DS-MS-TCN: Otago Exercises Recognition With a Dual-Scale Multi-Stage Temporal Convolutional Network

Meng Shang, Lenore Dedeyne, Jolan Dupont, Laura Vercauteren, Nadjia Amini, Laurence Lapauw, Evelien Gielen, Sabine Verschueren, Carolina Varon, Walter De Raedt, and Bart Vanrumste, *Senior Member, IEEE*

***Abstract*—The Otago Exercise Program (OEP) represents a crucial rehabilitation initiative tailored for older adults, aimed at enhancing balance and strength. Despite previous efforts utilizing wearable sensors for OEP recognition, existing studies have exhibited limitations in terms of accuracy and robustness. This study addresses these limitations by employing a single waist-mounted Inertial Measurement Unit (IMU) to recognize OEP exercises among community-dwelling older adults in their daily lives. A cohort of 36 older adults participated in laboratory settings, supplemented by an additional 7 older adults recruited for at-home assessments. The study proposes a Dual-Scale Multi-Stage Temporal Convolutional Network (DS-MS-TCN) designed for two-level sequence-to-sequence classification, incorporating them in one loss function. In the first stage, the model focuses on recognizing each repetition of the exercises (micro labels). Subsequent stages extend the recognition to encompass the complete range of exercises (macro labels). The DS-MS-TCN model surpasses existing state-of-the-art deep learning models, achieving f1-scores exceeding 80% and Intersection over Union (IoU) f1-scores surpassing 60% for all four exercises evaluated. Notably, the model outperforms the prior study utilizing the sliding window technique, eliminating the need for post-processing stages and window size tuning. To our knowledge, we are the first to present a novel perspective on enhancing Human Activity Recognition (HAR) systems through the recognition of each repetition of activities.**

***Index Terms*—Human activity recognition (HAR), wearable sensors, temporal convolutional network, otago exercises, micro activities.**

Received 18 February 2024; revised 20 July 2024; accepted 29 August 2024. Date of publication 6 September 2024; date of current version 6 December 2024. This work was supported in part by the China Scholarship Council (CSC) with the application number of 202007650007 and Province Flemish Brabant Smart Region Project: AI@WZC under Grant 3E230623, in part by ENHANce Project under Grant S60763, and in part by the Research Foundation Flanders (FWO) under Grant G099721N. The funding provider did not contribute or influence the design of the study and data collection, analysis and interpretation in writing this manuscript. *(Corresponding author: Meng Shang.)*

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was grated by Ethics Committee Research UZ/KU Leuven under Application No. S59660 and S60763.

Meng Shang is with the KU Leuven, STADIUS, Department of Electrical Engineering, 3000 Leuven, Belgium, also with the Imec, 3001 Leuven, Belgium, and also with the KU Leuven, e-Media Research Lab, 3000 Leuven, Belgium (e-mail: meng.shang@kuleuven.be).

Lenore Dedeyne, Laura Vercauteren, Nadjia Amini, and Laurence Lapauw are with the Geriatrics & Gerontology, Department of Public Health and Primary Care, KU Leuven, 3000 Leuven, Belgium.

Jolan Dupont and Evelien Gielen are with the Department of Geriatric Medicine, UZ Leuven, 3000 Leuven, Belgium, and also with the Geriatrics & Gerontology, Department of Public Health and Primary Care, KU Leuven, 3000 Leuven, Belgium.

Sabine Verschueren is with the Musculoskeletal Rehabilitation Research Group, Department of Rehabilitation Sciences, KU Leuven, 3000 Leuven, Belgium.

Carolina Varon is with the KU Leuven, STADIUS, Department of Electrical Engineering, 3000 Leuven, Belgium.

Walter De Raedt is with the Imec, 3001 Leuven, Belgium.

Bart Vanrumste is with the KU Leuven, STADIUS, Department of Electrical Engineering, 3000 Leuven, Belgium, and also with KU Leuven, e-Media Research Lab, 3000 Leuven, Belgium.

Digital Object Identifier 10.1109/JBHI.2024.3455426

I. Introduction

HUMAN activity recognition (HAR) has emerged as a significant research area in recent years due to its wide range of applications in various domains, including healthcare [1], sports monitoring [2], and smart environments [3]. Wearable sensors, such as inertial measurement units (IMUs), are popular for HAR since they provide a non-invasive and unobtrusive means of capturing human motion and activity patterns. HAR systems focus on developing machine learning algorithms and systems capable of automatically identifying and classifying human activities based on sensor data.

The Otago Exercise Program (OEP) comprises a set of activities designed to enhance balance, strength, and walking abilities, thereby mitigating the risk of falls in older adults [4]. Participants are instructed to engage in these exercises two or three times a week over consecutive weeks [5], [6]. To overcome the drawbacks of self-reports (diaries) such as high recall errors and substantial workload, researchers have proposed HAR systems using wearable IMUs. In our previous study [7], a hierarchical system was proposed and two tasks were performed: first, the OEP was recognized from Activities of Daily Life (ADLs) with f1-scores over 0.95, and second, some OEP sub-classes were recognized with f1-scores over 0.8. However, the systems required post-processing stages to smooth the output. Also, the model applied two stages with different sliding window sizes for signal segmentation and feature extraction, which introduced more hyperparameters to tune.

2168-2194 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on January 29,2025 at 10:01:05 UTC from IEEE Xplore. Restrictions apply.

Many dynamic human activities consist of repetitive movements (e.g., walking, eating, cycling). To date, much of the research up to now has annotated the daily activities and repetitive exercises as a whole [1], [8]. In other words, all movements belonging to the same activities were given the same ground truth label. However, this method also annotates some irrelevant movements/gestures of the activity. For example, *chair rising* as a common rehabilitation exercise, requires the repetition of sitting down and standing up. However, there are some in-between static sitting and standing gestures that are irrelevant. These irrelevant movements/gestures are also labeled as *chair rising* although they are actually not distinguishable. For food intake recognition [9], it has been proven possible to only annotate and recognize each cycle of a single gesture (micro movements) based on seq-to-seq classification. However, for HAR, such an annotation method has not been explored much.

The aim of this study is to develop a robust HAR system for OEP recognition for community-dwelling older adults. This study proposes a Dual-Scale MS-TCN (DS-MS-TCN) model with an innovative method to label and classify sensor data, called micro labeling. During annotation, each sample was given two labels: a label for each repetition of activities (micro label) and a label for the whole period of activities (macro label). Then the proposed DS-MS-TCN was applied to classify the labels. The first stage of the DS-MS-TCN learns from the raw sensor data and classifies micro labels. The following stages only learn from the output probability of the previous stage and classify macro labels. The system was applied to predict OEP for older adults using a single IMU on the waist. As an extended study of the previous work [7], the proposed system showed outstanding results compared to traditional HAR systems. Compared with the results of the baseline models, it was able to recognize some Otago exercises for older adults in their daily lives with higher f1-scores, fewer hyperparameters, and post-processing stages were not needed anymore.

The contributions of this study are:

- This study proposes a micro-labeling method to annotate sensor data for HAR. This method annotates each repetition of the activities (micro labels), which offers less variance for the model to converge efficiently. Based on the micro labels, the whole period of activities (macro labels) could be recognized with higher accuracy compared with the same model without micro labels.
- This study applies a DS-MS-TCN model. The first stage classifies the micro activities while the following stages learn from the output of the previous stage and classify the macro activities. The models outperformed the other state-of-the-art deep learning models with higher f1-scores.
- This study proves the possibility of using a single IMU to recognize OEP in daily life. By the proposed seq-to-seq system, four OEP exercises were classified with high f1-scores and low over-segmentation error. The system could be generalized to the home settings. Compared with the previous study [7], the system did not require sliding windows or any post-processing stages.

The paper is structured as follows: Section II reviews the state of the art of HAR systems. Section III outlines the datasets and describes the implementation and validation of the proposed system. Section IV presents the experimental findings, and Section V delves into the discussion of these results. Lastly, Section VI concludes the paper and suggests avenues for future research.

II. Related Works

The sliding window technique is commonly used. It segments the input sensors data into windows with equal size [10], [11], [12]. Each window with multiple samples is then classified as a certain class (window-wise classification). The drawback is that each window is independent of the others during training and the model thus has limited knowledge of the temporal dependencies. Also, the determination of sliding window sizes requires domain knowledge and experiments for hyperparameter tuning, as in our previous study for OEP recognition [7]. On the other hand, in the field of action recognition by videos, it has been proven efficient to train the models with a time series of frames [13]. Inspired by this, the sensor data could also be processed with sequence-to-sequence (seq-to-seq) systems [14], [15], which makes it possible that each sample serves as a training example (sample-wise classification). However, for rehabilitation programs, such systems still lack validation.

Traditional machine learning approaches have long been the cornerstone of many data analysis tasks, including HAR [1]. These methods typically involve the extraction of handcrafted features from raw sensor data, followed by the application of machine learning algorithms to classify activities. Feature engineering requires domain expertise and careful selection of relevant features, which can be time-consuming and may limit the model's ability to capture intricate patterns in complex datasets. In contrast, deep learning has emerged as a groundbreaking paradigm in artificial intelligence, revolutionizing various fields by automatically learning hierarchical representations directly from raw data. By leveraging architectures such as recurrent neural networks (RNNs) [16] and convolutional neural networks (CNNs) [17], [18], [19], [20], deep learning models can effectively learn hierarchical representations, automatically extract discriminative features, and model complex temporal dependencies in activity sequences. Recently, researchers have applied the combination of CNN and Long short-term memory (LSTM) cells (CNN-LSTM) [21], [22], which outperformed single CNN or LSTM network on many datasets. Such an architecture could represent both local spatial features and long-time temporal features. Besides, transformers are becoming popular for HAR systems and have demonstrated efficiency for seq-to-seq recognition [23]. However, transformers did not outperform CNN-LSTM according to the previous study for IMU-based activity recognition [24].

Recently, multi-stage temporal convolutional networks (MS-TCNs) were applied for action recognition by videos [13]. Then, there were several studies applying MS-TCNs for sensor data on food intake prediction [14] and physical activity recognition [15]. All of these studies reported the advantages of MS-TCNs over the state-of-the-art CNN-LSTM. MS-TCNs include dilated filters to extract features over a longer time without

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on January 29,2025 at 10:01:05 UTC from IEEE Xplore. Restrictions apply.

TABLE I
INFORMATION OF THE SUBJECTS

	number	age	gender & sarcopenia
Lab	36	79.33± 5.73	17 females (7 (pre-)sarcopenia), 19 males (11 (pre-)sarcopenia)
Home	7	69.43± 2.92	4 females (2 (pre-)sarcopenia), 3 males (2 (pre-)sarcopenia)

increasing computational cost. Besides, by stacking multiple stages of the same architecture, the sample-wise classification could be refined to reduce over-segmentation error. The issue of over-segmentation errors in HAR systems has not gained widespread attention.

III. Materials and Methodology

A. Data Collection

This study received approval from the Ethics Committee Research UZ/KU Leuven (S59660 and S60763). Written informed consent was obtained from all participants prior to study participation.

In the experiments, community-dwelling older adults (aged 65 and older) performed modified OEP [25] while equipped with a McRoberts MoveMonitor+ (McRoberts B.V., Netherlands) featuring a 9-axis Inertial Measurement Unit (IMU). Participants were instructed to wear the device on the waist. The original program includes multiple sub-classes, as the detailed information discussed in [7]. Subjects also performed other ADLs. Two datasets were collected in the study in different scenarios:

1) Lab: The dataset was recorded in a laboratory setting, where participants followed the OEP and ADLs under the guidance of researchers certified as OEP leaders. Data for both OEP and ADLs were gathered either on two distinct days or consecutively within a single day. The ADLs encompassed walking, climbing stairs, sitting, standing, and indoor cycling.

2) Home: The dataset was gathered in a home environment through recorded videos. Participants independently wore the device with the camera turned on and followed a booklet containing instructions for performing the OEP. Additionally, both before and/or after the OEP sessions, subjects randomly performed ADLs out of the camera's view. Consequently, while the ADLs were not directly observed, they were still recorded as the subjects wore the device.

The recruited older adults were (pre-)sarcopenic or non-sarcopenic (defined by EWGSOP1 [26]). The detailed information is shown in Table I. The recruited subjects of the two datasets were different. During intervals between exercises, participants were neither instructed nor monitored. They had the flexibility to practice the exercises, take a seated rest, or walk out of the camera view.

B. Micro Labels

For most human activities, each activity consists of a series of sub-movements (i.e. micro activities). Most existing studies for HAR systems annotate the full range of each activity (macro labels), i.e. from the start of the first movement to the end of the last movement [1], [7]. However, there are still intervals between each two micro activities. For many older adults and patients, the activities were performed at a low speed, which means that there is a break after each movement. This introduces irrelevant information for the models to learn.

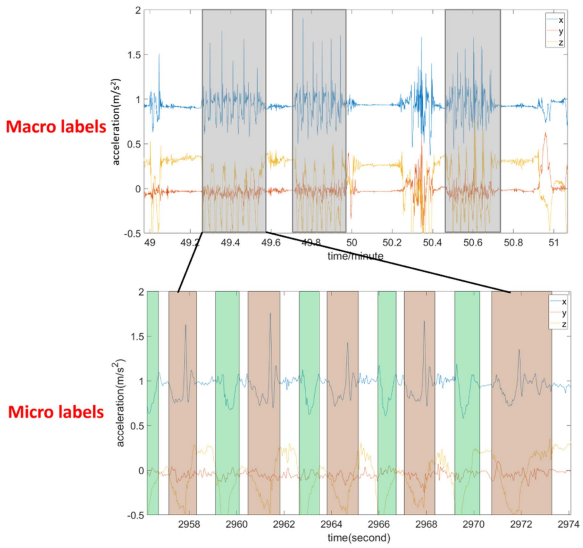


Fig. 1. The accelerometer signals and annotation of macro labels and micro labels for *chair rising*, where each segment of macro activity consisted of five repetitions of *sit-to-stand* (green blocks) and *stand-to-sit* (red blocks).

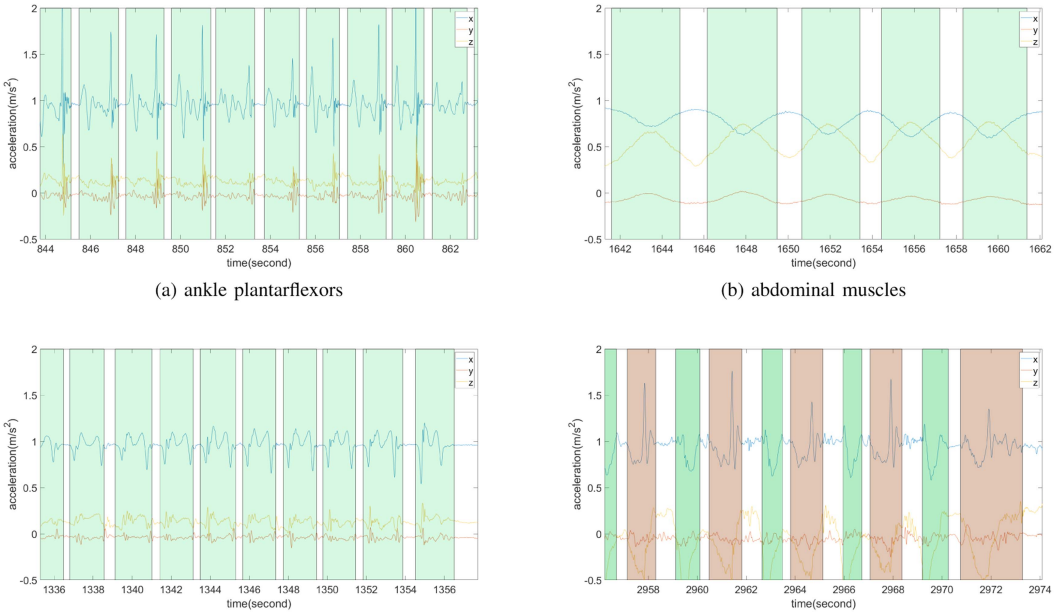
In this study, a method called micro-labeling was applied for more accurate information. Fig. 1 shows an example of the micro-labeling method for *chair rising*, which is an Otago Exercise. There are two types of micro activities in this activity: *sit-to-stand* and *stand-to-sit*. In this rehabilitation program, the subjects have to repeat these movements five times. The micro-labeling method annotates each *sit-to-stand* and *stand-to-sit* individually, without annotating the intervals between each repetition. In other words, a segment of macro labels consists of a set of micro labels.

Similarly, each movement of *ankle plantarflexors*, *knee bends*, and *abdominal muscles* was annotated as micro labels, as shown in Fig. 2. Specifically, each segment of the micro labels was annotated according to the following observations:

- *ankle plantarflexors:* start when the heels leave the ground and end when the heels fall back to the ground.
- *knee bends:* start when the knees bend and end when the knees are straight again.
- *abdominal muscles:* start when the trunk leans back and end when the trunk goes back to vertical.
- *chair rising:*
  - *sit-to-stand:* start when the hip leaves the chair and end when standing straight.
  - *stand-to-sit:* start when the knees bend and end when taking the seat.

As a result, there were five classes of macro labels: *ankle plantarflexors*, *knee bend*, *abdominal muscles chair rising*, and *others*. For micro labels, correspondingly, there were six classes:

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on January 29,2025 at 10:01:05 UTC from IEEE Xplore. Restrictions apply.



(a) ankle plantarflexors (b) abdominal muscles

(c) knee bends (d) chair rising (green blocks = *sit-to-stand*, red blocks = *stand-to-sit*)

Fig. 2. Examples of each repetition of micro activities and corresponding acceleration. The shadowed area illustrates the positive micro labels and the unshadowed area illustrates *other* labels.

TABLE II
THE NUMBER OF MICRO/MACRO LABELED SEGMENTS AND THE DURATION OF THE MACRO ACTIVITIES

	micro labels	macro labels	duration (min)
ankle plan	970	75	53.09
abdominal	202	35	16.94
knee bends	748	75	33.92
chair rising	369*	72	24.11
other	/	/	2483.90

*This values is the same for *sit-to-stand* and *stand-to-sit*

micro ankle plantarflexors, *micro knee bend*, *micro abdominal muscles sit-to-stand*, *stand-to-sit*, and *others*. The *others* type included all other labels: ADLs, other OEP exercises, unknown activities between OEP exercises, etc. For micro labels, the intervals between repetitions were also labeled as *others*. Table II shows the number of annotated micro and macro labeled segments duration of each class.

C. Architecture Overview: DS-MS-TCN

Traditionally, HAR systems apply the sliding window technique that takes a window of consecutive samples as input for classification. A sliding window is used to segment the continuous data stream into fixed-size windows, which are then processed as a unit. This approach then aggregates multiple samples together to extract features either manually or using deep learning models. The determination of window size is challenging because it is related to the activities and subjects.

With the development of deep learning models, HAR systems could apply seq-to-seq operations where the sliding window technique is not necessary. This technique is sample-wise because each sample from the sensors generates an individual output.

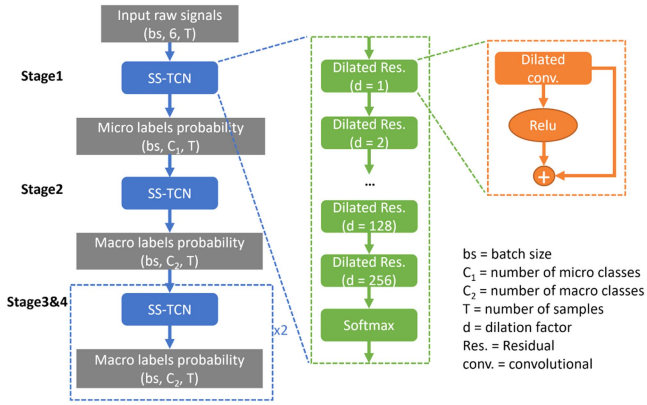


Fig. 3. The overview of the proposed DS-MS-TCN. Four stages of SS-TCN were applied to comply to classify micro labels and macro labels. Each SS-TCN composed of multiple residual dilated layers.

In this study, a seq-to-seq Dual-scale Multi-Stage Temporal Convolutional Network (DS-MS-TCN) was proposed based on both micro labels and macro labels. As shown in Fig. 3, the proposed deep learning model consisted of three parts: micro labels classification, macro labels classification, and macro labels refinement. The micro labels were first classified by the first stage of the model. Then the second stage generated macro labels from the micro labels. Finally, the third stage and the fourth stage refined the macro labels to reduce over-segmentation errors. The details of each stage are explained in the following sections.

Each stage of the model employed a Single-stage Temporal Convolutional Network (SS-TCN) architecture, comprising a

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on January 29,2025 at 10:01:05 UTC from IEEE Xplore. Restrictions apply.

stack of multiple residual dilated convolution layers. This design facilitated the model's ability to learn from broader time intervals without incurring additional computational costs. Following each dilated convolutional layer, a RELU operation is applied. To mitigate issues related to vanishing or exploding gradients, a residual connection is employed at the output, as depicted in the right part of Fig. 3. The input signals only included the accelerometer and gyroscope, as in our previous study [7]. Therefore, there were six channels in the input.

To expand the receptive field, a series of individual residual dilated convolution layers were stacked. This stacking process initiated with the first layer employing a dilation factor of 1 and then doubled the dilation factors for each subsequent layer, progressively using 2, 4, 8, 16, and so on. To simplify computations, each dilated layer was equipped with a filter of size $3 \times D$, where D represents the number of filters. By elevating the dilation factors, the receptive field can be calculated as follows:

$$\text{receptive field} = 2^{L+1} - 1. \tag{1}$$

Finally, a 1×1 convolutional layer with softmax function was applied to generate the probability for each sample. Therefore, the output was generated as follows:

$$Y_t = softmax(Wh_t + b), \tag{2}$$

where Y_t is the output possibility at timestamp t in the shape of (1,number of classes), and h_t is the output value of the last layer, in the shape of (1,number of features). Each output layer contained weights W and bias b.

The number of layers was tuned by comparing the f1-scores, and the results showed that nine layers resulted in the highest f1-scores as proved in [13]. These layers offered a large receptive field that considered more repetitions of the exercises. The receptive field was hence 1023 samples (10.23 seconds). Non-causal kernels were applied in the convolutional layers [27], resulting in a latency of 5.11 seconds for the system.

D. Stage 1: Micro Labels Classification

Using an SS-TCN, the first stage recognized micro activities from the input IMU signals, as follows:

$$\hat{Y}_{micro} = Z_1\{X\}, \tag{3}$$

where X is the input signals, and $\hat{Y}_{micro}$ is the micro labels predicted by the first stage Z_1.

The classification stage of micro activities was not expected to generate clinically applicable results, since these micro activities happened frequently in daily life. For example, the first stage cannot recognize whether a repetition of *chair rising* happened in OEP or in ADLs.

E. Stage 2: Macro Labels Classification

The first stage produced the classification probabilities for micro labels. Subsequently, another stage of SS-TCN was employed to perform the classification of macro labels. The second stage had the same structure as the first stage, consisting of multiple residual dilated convolution layers. The micro labels probability served as the input of the second stage to generate macro labels probability as follows:

$$\hat{Y}^{[2]}_{macro} = Z_2\{\hat{Y}_{micro}\}, \tag{4}$$

where $\hat{Y}^{[2]}_{macro}$ is the predicted labels by the second stage, noted by Z_2.

This stage could generate the clinically applicable macro activities recognition. By considering the temporal relationship of the micro activities, this stage can recognize which micro activities were within OEP and which were not. For example, if a single *chair rising* happened in daily life, the second stage would classify this event as ADLs. On the other hand, if five consecutive repetitions of *chair rising* were recognized in the first stage, the second stage would classify this event as *chair rising*.

F. Stage 3-4: Macro Labels Refinement

The macro labels were expected to perform low over-segmentation errors (explained in Section III-K). Therefore, another two stages of SS-TCN were applied to further refine the sample-wise macro labels as a time series. For both stages, the input and the output were both the macro labels, as follows:

$$\hat{Y}^{[s]}_{macro} = Z_s\{\hat{Y}^{[s-1]}_{macro}\}, \ for\ s = 3, 4, \tag{5}$$

where $\hat{Y}^{[s]}_{macro}$ denotes the refined macro labels from stage Z_s. The final stage generated the classified and refined macro labels.

G. Loss Function

There are two parts of the loss function: classification loss L_{CE} and refinement loss L_{TMSE}.

The first stage of DS-MS-TCN performed a classification task regarding the micro activities using cross entropy (CE) loss:

$$L^{[1]}_{CE} = \frac{1}{TC_1} \sum_{t,c_1} -y_{\text{micro},t,c_1} \log \hat{y}_{\text{micro},t,c_1}, \tag{6}$$

where y_{micro,t,c_1} and $\hat{y}_{\text{micro},t,c_1}$ denote the true and predicted micro labels at timestamp t for micro class c_1 in the first stage. C_1 is the number of micro classes, and T is the number of samples in the time series.

For the following stages, the classification was performed based on the macro activities

$$L^{[s]}_{CE} = \frac{1}{TC_2} \sum_{t,c_2} -y_{\text{macro},t,c_2} \log \hat{y}^{[s]}_{\text{macro},t,c_2}, \ for\ s > 1, \tag{7}$$

where y_{macro,t,c_2} denotes the true macro labels and $\hat{y}^{[s]}_{\text{macro},t,c_2}$ denotes the predicted macro labels from stage s for macro class c_2. C_2 is the number of macro classes.

To improve the classification performance, an additional Truncated Mean Squared Error (TMSE) function was incorporated for refinement, inspired by Farha et al. [13]. For each stage s (except the first stage), the function $L^{[s]}_{TMSE}$ is defined as

$$L^{[s]}_{TMSE} = \frac{1}{TC_2} \sum_{t,c_2} (\tilde{\Delta}^{[s]}_{t,c_2})^2, \ for\ s > 1, \tag{8}$$

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on January 29,2025 at 10:01:05 UTC from IEEE Xplore. Restrictions apply.

where $\tilde{\Delta}^{[s]}_{t,c_2}$ is the truncated difference between the logarithm probabilities of two adjacent samples

$$\tilde{\Delta}^{[s]}_{t,c_2} = \begin{cases} \Delta^{[s]}_{t,c_2}, & \text{if } \Delta^{[s]}_{t,c_2} \leq \tau \\ \tau, & \text{otherwise} \end{cases}, \tag{9}$$

$$\Delta^{[s]}_{t,c_2} = |log\hat{y}^{[s]}_{\text{macro},t,c_2} - log\hat{y}^{[s]}_{\text{macro},t-1,c_2}|. \tag{10}$$

L_{TMSE} was applied to reduce the segmentation error. It was thus not necessary for the first stage applied for classifying the micro labels. The reason was that the micro labels were short, and refining these labels would result in some lost information for the following stages.

The final loss function is computed as

$$L = \eta L^{[1]}_{CE} + L^{[s]}_{CE} + \lambda L^{[s]}_{TMSE},\ for\ s > 1. \tag{11}$$

The values of λ and τ have been explored by comparing their impact on the f1-scores and their values did not cause a significant impact on the results. Therefore, the values of λ and τ were set as 0.15 and 0.4 as in the previous study [13]. Therefore, this study kept the optimal values as 0.15 and 4, respectively. On the other hand, since η was originally proposed in this study, the values were explored in the following experiments.

H. Baseline Methods

1) Hierarchical System Based on Sliding Window: In our previous study [7], a hierarchical system was proposed, including two-step tasks based on sliding window techniques. In the previous study, we applied two sliding windows with different sizes to comply with two classification tasks. First, a large window size (10 minutes) was applied for classifying general Otago exercises and Daily activities (i.e. binary classification). Then, a small window size (6 seconds) was applied to classify the sub-classes. Besides, since the windows were classified individually, additional post-processing was needed to smooth the output labels as a time series. This system based on the sliding window technique was involved as a baseline method to compare with the proposed method. To make it comparable with the seq-to-seq methods, the classified windows were reconstructed as time series.

2) CNN: A typical CNN model proposed by [20], where three convolutional layers with dropout and pooling layers were applied to extract features. After flattening these features, they were classified by a fully connected layer. Finally, an output layer with a softmax function was applied to generate the final prediction. This model was implemented as a seq-to-seq model.

3) Transformer: The model was proposed by [23]. After the initial steps of normalization and position embedding, the architecture employed three encoder blocks. Each of these blocks included a multi-head attention layer and a fully-connected layer, interspersed with dropout layers for regularization. Following these blocks, classification outcomes were produced using an additional fully-connected layer. This model was implemented as a seq-to-seq model.

4) CNN-LSTM: The architecture and hyperparameters of the CNN-LSTM model used in this study were proposed by [22]. Therefore, the strides of the convolutional kernels were all set to one. And the LSTM layer returned the sequence rather than only the last output. Then dense layer was applied to generate the probabilities of each sample. This model was implemented as a seq-to-seq model.

5) MS-TCN (Without Micro Labels): To explore the influence of the micro labels, as a benchmark, a 4-stage MS-TCN model was applied without micro labels. Therefore, the first stage of this MS-TCN model also complied with macro activity classification and refinement. For this model, η was set to 1in the loss function.

I. Training Details

Since the signals of Otago exercises and Daily activities were long time series, the models could not converge efficiently. Therefore, the signals were cut into slices with an equal length of 40 seconds, with 50% overlap. This operation was different from the traditional sliding window techniques where the output was a single value. In this proposed sequence-to-sequence technique, theoretically, no windows were needed for training and the whole time series can be fed into the model. However, to make use of batch training for better efficiency, the input with the same length was applied in this study. The length of the input did not impact the results as long as it was longer than the receptive field (i.e. 10.23 seconds), since the model cannot learn from samples outside of this range. The overlap was applied to mitigate information loss at the edges of non-overlapping windows. Then batch learning was applied for training with a batch size of 32.

The deep learning models were developed using Pytorch and trained on an NVIDIA P100 SXM2 GPU. The Adam optimization algorithm was utilized to minimize the loss function.

To make the deep learning models (CNN, Transformer, CNN-LSTM) comparable with the proposed model, they also complied with the seq-to-seq classification task. The CNN model and CNN-LSTM model were implemented by padding the convolutional layers to the same length as the input.

J. Data Split

The Leave-One-Subject-Out Cross-Validation (LOSOCV) method was applied to the data of the subjects from the lab. The following impacts on the model were explored by comparing the results:

- Using two stages for micro label classification, compared with using one stage. Correspondingly, another term $\eta L^{[2]}_{CE}$ was added to the loss function as in (11), where

$$L^{[2]}_{CE} = \frac{1}{TC_1} \sum_{t,c_1} -y_{\text{micro},t,c_1} \log \hat{y}^{[2]}_{\text{micro},t,c_1}, \tag{12}$$

  where $\hat{y}^{[2]}_{\text{micro},t,c_1}$ denotes the predicted micro labels by the additional stage. The results were compared with the proposed model using one stage for micro labels classification.
- Different values of loss weight of the first stage (η).
- Different numbers of training examples for micro labels.
- Different numbers of stages for macro labels classification.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on January 29,2025 at 10:01:05 UTC from IEEE Xplore. Restrictions apply.

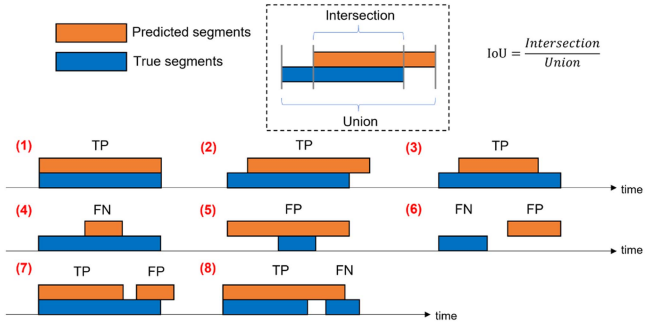


Fig. 4. The definition of IoU and some examples of TP, FP, and FN. [7].

Then, the proposed DS-MS-TCN was compared to the baseline models. Finally, the model was generalized on the home-based data using LOSOCV, with the lab-based data also in the training set.

K. Evaluation

1) Sample-Wise Evaluation: For seq-to-seq classification, each input sample yielded a corresponding output. As a result, the sample-wise F1-score was employed to assess the classification performance. The F1-score for class c was calculated as follows:

$$f1_c = 2 * \frac{precision_c * recall_c}{precision_c + recall_c}, \tag{13}$$

where

$$precision_c = \frac{TP_c}{TP_c + FP_c}, and \tag{14}$$

$$recall_c = \frac{TP_c}{TP_c + FN_c}. \tag{15}$$

where TP_c, FP_c, and FN_c denote the numbers of true positive, false positive, and false negative labels from the sample-wise classification output.

2) Segment-Wise Evaluation: Given that seq-to-seq classification assigns a label to each sample, the potential for over-segmentation errors exists. This occurs when a continuous activity segment is classified as several smaller segments, with misclassified labels interspersed. As an evaluation metric, segmental F1-scores based on Intersection over Union (IoU) were computed for each class based on Lea et al. [28] and Sarapata et al. [29]. The Intersection over Union (IoU) matrix was defined as the ratio of the intersection to the union of the predicted and true segments. Initially, a threshold for the IoU values was established. Subsequently, segment-wise true positive ($TPseg$), false positive ($FPseg$), and false negative ($FNseg$) were defined:

- $TPseg$: IoU $\geq$ threshold
- $FPseg$: IoU $<$ threshold, true segments shorter than predicted segments
- $FNseg$: IoU $<$ threshold, true segments longer than predicted segments

To better illustrate the definitions, Fig. 4 shows some examples [7]. According to (13), (14), and (15), the segment-wise evaluation matrices could be calculated. Compared with the sample-wise f1-score, this method could also evaluate the over-segmentation errors. In this study, the threshold of 0.5 was applied.

TABLE III
F1-SCORES AND IOU F1-SCORES UTILIZING ONE STAGE AND TWO STAGES FOR MICRO LABELS CLASSIFICATION

	f1			
	ankle plan	abdominal	knee bends	chair rising
1 stage	89.777	85.363	88.298	90.028
2 stages	88.877	77.51	82.915	84.483
	IoU f1			
	ankle plan	abdominal	knee bends	chair rising
1 stage	66.667	68.852	63.333	92.308
2 stages	64.865	60.674	57.803	73.446

IV. RESULTS

A. Influence of Micro Labels Stage

a) Number of stages for micro labels: As a comparison, two models were compared: a DS-MS-TCN with two stages for micro labels classification and a DS-MS-TCN with one stage for micro labels classification. Both models applied three stages for macro labels classification. The results are shown in Table III. It is observed that both f1-scores and IoU f1-scores were decreased compared with the model using one stage for micro labels classification.

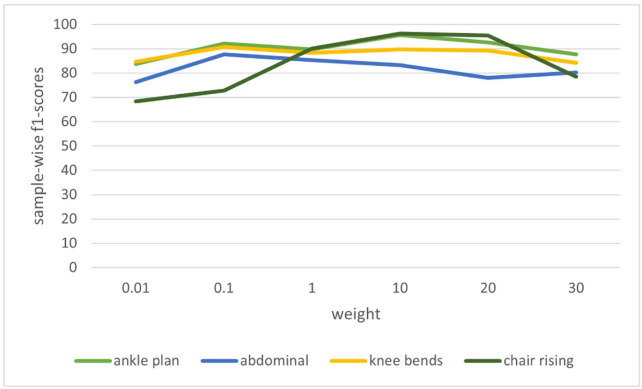
b) Loss weight of micro labels: The influence of loss weight of the first stage, i.e. η in formula (11), is shown in Fig. 5. The results show that the classification performance for each activity was maximized by different values of η. For example, the IoU f1-scores reached the maximum at $\eta = 0.1$ for *abdominal muscles* and *knee bends*, whereas for *chair rising* the optimal value for η was 10. Besides, the η value resulting in the highest f1-scores might not obtain the highest IoU f1-scores. For example, for *chair rising*, the f1-score reached the highest at $\eta = 20$ whereas the IoU f1-score reached the highest at $\eta = 10$. In general, the classification performance was decreased when η was lower than 0.1 or higher than 20. Therefore, in the following experiments, the value of η was set as default as 1.

c) Number of trained micro labels: Since manually annotating the micro labels required a huge workload, an experiment was conducted to investigate the impact of the number of micro labels in the training set. Different numbers of micro labels segments were randomly selected to train the model. As shown in Fig. 6, the f1-scores increase with more percentages of micro labels in the training set. All activities could reach the clinically applicable threshold of 80% f1-scores when 80% of micro labels were trained. For *chair rising*, the f1-score could reach the threshold with 40% micro labels in the training set. However, the bars indicate a continual rise in f1-scores, even when incorporating all micro labels.

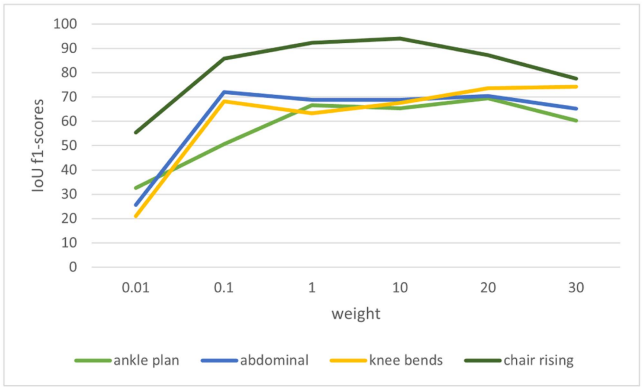
B. Influence of Macro Labels Stages

Fig. 7 shows the f1-scores and IoU f1-scores of the macro labels applying one, two, three, and four stages in DS-MS-TCN. Fig. 7(a) shows that the improvement in f1-scores was not

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on January 29,2025 at 10:01:05 UTC from IEEE Xplore. Restrictions apply.

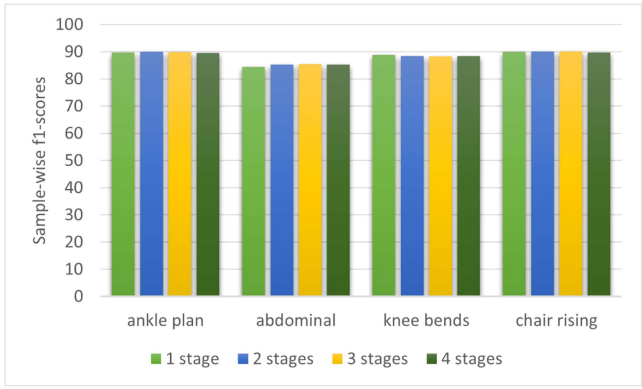


(a) f1-scores

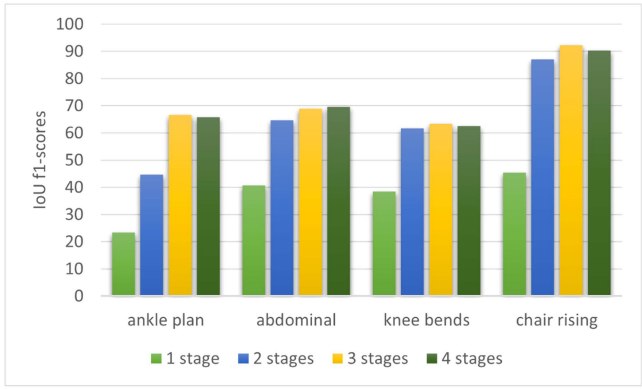


(b) IoU f1-scores

Fig. 5. F1-scores and IoU f1-scores applying different loss weight for the first stage (i.e. η).



(a) f1-scores



(b) IoU f1-scores

Fig. 7. F1-scores and IoU f1-scores applying different number of stages for macro labels classification.

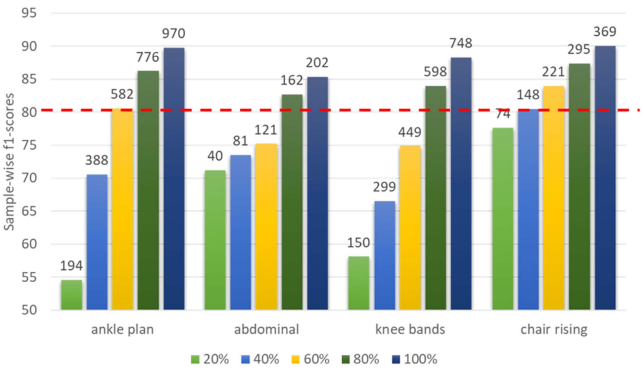


Fig. 6. F1-scores with different percentages of micro labels in the training set. The red dashed line illustrates a clinically applicable threshold of 80%. The absolute numbers of labels were shown outside the end of bars.

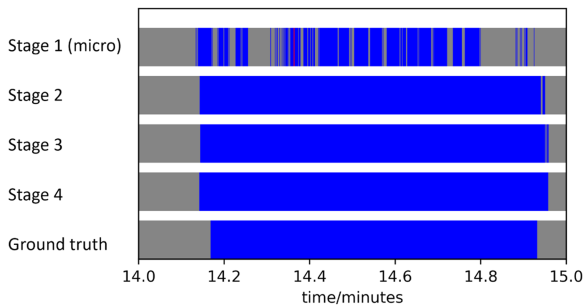


Fig. 8. The output of each stage from DS-MS-TCN: an example of *ankle plantarflexors*.

significant by adding the number of stages. On the other hand, the IoU f1-scores were substantially increased for all activities until the third stage. The second stage led to more improvement compared with the third stage, since the third stage did finer tuning of the series, as shown in Fig. 7(b). From the fourth stage on, there was no significant improvement anymore.

To visualize the influence of each stage of DS-MS-TCN, Fig. 8 shows an example of *ankle plantarflexors* predicted by each stage. The first stage produced micro labels and generated high over-segmentation errors. The second and the third stage produced the macro labels with smaller over-segmentation errors. The final stage then refined the macro labels with the errors removed.

C. Model Comparison

The f1-scores and IoU f1-scores of both the baseline methods and the proposed method are presented in Table IV. The hierarchical system, utilizing the sliding window technique, incorporated a post-processing stage to enhance the smoothness of the output sequence [7]. Therefore, the IoU f1-scores were

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on January 29,2025 at 10:01:05 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
THE EVALUATION (F1-SORES AND IOU F1-SCORES) OF THE PROPOSED DS-MS-TCN AND OTHER BASELINE MODELS, AND THEIR NUMBER OF PARAMETERS AND AVERAGE TRAINING TIME FOR EACH PARTICIPANT

			evaluation				number of parameters	avg. training time (h)
			ankle plan	abdominal	knee bends	chair rising		
proposed	DS-MS-TCN	f1	89.777	85.363	88.298	90.028	134580	17.42
		IoU f1	66.667	68.852	63.333	92.308		
baseline	MS-TCN	f1	72.088	66.855	44.917	83.686	134580	18.21
		IoU f1	9.871	23.377	11.232	70.27		
	Transformer	f1	74.429	72.63	59.205	84.212	1780742	52.26
		IoU f1	5.623	14.395	14.853	22.402		
	CNN-LSTM	f1	64.42	74.745	74.918	81.817	187975	23.50
		IoU f1	3.659	16	3.902	6.015		
	CNN	f1	63.747	72.059	48.74	80.028	19,077	1.95
		IoU f1	8.791	2.492	0	0.442		
	Hierarchical system	f1	78.028	85.35	77.243	84.878	204,839	27.84
		IoU f1	13.084	40.549	33.333	51.25		

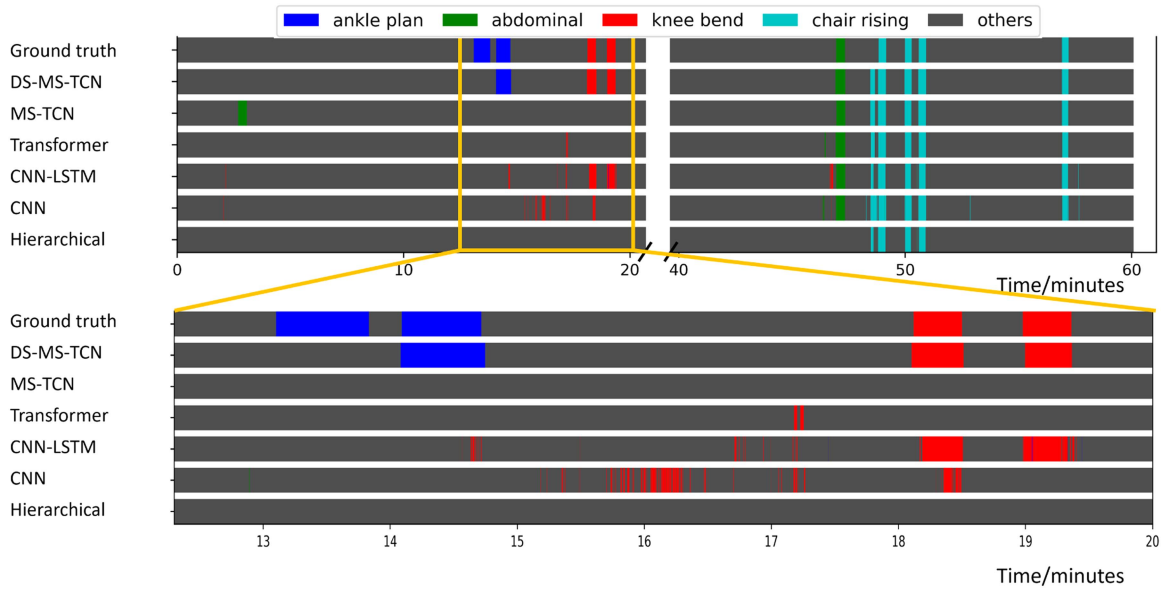


Fig. 9. The recognized macro labels from a subject. The upper part shows the time series of the whole OEP program. The lower part zooms in the results of *ankle plantarflexors* and *abdominal muscles*.

higher than the other seq-to-seq baseline models, except the MS-TCN.

All baseline models achieved f1-scores exceeding 80% for *chair rising*. The MS-TCN model (without micro labels) achieved the highest IoU f1-score of 70.27% for *chair rising*. However, for the other three activities, the IoU f1-scores of the baseline models did not surpass 50%, despite some reporting f1-scores higher than 80%.

The proposed DS-MS-TCN showed the highest f1-scores over 85% and IoU f1-scores over 60% for all activities. It could recognize *chair rising* with both f1-scores and IoU f1-scores over 90%.

Fig. 9 shows the recognized macro labels by different models. The upper part of the figure illustrates the classified time series from a complete OEP program. Some baseline models obtained predicted segments matching the ground truth (e.g. MS-TCN and CNN-LSTM). On the other hand, the other part of the figure at the bottom zooms in the recognition of *ankle plantarflexors* and *abdominal muscles*. The baseline models based on seq-to-seq classification showed many false positive and false negative segments, whereas the proposed DS-MS-TCN model could recognize most segments with only one false positive segment for *chair rising* and one false negative segment for *ankle plantarflexors*. Fig. 10 shows the confusion matrix generated from the DS-MS-TCN model.

Fig. 11 shows two examples of convergence using DS-MS-TCN with micro labels and MS-TCN without micro labels. The DS-MS-TCN showed faster convergence speed and resulted in lower loss values for the final output (stage 4). From stage 2 to stage 4, the difference in loss values was not large for the small weight of refinement loss.

D. Generalization on Subjects At Home

After exploring the optimal model based on the dataset collected in the lab, the home-based dataset was tested using LOSOCV with the lab-based data also in the training set. The f1-scores and IoU f1-scores are shown in Table V. As a comparison, the results of the lab-based dataset by LOSOCV were also in the table. It was observed that both f1-scores and IoU f1-scores

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on January 29,2025 at 10:01:05 UTC from IEEE Xplore. Restrictions apply.

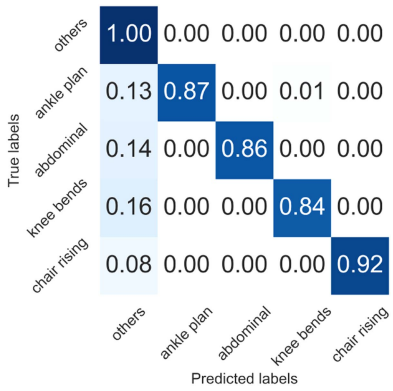
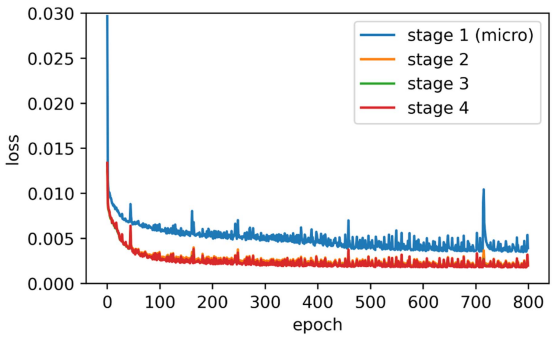
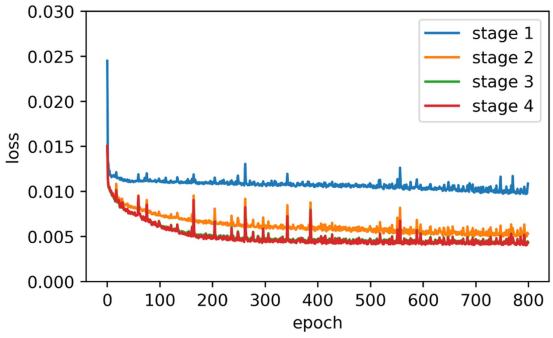


Fig. 10. The confusion matrix of the proposed DS-MS-TCN model.



(a) DS-MS-TCN



(b) MS-TCN

Fig. 11. Loss values of each stage from DS-MS-TCN with micro labels and MS-TCN without micro labels.

TABLE V
F1-SCORES AND IOU F1-SCORES OF THE LAB-BASED DATA AND HOME-BASED DATA

	Lab			
	ankle plan	abdominal	keen bends	chair rising
f1	89.777	85.363	88.298	90.028
IoU f1	66.667	68.852	63.333	92.308
	Home			
	ankle plan	abdominal	keen bends	chair rising
f1	63.401	87.377	67.570	85.909
IoU f1	29.412	44.444	16.667	50.000

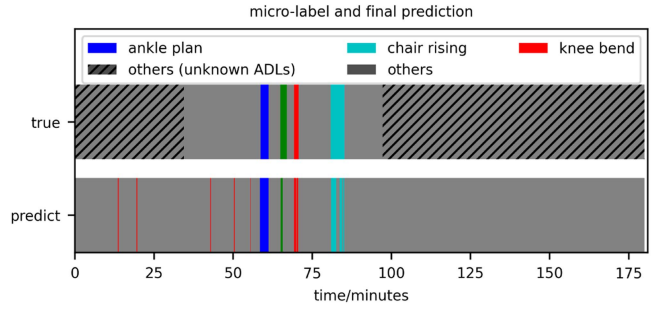


Fig. 12. The recognized macro labels from a subject at home. The shadowed area illustrates the ADLs that were not observed in daily life due to the limitation of camera installation.

decrease when generalizing on the subjects at home. Among the four activities, *ankle plantarflexors* obtained the highest drop, from 89.8% to 63.4% for f1-scores, and from 66.7% to 29.4% for IoU f1-scores. The IoU f1-scores of *chair rising* also drastically decrease from 92.3% to 50%.

An example of predicted labels for one subject at home is shown in Fig. 12. The shadowed blocks highlight the labels of ADLs that were not observed by the camera. The results show that the model generated several false positive segments for *knee bends*, whereas the other labels of ADLs were correctly predicted. However, there were more over-segmentation errors during the OEP. On the one hand, a single segment of macro activity was classified as more smaller segments. On the other hand, some other activities were classified as short false positive macro labels.

E. Comparison With the Previous Studies

This study is a follow-up study of the previous work [7], which has been discussed in Section IV-C. Besides, there were only two studies applying wearable IMUs for Otago exercises recognition. One study [25] predicted ADLs and OEP (binary classification) and obtained the f1-score of 77.7%. The other study [30] explored most Otago sub-exercises using different sensor combinations, as shown in Table VI. To compare with this study, the table only presents the results including the waist-mounted IMU. The previous study was performed in a highly controlled environment (young healthy adults in the lab). It applied CNN based on the sliding window technique (seq-to-one) evaluated based on 5-fold cross-validation. This study reported higher f1-scores for *knee bends* and *chair rising* even without the IMUs on the shank and feet. However, the previous study reported higher f1-scores for *ankle plantarflexors* with two more IMUs.

V. DISCUSSION

A. Seq-to-Seq HAR

Seq-to-seq models for activity recognition have not been explored much compared with the sliding window technique. In the previous study [7], a large sliding window was first applied to distinguish the ADLs and OEP (binary classification). Then, a small sliding window was applied to classify which exercises

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on January 29,2025 at 10:01:05 UTC from IEEE Xplore. Restrictions apply.

TABLE VI
COMPARISON WITH A PREVIOUS STUDY

Study	sensors	subjects	scenario	method	f1-scores		
					ankle plan	knee bends	chair rising
[30]	1 IMU (waist)	19 healthy adults	lab	CNN, seq-to-one	32	31	49
[30]	3 IMU (right shank + right feet + waist)	19 healthy adults	lab	CNN, seq-to-one	**98**	84	67
[30]	3 IMU (left shank + left feet + waist)	19 healthy adults	lab	CNN, seq-to-one	**98**	85	71
This study	1 IMU (waist)	35 older adults, some of them sarcopenic	lab	DS-MS-TCN, seq-to-seq	90	**88**	**90**
This study	1 IMU (waist)	7 older adults, some of them sarcopenic	home	DS-MS-TCN seq-to-seq	49	83	80

The f1-scores of the previous study was based on 5-fold cross-validation, whereas this study used LOSOCV. *Abdominal muscles* was not collected in the previous study.
The bold values highlight the highest f1-score and IoU f1-score for each class.

from OEP were performed. The drawback of this technique was that the sizes of sliding windows had to be determined and the post-processing stages were necessary. Also, post-processing was necessary to improve the smoothness of the output, because the predicted segments were independent from each other. It is also notable that this study only explored four types of OEP activities. This decision was influenced by the fact that some exercises lacked clinical significance, and some were challenging to identify due to limitations in sensor capabilities.

With seq-to-seq techniques, the models could learn temporal dependencies from a longer time series. Therefore, it could recognize the specific exercises using one model, rather than two hierarchical models. MS-TCN and DS-MS-TCN applied multiple convolutional layers with different dilation factors. Therefore, both short-term and long-term information could be captured. Also, with the TMSE functions, the output was smoothed by considering the adjacent samples. The results in Fig. 12 show limited confusion between ADLs and Otago exercises, despite most training samples belonging to *others*, as shown in Table II.

B. DS-MS-TCN Based on Micro Labels

This study proposes an innovative annotation method for activity recognition using DS-MS-TCN. In the first stage of the model, each repetition of the activities (micro activity) was recognized. The reason for using micro labels was to bring less irrelevant information to the model. Hence, micro labels contained lower variance. For example, when performing *ankle plantarflexors*, subjects paused between each two repetitions. During these pauses, the sensors did not capture any information whereas they were still labeled as positive. Although the pauses were short (no more than one second), for seq-to-seq models, it was harder to learn and generalize. Therefore, DS-MS-TCN based on micro labels outperformed MS-TCN with the same structure.

Due to the lower variance of micro labels, smaller training sets were needed to get good performance. Compared with micro labels that were abundant, macro labels are less prominent in the data. Using the duel-scale model, from micro to macro has to do with distributions of micro labels rather than the underlying sensor signals. Therefore, the performance of macro activity classification was improved using the same dataset.

The recognition performance of micro labels was not expected to be perfect because each micro labels segment was too short. However, introducing an additional stage for micro labels classification did not yield improvements in the results. This can be attributed to the refinement process, which resulted in the elimination of certain short micro labels and contributed to the loss of information for macro label classification.

The loss weight of the first stage (η) had varying effects on the final f1-scores for different activities. Therefore, it was difficult to determine η that led to the highest overall f1-scores and IoU f1-scores. However, the influence of η was rather constant within a certain range (between 0.1 and 20). It was thus set as one in this study.

Some micro activities happen in both Otago exercises and other ADLs. However, the activities happening in ADLs should not be classified as Otago exercises. The difference among both was that Otago exercises were always performed consistently with multiple repetitions whereas the ADLs were performed for a single repetition. Therefore, the first stage could not recognize the difference and the second stage was necessary to remove over-segmentation error. If there was only a single repetition of micro activities, it was excluded from the Otago exercises by the higher stages. For similar reasons, the third stage and fourth stage were also applied, since the second stage was not sufficient to completely remove the over-segmentation error.

Although Fig. 6 shows that 80% of the micro labels were enough to reach the f1-scores of 80%, the f1-scores were not maximized even with all micro labels used. With the more variance of the activities, the number of micro labels had more influence on the results. For example, *chair rising* and *abdominal muscles* had more distinguishable and consistent IMU signals and hence less variance. Thus they required fewer micro labels to reach the threshold. On the other hand, *ankle plantarflexors* and *knee bends* required more micro labels to cover the higher variance of the IMU signals.

The annotation of micro labels required a substantial workload. Consequently, it becomes crucial to determine the minimum number of micro labels that result in optimal classification performance. In Fig. 6, theoretically, the F1-scores should

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on January 29,2025 at 10:01:05 UTC from IEEE Xplore. Restrictions apply.

plateau after involving a certain quantity of micro labels. The limitation of this study lies in the scarcity of micro labels, preventing exploration of the optimal number for achieving peak performance.

C. Otago Exercises Recognition

This study proposes a system with the simplest and most user-friendly implementation using only one wearable IMU on the waist. Therefore, only the activities involving trunk movement could be recognized. The four activities explored in this study were all important for strength (*ankle plantarflexors* and *abdominal muscles*)and balance (*knee bends* and *chair rising*). Compared with the previous study [30], the proposed system was validated in a less controlled environment focusing on older adults. When comparing the results using one IMU on the waist, the proposed system reported higher f1-scores, proving the advantages of the seq-to-seq DS-MS-TCN model. This study also proved the possibility of recognizing these activities without additional IMU.

The previous study [30] reported lower f1-scores for *chair rising* because it was confused with *knee bends*. For younger healthy adults, these two activities are similar regarding the changing of vertical acceleration. However, in this study, these two activities were more distinguishable for older adults because they did not have support for *knee bends* whereas they had a chair to sit on for *chair rising*. The vertical acceleration was thus different for these two activities.

Generalization on the subjects at home showed lower f1-scores than the subjects in the lab. One of the reasons might be that subjects at home only followed a booklet and did not receive instructions from the researchers. Therefore, the exercises that they performed were not as standard as the subjects in the lab. Another reason was that the other activities were more complicated and could lead to confusion. In the future study, more data will be collected in daily life to capture more variance of the signals. And the model should be improved regarding generalization ability.

The proposed model applied non-causal kernels which led to a latency of 5.11 seconds. The application focused on clinical usage where the data were expected to be collected for a long time (e.g. several weeks) for analysis. Therefore, an online system was not necessary and a latency of 5.11 was acceptable. In the future, some self-evaluation applications might require online recognition of OEP. In those cases, causal kernels can be applied for online applications although the recognition performance will be decreased.

VI. Conclusion

This paper proposes a system designed for older adults to recognize OEP in their daily lives. The DS-MS-TCN model demonstrates the efficacy of micro labels classification in enhancing overall classification performance. This dual-scale system contributes novel perspectives to the field of human activity recognition.

Future research directions include improving the recognition of micro labels to enable quantitative analysis of exercise repetitions within each OEP routine. Considering the substantial workload for micro labels annotation, transfer learning or semi-supervised learning could be applied with the limited number of labels [31]. Additionally, the integration of more tools, such as video-based classification systems, could contribute to enhancing the overall classification performance.

References

[1] Y. Wang, S. Cang, and H. Yu, "A survey on wearable sensor modality centred human activity recognition in health care," *Expert Syst. Appl.*, vol. 137, pp. 167–190, Dec. 2019.

[2] S. Mekruksavanich and A. Jitpattanakul, "Multimodal wearable sensing for sport-related activity recognition using deep learning networks," *J. Adv. Inf. Technol.*, vol. 13, no. 2, pp. 132–138, 2022.

[3] V. Bianchi, M. Bassoli, G. Lombardo, P. Fornacciari, M. Mordonini, and I. D. Munari, "IoT wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8553–8562, Oct. 2019.

[4] S. Thomas, S. Mackintosh, and J. Halbert, "Does the 'Otago exercise programme' reduce mortality and falls in older adults?: A systematic review and meta-analysis," *Age Ageing*, vol. 39, pp. 681–687, Nov. 2010.

[5] S. Mat et al., "Effect of modified Otago exercises on postural balance, fear of falling, and fall risk in older fallers with knee osteoarthritis and impaired gait and balance: A secondary analysis," *PMR*, vol. 10, no. 3, pp. 254–262, 2018.

[6] R. Almarzouki et al., "Improved balance in middle-aged adults after 8 weeks of a modified version of otago exercise program: A randomized controlled trial," *PLoS One*, vol. 15, no. 7, 2020, Art. no. e0235734.

[7] M. Shang et al., "Otago exercises monitoring for older adults by a single IMU and hierarchical machine learning models," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 462–471, 2024.

[8] L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognit.*, vol. 108, 2020, Art. no. 107561.

[9] K. Kyritsis, C. Diou, and A. Delopoulos, "Food intake detection from inertial sensors using LSTM networks," in *Proc. New Trends Image Anal. 2017: ICIAP Int. Workshops, WBICV, SSPandBE, 3AS, RGBD, NIVAR, IWBAAS, and MADiMa 2017, Catania, Italy,* 2017, pp. 411–418.

[10] H. Bi, M. Perello-Nieto, R. Santos-Rodriguez, and P. Flach, "Human activity recognition based on dynamic active learning," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 4, pp. 922–934, Apr. 2021.

[11] Y. Li and L. Wang, "Human activity recognition based on residual network and bilstm," *Sensors*, vol. 22, no. 2, 2022, Art. no. 635.

[12] K. Ellis, J. Kerr, S. Godbole, and G. Lanckriet, "Multi-sensor physical activity recognition in free-living," in *Proc. 2014 ACM Int. Joint Conf. Pervasive Ubiquitous Comput. Adjunct Pub., (Seattle Washington)*, ACM, 2014, pp. 431–440.

[13] Y. A. Farha and J. Gall, "MS-TCN: Multi-stage temporal convolutional network for action segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 3575–3584.

[14] C. Wang, T. S. Kumar, W. D. Raedt, G. Camps, H. Hallez, and B. Vanrumste, "Eat-radar: Continuous fine-grained eating gesture detection using FMCW radar and 3D temporal convolutional network," in *IEEE J. Biomed. Health Inform.*, vol. 28, no. 2, pp. 1000–1011, Feb. 2024.

[15] M. Shang et al., "A multi-stage temporal convolutional network for volleyball jumps classification using a waist-mounted IMU," 2023, *arXiv:2310.13097*.

[16] Y. Zhao, R. Yang, G. Chevalier, X. Xu, and Z. Zhang, "Deep residual bidir-LSTM for human activity recognition using wearable sensors," *Math. Prob. Eng.*, vol. 2018, Dec. 2018, Art. no. e7316954.

[17] J. Huang, S. Lin, N. Wang, G. Dai, Y. Xie, and J. Zhou, "TSE-CNN: A two-stage end-to-end CNN for human activity recognition," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 1, pp. 292–299, Jan. 2020.

[18] S.-M. Lee, S. M. Yoon, and H. Cho, "Human activity recognition from accelerometer data using convolutional neural network," in *2017 IEEE Int. Conf. Big Data Smart Comput.*, Feb. 2017, pp. 131–134.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on January 29,2025 at 10:01:05 UTC from IEEE Xplore. Restrictions apply.

[19] D. Wagner, K. Kalischewski, J. Velten, and A. Kummert, "Activity recognition using inertial sensors and a 2-D convolutional neural network," in *2017 10th Int. Workshop Multidimensional Syst.*, Sep. 2017, pp. 1–6.

[20] W. Tao, Z.-H. Lai, M. C. Leu, and Z. Yin, "Worker activity recognition in smart manufacturing using IMU and sEMG signals with convolutional neural networks," *Procedia Manuf.*, vol. 26, pp. 1159–1166, 2018.

[21] R. Mutegeki and D. S. Han, "A CNN-LSTM approach to human activity recognition," in *2020 Int. Conf. Artif. Intell. Inf. Commun.*, 2020, pp. 362–366.

[22] S. Mekruksavanich and A. Jitpattanakul, "Smartwatch-based human activity recognition using hybrid LSTM network," in *Proc. IEEE SENSORS*, Oct. 2020, pp. 1–4.

[23] I. Dirgová Luptáková, M. Kubovčík, and J. Pospíchal, "Wearable sensor-based human activity recognition with transformer model," *Sensors*, vol. 22, no. 5, 2022, Art. no. 1911.

[24] M. F. Trujillo-Guerrero, S. Román-Niemes, M. Jaén-Vargas, A. Cadiz, R. Fonseca, and J. J. Serrano-Olmedo, "Accuracy comparison of CNN, LSTM, and transformer for activity recognition using IMU and visual markers," *IEEE Access*, vol. 11, pp. 106650–106669, 2023.

[25] L. Dedeyne, J. A. Wullems, J. Dupont, J. Tournoy, E. Gielen, and S. Verschueren, "Exploring machine learning models based on accelerometer sensor alone or combined with gyroscope to classify home-based exercises and physical behavior in (PRE) sarcopenic older adults," *J. Meas. Phys. Behav.*, vol. 4, no. 2, pp. 174–186, 2021.

[26] A. J. Cruz-Jentoft et al., "Sarcopenia: European consensus on definition and diagnosisreport of the European working group on sarcopenia in older peoplea. j. cruz-gentoft," *Age Ageing*, vol. 39, no. 4, pp. 412–423, 2010.

[27] D. Rethage, J. Pons, and X. Serra, "A wavenet for speech denoising," in *2018 IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 5069–5073.

[28] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in *2017 IEEE Conf. Comput. Vis. Pattern Recognit.*, (Honolulu, HI), 2017, pp. 1003–1012.

[29] G. Sarapata et al., "Video-based activity recognition for automated motor assessment of Parkinson's disease," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 10, pp. 5032–5041, Oct. 2023.

[30] A. Bevilacqua, K. MacDonald, A. Rangarej, V. Widjaya, B. Caulfield, and T. Kechadi, "Human activity recognition with convolutional neural networks," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discov. Databases*, 2019, pp. 541–552.

[31] S. Oh, A. Ashiquzzaman, D. Lee, Y. Kim, and J. Kim, "Study on human activity recognition using semi-supervised active transfer learning," *Sensors*, vol. 21, no. 8, 2021, Art. no. 2760.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on January 29,2025 at 10:01:05 UTC from IEEE Xplore. Restrictions apply.