BALNEO RESEARCH Balneo and PRM Research Journal
English Edition
e ISSN 2734 - 8458 p ISSN 2734 - 844X ROMANIAN ASSOCIATION OF BALNEOLOGY

Research article

A Deep Learning Approach for Classification of Physiotherapy Exercises Using Segmentation of Techniques

Marius Turnea [1], Andrei Gheorghita [1,*], Mariana Rotariu [1], Mihai Ilea [1], Dragos Arotaritei [1], Irina Duduca[1], Iustina Condurache [1]

1 University of Medicine and Pharmacy "Grigore T. Popa" Iasi;

* Correspondence: Andrei Gheorghita, andu_ghe@yahoo.com

Citation: Turnea M., Gheorghita A., Rotariu M., Ilea M., Arotaritei D., Duduca I., Condurache I. - A Deep Learning Approach for Classification of Physiotherapy Exercises Using Segmentation of Techniques
Balneo and PRM Research Journal **2024**, *15(2): 708*

Academic Editor(s):
Constantin Munteanu

Reviewer Officer:
Viorela Bembea

Production Officer:
Camil Filimon

Received: 18.04.2024
Published: 21.06.2024

Reviewers:
Alin Mălăelea
Codrut Carstoaia

Publisher's Note: Balneo and PRM Research Journal stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: Physiotherapy exercises are necessary to patients to restore their functional abilities in many cases as disabilities, injury, or basic with complementary approach as balneotherapy. Different type of exercised and different template sessions are used depending on the medical diagnostics. The evaluation of effectiveness of these exercises are important for patient's rehabilitation process as time and level of recovery of locomotor skills. A dataset publicly available (Physical Therapy Exercises) is used for classification of session of repeated exercises that includes movement executed correct (C), fast execution (F) and low-amplitude execution (L). A novel approach is proposed by using segmentation of signal using deep learning neural network followed by a convolutional neural network for classification of sequence of the labeled classes L,C, F, and N (a new class introduced to label the noise of sensor of exercised or incorrect movement of the patient. The signal is extensively analyzed in order to made and corresponding labeling for analyzing using sliding window with a drive user selected length. The accuracy of classification is greater than 96% and sensitivity is greater than 95% but the results can be better if the labelling of N class is more restrictive and the effect of imbalanced dataset is reduced.

Keywords: physiotherapy exercises; segmentation techniques; deep learning neural networks, classification; imbalanced dataset

1. Introduction

The physical activities and complementary therapies as balneotherapy [1] play an important role in patient's rehabilitation after trauma, muscle diseases or muscle atrophy due to bed rest. A common approach to evaluate the outcomes of physical rehab and the treatment success (that can take several week of even months) is the patient-reported outcome measures (PROMs) from questionnaires [2]. The method of questionary is inherently subjective and the results can create controversial debates by medical point of view. A more objective measure computational useful can be used in order to obtain accurate data: motion measure capture (by video images) [3-4] and positional ones based on inertial sensors [5-9].

A benchmark dataset, KIMORE is used in [3]. The features are extracted from Kinect V2 sensor as video frame and the skeleton parameters fed a Deep Density Neural network to predict the exercise score. In [4], the authors used You-Tube video dataset (physiotherapy video dataset), a set of various actions of human beings (classified a sub-activities). As classifier, the authors used Strawberry-based recurrent neural framework (SbRNF) algorithm for classification with a complex pre-processing data algorithm [4].

It is enough hard to extract the information from video images human activity recognition. A more effective methods is to use a set of inertial measurement units (IMUs) places on the human body and collect the information from accelerometers, gyroscopes and magnetometers from each IMUs in real-time 1D signals.

A measurement system consisting of 17 inertial measurement units is proposed in [5]. The dataset of four Functional Movement Screening exercises is used for 3374 usable repetitions. The human body is divided into 17 segments, to each of which an IMU is attached. The signals are used as input for CNN-LSTM deep learning architecture (Convolutional Neural Network - Long-Short Term Memory) with four labeled classes: “1” (incomplete execution, even with compensation movements), “2” (complete execution with compensation movements) and “3” (perfect execution) are assigned, and “0” when pain occurs.

Raw data from IMUs are used to classify the human activities dataset into four actions [6]. The paper has a large par allocated to mathematical part, a dynamic convolutional neural network (D-CNN) and a state transition probability CNN (S-CNN) for classifier along with modified Lempel–Ziv–Welch (LZW) algorithm from information theory to extract the transition probabilities and discriminate features of movements. The computation effort is an important one (as the coding of entire system) with an accuracy between 98.50%-99.67% for training stage and between 84.65%-98.73% for test stage [6]. A close related approach is used in [7] with the main modification is that the method is applied for more than four classes of action.

The approach from this paper is focused on the results obtained in [8-9] using the same dataset [10]. A multi-template multi-match dynamic time warping(MTMM-DTW) algorithm is proposed in [8], to detect multiple occurrences in exercise type in the recording from sensors of a physical therapy session. In [9], the authors used the same dataset with deep learning approach for classifier: DNN (Deep Learning Neural Network) based on autoencoders; CNN (Convolutional Neural Network) based on LSTM; CNN with transfer learning over Alexnet; Bidirectional LSTM and Muli-Layer LSTM. The accuracy obtained is greater than 95% and the average sensitivity is greater than 95% using recurrent neural networks (RNN).

2. Results

The dataset used in this paper is [10], the same as the one used in [8-9]. The dataset is organized in text files 5 subjects, 8 types of exercises and 5 sensor units (each of them having accelerometer-acc, gyroscope-gyr and magnetometer-mag on X,Y and Z axis). The records for template session have a set of repetition of each three types (C, F, L) meanwhile the test session has 10 repetitions for each execution type. The sampling rate fs=25 Hz (T=1/fs=0.04 s).

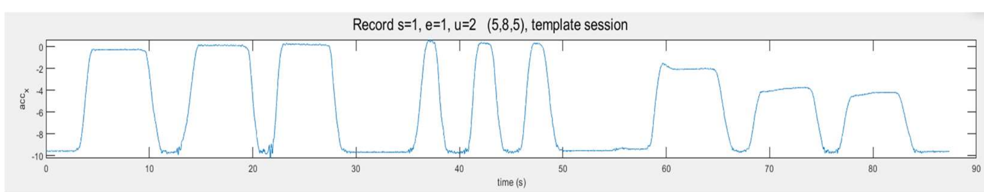

Figure 1. The record with correct template of the three movements (C, F, L)

The set of time series is imbalanced but in the papers that used this database [8-9], the set is uses as it is. Moreover, a closer analysis of signals reveals that this pattern of 3 repetition of type of exercise that are clearly well executed (Figure 1) can have only 2 repetitions on of sometime parts of exercises are wrong executed or in an incomplete form that visually inspection cannot identify one of the tree classes (C,L,F) to them belong the window of signal (Figure 2, Figure 3).

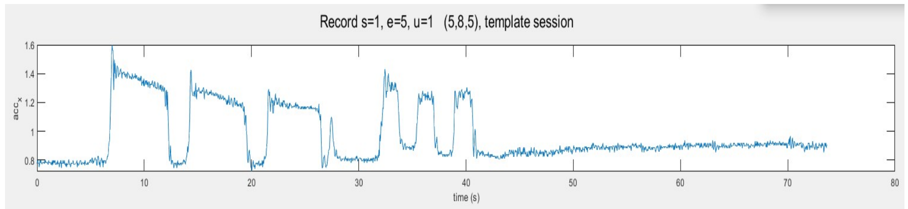

Figure 2. A record from template set that has an incomplete exercise set (C and L only)

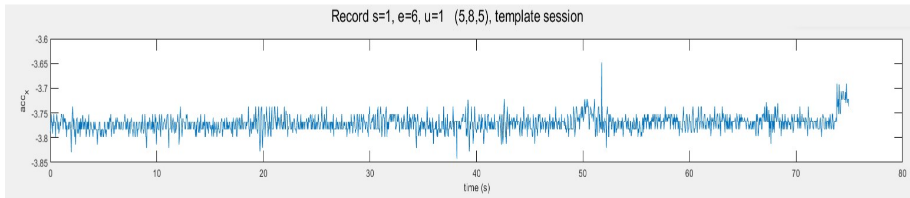

Figure 3. An example of template set with practically no movement in exercises (the class labeled as N in proposed application).

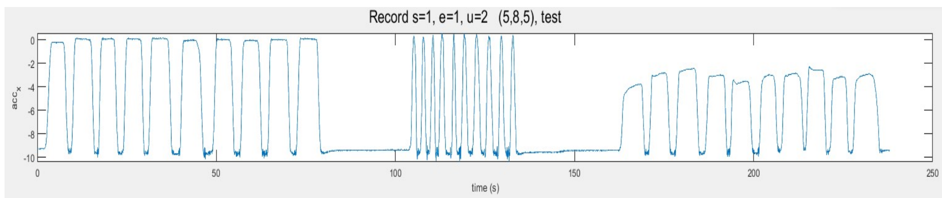

Figure 4. The record with correct correct of three set or repeated movements in test group (C, F, L).

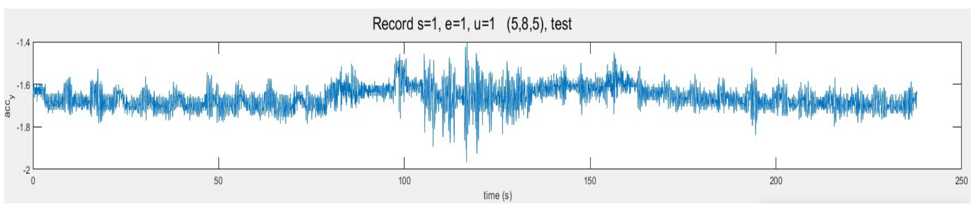

Figure 5. An example of template set with practically no movement in exercises in test group (the class labeled as N in proposed application).

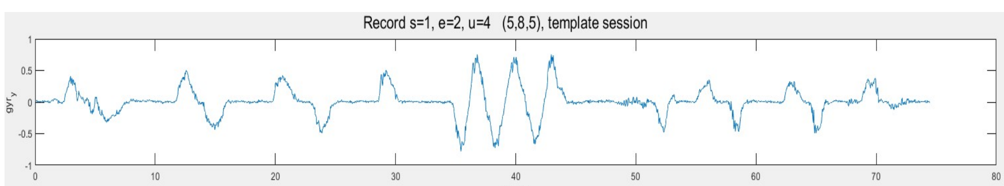

Figure 6. An example of template set for gyroscope sensor, template case (the classes C,F, L can be clearly segmented).

The authors split from start the records in template set and in test set. The situation described in the previous section occurs sometime in the test set (Figure 4, Figure 5). The shape of the single exercise for magnetometer type is close somewhat to the accelerometer one meanwhile the shape of gyroscope, when it is correct is very different one (Figure 6, Figure 7).

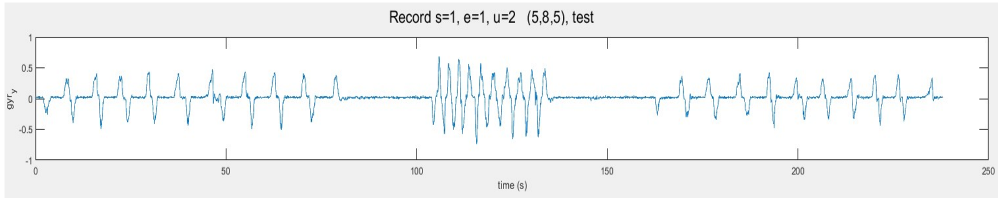

Figure 7. An example of template set for gyroscope sensor, test case (the classes C,F, L can be clearly segmented).

The test records for gyroscope sensors have many shapes that look like a noise but the distribution of the value is not close to a normal, Gaussian or Poisson one.

By this point of view, analyzing the records from dataset one by one in order to perform amanually labelling of classed as time series for segmentation there is a clearly imbalanced database, as the authors mentioned in [8-9]. They are commonly to main methods to use to generate suitable training data sets for imbalances set of records: oversampling and downsampling. While oversampling adds new samples of the minority class, undersampling (or downsampling) reduces the number of samples in the majority class [11-12]. In order to avoid the reduce the dataset, the oversampling method is preferred, the T-SMOTE as easiest to be implementable with good performances.

The waveforms are classes are labelled manually with a letter for each point in dataset. There are repeated exercises in group of three or ten in alternating classes in order C, F and L with not automatically sliced in records with one single exercise repeated. This complicated the problem so we take as sliding window the length of time series from template case w = 854. The records from template has the length in the range [854, 945] (the range is [5053, 502] for test case) so we will use a sliding window for take all in account all the values of dataset.

After segmentation the signal use a classifier (deep neural network - DNN) to predict the four classes: C, F, L and N. The accuracy (acc) for multiclass classification in this paper is defined simply as:

$$\text{acc= correct predictions/all predictions} \quad (1)$$

The accuracy obtained is greater than 95% and the average sensitivity is greater than 95%.

3. Discussion

Accuracy when is applied multiclass as metric of performance to reflects the overall correctness of the model [13]. However, the macro-averaged F1 score is the most suitable measure for multiclass classification. Cohen's Kappa Score can be sometime contradictory so it is avoided in this paper [14].

The results are clearly dependent of how was labelled the segments in template set. There are three specialist that made this but it is not official one. The concordant opinion was chosen rho represent the final labels for template and test case. So, it is clear that a subjective part (even in a very small percent) is present in the method.

Counterpart, the proposed method that uses a sliding window and segmentation doesn’t depend of how many repetitions occurs and how were ordered the sequence of groups of repeated exercise.

The accuracy of results in deep learning methods is depending on the training set, and the selection of what DNN learning in order to have good accuracy in the test case. By solving the imbalance problem of this dataset, the learning set offers the maximum possibility to DNN to learn a large types of dataset waveforms in order to perform good classification.

The results are very good, in some cases better the results presented in literature the justify the proposed algorithms.

4. Materials and Methods

The block diagram used in this paper is represented in Figure 8. The Segmentation Block (SB) has taken into account three DNN: LSTM, U-NET and U-NET++. The classifier block (CB) with four classes at outputs, softmax layer for classification in order to permit a Bayesian multi-class approach has three options: dense net (multilayer perceptron), AlexNet, GRU (Gated Recurrent Unit) and 1D-ResNet [15-16]. The best result is achieved with the combination U-NET and 1D-ResNet.

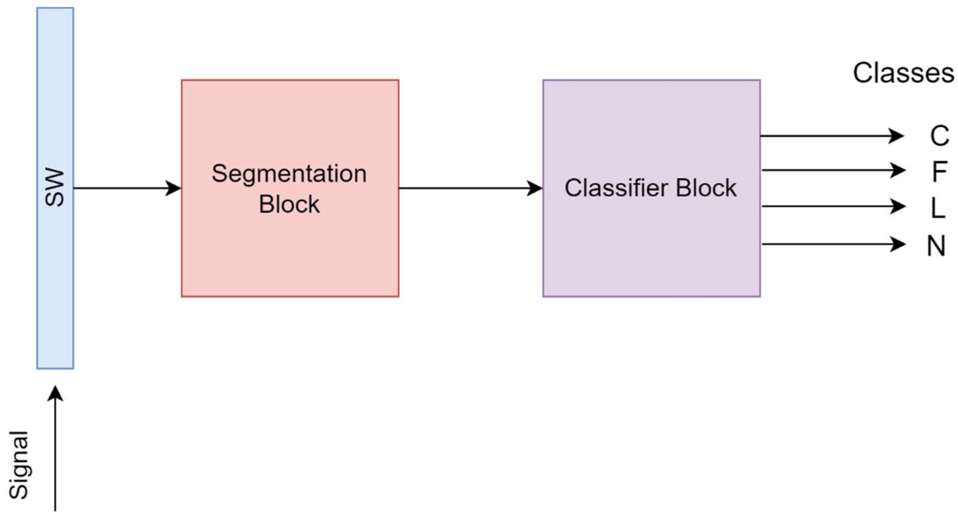

Figure 8. The block schema of the proposed method.

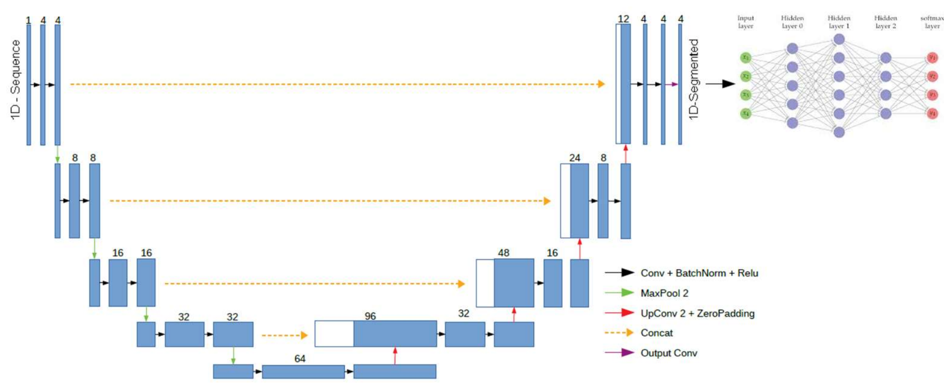

Figure 9. . The block schema of the proposed method: Segmentation block – U-net and Classifier Block – dense network.

The learning stage is split in two: SB is trained first to perform the maximum accuracy in the segmentation of 1D record. Thereafter the CB is trained using as input the segmented signals. The final decision is taken window with window as label of the maximum output softmax node. In the case of two ore more equal values, the decision is N, take by empirical analysis of signals that give this result.

In the further research, the approach is to use a fusion of SB and CB and perform a train on the fused network. Also the segmentation will be refined in order to have a more accurate representation by human specialists.

The method works actually for the dataset as it is available in [9]. If a person has some neuromuscular disease the record are clearly more complex and a more sophisticated analysis is necessary. Also, during rehabilitation, for the same patient will bet a sequential record from sensors, they can give a view over efficacy of therapy methods. But this need a set of new records at different moment of time (monthly or weekly, e.g.) and also specialist that evaluate the percent of patients' recovery. This approach is a target for future research.

Supplementary Materials: Not applicable.

Author Contributions:"Conceptualization, Mariana Rotariu.; methodology, Gheorghita Andrei.; software, Marius Turnea.; validation, Mihai Ilea; formal analysis, Iustina Condurache.; resources, Mariana Rotariu..; data curation, Mihai Ilea; writing—original draft preparation, Dragos Arotaritei; writing—Gheorghita Andrei; visualization, Marius Turnea.; supervision Marius Turnea; project administration, Marius Turnea.; funding acquisition, Mariana Rotariu All authors have read and agreed to the published version of the manuscript." Please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References:

1. Tognolo, L.; Daniele Coraci, D.; Fioravanti, A.; Tenti, S.; Scanu, A.; Magro, G.; Maccarone, M.C.; Masiero, S. Clinical Impact of Balneotherapy and Therapeutic Exercise in Rheumatic Diseases: A Lexical Analysis and Scoping Review. Appl. Sci. 2022, 12, 7379, doi: 10.3390/app12157379.
2. Tschuggnall, M.; Grote, G.; Pirchl, M.; Holzner, B.; Rumpold, G.; Fischer, M.J. Machine learning approaches to predict rehabilitation success based on clinical and patient-reported outcome measures. Informatics in Medicine Unlocked 2021, 24(100598), doi: doi.org/10.1016/j.imu.2021.100598.
3. Mottaghi, E.; Akbarzadeh-T, M.-R. Automatic Evaluation of Motor Rehabilitation Exercises Based on Deep Mixture Density Neural Networks. Journal of Biomedical Informatics 2022, 130(104077), doi: 10.1016/j.jbi.2022.104077.
4. Deotale, D.; Madhushi Verma, M.; Suresh, P.; Kumar, N. Physiotherapy-based human activity recognition using deep learning. Neural Computing and Applications 2023, 35, pp. 11431–11444, doi: 10.1007/s00521-023-08307-4.
5. Spilz, A.; Munz, M. Automatic Assessment of Functional Movement Screening Exercises with Deep Learning Architectures. Sensors 2023, 23(5), doi: 10.3390/s23010005.
6. Zhu, Z.-A.; Lu, Y.-C.; You, C.-H.; Chiang, C.-K.. Deep Learning for Sensor-Based Rehabilitation Exercise Recognition and Evaluation. Sensors 2019, 19(887), doi:10.3390/s19040887.
7. Zhang, W.; Su, C.; Chuan He, C. Rehabilitation Exercise Recognition and Evaluation based on smart Sensors with Deep Learning Framework. IEEE Access 2019. doi: 10.1109/ACCESS.2017.
8. Carrera, E.V.; Arequipa, X.; Hernández, B. Automatic Evaluation of Physiotherapy Activities Using Deep Learning Techniques. In: Botto-Tobar, M., Zambrano Vizuete, M., Montes León, S., Torres-Carrión, P., Durakovic, B. (eds) Applied Technologies. ICAT 2022. Communications in Computer and Information Sciencel 2023, vol 1755. Springer, Cham, pp 143–154, doi: 10.1007/978-3-031-24985-3_11.
9. Yurtman, A.; Barshan, B. Automated evaluation of physical therapy exercises using multi-template dynamic time warping on wearable sensor signals. Computers and Methods in Biomedicine 2014, 117(2), pp. 189-207, doi: 10.1016/j.cmpb.2014.07.003.
10. https://archive.ics.uci.edu/dataset/730/physical+therapy+exercises+dataset
11. Zhao, P.; Luo, C.; Qiao, B.; Wang, L.; Rajmohan, S.; Lin, Q.; Zhang, D. T-SMOTE: Temporal-oriented Synthetic Minority Oversampling Technique for Imbalanced Time Series Classification. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence IJCAI-22, Messe Wien, Vienna, Austria, July 23-29, 2022, pp. 2406-2412.

12. Zhu, T.; Luo, C.; Zhang, Z.; Jing Li, J.; Ren, S.; Zeng, Y. Minority oversampling for imbalanced time series classification. Knowledge-Based Systems 2022. 247(108764), doi: 10.1016/j.knosys.2022.108764.
13. Grandini, M.; Bagli, E.; Visani, G. Metrics for Multi-Class Classification: an Overview. https://arxiv.org/abs/2008.05756
14. Delgado, R.; Xavier-Andoni Tibau, X.A. Why Cohen's Kappa should be avoided as performance measure in classification. PLoS ONE 14(9):e022291 2019. doi: 10.1371/journal.pone.0222916.
15. Theisser, A.; Spinnato, F.; Schegel, U.; Guidotti, R. Explainable AI for Time Series Classification: A Review, Taxonomy and Research Directions. IEEE Access 10 2022. pp. 100700. 10.1109/ACCESS.2022.3207765
16. Fawaz, H.I.; Forestier, G.; Weber, J; et al. Deep learning for time series classification: a review. Data Min Knowl Disc 33 2019. pp. 917–963. doi: 10.1007/s10618-019-00619-1.