2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Classifying Unimanual and Bimanual Upper Extremity Tasks in Individuals Post-Stroke

Aaron Miller[1], *Student Member, IEEE* and Eric Wade[2], *Member, IEEE*

***Abstract*— After stroke, many individuals develop impairments that lead to compensatory motions. Compensation allows individuals to achieve tasks but has long-term detrimental effects and represents maladaptive motor strategies. Increased use of bimanual motions may serve as a biomarker for recovery (and the reduction of reliance on compensatory motion), and tracking such motion using sensor data may provide critical data for health care specialists. However, past work by the authors demonstrated individual variation in motor strategies results in noisy and chaotic sensor data. The goal of the current work is to develop classifiers capable of differentiating unimanual, bimanaual asymmetric, and bimanual symmetric gestures using wearable sensor data. Twenty participants post-stroke (and 20 age-matched controls) performed a set of tasks under the supervision of a trained occupational therapist. Sensor data were recorded for each task. Classifiers were developed using artificial neural networks (ANNs) as a baseline, and the echo state neural network (ESNN) which has demonstrated efficacy with chaotic data. We find that, for control and post-stroke participants, the ESNN results in improved testing accuracy performance (91.3% and 80.3%, respectively). These results suggest a novel method for classifying gestures in individuals post-stroke, and the developed classifiers may facilitate longitudinal monitoring and correction of compensatory motion.**

I. INTRODUCTION

Every year, almost 800,000 people in the US suffer a stroke, and stroke is now the leading cause of serious long-term disability in the US with as many as three-quarters of survivors reporting impairment in some studies [1]. Hemiparesis (one side of the body presenting with weakness or paralysis) is one such impairment. Stroke survivors with hemiparesis often employ movement strategies that are different than those that were used before the onset of stroke known as compensatory strategies. Common compensatory strategies include non-use of the affected limb or atypical movement or coupling of the limb(s) to compensate for weakness, rigidity, and limited motor control [2]–[4]. For patients with severe impairment, compensation may be ideal; however, long term use of these compensatory behaviors has been shown to decrease quality of life and life span, and to reduce capability and cortical representation of the more affected limb [5], [6].

Stroke survivors are developing, performing, and reinforcing compensatory behaviors in the ambient setting while they are unmonitored by a healthcare specialist [6]. Because of this, enthusiasm has increased for remote assessment and rehabilitation [7]–[9]. Although compensatory strategies are well defined, the specific behaviors that individuals employ can be more difficult to characterize and quantify. Given the task being performed, it is also difficult to autonomously determine whether a movement strategy is compensatory because some behaviors are appropriate while performing a certain type of task but inappropriate while performing others [10]. Thus, we believe the first step in autonomous treatment of compensatory strategies is the capability to distinguish between the common categories (or types) of tasks that people post-stroke will often attempt to perform in their daily life.

Miller et al. [11] investigated this phenomenon using a selected set of upper extremity (UE) tasks. They demonstrated that wearable sensor data are sensitive to differences in task types as well as the differences in task performance by post-stroke and control groups. Miller et al. [10] also showed that a select set of features extracted from overlapping data windows can be used to train a supervised machine learning model to distinguish between task types. However, given the complexity and variability of UE motion, achieving high accuracy required classification techniques trained and evaluated on the same datasets (i.e., user–dependent models). The goal of the current work is to determine if other machine learning approaches are capable of creating a generalized, universally applicable (i.e., user–independent) model capable of distinguishing between task types for post-stroke and control individuals using windowed motion data.

Machine learning methods have been used in many applications for people post-stroke. Health outcomes [12], therapy stage [13], and Fugl-Meyer Assessment (FMA) scores [14] have been predicted using supervised machine learning techniques (random forest, linear regression, and Support Vectors Machines (SVM)) and deep neural networks. Hand gestures have been classified using supervised techniques such as linear discriminant analysis (LDA), SVM, and k-Nearest Neighbors (KNN) [15], [16], and gait and walking conditions have been characterized by SVM and neural networks [17], [18]. To date, machine learning methods have not been extensively used to classify upper extremity (UE) gestures. This is likely because time series data that present with seasonal cycles combined with linear trends or other randomness, such as UE motion data, often appear chaotic and can be difficult to classify using a time series model. However, the echo state neural network (ESNN) is a recurrent neural network that uses a loosely connected hidden layer (known as a reservoir) and works well with chaotic data [19], [20]. To investigate the utility of this approach with UE data we use a multilayer perceptron (MLP) artificial neural network (ANN) to

[1] Department of Mechanical, Aerospace, and Biomedical Engineering, University of Tennessee, Knoxville, TN, USA. [2] Department of Mechanical Engineering, California State Polytechnic University, San Luis Obispo, CA, USA.

establish baseline accuracy. We also investigate the performance of neural network–based task classification, relative to the previously employed machine learning methods. We then test an ESNN to evaluate model accuracy given complex, highly variable, windowed time-series data from post-stroke and control groups performing UE tasks.

II. METHODS

A. Design

Individuals post stroke and healthy-age matched controls were recruited from two clinical sites: Columbia University Irving Medical Center/Teachers College, Columbia University and Chapman University. In this cross-sectional study, participants attended a single session where they were asked to perform a series of tasks, while both seated and standing, that are representative of the types of tasks that an individual would be performing in an ambient setting. These included: $\mathbf{U_L}$, unimanual tasks performed with the dominant (control) or less-affected (post-stroke) limb (e.g., reaching for and picking up a spoon or a bottle); $\mathbf{U_M}$, unimanual tasks performed with the non-dominant or more affected limb; $\mathbf{B_L}$, bimanual asymmetric tasks where both limbs were expected to be engaged with differing intensity and the dominant or less affected limb is more actively being used (e.g., stirring a bowl or unscrewing the lid of a bottle); $\mathbf{B_M}$, bimanual asymmetric tasks where the non-dominant or more affected limb is more actively being used; and $\mathbf{B_S}$, bimanual symmetric tasks where both limbs were expected to be engaged an equal amount (e.g., donning a hat or folding a towel). This study was approved by the Institutional Review Boards at Columbia University Irving Medical Center, Teachers College, Columbia University and Chapman University. All participants provided written informed consent.

B. Procedure

Participants wore five APDM Opal inertial measurement units (IMUs) on the sternum, both wrists, and both upper arms during task performance. Each sensor recorded three sensor modalities (accelerometer, angular rate of change, and magnetic field strength) on three axes (x, y, and z) at 128 Hz (www.apdm.com). Participants were also filmed to obtain ground-truth task performance times. These data were filtered using a Butterworth bandpass filter with cutoff frequencies of 0.1 Hz–2 Hz (this filter was determined to be appropriate for signal power of volitional movement common to activities of daily living, (ADLs) [21]). Finally, data were detrended to remove effects of drift. Following task completion, video data were reviewed by a clinician who manually determined start and stop times for each task resulting in a dataset consisting of twelve tasks (2 $\mathbf{U_L}$, 2 $\mathbf{U_M}$, 2 $\mathbf{B_L}$, 2 $\mathbf{B_M}$, 4 $\mathbf{B_S}$), 40 participants (20 post-stroke and 20 control), and 45 signals (5 sensors$\times$3 modalities$\times$3 axes). Each task was then divided into two second windows, with 50% overlap.

C. Feature Extraction

For each window, data were downsampled to a sampling rate of 64 Hz (from 128 Hz) to increase computational speed. We then extracted various features useful for differentiating and classifying human motion [22] including: area under the curve (AUC); root mean square (RMS); signal magnitude area (SMA); signal vector magnitude (SMV); energy; entropy; fast Fourier transform peak (FFTP); mean; and standard deviation (SD). Each two second window resulted in a matrix of 45 values (5 locations $\times$ 3 modalities $\times$ 3 axes) by 128 (two seconds $\times$ 64 Hz).

Beginning with the second data point (2/64$^{\text{th}}$ seconds into the window), features were extracted from subvectors of the raw data that increased in size until the final extracted feature was derived using raw data from the entire window. For features such as mean, methods like this are already commonly employed (e.g. cumulative moving average [23]) except, in our case, each new feature is extracted from the previous subvector of raw data appended by the latest datum point. For the first instance of data in the window (1/64$^{\text{th}}$ of a second into the window) there are insufficient data points to create a vector from which features can be extracted. To counteract this and ensure that feature vectors have the same length as the raw data, a new data point was approximated from the raw data. The new data point is assumed to happen at time t_0 = 0s and the data point itself is approximated by the formula below where d_1 and d_2 are the data points that are found at time t_1 = 1/64$^{\text{th}}$ (s) and t_2 = 2/64$^{\text{th}}$ (s) respectively:

$$d_0 = d_1 - [d_2 - d_1] = 2 \times d_1 - d_2$$

This formula assumes equal spacing between data points and allows for the generation of a feature vector of the same length as raw data. Table I illustrates how values are generated for **one** feature using this method: This approach

n	1	2	...	128
Feature	f_1	f_2	...	f_{128}
Time	$[t_0, t_1]$	$[t_1, t_2]$	...	$[t_1, t_2, ..., t_{128}]$
Data	$[d_0, d_1]$	$[d_1, d_2]$	...	$[d_1, d_2, ..., d_{128}]$

TABLE I
DATA SYNTHESIS APPROACH

was used to ensure that all features were appropriately formatted as inputs for MLP and ESNN models.

D. Data Preprocessing and Shaping

Each feature was standardized (using the appropriate means and standard deviations for each modality) to account for relative differences in magnitude. Because the windowing technique was used to generate input data, tasks that took longer to perform (e.g., folding a towel) were overrepresented while other tasks (e.g., reaching for a spoon) were underrepresented. To account for this, participants were randomly combined in groups and their data were balanced using the Synthetic Minority Oversampling Technique

(SMOTE) [24] combined with Edited Nearest Neighbors (ENN) [25]. Both techniques are used to synthetically balance unbalanced datasets. They also removed data points at the overlap between classes, resulting in better accuracy and reduced computational costs. Sixty percent of the data from each participant were randomly selected to train the classifier and the other forty percent were held for validation and testing.

Finally data were reshaped into a three-dimensional matrix where each row corresponds to a single feature for a single sensor, each column represents a time step, and each layer represents a modality. Input data were of dimension $15N$ (where N is the number of included features) by 128 by 3.

E. Model Structure

Data were classified using two different neural networks; a simple multi-layer perceptron (MLP) and an Echo State Recurrent Neural Network (ESNN). The MLP was chosen because of its relatively low computational cost and its frequent use for nonlinear time series prediction [26]–[29], and the MLP has an input layer, a hidden layer consisting of a flattened layer and 5 dense layers (4 with the rectified linear unit activation and a final dense layer with a number of nodes corresponding to the number of task types (5) with a SoftMax activation), and an output layer. The number of dense layers in the hidden layer and the number of neurons per dense layer were empirically determined by systematically iterating numbers of layers and neurons. The MLP was then compiled using the Adam optimizer which is computationally efficient and works for classification problems with large amounts of data [30]. The loss function used was categorical cross-entropy which computes the cross-entropy loss between the labels and predictions. The model was then fit to the training set and classification performance was improved by minimizing the cross-validation error over several iterations.

The hyper-parameters of the ESNN were also empirically determined to maximize classification accuracy. Using a local, random search, the selected hyper-parameter values were similar to those utilized in [31] (Table II).

Hyper-parameters of the Reservoir	
Size of the Reservoir	495
Largest Eigenvalue of the Reservoir	0.53
Leakage in the Reservoir State Update	0.66
% of Nonzero Connections in the Reservoir	25
Scaling of the Input Weights	0.0878
Noise in the Reservoir State Update	0.01
Transient States to be Dropped	5
If True, Use Bidirectional Reservoir	False
Use Reservoir with Circle Topology	False
Dimensionality Reduction Hyper-parameters	
Dimensionality Reduction Method	PCA
Number of Resulting Dimension After PCA	168
Multivariate Time Series (MTS) Representation	
MTS Representation	`reservoir`
Parameter of the Ridge Regression	11
Type of Readout (Linear, SVM, MLP)	
Readout used for Classification	`linear`
Linear Readout Hyper-parameters	
Reg. of the Ridge Regression Readout	4
SVM Readout Hyper-parameters	
Bandwidth of the RBF Kernel	0.0055
Reg. of the SVM Hyperplane	5.5
MLP Readout Hyper-parameters	
Neurons in each MLP Layer	9
Number of Epochs	2000
Weight of the L2 Regularization	0.001
Type of Activation Function	`tanh`

TABLE II
MLP HYPER-PARAMETERS

III. RESULTS

To obtain baseline model performance, the MLP was first evaluated using preprocessed raw data only (RAW). We then obtained a model using preprocessed raw data and the extracted area under the curve features (RAW+AUC). Next, a third feature was added (RAW+AUC+f_3). The third feature was one of the list from Section II-C. The features for which highest accuracy was achieved (RMS, SMA, SVM, FFTP) were then used in combination with RAW+AUC data, as well as each other, to evaluate the model for both post-stroke and control participants. Only the highest performing features were used to increase computational efficiency and decrease time spent on analysis. Finally, a fourth feature was added (RAW+AUC+f_3+f_4). No additional features were added to the input dataset because continuing to increase the number of features increases the chance or correlated features, which can lead to model over-fitting and decreased validation and testing accuracy [32], [33]. Results are depicted in **Figure 1**.

We used the Mann-Whitney U test to determine if adding the fourth feature improved accuracy (when compared to RAW+AUC+f_3). For control participants, continuing to add features after RAW+AUC did not increase model accuracy and as the number of features increased from three to four, testing accuracy significantly decreased (accuracy $p = 0.071$; validation accuracy $p = 0.439$; testing accuracy $p <= 0.003$). The highest testing accuracy for controls (86.7%) was achieved using the RAW+AUC+SMA features. For participants post-stroke, although the average testing and validation accuracy continued to increase as more features were added, improvements were not significant (accuracy $p = 0.156$; validation accuracy $p = 0.156$; testing accuracy $p = 0.197$). The highest testing accuracy for participants post-stroke (64%) used RAW+AUC+RMS features.

The ESNN was evaluated using all three readouts used for classification. For the control participants, testing accuracy was 91.3%, 86.5%, and 87.1% (when using the linear, support vector machine, and multi-layer perceptron readout,

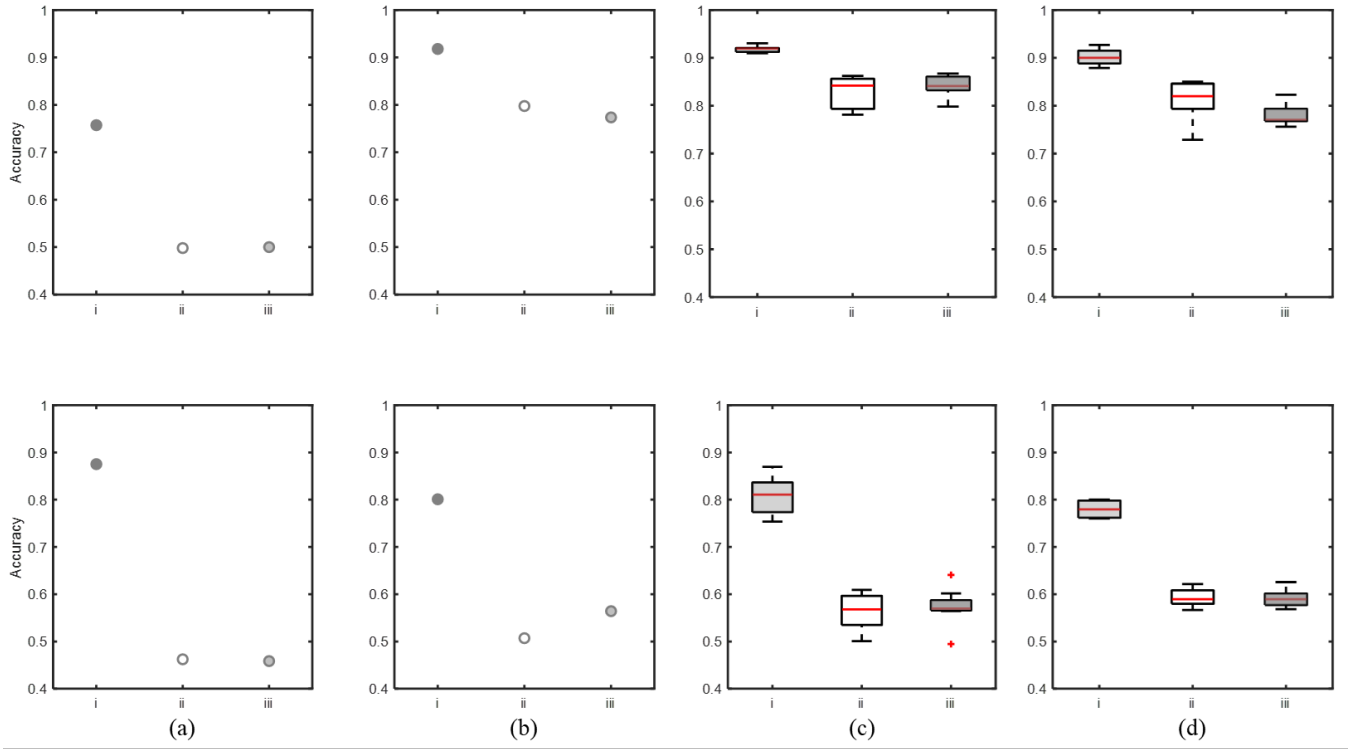

Fig. 1. Accuracy (i), validation accuracy (ii), and test accuracy (iii) of the MLP classifier when using: (a) raw data; (b) raw data and the derived AUC feature; (c) raw data, AUC, and one other derived feature; or (d) raw data, AUC, and two other derived features for both the Control (**row 1**) and Post-Stroke (**row 2**) groups

respectively). For participants post-stroke, the testing accuracy was 80.3%, 74.1%, and 76.7% (when using the linear, support vector machine, and multi-layer perceptron readout, respectively).

IV. Discussion

The purpose of this analysis was to determine if the application of novel classification approaches (including the multi-layer perceptron and echo state neural network) would improve our ability to classify noisy, chaotic, human-derived, time series data. According to our prior results [10], these machine learning techniques are a marked improvement. Our previous analyses demonstrated data collected from each participant were necessary to train accurate classification models prior to real time use (i.e., user dependent models). Use of the MLP and ESNN will facilitate real time, user independent models.

Using the simple MLP, we found that, although the validation and testing accuracy continued to improve for participants post-stroke as feature were added, adding features did not significantly increase the accuracy of the classifier. Significant decreases in testing accuracy with increased number of features for control participants was likely due to model over-fitting. Given the risk of over-fitting and the increased computational cost of adding features, we conclude that it is best to let the neural network extract relevant features.

Finally, when using the ESNN, classification accuracy greatly increased relative to MLP performance. This analysis has shown that using a complex RNN such as the echo state neural network can be useful in the real-time classification of tasks in both control participants and participants post-stroke using minimally intrusive sensing techniques that do not impede the function of a participant.

As mentioned in the introduction, the first step in autonomous recognition and reversal of compensatory movement strategies is the capability to differentiate between the types of tasks that people post-stroke perform in the ambient setting. This classification can then be further used to identify, address, and decrease the use of compensatory behaviors that can lead to a decreased quality of life or lifespan. In the future, work can be done to improve the results of the analysis using different preprocessing techniques to limit the likelihood of overfitting, to determine the impact that data from each sensor location has on the outcome of the models, and to implement the system in real-time to validate the models on a new group of participants.

V. Acknowledgement

The authors thank Dr. Susan Duff, EdD, PT, OT/L, CHT, for her ongoing collaboration in collection, analysis, and contextualization of data. This work was supported by NSF Award #2054191.

References

[1] A. H. Virani SS, Alonso A *et al.*, "American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association.," *Circulation*, vol. 143, pp. e254–e743, Feb 2021.

[2] M. F. Levin, J. A. Kleim, and S. L. Wolf, "What do motor "recovery" and "compensation" mean in patients following stroke?," *Neurorehabil Neural Repair*, vol. 23, pp. 313–319, May 2009.

[3] J. van Kordelaar, E. E. van Wegen, and G. Kwakkel, "Unraveling the interaction between pathological upper limb synergies and compensatory trunk movements during reach-to-grasp after stroke: a cross-sectional study," *Exp Brain Res*, vol. 221, pp. 251–262, Sep 2012.

[4] T. Merdler, D. G. Liebermann, M. F. Levin, and S. Berman, "Arm-plane representation of shoulder compensation during pointing movements in patients with stroke," *J Electromyogr Kinesiol*, vol. 23, pp. 938–947, Aug 2013.

[5] E. Taub, G. Uswatte, V. W. Mark, and D. M. Morris, "The learned nonuse phenomenon: implications for rehabilitation," *Eura Medicophys*, vol. 42, pp. 241–256, Sep 2006.

[6] M. E. Michielsen, R. W. Selles, H. J. Stam, G. M. Ribbers, and J. B. Bussmann, "Quantifying nonuse in chronic stroke patients: a study into paretic, nonparetic, and bimanual upper-limb use in daily life," *Arch Phys Med Rehabil*, vol. 93, pp. 1975–1981, Nov 2012.

[7] G. Uswatte, C. Giuliani, C. Winstein, A. Zeringue, L. Hobbs, and S. L. Wolf, "Validity of accelerometry for monitoring real-world arm activity in patients with subacute stroke: evidence from the extremity constraint-induced therapy evaluation trial," *Arch Phys Med Rehabil*, vol. 87, pp. 1340–1345, Oct 2006.

[8] W. J. Teskey, M. Elhabiby, and N. El-Sheimy, "Inertial sensing to determine movement disorder motion present before and after treatment," *Sensors (Basel)*, vol. 12, no. 3, pp. 3512–3527, 2012.

[9] E. W. Parnandi, A.R. and M. Matarić, "Motor function assessment using wearable inertial sensors," *IEEE EMBC*, 2010.

[10] D. S. W. E. Miller A, Quinn L, "Comparison of Machine Learning approaches for Classifying Upper Extremity Tasks in Individuals Post-Stroke.," *Annu Int Conf IEEE Eng Med Biol Soc.*, vol. 2020, pp. 4330–4336, Jul 2020.

[11] A. Miller, S. V. Duff, L. Quinn, L. Bishop, G. Youdan, H. Ruthrauff, and E. Wade, "Development of Sensor-Based Measures of Upper Extremity Interlimb Coordination," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2018, pp. 3160–3164, Jul 2018.

[12] J. Heo, J. G. Yoon, H. Park, Y. D. Kim, H. S. Nam, and J. H. Heo, "Machine learning–based model for prediction of outcomes in acute stroke," *Stroke*, vol. 50, no. 5, p. 1263–1265, 2019.

[13] R. Mohanty, A. M. Sinha, A. B. Remsik, K. C. Dodd, B. M. Young, T. Jacobson, M. Mcmillan, J. Thoma, H. Advani, V. A. Nair, and et al., "Machine learning classification to identify the stage of brain-computer interface therapy for stroke rehabilitation using functional connectivity," *Frontiers in Neuroscience*, vol. 12, 2018.

[14] C. Tozlu, D. Edwards, A. Boes, D. Labar, K. Z. Tsagaris, J. Silverstein, H. P. Lane, M. R. Sabuncu, C. Liu, A. Kuceyeski, and et al., "Predicting response to motor therapy in chronic stroke patients using machine learning," 2018.

[15] X. Jiang, L.-K. Merhi, Z. G. Xiao, and C. Menon, "Exploration of force myography and surface electromyography in hand gesture classification," *Medical Engineering Physics*, vol. 41, p. 63–73, 2017.

[16] W.-J. Li, C.-Y. Hsieh, L.-F. Lin, and W.-C. Chu, "Hand gesture recognition for post-stroke rehabilitation using leap motion," *2017 International Conference on Applied System Innovation (ICASI)*, 2017.

[17] K. Kaczmarczyk, A. Wit, M. Krawczyk, and J. Zaborski, "Gait classification in post-stroke patients using artificial neural networks," *Gait Posture*, vol. 30, no. 2, p. 207–210, 2009.

[18] H.-Y. Lau, K.-Y. Tong, and H. Zhu, "Support vector machine for classification of walking conditions of persons after stroke with dropped foot," *Human Movement Science*, vol. 28, no. 4, p. 504–514, 2009.

[19] P. J. Ozturk MC, Xu D, "Analysis and design of echo state networks.," *Neural Comput.*, vol. 19, pp. 111–38, Jan 2007.

[20] L. Wang, Z. Wang, and S. Liu, "An effective multivariate time series classification approach using echo state network and adaptive differential evolution algorithm," *Expert Systems with Applications*, vol. 43, pp. 237–249, 2016.

[21] T. E. Nightingale, J. P. Walhin, D. Thompson, and J. L. Bilzon, "Influence of accelerometer type and placement on physical activity energy expenditure prediction in manual wheelchair users," *PLoS ONE*, vol. 10, no. 5, p. e0126086, 2015.

[22] S. D. Bersch, D. Azzi, R. Khusainov, I. E. Achumba, and J. Ries, "Sensor data acquisition and processing parameters for human activity classification," *Sensors (Basel)*, vol. 14, no. 3, pp. 4239–4270, 2014.

[23] D. Afolabi, S.-U. Guan, K. L. Man, P. W. Wong, and X. Zhao, "Hierarchical meta-learning in time series forecasting for improved interference-less machine learning," *Symmetry*, vol. 9, no. 11, p. 283, 2017.

[24] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, p. 321–357, 2002.

[25] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, p. 20–29, 2004.

[26] Q. Ma, Q. Zheng, H. Peng, T. Zhong, and L. Xu, "Chaotic time series prediction based on evolving recurrent neural networks," in *2007 International Conference on Machine Learning and Cybernetics*, vol. 6, pp. 3496–3500, 2007.

[27] R. Bakker, J. C. Schouten, C. L. Giles, F. Takens, and C. M. Van Den Bleek, "Learning chaotic attractors by neural networks," *Neural Computation*, vol. 12, no. 10, pp. 2355–2383, 2000.

[28] K. K. Teo, L. Wang, and Z. Lin, "Wavelet packet multi-layer perceptron for chaotic time series prediction: effects of weight initialization," in *International Conference on Computational Science*, pp. 310–317, Springer, 2001.

[29] H. Inoue, Y. Fukunaga, and H. Narihisa, "Efficient hybrid neural network for chaotic time series prediction," in *International Conference on Artificial Neural Networks*, pp. 712–718, Springer, 2001.

[30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[31] M. Stewart, "Predicting stock prices with echo state networks.," 2019.

[32] A. S. Weigend, M. Mangeas, and A. N. Srivastava, "Nonlinear gated experts for time series: Discovering regimes and avoiding overfitting," *International journal of neural systems*, vol. 6, no. 04, pp. 373–399, 1995.

[33] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 1578–1585, 2017.