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

A comprehensive evaluation of state-of-the-art time-series deep learning models for activity-recognition in post-stroke rehabilitation assessment

Issam Boukhennoufa*, Xiaojun Zhai*, Victor Utti**, Jo Jackson** and Klaus D. McDonald-Maier*
**School of Computer Science and Electronic Engineering*
***School of Sport, Rehabilitation and Exercise Sciences*
University of Essex
Colchester, United Kingdom

***Abstract*—The recent COVID-19 pandemic has further highlighted the need for improving tele-rehabilitation systems. One of the common methods is to use wearable sensors for monitoring patients and intelligent algorithms for accurate and objective assessments. An important part of this work is to develop an efficient evaluation algorithm that provides a high-precision activity recognition rate. In this paper, we have investigated sixteen state-of-the-art time-series deep learning algorithms with four different architectures: eight convolutional neural networks configurations, six recurrent neural networks, a combination of the two and finally a wavelet-based neural network. Additionally, data from different sensors' combinations and placements as well as different pre-processing algorithms were explored to determine the optimal configuration for achieving the best performance. Our results show that the XceptionTime CNN architecture is the best performing algorithm with normalised data. Moreover, we found out that sensor placement is the most important attribute to improve the accuracy of the system, applying the algorithm on data from sensors placed on the waist achieved a maximum of 42% accuracy while the sensors placed on the hand achieved 84%. Consequently, compared to current results on the same dataset for different classification categories, this approach improved the existing state of the art accuracy from 79% to 84%, and from 80% to 90% respectively.**

***Index Terms*—Deep learning, Tele-rehabilitation, Wearable sensors, Time-series, Stroke.**

I. INTRODUCTION

Stroke can be a consequence of cerebrovascular disease, which is one of the leading causes of death in the world. It accounts for roughly 75% of deaths from cerebrovascular diseases [1]. If not fatal, stroke can lead to temporary or permanent paralysis [2], limiting the ability to perform Activities of Daily Living (ADL) and thus impacting on the patient's quality of life [3].

To recover their lost skills, patients are required to undertake rehabilitation and an important part of it is performed in an outpatient environment [4]. At this stage therapists prescribe the necessary rehabilitation exercises based on their assessment of the patients [5]. Some of the research evidence, however shows that functional improvements observed in the hospital do not necessarily translate to the home setting [6].

To tackle these issues technology-based systems have been utilised and wearable sensor-enabled technologies constitute one common solution. This is due to their high portability combined with their low costs [7]. Moreover, these devices can assist therapists in monitoring their patients and coupled with intelligent algorithms, they can provide an objective assessment to track their patients' progress [8], [9].

In order to implement an effective assessment technology, the system needs to accurately detect the exercises being executed, which relate in most cases to ADL. To do so, many approaches have been utilised ranging from a conventional signal processing modelling approach that seeks a mathematical relationship between an activity and the different modelling parameters, to machine learning algorithms, that extract pertinent features to allow the model to differentiate and recognise the different activities, to more recently deep learning algorithms that can automatically extract features and learn to distinguish between the activities [10], [11].

Deep learning achieved significant results in the computer vision domain [12]–[15] and has started to outperform traditional machine learning techniques in Human Activity Recognition (HAR) . In addition, it is possible to achieve state-of-the-art accuracy without the need to re-train the entire model using transfer learning. Transfer learning is the process of training the model on a source dataset, and then transferring the learned features to be used on a new target dataset [16].

In this paper, sixteen different state of the art time-series deep learning models are evaluated on a complex ADL benchmark dataset, that includes eighteen different activities (e.g. walking, jogging, climbing stairs, sitting, standing, typing, brushing teeth, eating soup, eating chips, eating pasta, drinking from a cup, eating a sandwich, kicking a soccer ball, playing catch with a tennis ball, dribbling a basketball, writing, clapping, and folding clothes). The sixteen classification models belong to four distinct architectures/ (1) Convolutional Neural Networks (CNN), (2) Recurrent Neural Networks (RNN), (3) a combination of the two and (4) a wavelet-based neural network with the objective to determine which configuration gives a higher accuracy on HAR. In addition, data collected from different sensors and mounted on different body locations are pre-processed using three different algorithms, which are also analysed and compared to find out the best source for distinguishing the different activities.

Results obtained in this paper show an increase of 10% in accuracy from current state of the art classifications. Indeed, data from a single sensor could identify up to 90% of 18 different complex ADLs using time-series deep learning algorithms.

The remainder of the paper is organised as follows: in section II, the utilised dataset is presented with an explanation of the different preparation steps for the data used in the classification algorithms. After that in section III, experimental results achieved for the best performance are presented in the first sub-section III-A, the used models are evaluated with different data sources in sub-section III-B, and finally in section III-C fine tuning and post-processing the models are also elaborated. Section IV concludes the paper.

II. Dataset description and preparation

In this paper the WISDM Smartphone and Smartwatch Activity and Biometrics Dataset [17] was utilised. This dataset was actualised in late 2019. It includes diverse and complex ADL and this makes it a good candidate for evaluating the algorithms. It consists of 18 activities (Table I) performed by 51 different participants for three minutes. Two Inertial measurement unit (IMU) sensors (triaxial accelerometer and triaxial gyroscope) from a smartwatch and a smartphone were utilised respectively to collect the data. The smartwatch was mounted on the participant's dominant hand, and the smartphone is placed on the waist, with each using a frequency of 20 Hz. Hierarchically, the dataset is divided into two folders, phone and watch, each folder is sub-divided into two sub-folders accelerometer and gyroscope, each containing 51 files corresponding to the different participants IDs. Each file contains the following information: subject-ID, activity-code (character between 'A' and 'S' no 'N' that identifies the activity), timestamp, x, y, z sensors' readings (i.e. accelerometer or gyroscope).
Table I shows the different activities involved and their labels.

TABLE I: Dataset activities and their labels.

Activity orientation	Activities
Non-hand-oriented activities	Walking (A), Jogging (B), Stairs (C), Sitting (D), Standing (E), Kicking (M)
Hand-oriented activities (Eating)	Eating soup (H), Eating chips (I), Eating pasta (J), Drinking (K), Eating sandwich (L)
Hand-oriented activities (General)	Typing (F) , Playing catch (O) , Dribbling (P), Writing (Q), Clapping (R), Brushing teeth (G) , Folding clothes (S)

The data obtained from the different sensors were investigated to determine the most suitable way for recognising the activities i.e: accelerometer from the phone, gyroscope from the phone, accelerometer from the watch, gyroscope from the watch, phone (accelerometer + gyroscope) and watch (accelerometer + gyroscope). The accelerometer-gyroscope were merged using the timestamp provided for each reading, and the number of readings for each sensor is provided in Table II.

TABLE II: Total number of readings for each sensor

Sensor	Number of readings
Accelerometer phone	4,804,403
Gyroscope phone	3,608,635
Accelerometer watch	3,777,046
Gyroscope watch	3,440,342
Phone (Gyro + acc)	2,909,149
Watch (Gyro + acc)	3,370,861

The dataset is segmented into multiple data windows of 10 s corresponding to the 200 readings, every window of data is labeled with the most recurrent activity in that window.

III. Experimental results

The selection of the model is performed in three stages, 1) Sixteen different state-of-the art time series algorithms are tested in order to choose the best classifier. 2) The chosen model is investigated further by using three different pre-processing algorithms (i.e. feeding raw data, feeding standardised data, and feeding normalised data) on the six sensors' data sources from the phone and watch (i.e. Gyroscope, accelerometer and combination of both for each device). 3) The data is further tweaked and the model is then fine-tuned to improve the final performance.

A. First stage: model selection

In order to find the best performing algorithm, sixteen state of the art deep learning classifiers are employed. The dataset is divided into 80% for training and 20% for testing (not used for training). The Tsai [18] library from fastai is utilised in Python, which is a deep learning library for time-series model based on Fastai [19].
Different architecture are used namely CNN, RNN, a combination of the two (CNN-RNN) and Wavelet-based neural network, as previously identified.
A brief description of each algorithm is given as follows:

1) RNN models: RNNs are a class of neural networks that allow previous outputs to be used as inputs while having hidden states. In this paper, six RNNs with different Long Short Time Memory (LSTM) are used for HAR, and the main difference between each of them is the number of layers (1, 2, 3) as well as using the bidirectional or non bidirectional architectures.

TABLE III: Accuracies of the different models on the different sensors' raw data.

Model	Acc watch	Acc phone	Gyro watch	Gyro phone	watch	phone
XceptionTime	0.776013	0.275341	0.707558	0.4087	0.697626	0.299416
ResNet	0.752714	0.224783	0.707558	0.353006	0.704748	0.294947
InceptionTime	0.751655	0.264684	0.683721	0.389305	0.709792	0.275696
ResCNN	0.748478	0.244114	0.680523	0.327237	0.709496	0.27879
LSTM FCN	0.73471	0.225774	0.681105	0.349958	0.707122	0.291165
OmniScaleCNN	0.730209	0.260719	0.680814	0.346911	0.657665	0.263445
LSTMFCN	0.726238	0.245353	0.681105	0.349958	0.697626	0.246476
FCN	0.721472	0.263445	0.677616	0.348573	0.711573	0.289447
xresnet1d34	0.709293	0.253779	0.67907	0.373234	0.680119	0.279821
mWDN	0.67964	0.238662	0.587209	0.313106	0.605242	0.275696
LSTM3	0.666667	0.203965	0.624419	0.365475	0.669436	0.258508
LSTM3bi	0.661371	0.203965	0.612791	0.353006	0.65905	0.237882
LSTM2	0.657665	0.199009	0.604942	0.343585	0.623739	0.232726
LSTM2bi	0.656606	0.17596	0.603198	0.337213	0.619881	0.217944
LSTM1	0.605242	0.174226	0.573837	0.326129	0.60178	0.202475
LSTM1bi	0.587503	0.163569	0.556395	0.321142	0.588131	0.196631

2) 1D-CNN based models: Seven different CNN models are explored:

- Fully Connected Neural networks (FCN): Inspired by the work introduced by Wang et al [20], it consists of CNNs that do not contain any local pooling layers, meaning that the length of a time series is kept unchanged throughout the layers of convolutions.
- XceptionTime: A novel time series architecture designed by Elahe Rahimian et.al in [21] by the integration of depth-wise separable convolutions, adaptive average pooling, and a novel non-linear normalisation technique.
- InceptionTime: An ensemble of deep Convolutional Neural Network models, inspired by the Inception-v4 architecture for computer vision introduced by Fawaz et.al [22].
- ResNet: Convolutional layers that stack residual blocks on top of each other to form a network, very popular in the computer vision domain introduced by Kaiming He in [23].
- ResCNN: Applies the residual block to overcome the vanishing gradient problem. It is additionally enhanced by using the k-fold ensemble method.
- XResNet: A modification of the traditional ResNet architecture suggested by Tong He in [24].
- OmniscaleCNN: A CNN architecture whose specificity is to concatenate the outputs of several convolution filters whose length is one plus all the prime numbers between two and a quarter of the time series length proposed by Tang et al [25].

lesl

3) RNN-CNN models: A combination of CNN and RNN architectures was investigated, it consists of LSTM layers and convolution layers for feature extraction with different pooling layers.

4) Wavelet-based neural network: This model consists of the Multilevel Wavelet Decomposition Network for Interpretable Time Series Analysis (mWDN) Algorithm introduced by Wang et. al in [26]. The particularity of this model is that it preserves the advantage of multilevel discrete wavelet decomposition in frequency learning while enabling the fine-tuning of all parameters under a deep neural network framework.

5) Results: Table III shows the results of training each classifier for 100 epochs (training them for more epochs did not improve accuracy) and taking the maximum validation accuracy obtained for each one. The XceptionTime gave the best overall results for most of the data sensors and is therefore the one that has been selected to carry out further analysis and fine tuning to improve the performance.

B. Second stage: sensor and pre-processing algorithm selection

As explained in section III-A5, the XceptionTime model outperformed the other algorithms and is therefore selected to be used as our activity recognition algorithm. A 5-fold cross-validation test is used to assess the accuracy of the model in recognising the different activities. 80% of the data are used to train the model, and the rest are used to determine the accuracy of the model. We want our model to generalise well with new participants, so the windows of data are not shuffled, but in each one of the five iterations 10 participants (which corresponds to 20% of the data) among the 51 are chosen to be the validation data. In addition, the accuracy between models are compared based on each sensor's data (i.e. accelerometer from the phone only, gyroscope from the phone only and accelerometer and gyroscope combined from the phone and same thing for the sensors in the watch) in order to find out which sensor's data are most accurate for the activity recognition. Moreover, three different pre-processing methods are employed named: (1) feeding raw data from the sensors to the classification model, (2) feeding normalised data and (3) feeding standardised data. Normalisation typically means re-scaling the values into a range of [0, 1] while standardisation means re-scaling data to have a mean of 0 and a standard deviation of 1.

The accuracy was computed using the formula:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) * 100$$

Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

For all the models, in order to select the best Learning Rate

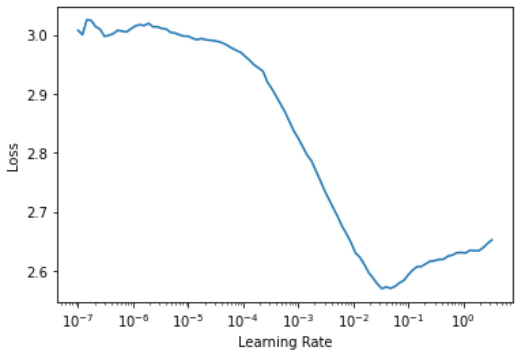


Fig. 1: Loss Vs learning rate.

(LR) to use we followed the approach taken by Leslie N. Smith [27], where the loss is computed and plotted using different LRs and the latter is chosen to be in the interval between the point were the loss is minimal and a factor of 10 smaller. An example is presented in Figure 1, where the LR is chosen to be in the interval [2e-3,2e-2].

1) Phone data: The data collected from the phone are the first to be presented, accuracy results from the accelerometer, gyroscope and the combination of both are compared using the three different pre-processing algorithms discussed earlier. Table IV shows the accuracy rates obtained for each configuration. The accuracy rates obtained are very low, with a maximum of 42% obtained when using standardised gyroscope data.

TABLE IV: Accuracy rates of the different phone sensors' configurations

	Raw	Standardised	Normalised
Accelerometer	29%	35%	30%
Gyroscope	36%	42%	41%
Combination	37%	33%	28%

Accelerometer data as well as the combination of both give very poor accuracy rates. Since the number of models is high (i.e. three data processing methods and three different sensors' sources which makes a total of nine different configurations) presenting further results of all of them could be cumbersome. Subsequently, only the one with the maximum accuracy (i.e Standardised gyroscope data) is presented in Figure 2.

Based on the phone sensors' data, the algorithm has a high difficulty differentiating between the activities, with a maximum accuracy of 42% reached when using the gyroscope and the standardised data. We can also see from the associated confusion matrix (classes names are in Table I), that the hand-associated activities (especially the eating activities) are the ones being mostly confused, the reason is that the phone is placed on the waist and therefore the sensors cannot accurately recognise the hand activity.

2) Watch data: Similarly to the procedure followed in section III-B1 we find the appropriate LR range to use, and train the model using 100 epochs. The accuracy rates obtained for the different configurations are given in Table V.

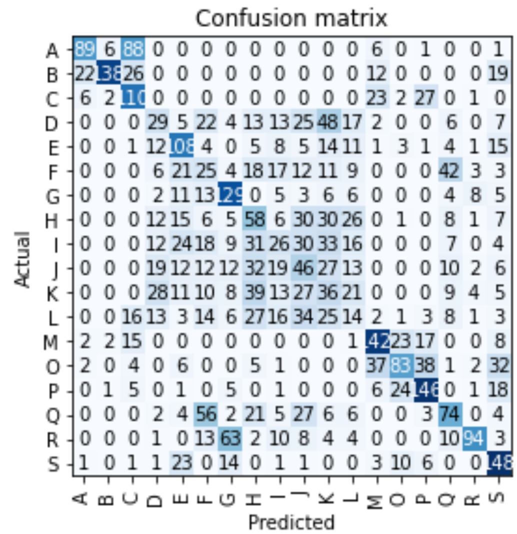


Fig. 2: Confusion matrix of the standardised data from the phone.

TABLE V: Accuracy rates of the different watch sensors' configurations

	Raw	Standardised	Normalised
Accelerometer	82%	81%	69%
Gyroscope	74%	73%	74%
Combination	72%	84%	82%

We see that the accuracy improves significantly from the phone data, the standardised data from the combined accelerometer-gyroscope is the configuration that provides the highest accuracy 84%. We investigate the model performance further, using the associated confusion matrix in Figure 3.

The data from the watch sensors are more successful in differentiating between the activities by achieving near perfect classifications for non-hand oriented activities and hand oriented activities (general). It still confuses eating related activities especially eating sandwich (L) eating chips (I) and eating pasta (J), which is not surprising. Additionally, combining the data from the accelerometer and gyroscope improves relatively the classification and the normalisation pre-processing gives a better accuracy.

C. Third step: dataset tweaking and discussion on the results

In this section we try to increase the accuracy even further by tweaking the features using the information we identified in the previous sections. The model used is the combined standardised watch data from section III-B2. Two approaches are listed as follows:

- The two activities with the highest miss-classifications (i.e. eating a sandwich and eating pasta) are merged together into one class (we can call it eating sandwich or pasta). Taking into account that adding two classes

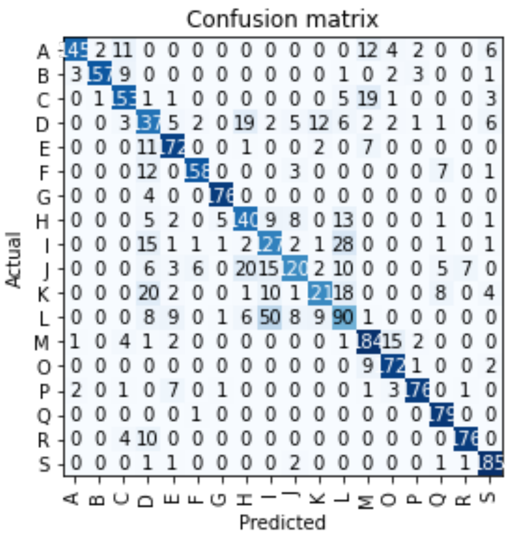


Fig. 3: Confusion matrix of the normalised data from the watch.

together will increase the size of the data in that new class, and thus creates a bias toward that activity. Therefore, the two merged activities were randomly split in two halves and one half of the first activity is merged with one half of the other, which creates at the end a new class that has approximately the same number of data as other classes.

- All the activities related to eating activities are merged together which are a total of five (eating: soup chips pasta sandwich and drinking) into one class only (eating). This is done by following the same procedure described previously to avoid an unbiased dataset (including the phone data did not improve accuracy).

The accuracy has increased significantly from our previous best performing model with an accuracy of 86% for our first merged model and 90% for the second one. The confusion matrices for these two models are shown in Figure 4.

The confusion matrices show that the models are more successful in differentiating between the different activities, most of the miss-classifications in the first merged model Figure 4a still lie in the eating activities, while in the second model (Figure 4b) the highest miss-classification is between the eating activities and sitting.

The results obtained in this paper show that deep learning provides state-of-the-art results in the sensor based HAR. We have tested numerous time-series models that gave outstanding accuracies on a complex dataset that contains eighteen different activities and this using only a single sensor. We are able to significantly increase the performance obtained in [28] that used some classical deep learning approaches and [29] that used KNN, Random forest and SVM on a tweaked dataset (similar to our tweaking done in section III-C) (Table VI). This demonstrates that time-series deep learning models are good candidates to be employed for the tele-rehabilitation processing algorithms.

Moreover, sensor placement is a highly important attribute for the success of the model, our results show that there is a huge difference of 40% between the sensor placed on the waist and the one placed on the hand. The reason for that is the sensor placed on the hand is more sensitive to some activities than the other sensors for this specific dataset.

TABLE VI: Comparison between best accuracies obtained in different papers

	Original dataset	Tweaked (merging eating activities)
[28]	79%	NA
[29]	NA	80%
This paper	84%	90%

IV. Conclusion and Future Work

In this paper we investigated the performance of numerous state-of-the-art time-series deep learning algorithms on a complex HAR dataset, as well as the effect of the sensors utilised, their placement and the pre-processing algorithms employed on the data. Results showed that the different algorithms gave relatively close accuracy rates with the Xceptiontime outperforming the others slightly. Additionally, sensor placement plays a significant role in accurately recognising the activities, some placements are more sensitive to specific activities than others. In fact, applying the algorithm on data from sensors placed in the waist achieves a maximum of 42% accuracy, while the sensors placed on the hand achieved 84% accuracy. Moreover merging the eating activities increased the accuracy from 84% to 90% further on the validation data using a five-fold cross-validation test. As future work, we aim to increase the accuracy rates obtained further by encoding the time-series data into images and feeding it to popular computer vision classifiers.

V. Acknowledgements

This work is supported by the UK Engineering and Physical Sciences Research Council through grants EP/R02572X/1, EP/P017487/1, EP/P015956/1, EP/V034111/1 and EP/V000462/1.

References

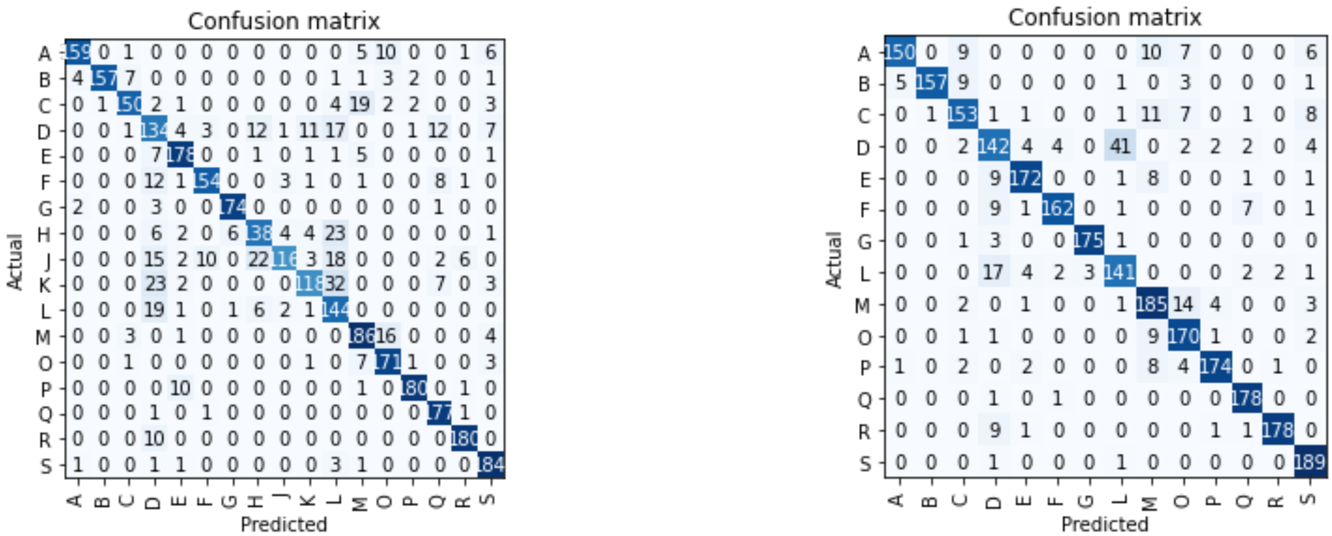
[1] J. Kim, T. Thayabaranathan, G. A. Donnan, G. Howard, V. J. Howard, P. M. Rothwell, V. Feigin, B. Norrving, M. Owolabi, J. Pandian *et al.*, "Global stroke statistics 2019," *International Journal of Stroke*, vol. 15, no. 8, pp. 819–838, 2020.

[2] P. Pound, P. Gompertz, and S. Ebrahim, "Illness in the context of older age: the case of stroke," *Sociology of health & illness*, vol. 20, no. 4, pp. 489–506, 1998.

[3] L. Santisteban, M. Térémetz, J.-P. Bleton, J.-C. Baron, M. A. Maier, and P. G. Lindberg, "Upper limb outcome measures used in stroke rehabilitation studies: a systematic literature review," *PloS one*, vol. 11, no. 5, p. e0154792, 2016.

[4] C. for Disease Control, P. (CDC *et al.*, "Outpatient rehabilitation among stroke survivors–21 states and the district of columbia, 2005," *MMWR. Morbidity and mortality weekly report*, vol. 56, no. 20, pp. 504–507, 2007.

[5] T. Platz, C. Pinkowski, F. van Wijck, I.-H. Kim, P. Di Bella, and G. Johnson, "Reliability and validity of arm function assessment with standardized guidelines for the fugl-meyer test, action research arm test and box and block test: a multicentre study," *Clinical rehabilitation*, vol. 19, no. 4, pp. 404–411, 2005.



(a) Two activities merged.

(b) All eating activities merged.

Fig. 4: Confusion matrices of the two improved models.

[6] D. Rand and J. J. Eng, "Disparity between functional recovery and daily use of the upper and lower extremities during subacute stroke rehabilitation," *Neurorehabilitation and neural repair*, vol. 26, no. 1, pp. 76–84, 2012.

[7] S. Díaz, J. B. Stephenson, and M. A. Labrador, "Use of wearable sensor technology in gait, balance, and range of motion analysis," *Applied Sciences*, vol. 10, no. 1, p. 234, 2020.

[8] A. Kaku, A. Parnandi, A. Venkatesan, N. Pandit, H. Schambra, and C. Fernandez-Granda, "Towards data-driven stroke rehabilitation via wearable sensors and deep learning," *arXiv preprint arXiv:2004.08297*, 2020.

[9] C. Adans-Dester, N. Hankov, A. O'Brien, G. Vergara-Diaz, R. Black-Schaffer, R. Zafonte, J. Dy, S. I. Lee, and P. Bonato, "Enabling precision rehabilitation interventions using wearable sensors and machine learning to track motor recovery," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–10, 2020.

[10] O. D. Incel, M. Kose, and C. Ersoy, "A review and taxonomy of activity recognition on mobile phones," *BioNanoScience*, vol. 3, no. 2, pp. 145–171, 2013.

[11] X. Su, H. Tong, and P. Ji, "Activity recognition with smartphone sensors," *Tsinghua science and technology*, vol. 19, no. 3, pp. 235–249, 2014.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[13] G. Lindsay, "Convolutional neural networks as a model of the visual system: past, present, and future," *Journal of Cognitive Neuroscience*, pp. 1–15, 2020.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[15] I. Boukhennoufa, X. Zhai, K. D. McDonald-Maier, V. Utti, and J. Jackson, "Improving the activity recognition using gmaf and transfer learning in post-stroke rehabilitation assessment," in *2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. IEEE, 2021, pp. 000 391–000 398.

[16] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Transfer learning for time series classification," in *2018 IEEE international conference on big data (Big Data)*. IEEE, 2018, pp. 1367–1376.

[17] G. M. Weiss, "Wisdm smartphone and smartwatch activity and biometrics dataset," *UCI Machine Learning Repository, WISDM Smartphone and Smartwatch Activity and Biometrics Dataset Data Set*, 2019.

[18] I. Oguiza, "tsai - a state-of-the-art deep learning library for time series and sequential data," Github, 2020. [Online]. Available: https://github.com/timeseriesAI/tsai

[19] J. Howard and S. Gugger, "Fastai: A layered api for deep learning," *Information*, vol. 11, no. 2, p. 108, 2020.

[20] 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)*. IEEE, 2017, pp. 1578–1585.

[21] E. Rahimian, S. Zabihi, S. F. Atashzar, A. Asif, and A. Mohammadi, "Xceptiontime: A novel deep architecture based on depthwise separable convolutions for hand gesture classification," *arXiv preprint arXiv:1911.03803*, 2019.

[22] H. I. Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P.-A. Muller, and F. Petitjean, "Inceptiontime: Finding alexnet for time series classification," *Data Mining and Knowledge Discovery*, vol. 34, no. 6, pp. 1936–1962, 2020.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[24] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 558–567.

[25] W. Tang, G. Long, L. Liu, T. Zhou, J. Jiang, and M. Blumenstein, "Rethinking 1d-cnn for time series classification: A stronger baseline," *arXiv preprint arXiv:2002.10061*, 2020.

[26] J. Wang, Z. Wang, J. Li, and J. Wu, "Multilevel wavelet decomposition network for interpretable time series analysis," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 2437–2446.

[27] L. N. Smith, "A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay," *arXiv preprint arXiv:1803.09820*, 2018.

[28] S. Benavidez and D. McCreight, "A deep learning approach for human activity recognition project category: Other (time-series classification)," 2019.

[29] S. J. Saleh, S. Q. Ali, and A. M. Zeki, "Random forest vs. svm vs. knn in classifying smartphone and smartwatch sensor data using crisp-dm," in *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*. IEEE, 2020, pp. 1–4.