabilitation robotics have been proposed to enable the patients recover their lost limb functions. Besides, a large proportion of these robots function in passive mode that only allow users to navigate trajectories that rarely align with their limb movement intent, thus precluding full functional recovery. A potential solution would be to explore utilizing an efficient Transfer Learning based Convolutional Neural Network (TL-CNN) to decode multiple classes of post-stroke patients' motion intentions towards realizing dexterously active robotic training during rehabilitation. In this regard, we propose and examined for the first time, the use of Spatial-Temporal Descriptor based Continuous Wavelet Transform (STD-CWT) as input to TL-CNN to optimally decode limb movement intent patterns of stroke patients to provide adequate input for active motor training in rehabilitation robots. Importantly, we examined the proposed (STD-CWT) method on three distinct wavelets including the *Morse*, *Amor*, and *Bump*, and compared their decoding outcomes with those of the commonly adopted CWT technique under similar experimental conditions. Our method was validated using electromyogram signals of five stroke survivors who performed up to twenty-two distinct limb motions. The obtained results showed that the proposed technique recorded a significantly higher decoding ($p<0.05$) and converges faster compared to the commonly adopted method. The proposed method equally recorded obvious class separability for individual movement classes across the stroke patients. Findings from this study suggest that the STD-CWT Scalograms would provide potential inputs for robust decoding of motor intent that may facilitate intuitively active motor training in stroke rehabilitation robots.**

***Clinical Relevance*— The study demonstrated the potential of Spatial Temporal based Scalograms in aiding precise and robust decoding of multi-class motor intents upon which dexterously active rehabilitation robotic training for full motor function restoration could be realized.**

I. INTRODUCTION

The human upper limb is an integral part of the body that aids execution of various daily life arm-related tasks. However, post-stroke survivors are often hindered from full usage of their arms during activities of daily living due to motor function loss [1-2]. Stroke is ranked among the leading causes of mortality in America (USA) and Canada, killing over one hundred thousand persons per year with more than two million survivors in China [3], reinforcing the need for proper stroke preventive and rehabilitation technology [4]. Toward restoring the lost motor function in stroke patients, traditional physiotherapy and advanced rehabilitation robotic based approaches have been mostly considered. Off these approaches, the rehabilitation robotic strategy is reported to have more clinical efficacy, requiring lesser resources to administer qualitative rehabilitation exercises to a large number of post-stroke patients per time.

Towards advancing stroke rehabilitation technology, adaptive robotic training strategies that could identify and respond to patients' motor intents have been proposed [5-7]. Thus, pattern recognition (PR) of myoelectric signals (that houses motor information) have been explored to provide active and accurate robotic training for stroke patients [6, 9-10]. In this direction, various machine learning-driven methods have been proposed for deciphering upper limb movement intention of stroke patients upon which active rehabilitation robotic training could be administered. For instance a temporal-driven feature set that incorporates myoelectric signals' amplitude for decoding motion intention in stroke survivors was proposed [11]. In another study, time-domain based descriptors extracted from sEMG were utilized to predict normal or abnormal muscle activation patterns to aid provision of information for requisite muscle recruitment in stroke rehabilitation robots [12]. Furthermore, a recent study proposed a machine learning models that integrated traditional hand crafted feature that is robust to White Gaussian Noise for the task of decoding multiple classes of limb movement intentions of stroke-survivors.

Importantly, the above studies and other related works have made meaningful contributions towards advancing research and development in the field of stroke rehabilitation robotics. However, these studies are limited because they mostly employ temporal information without taking the spatial neuromuscular characteristics (that are equally essential for motor intent decoding especially in severely impaired patients) into consideration. Additionally, they only focus on utilizing the traditional machine learning approaches that would normally require engineered features. And such hand crafted features often require expert input to construct and

The research work was supported in part by the National Natural Science Foundation of China under grants (#82161160341, #82050410452, #62150410439) and Shenzhen Governmental Collaborative Innovation Program (#SGLH20180625142402055).

O.W. Samuel, M.G. Asogbon, F. Kulwa, T.T. Oyemankinde, Y. Li, and G. Li are with the CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), Shenzhen, Guangdong 518055, China.

O.W. Samuel and A.A. McEwan are with School of Computing and Engineering, University of Derby, Derby, DE22 3AW, United Kingdom.

F. Kulwa and T.T. Oyemankinde are also with the Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, Guangdong 518055, China.

T. Igbe is with the Center for Diabetes Technology, School of Medicine, University of Virginia University, Virginia, VA22903, United States of America

(Corresponding authors: Oluwarotimi Williams Samuel and Guanglin Li, Email: o.samuel@derby.ac.uk and gl.li@siat.ac.cn).

may not be robust for characterizing multiple patterns of motor tasks associated with the paralyzed limbs of severely impaired patients. To resolve these limitations, deep neural networks that automatically learns and extract high/low level features (that houses temporal and spatial information) for adequate characterization of EMG signals have been proposed [13-15]. Further, building and training a deep learning model such as the deep convolutional neural network (CNN) from scratch is computationally complex and would require huge amount of training data, which may not be readily available particularly in the case of stroke patients' data. Thus, exploring benchmark deep CNNs that have been built on large data sets for similar tasks would be a viable option. Technically, this concept is referred to as Transfer Learning (TL) and the resulting networks are termed TL-CNNs. Typically, two time-frequency data representation techniques including Short-Time Fourier Transforms (STFT yield ***Spectrogram*** images) and Continuous Wavelet Transform (CWT yield ***Scalogram*** images) [16], have been broadly utilized as inputs to TL-CNNs for EMG signal pattern characterization. It is worth mentioning that CWT-based Scalogram integrates multiple scales which allows them to reach the optimal time-frequency resolution, aligning well with the Heisenberg uncertainty concept [17-18]. Unlike Spectrograms that offer only fixed time-frequency resolution, Scalograms support variable resolution in both space, offering better representation of the EMG signal characteristics with less computational requirement that could facilitate practical deployment. Furthermore, constructing spatio-temporal descriptors from the EMG signals prior to generating Scalogram images via CWT technique would help preserve requisite motor information for characterizing multiple classes of limb movements especially for severely impaired stroke survivors, which is essential for imitating active motor training in stroke rehabilitation robotic systems.

Therefore, for the first time, this study proposed and examined the use of spatio-temporal based Scalograms as inputs to TL-CNNs (GoogleNet deep transfer learning model) to characterize multiple patterns of targeted limb motions of stroke patients. With the aid of a 56 electrode-channel sEMG recordings obtained from post-stroke survivors, the decoding accuracy and consistency of the proposed method was investigated compared to conventional methods using various evaluation metrics.

II. METHODS

A. Subjects Information

A total of five stroke patients were employed to participate in the study. Before that, their level of impairment was assessed using the popular Fugl-Meyer scale by a physiotherapist. And they were found to be in stages 4-5 the Brunnstrom scale and got scores in the range of 35-61 for Fugl-Meyer scale. The patients agreed to the study's goal and gave informed consent as well as permission to publish their data for scientific benefits.

B. Data collection and preprocessing

To record the data, a high density myoelectric signal recording system was used. 56 monopolar surface electrodes with each having a diameter of 5mm was fixed on the arm muscles of the stroke patients (Figure 1). After carefully examining the limb status of the patients, we placed 48 electrodes on the arm muscles in a grid of 8×6 such that it is proximal by 1cm to elbow and distal to joint of the wrist Besides, an inter-electrode distance of 2mm was maintained throughout. And the remaining eight electrodes are fixed on the first dorsal muscles while a reference electrode is placed on the wrist.

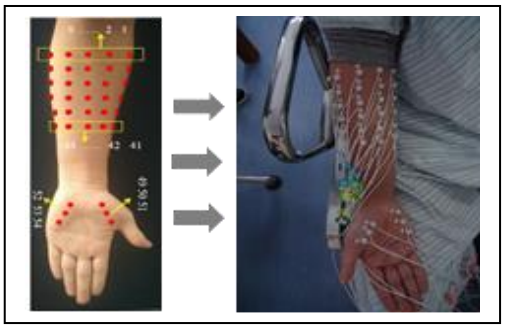


Fig. 1. Electrode configuration adopted for EMG signal conllection.

Having successfully placed the sensors on the arm of the subject, they were told to observe about twenty two different upper limb motions. And motion class lasted up to six seconds while a moderate contraction force was utilized. In addition, each motion was repeated six times per trial while a rest period of about eight seconds was allowed between consecutive trials each time. Then, the myoelectric signals were obtained with a 1024 Hz sampling frequency. The data was preprocessed using 50Hz power-line attenuation filter and a band pass filter with frequencies of 10 to 500Hz utilized to reduce the impact artifacts on the data. Then, the data was segmented via overlapping window of length 150ms and increment of 100ms in preparation for the spatio-temporal descriptor construction described in section II.C.

C. Construction of the *Spatial-Temporal Based Scalogram*

To construct a highly informative Scalogram inputs to the TL-CNN, that would properly characterize limb movement intent, an integrated framework was utilized. In the first stage of the framework, a spatio-temporal descriptor matrix is constructed directly from the preprocessed HD-sEMG signal as follows: The energy information designated as weighted integral square (*WISx*) is obtained from the windowed signals (1).

$$WISx = \sum_{i=0}^{n-1} x[i]^2 * wfi \qquad (1)$$

Here, *x* represent the windowed signal per time, *n* denote the signal length, and wfi is the weighted factor (which normalizes the descriptor). Further, closely correlated spectral moments from which spatial neuromuscular information are extracted for each movement pattern were obtained based on the differential derivatives of 1st (f') and 2nd (f'') order of the windowed signal per time, the corresponding normalized root squared coefficients (NormRSC) was obtained as in equations (2) and (3).

$$NormRSCx_1 = \frac{1}{\emptyset}\sum_{i=0}^{n-1} f^{'} x_1[i]^2 \quad (2)$$

$$NormRSCx_2 = \frac{1}{\emptyset}\sum_{i=0}^{n-1} f^{''} x_2[i]^2 \quad (3)$$

Besides, the optimal value of $\emptyset$ is obtained by investigating a set of values in the range of [-1 to +1] using a step size of 0.01. Moreover, $\emptyset$ represent the standardization factor. Further, the variations among $WISx$, $NormRSCx_1$ and $NormRSCx_2$ are obtained. Further, a variant of the non-linear log-detector (*NL_LogKernek*) employed in [14], was computed to estimate the corresponding contraction force of the muscles for each class of motion (4).

$$NL\_LogKernel = Reg(e^{\frac{1}{n}\sum_{i=0}^{n-1}\log(x[i])}) \quad (4)$$

And Reg represents the regularization form of the non-linear log detector. Furthermore, the average of the root squared (ARS) feature was computed to capture the full-wave rectification that reflects the full muscular activities in the given windowed signals. Finally, the above descriptors are combined into a feature vector that represent the spatio-temporal descriptor (p) which serves as input to the CWT module upon which the STD-Scalogram is generated, described as follows.

The time-frequency representation (transformation) of the spatio-temporal descriptor (p) is obtained by applying the CWT technique as shown in equation (5).

$$C(b,\tau) = \int \frac{1}{\sqrt{b}}\psi\left(\frac{t-\tau}{b}\right)p(t)dt \quad (5)$$

Where C denote the transformation, b represent the scaling factor, τ represent the time for translation. Besides, Ψ represent the employed mother wavelet function and $p(t)$ is the input signal at specific time intervals *t*. It is worth mentioning that three distinct wavelet functions were considered in generating the STD-Scalogram images. Notably, the wavelet is contracted (b<1), this leads to high spectral resolution and when the wavelet is enlarged (b>1) this results in high temporal resolution. The former scenario depicts the transient events while the latter case reflects frequencies in the steady-state.

$$Eng = \int_{-\infty}^{\infty} |\psi(t)|^2 dt \quad (6)$$

$$D = \int_{-\infty}^{\infty} \psi(t)dt = 0 \quad (7)$$

Where Eng represents the energy of the wavelet and Ψ depicts the wavelet mother function. Thus, this leads to the following dual properties of the wavelet associated with equation (5) which are the finite energy denoted in equation (6) and its admissibility depicted in equation (7).

$$S(b,\tau) = |C(b,\tau)|^2 \quad (8)$$

Furthermore, we obtain the graphical representation of the correlation that exist between the transformed signals and the wavelets scaled over time to obtain the Scalogram via equation (8) which serves as input to the TL-CNN.

D. Performance evaluation of the spatial-temporal descriptor

The performance of the proposed method was validated in comparison with notable existing approaches in terms of accuracy and robustness for characterizing multiple patterns of limb movements of the post-stroke survivors. As mentioned in section II.C, the Scalogram images were generated based on the conventional CWT technique and the proposed STD method based on the three different wavelet functions which includes: 'Morse', 'Amore', and 'Bump'. In order words, we investigated a total of six variants of Scalograms construction techniques: three based on the conventional CWT technique ($CWT_{morse,}$ $CWT_{amor,}$ and CWT_{bumb}) and three based on the proposed STD method ($STD\_CWT_{morse,}$ $STD\_CWT_{amor,}$ and STD_CWT_{bumb}). Further, the movement-intent of the post-stroke survivors were decoded by the TL-CNN built based on each of the above six variants of Scalograms while their performances were assessed using classification accuracy (CA) metric shown in equation 9.

$$CA = \frac{No.\ of\ Correctly\ Classified\ Samples}{Total\ No.of\ Samples} * 100\% \quad (9)$$

Also, the means of the proposed method were compared with those of the other methods with paired t-test statistical analysis technique at *p-value* of $p<0.05$ as the significance level. Meanwhile, the TL-CNN comprise of a total of 144 layers and requires RGB version of the constructed Scalogram images of size 224-by-224-by-3 as input. The earlier layers attempt to identify low-level features while the latter (dipper) layers attempt to extract high-level/specific features from Scalogram images to ensure proper decoding of the inherent motor task. A dropout layer that randomly configures the inputs to zero was introduced to avoid overfitting. Afterward, the network was trained with 80% of the data while the remaining 20% was used for testing; MiniBatchSize: 20; MaxEpoch: 10; LearningRate: 0.0001; and loss function: gradient descent algorithm.

III. Results and Discussion

A. The TL-CNN's limb Movement-intent Decoding Outcomes

In this section, we first present the results obtained after training the TL-CNN using Scalograms obtained based on the proposed method ($STD\_CWT_{morse,}$ $STD\_CWT_{amor,}$ and STD_CWT_{bumb}) and those obtained based on the conventional CWT technique ($CWT_{morse,}$ $CWT_{amor,}$ and CWT_{bumb}), and the results are shown in Fig. 2. This results mainly depicts the extent to which the proposed method enabled the *TL-CNN* model learn motor information which is essential for the classification task.

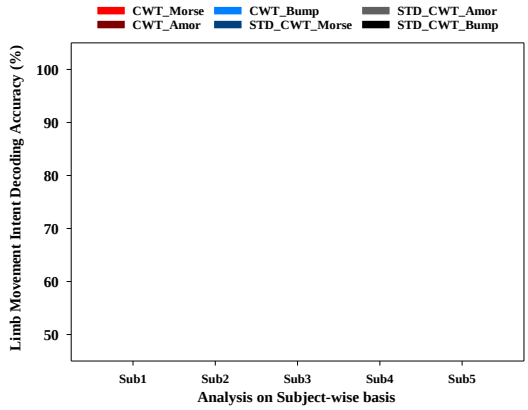


Fig. 2: The TL-CNN model training results for the proposed and existing methods with the three distinct wavelets (Morse, Amor, and Bulp)

From the results in Fig. 2, it can be clearly seen that the three variants of the proposed method ($STD\_CWT_{morse,}$ $STD\_CWT_{amor,}$ and STD_CWT_{bumb}) enabled the TL-CNN model to achieve significantly higher decoding accuracies for all the stroke-survivors in comparison to the commonly applied method ($CWT_{morse,}$ $CWT_{amor,}$ and CWT_{bumb}). Overall, the proposed method recorded the best training accuracy across subjects when the STD_CWT_{amor} was employed in building the TL-CNN model, and similar trend could be observed for the conventional method when the CWT_{amor} was employed. This suggest that the wavelet type of *amor* may lead to better learning of important features for both the proposed and conventional CWT approach for Scalogram construction. In addition, classification/prediction results obtained after training the TL_CNN model using each of the above constructed Scalogram as input is shown in Fig. 3.

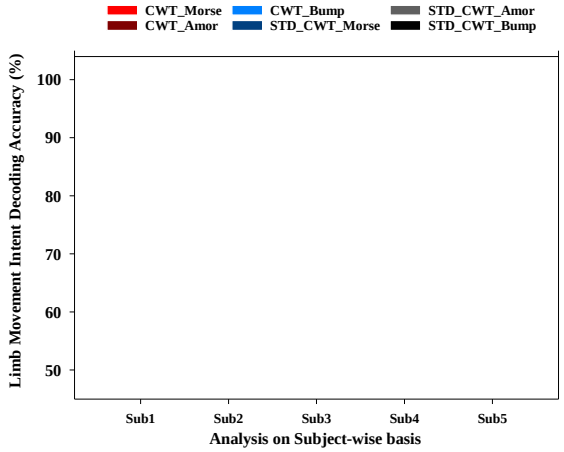


Fig. 3: The TL-CNN model testing results for the proposed and existing methods with the three distinct wavelets (Morse, Amor, and Bulp).

Again, it can be observed that the proposed method ($STD\_CWT_{morse,}$ $STD\_CWT_{amor,}$ and STD_CWT_{bumb}) led to substantially higher decoding accuracies (at $p<0.05$) across subjects compared to the existing CWT approach ($CWT_{morse,}$ $CWT_{amor,}$ and CWT_{bumb}). Figure 3 suggests the potential of adopting this method in practical settings.

Furthermore, we present the mean classification results for both the trained and tested TL_CNN based on the proposed method and the existing method as shown in Table I.

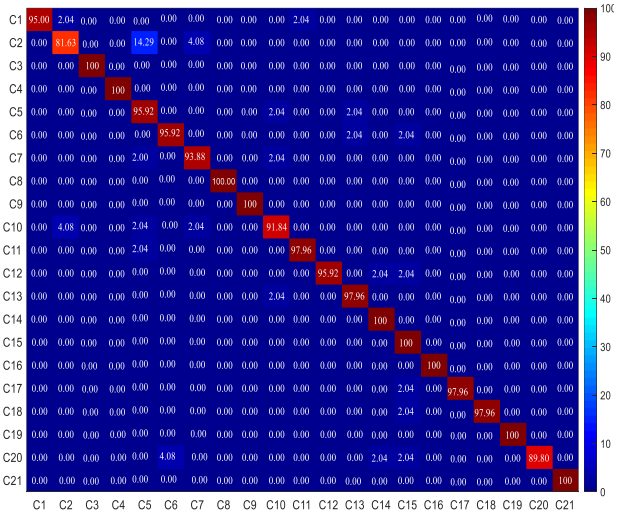
Table I: Average Decoding Performance of the TL_CNN model across subjects based on the proposed and the existing method.

	TL-CNN's Training Result		TL-CNN's Test Result	
	ACC (%)	**STD (%)**	**ACC (%)**	**STD (%)**
CWT_Morse	79.00	15.17	77.91	10.56
STD_CWT_Morse	**94.00**	**6.52**	**94.28**	**5.53**
CWT_Amor	78.00	10.37	77.62	10.68
STD_CWT_Amor	**100.00**	**0.00**	**95.07**	**4.67**
CWT_Bump	76.00	13.42	79.95	8.56
STD_CWT_Bump	**98.00**	**2.74**	**94.34**	**10.62**

B. Analyzing TL-CNN Model for Individual Movement Decoding

Given the relatively high number of upper limb movements involved in the experiments, it is important to investigate the extent to which the proposed method can characterize all classes of movement tasks compared to existing methods. Therefore, two confusion matrices were computed by employing data of a representative subject. Then the proposed method and existing method were applied to classify the inherent motor tasks with the obtained results shown in Fig. 4.

(A)



(B)

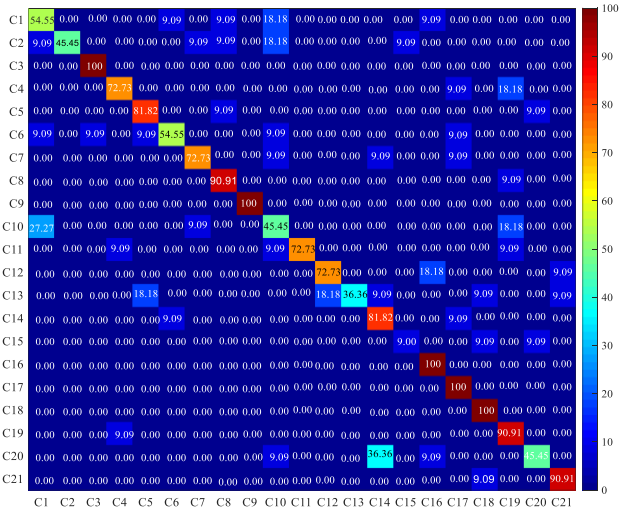


Fig. 4. Decoding of Individual limb movement with the proposed method (A) and conventional method (B) for the 21 classes of motion. It should be noted that only Amor wavelet based techniques (**A:** STD_CWT_Amor for the proposed method and **B:** CWT_Amor for the conventional method) were considered due page limitation.

By analyzing the entries in the diagonal of the confusion matrices (class-wise decoding outcomes) in Fig. 4, it can be observed that the proposed method (*STD_CWT_Amor* in Fig. 4a) achieved significantly higher decoding results for most of the 21 classes of motion in comparison to the commonly applied CWT method (*CWT_Amor* in Fig. 5b). Thus, this results further corroborate the potential of the proposed method in enabling the TL-CNN model to achieve precise and robust decoding of post-stroke patients' limb movement intent which is essential for intuitive control of rehabilitation robots.

IV. Conclusion

The need for intelligent and adaptive rehabilitation robots for restoration of motor dysfunction stroke survivors is on the rise. A core characteristic of such rehabilitation robotic technology is their ability to adequately decode limb movement intents of stroke survivors from bio-signals. And such decoded movements often serve as control input to enable active motor training that are beneficial for motor function restoration. Besides, a core drawback is that existing transfer learning based deep neural network model employed in such stroke rehabilitation robots are unable to guarantee adequate limb motion decoding especially for severely impaired stroke patients, thus limiting the clinical efficacy of the existing robots [19-20]. To resolve this issue, we propose and investigated the use of spatio-temporal based Scalograms as inputs to TL-CNNs (GoogleNet deep transfer learning model) to characterize multiple patterns of targeted limb motions of post-stroke patients. Furthermore, we implemented the proposed method and then compared its performance with that of conventional method that only employ CWT technique in constructing the Scalogram inputs to the network, using sEMG dataset obtained from post stroke patients who performed up to 22 classes of upper limb motor tasks. In terms of decoding accuracies, our experiment results showed that the proposed method could consistently aid accurate and robust decoding of motor tasks inherent in the sEMG recordings, and it demonstrated significantly higher accuracy compared to conventional method. Findings from this study suggest that the spatio-temporal based Scalograms would provide potential inputs for robust decoding of motor intent which may facilitate intuitively active motor training in stroke rehabilitation robotic systems and related robotic systems.

References

[1] Trujillo, P., Alfonso, M., Alessandro, S., et al. "Quantitative EEG for Predicting Upper Limb Motor Recovery in Chronic Stroke Robot-Assisted Rehabilitation," IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25(7):1058-1067, 2017.

[2] Li, X. et al. "A motion-classification strategy based on sEMG-EEG signal combination for upper-limb amputees," Journal of NeuroEngineering & Rehabilitation, 14(1), pp.1-13, 2017.

[3] Feigin, V. L., et al. "Global burden of stroke and risk factors in 188 countries, during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013," The Lancet Neurology, 15(9), 913-924, 2016.

[4] Béjot, Y., Bailly, H., Durier, J. and Giroud, M., "Epidemiology of stroke in Europe and trends for the 21st century," La Presse Médicale, 45(12):e391-e398, 2016.

[5] Blank, A. A., French, J. A., Pehlivan, A. U., & O'Malley, M. K. "Current trends in robot-assisted upper-limb stroke rehabilitation: promoting patient engagement in therapy," Current physical medicine and rehabilitation reports, 2(3), 184-195, 2014.

[6] Krebs, H. I., Volpe, B. T., Williams, D., et al. "Robot-aided neurorehabilitation: a robot for wrist rehabilitation," IEEE Transactions on Neural Systems and Rehabilitation Engineering, 15(3):327-335, 2007.

[7] Lu, Z., Tong, K. Y., Zhang, X., Li, S., & Zhou, P. "Myoelectric pattern recognition for controlling a robotic hand: a feasibility study in stroke," IEEE Transactions on Biomedical Engineering, 66(2), 365-372, 2018.

[8] Samuel, O.W., et al., "Resolving the adverse impact of mobility on myoelectric pattern recognition in upper-limb multifunctional prostheses, Computers in Biology & Medicine, 90, pp. 76-87, 2017.

[9] Samuel, O.W. et al. "Examining the effect of subjects' mobility on upper-limb motion identification based on EMG-pattern recognition," In IEEE Intelligent Robot Systems (ACIRS), Asia-Pacific Conference, 137-14, 2016.

[10] Hu, X. L., Tong, K. Y., Wei, X. J., et al. "The effects of post-stroke upper-limb training with an electromyography (EMG)-driven hand robot," Journal of Electromyography and Kinesiology, 23(5):1065-1074, 2013.

[11] Song, R., Tong, K.Y., Hu, X. and Zhou, W., "Myoelectrically controlled wrist robot for stroke rehabilitation," Journal of Neuroengineering and Rehabilitation, 10(1):52, 2013.

[12] Cesqui, B., Tropea, P., Micera, S. and Krebs, H. I., "EMG-based pattern recognition approach in post stroke robot-aided rehabilitation: a feasibility study," Journal of NeuroEngineering and Rehabilitation, 10(1):75, 2013.

[13] Ameri, A., Akhaee, M. A., Scheme, E., & Englehart, K. (2019). A deep transfer learning approach to reducing the effect of electrode shift in EMG pattern recognition-based control. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(2), 370-379.

[14] Ma, C., Lin, C., Samuel, O. W., Xu, L., & Li, G. (2020). Continuous estimation of upper limb joint angle from sEMG signals based on SCA-LSTM deep learning approach. Biomedical Signal Processing and Control, 61, 102024.

[15] [15] Ma, C., Lin, C., Samuel, O. W., Guo, W., Zhang, H., Greenwald, S., ... & Li, G. (2021). A bi-directional LSTM network for estimating continuous upper limb movement from surface electromyography. IEEE Robotics and Automation Letters, 6(4), 7217-7224.

[16] Oh, D. C., & Jo, Y. U. (2021). Classification of hand gestures based on multi-channel EMG by scale Average wavelet transform and convolutional neural network. International Journal of Control, Automation and Systems, 19(3), 1443-1450.

[17] Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. IEEE transactions on information theory, 36(5), 961-1005.

[18] Ren, Z., Qian, K., Zhang, Z., Pandit, V., Baird, A., & Schuller, B. (2018). Deep scalogram representations for acoustic scene classification. IEEE/CAA Journal of Automatica Sinica, 5(3), 662-669.

[19] Li, L., Fu, Q., Tyson, S., Preston, N., & Weightman, A. (2022). A scoping review of design requirements for a home-based upper limb rehabilitation robot for stroke. Topics in Stroke Rehabilitation, 29(6), 449-463.

[20] Asogbon, M.G., et al. "Towards resolving the co-existing impacts of multiple dynamic factors on the performance of EMG-pattern recognition based prostheses," Computer Methods and Programs in Biomedicine, 184, 105278, 2020.

.