Enhancing Lower Limb Activity Recognition Through Multi-Sensor Data Fusion in Tele-Rehabilitation

Alireza Ettefagh (✉ alireza.ettefagh@mail.utoronto.ca)
KITE Research Institute

Atena Roshan Fekr
KITE Research Institute

Research Article

Keywords: Tele-rehabilitation, Exercise recognition, Classification, Data fusion, Convolutional Neural Network, Deep learning

Posted Date: September 6th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3318478/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.
Enhancing Lower Limb Activity Recognition Through Multi-Sensor Data Fusion in Tele-Rehabilitation

Alireza Ettefagh\textsuperscript{1,2}\textsuperscript{*} and Atena Roshan Fekr\textsuperscript{1,2}

\textsuperscript{1}\textsuperscript{*}KITE Research Institute, Toronto Rehabilitation Institute, University Health Network, 550 University Ave., Toronto, M5G 2A2, Ontario, Canada.
\textsuperscript{2}Institute of Biomedical Engineering, University of Toronto, 164 College St., Toronto, M5S 3E2, Ontario, Canada.

*Corresponding author(s). E-mail(s): alireza.ettefagh@mail.utoronto.ca; atena.roshanfekr@uhn.ca;
Contributing authors: atena.roshanfekr@uhn.ca;

Abstract

Background: Tele-rehabilitation is the provision of physiotherapy services to individuals in their own homes. Activity recognition plays a crucial role in the realm of automatic tele-rehabilitation. By assessing patient movements, identifying exercises, and providing feedback, these platforms can offer insightful information to clinicians, thereby facilitating an improved plan of care. This study introduces a novel deep learning approach aimed at identifying lower-limb rehabilitation exercises. This is achieved through the integration of depth data and pressure heatmaps. We hypothesized that combining pressure heatmaps and depth data could improve the model’s overall performance.

Methods: In this study, depth videos and body pressure data from an accessible online dataset were used. This dataset comprises data from 30 healthy individuals performing 7 lower-limb rehabilitation exercises. To accomplish the classification task, three deep learning models were developed, all based on a 3D-CNN architecture. The models were designed to classify the depth videos, sequences of pressure data frames, and combination of depth videos and pressure frames.
The models’ performance was assessed through leave-one-subject-out and leave-
multiple-subjects-out cross-validation methods. Performance metrics, including
accuracy, precision, recall, and F1-score, were reported for each model.

**Results:** Our findings indicated that the model trained on the fusion of depth
and pressure data showed the highest and most stable performance when com-
pared with models using individual modality inputs. This model could effectively
identify the exercises with an accuracy of 95.71%, precision of 95.83%, recall of
95.71%, and an F1-score of 95.74%.

**Conclusion:** Our results highlight the impact of data fusion for accurately classi-
ifying lower-limb rehabilitation exercises. We showed that our model could capture
different aspects of exercise movements using the visual and weight distribution
data from the depth camera and pressure mat, respectively. This integration of
data provides a better representation of exercise patterns, leading to higher clas-
sification performance. Notably, our results indicate the potential application of
this model in automatic tele-rehabilitation platforms.

**Keywords:** Tele-rehabilitation, Exercise recognition, Classification, Data fusion,
Convolutional Neural Network, Deep learning

## 1 Introduction

Regular rehabilitation services are essential for patients who suffer from Musculoskele-
tal Disorders (MSDs). MSDs encompass a wide range of conditions that can cause
chronic pain, mobility impairment, falls, and a decreased quality of life. These disor-
ders primarily affect the muscles, tendons, nerves, ligaments, and other tissues of the
body, often leading to inflammation, pain, discomfort, or tingling sensations. Among
the various types of MSDs, Lower Limb Disorders (LLDs) specifically target different
regions of the lower body, including the hip, thigh, knee, calf, ankle, and foot [1]. These
LLDs negatively impact an individual’s ability to move and perform daily activities
of daily living.

Following cancer and cardiovascular diseases, MSD stands as the third leading
cause of disease burden in Canada [2]. According to [2], the all-age prevalence of
various MSD conditions increased from 23% in 1990 to 27.8% in 2017. As a result, in
2017, Canada ranked among the top 10 countries globally for the prevalence of several
MSDs, such as osteoarthritis and gout. Regular exercise in rehabilitation programs
plays a vital role in the management of MSD conditions. This highlights the need for automatic rehabilitation solutions to address the consequences of this growing issue.

Tele-rehabilitation (tele-rehab) is the delivery of medical or rehabilitative services to patients using tele-communication or the internet [3]. Despite the existence of tele-rehab for several years, its adoption in clinical practice has been limited due to various factors. These include concerns regarding the costs, complexity of implementation, low accuracy, and high incidence of false alarms. These challenges have inhibited the widespread use of tele-rehab solutions and have prevented their full potential from being realized in healthcare settings.

One important feature of an automatic tele-rehab platform is activity recognition, which refers to the process of automatically identifying human activities based on sensor data or visual inputs. In essence, analyzing and understanding movements performed by individuals during their rehabilitation therapy offer valuable insights to clinicians for developing their care plans. These platforms should have the capability to recognize and evaluate different exercises and deliver meaningful feedback to help patients refine their movements and optimize their plan of care. To address the exercise recognition problem, in this paper, a novel deep learning approach is proposed to classify different lower-limb rehabilitation exercises using privacy-preserving depth information and pressure data.

Several studies have also employed different machine learning techniques to perform exercise recognition based on various input data. For instance, Anton et al. developed a system using Kinect technology to monitor and evaluate the type and quality of physical rehabilitation exercises in real time [4]. Their system employed two methods: posture classification and exercise recognition. By capturing the spatial coordinates of body joints, the algorithm calculated relative positions, joint angles, and limb angles. These measurements were used to create a posture descriptor consisting of 30 features. Posture classification was performed by comparing the captured
descriptor with prestored posture descriptors using Dynamic Time Warping (DTW).

For exercise recognition, the system identified the starting and ending postures of each exercise and utilized DTW-based trajectory recognition to assess the accuracy of movement patterns. The proposed algorithm was evaluated through clinical trials involving 15 patients with shoulder disorders. They obtained an accuracy of 95.16% in recognizing 4 different shoulder exercises.

Barriga et al. introduced a vision-based system for telecare and tele-rehabilitation using a depth camera and neural networks [5]. They claimed that their system has the capability to automatically classify 7 static postures and falls. The system’s performance was validated using data collected from 6 participants. The researchers also investigated various parameters, including the number of hidden neurons, maximum error, learning rate, and learning function, in the design of their neural network. Additionally, they explored the impact of distance from the camera and the angle between the camera and subjects in the skeleton tracking system. Through their experiments, they achieved an accuracy of 96% for classifying static postures and detecting falls.

Decroos et al. developed a machine learning pipeline using Kinect to monitor and assess the correctness of physiotherapy exercises performed by patients at home [6]. Their pipeline involved three main steps: identifying individual exercise repetitions, representing time-series data with statistical features about joint angles, and detecting the exercise’s type, correctness, and possible mistakes. To evaluate the performance of their method, they recorded 10 healthy participants performing 3 rehab exercises (squats, forward lunges, and side lunges) while tracking joint movements with Kinect. For exercise recognition, they used 5 learners, including Linear Regression, Naïve Bayes, Decision Tree, Random Forest, and XGBoost. The input feature vector to the learners consisted of 150 summary statistics (30 joint angles x 5 statistics - min, max, mean, median, std) for each exercise repetition. The best accuracy achieved was 99% using XGBoost algorithm with LOSO cross-validation.
Bijalwan et al. proposed a heterogeneous deep learning model to identify lower-limb rehab exercises [7]. To this end, they considered a total of 10 exercises involving abduction, flexion, rotation, and dorsi-flexion of the lower limb on both the left and right sides. These exercises were performed by 25 healthy and 10 crouch walking subjects. Depth data were collected from a Kinect v2 sensor. They employed CNN and CNN-LSTM models to classify these exercises. For validation, a hold-out validation approach was employed, with the dataset split into 50% for training, 20% for validation, and 30% for testing. Their experimental results demonstrated both accuracies and F1-scores of 96% for the CNN model and 98% for the CNN-LSTM model.

Barzegar et al. proposed a vision-based system to assess the quality of rehab exercises [8]. They used an open dataset consisting of 16 patients and 14 healthy participants performing 9 different rehabilitation exercises. Data were depth videos recorded from a Kinect 1 sensor. They used a pretrained 3D convolutional neural network to perform exercise recognition on correctly executed data as a part of their assessment system. They obtained average accuracies of 96.62% and 86.04% in identifying the exercises using 10-fold and Leave-One-Subject-Out (LOSO) cross validations, respectively.

Wijekoon et al. introduced the Multi-modal Exercises Dataset (MEx) as a multi-sensor Human Activity Recognition (HAR) dataset [9]. The data collection involved a pressure mat and a depth camera, both operating at 15 Hz, and two accelerometers operating at 100 Hz. One accelerometer was positioned on the thigh, while the other was placed on the wrist. The dataset includes 7 lower-limb exercises performed by 30 healthy participants. Through Leave-Multiple-Subjects-Out (LMSO) cross-validation, the average F1-scores for exercise recognition were 86.34%, 88.92%, 64.99%, and 71.95% using depth data, thigh accelerometer data, wrist accelerometer data, and pressure data, respectively. This study concluded that vision data such as...
depth provided better results than the time-series data from accelerometers. In subsequent work, the authors proposed a multi-modal Hybrid Attention Fusion (mHAF) deep learning architecture [10]. With a combination of pressure mat, depth camera, and thigh accelerometer data, they achieved an F1-score of 96.24% for exercise recognition using LOSO cross-validation. When pressure and depth data were used without the accelerometer data, the performance was reduced to 90.41%.

Wearable technology shows great potential for lower limb tele-rehab systems. For example, Lai et al. achieved 99% accuracy in recognizing 6 lower-limb exercises using one Inertial Measurement Unit (IMU). The IMU was attached to the knee for 4 exercises and instep for the other two [11]. García-de-Villa et al. classified 8 exercises (5 lower limbs) with 96.2% accuracy [12]. Kim et al. also detected Sarcopenia patients with 95% accuracy using IMUs mounted on the left and right feet [13]. Albeit useful, using wearables would be challenging for seniors. One primary obstacle in using wearable technology for seniors is the difficulty they may face in accurately positioning the sensors on their body. They may require external help to properly place the sensors at the appropriate location, angle, and direction. Additionally, research suggests that many older adults are not keen on using such technology. They prefer their usual routines without electronic devices [14]. As a result, they might hesitate to wear sensors on their bodies.

Vision-based technology as a contactless approach offers a great alternative to wearables. These systems often use skeleton tracking models to locate body joints. Such models require RGB data to capture body limb movements, which involves recording images or videos of users. This raises potential privacy concerns, as it involves capturing and processing visual information of individuals within their private living spaces. Patients may feel uncomfortable knowing that their movements and activities are being monitored through RGB cameras. This may lead to potential reluctance in using such technologies. Researchers have used depth cameras to mitigate this challenge.
The depth data captures only an outline of the body, ensuring complete anonymity. One challenge with vision-based systems is occlusion, where the joints and body parts are hidden from the camera [15]. This is even more likely to happen for exercises that should be performed in lying down positions, i.e., lower-limb exercises. The presence of occlusion can negatively impact the performance of exercise recognition models, leading to a decrease in accuracy and reliability.

Given the potential challenges discussed above, we aim to fill these gaps by using the fusion of depth and pressure heatmaps. Depth data can provide information about the pattern of body movements without the need for intrusive RGB visuals. Additionally, pressure data can offer insights into the patterns of body limbs and the force exerted on the ground by them during exercises. Our hypothesis is that the combination of pressure distribution data and depth data can enhance a deep learning model’s ability to differentiate between various types of exercises. By leveraging these alternative data sources, we strive to create a more user-friendly and privacy-conscious approach for exercise recognition in lower limb tele-rehabilitation.

2 Methods

2.1 Dataset

This study used depth video recordings and body pressure data from the dataset published by Wijekoon et al. [9]. The data were collected from 30 healthy participants, comprising 18 females and 12 males. Fourteen subjects were aged 18 to 24, while the rest of the individuals were aged 24 to 54. It is noteworthy that 8 participants had some background in physiotherapy, either as physiotherapists or physiotherapy students, thus having a good knowledge of the exercises. The participants performed the 7 different lower-limb rehabilitation exercises listed in Table 1. These exercises are frequently recommended by clinicians for the prevention or management of musculoskeletal pain [9]. The participants performed all exercises while lying down on the
<table>
<thead>
<tr>
<th>Exercise</th>
<th>Starting Position</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knee-rolling (KR)</td>
<td>Lying on back, knees bent</td>
<td>Roll knees side to side, keeping upper body still</td>
</tr>
<tr>
<td>Bridging (BG)</td>
<td>Lying on back, knees bent</td>
<td>Lift hips off floor, hold for 5 seconds, and lower</td>
</tr>
<tr>
<td>Pelvic tilt (PT)</td>
<td>Lying on back, knees bent</td>
<td>Tighten stomach muscles, press lower back to floor, rise bottom, hold for 5 seconds, relax</td>
</tr>
<tr>
<td>The Clam (TC)</td>
<td>Lying on side, knees bent</td>
<td>Rotate leg and open knee while keeping hips aligned, return to starting position</td>
</tr>
<tr>
<td>Repeated Extension in Lying (EL)</td>
<td>Lying face down, palms on floor</td>
<td>Straighten elbows, push upper body up for 2 seconds, and lower back down</td>
</tr>
<tr>
<td>Prone punches (PP)</td>
<td>On all 4s</td>
<td>Punch arms forward while keeping the core stable</td>
</tr>
<tr>
<td>Superman (SM)</td>
<td>On all 4s</td>
<td>Extend the opposite arm and leg for 5 seconds while keeping the core stable</td>
</tr>
</tbody>
</table>
Fig. 1 Seven exercises performed by Subject #1 and their corresponding depth frames [10]

Fig. 2 Seven exercises performed by Subject #1 and their corresponding pressure frames [10]
2.2 Methodology

We created three exercise recognition models to classify the following: 1- depth videos (DC), 2- sequences of pressure data frames (PM), and 3- concatenated depth videos and pressure frames (DC-PM). These models were developed to classify all 7 exercises in the dataset. We used a pretrained 3D CNN model proposed by Carreira et al. This state-of-the-art network, known as “Inflated 3D ConvNets” (I3D), was trained on the Kinetics dataset, which comprises a total of 240,000 training videos of 400 different human actions, including person actions, e.g., drawing; actions involving interactions between individuals and objects, e.g., washing dishes; and actions involving interactions between individuals, e.g., hugging. This model achieved an accuracy of 74.1% when applied to the RGB data from the Kinetics dataset. Also, after pretraining on both ImageNet and Kinetics, it demonstrated accuracies of 97.9% and 96.9% when tested on UCF-101 [16] and HMDB-51 [17] datasets, respectively [18].

The I3D model uses 3D convolution to learn spatiotemporal information directly from input videos [19]. More specifically, the architecture consists of a series of 3D Inception modules followed by 3D max pooling and batch normalization layers. The Inception module, as depicted in Figure 3, operates with parallel $1 \times 1 \times 1$ and $3 \times 3 \times 3$ 3D convolution kernels and a $3 \times 3 \times 3$ max pooling operation using the same input data, merging their outputs into a single output. Incorporating $1 \times 1 \times 1$ convolution layers reduces the dimensions of the input data within the network and therefore reduces the computational cost. The $3 \times 3 \times 3$ convolution layer enables the network to learn spatiotemporal features at a different scale. The dimensions of the input data are reduced by the $3 \times 3 \times 3$ max-pooling layer while allowing the extraction of different features simultaneously. Max-pooling is thus employed to extract more features from the input data [20]. A dropout layer was also used to prevent overfitting of the models.

The input for the 3D models consists of videos with a size of $N \times R \times C \times 3$, where $N$ represents the number of frames in the video. Each frame has a resolution of
R × C × 3, where R and C are the number of rows and columns, respectively. Also, 3 indicates the number of channels.

For preprocessing, each depth video was downsampled to 158 frames, which is the shortest length of depth videos in the dataset. Each frame was zero-padded to 32×32 pixels. Likewise, the pressure videos were downsampled to 252 frames, and each frame was zero-padded to 32×32 pixels. To create the concatenated input video, the pressure videos were also downsampled to 158 frames to be consistent with the depth data. Corresponding depth and pressure frames were zero-padded and concatenated next to each other to form 32 × 64 input frames. An example of input videos for PM (top frames) and DC of knee-rolling exercise is depicted in Figure 4. To classify the performed exercises, we used categorical cross entropy as the loss function in our classification task. The categorical cross-entropy loss function is defined as follows:

\[
f(x) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} 1_{y_i \in C_c} \log(p_{model}(y_i \in C_c))
\]  

(1)
The summation is performed over $N$ observations (the training sample size), where $i$ iterates over the observations and $c$ iterates over the number of classes (exercises). The term $1_{y_i \in C_c}$ represents an indicator function that equals 1 if the $i^{th}$ observation belongs to the $c^{th}$ category and 0 otherwise. The logarithm of the predicted probability by the model for the $i^{th}$ observation belonging to the $c^{th}$ category is calculated. The objective was to minimize this loss function during the training phase. To optimize the model, we employed the Adam optimizer \[21\]. The Adam optimizer adapts the learning rate for each weight of the neural network by using estimations of the first moment and the second moment of the gradient. This adaptive learning rate scheme aids in effectively updating the weights during the training process.

The models were validated using two cross-validation techniques: LOSO and LMSO with 6 groups of 5 individuals. LOSO cross-validation mimics the practical situation where our models encounter new individuals, one at a time, during its application. In addition, the LMSO cross-validation goes beyond LOSO by simulating scenarios where the model is exposed to completely new groups of subjects. To determine the optimal training hyperparameters, including batch size, learning rate, and number of epochs, we employed 5-fold cross-validation with grid search. The performance of the models was evaluated using Eq. (2-4) as follows:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \quad \text{Recall} = \frac{TP}{TP + FN}
\]

The macro F1-score is computed by taking the average of the F1-scores for each exercise. The F1-score for each exercise is determined by calculating the harmonic
mean of precision and recall:

\[
F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}
\]  

(4)

Where \(TP\), \(TN\), \(FP\), and \(FN\) are the number of true positives, true negatives, false positives, and false negatives in the classification of each exercise, respectively.

### 3 Results and Discussion

Among all models, a batch size of 4 yielded the best results. The learning rate was set to \(5 \times 10^{-5}\) for the DC and PM models and \(1 \times 10^{-4}\) for the DC-PM model. For the DC and PM models, the best epoch size was found to be 76, while this was 60 for the DC-PM model. These hyperparameter settings were found to be optimal for both LMSO and LOSO cross-validations. Table 2 presents the classification performance of all three models with both LOSO and LMSO. This table shows that the PM model consistently provides the lowest performance among the other models. When considering classification accuracy, there is a relatively large variance across different subjects. This difference might arise from the pressure mat’s capability to capture
individual characteristics, such as weight distribution and body shape \[9\]. This model identifies exercises by analyzing the pattern of body parts in contact with the ground and the force applied to the ground by the active limbs. It uses pressure patterns to determine the exercise type, focusing on how the body engages with the ground during the movement.

While pressure data could be indicative of exercise type based on pressure patterns, they do not capture the same level of detailed information about body movements as the depth camera. As shown in Figures 1 and 2, depth data provide a better view of the body during exercise and capture the entire movement sequence. It includes information about all body parts and their positions relative to the camera. As shown in Table 2, the DC model could better classify the 7 types of exercises with approximately 94% accuracy.

The DC-PM model, which combines both depth camera and pressure mat data, was the most accurate model in identifying the exercises in LOSO. In LMSO, all models experience a decrease in performance compared to LOSO, which is expected due to the reduced subject-specific data for training. Despite the drop in performance, the DC-PM model still provided the highest performance among the other two models. The DC-PM model also demonstrates the most consistent outcomes, as indicated by its low standard deviation in Table 2. This improvement in performance can be attributed to the complementary nature of the two data modalities and how they, as a group, address the limitations of the individual models. More specifically, the combination of visual and weight distribution information from the depth camera and pressure mat allows the model to capture different aspects of exercise movements. This fusion of data provides a richer representation of exercise patterns, leading to higher classification performance. Figure 5 (a-c, j-l) presents the confusion matrices for the DC, PM, and DC-PM models, respectively. The misclassified data by each model can be found in Figure 5 (d-f, m-o). The F1-scores for all exercises are displayed in Figure 5 (g-i, p-r).
Fig. 5  (a)-(c) Confusion matrices, (d)-(f) the proportion of misclassified labels, and (g)-(i) f1-score per exercise for the DC, PM and DC-PM models, respectively, considering the LMSO technique. (j)-(l) Confusion matrices, (m)-(o) the proportion of misclassified labels, and (p)-(r) f1-score per exercise for the DC, PM and DC-PM models, respectively, considering the LOSO technique.

In most cases, the Bridging (BG) and Pelvic Tilt (PT) exercises were misclassified by each other. This is likely due to their similar starting positions and body trajectories, as evident in Figures 1 and 2. Additionally, the Prone Punches (PP) and Superman (SM) exercises were almost perfectly classified when using depth data; however, considering the pressure data, they were misclassified by the other. Looking at Figures 1 and 2, it is evident that the Repeated Extension in Lying (EL) exercise has distinctive patterns in both depth and pressure data. This exercise, thus, had the lowest misclassification rate when considering the DC and PM models individually.
To further analyze the results, we used the Gradient-weighted Class Activation Mapping (Grad-CAM) [22] as a technique to visualize and understand the decision-making process of deep learning models. Figure 6(a-c) displays a sample Grad-CAM frame for EL, TC, and BG exercises, respectively. The more intense red colors represent the areas of the body heatmap that the model was more focused on and considered significant for the prediction.

For the EL exercise in Figure 6(a), the model predominantly focused on the depth part of the input to make a prediction. Conversely, in the BG exercise (Figure 5(c)), the model relied more on the pressure data for the prediction. For the TC exercise in Figure 5(b), the model’s attention was distributed across both the depth and pressure parts of the input data.

![GradCAM visualization for (a) EL, (b) TC, and (c) BG exercises](image)

Table 3 presents a comparison between the findings of previous studies and our own study.

<table>
<thead>
<tr>
<th>Model</th>
<th>Cross Validation</th>
<th>Data Modality</th>
<th>Network Architecture</th>
<th>F1-Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[9] and [10]</td>
<td>LOSO</td>
<td>DC-PM</td>
<td>Hybrid Attention Fusion</td>
<td>90.41</td>
</tr>
<tr>
<td></td>
<td>LMSO</td>
<td>PM</td>
<td>1D-CNN-LSTM</td>
<td>74.08</td>
</tr>
<tr>
<td>This study</td>
<td>LOSO</td>
<td>DC-PM</td>
<td>I3D</td>
<td>95.74</td>
</tr>
<tr>
<td></td>
<td>LMSO</td>
<td>PM</td>
<td>I3D</td>
<td>75.28</td>
</tr>
</tbody>
</table>
We found that the I3D model performs better than the 1D-CNN and 2D-CNN models in all situations. Given that 1D-CNN and 2D-CNN use flattened data, there is a potential for loss of spatial information. In contrast, the I3D model analyzes the data in both spatial and temporal dimensions, allowing it to effectively capture patterns and dynamics in the exercise sequences.

4 Conclusion

In this study, we present a state-of-the-art 3D-CNN model capable of recognizing lower-limb rehabilitation exercises using privacy-preserving depth information and pressure data from an available online dataset. The dataset consisted of a total of 210 videos of 30 healthy individuals performing 7 exercises. We evaluated the effectiveness of this model with three different inputs: depth data, pressure data, and concatenated depth and pressure data. With LOSO cross-validation, the model demonstrated macro F1-scores of 93.80%, 81.45%, and 95.74% for depth data, pressure data, and concatenated data, respectively. Similarly, with LMSO cross-validation, the performance was 90.83%, 75.28%, and 94.77% for depth data, pressure data, and concatenated data, respectively. This outcome highlights the impact of data fusion for accurately classifying the exercises, both in the LOSO and LMSO scenarios. The proposed 3D-CNN model outperforms the previous models as it can analyze data in both spatial and temporal dimensions. Due to its high accuracy, it can be readily applied to recognize lower limb exercises in automatic tele-rehabilitation applications.

Declarations

- Funding

  This research has been supported by Mitacs Accelerate Program.

- Competing interests

  The authors declare that they have no competing interests.
• Ethics approval
This study used an openly accessible dataset retrieved from https://archive.ics.uci.edu/dataset/500/mex

• Consent to participate
Not applicable.

• Consent for publication
Not applicable.

• Availability of data and materials
The datasets generated and/or analysed during the current study are available in the UCI Machine Learning repository, https://archive.ics.uci.edu/dataset/500/mex

• Code availability
Not applicable.

• Authors’ contributions
Conceptualization, A.E. and A.R.; methodology, A.E.; software, A.E.; validation, A.E.; formal analysis, A.E and A.R.; investigation, A.E and A.R.; writing—original draft preparation, A.E.; writing—review and editing, A.R.; visualization, A.E and A.R.; supervision, A.R.; funding acquisition, A.R. All authors have read and agreed to the published version of the manuscript.
References


