Comparison of Deep Learning and Machine Learning Approaches for the Recognition of Dynamic Activities of Daily Living

Cassandra Krause[1,⋆], Lena Harkämper[1,⋆], Gabriela Ciortuz[1][0000−0001−9443−7825], and Sebastian Fudickar[1][0000−0002−3553−5131]

Institute of Medical Informatics, University of Lübeck, Lübeck, Germany
{gabriela.ciortuz,sebastian.fudickar}@uni-luebeck.de

Abstract. As a consequence of demographic shifts, the proportion of the older population is growing at an accelerated pace, resulting in a notable decline in cognitive and motor functions. This study examines the potential of wearables to monitor activities of daily living (ADL) and identify changes in behavior, thereby enabling early intervention to maintain the independence of the person. Eight dynamic ADLs were analyzed using data collected from eight subjects who were wearing a sensor belt that includes an accelerator and a gyroscope. The data were preprocessed and employed to train and evaluate two distinct types of classifiers: a deep learning and several machine learning approaches. Two data splits were considered: a subject-dependent model, which utilized data from all subjects for training and testing, and a Leave-One-Subject-Out Cross-Validation (LOSO-CV) subject-independent model, which excluded one subject from the training set for validation. The subject-dependent approach yielded high accuracies of 99.5% and 99.8% for the classification network and the best support vector machine, respectively. The LOSO-CV yielded accuracies of 77.8% for the convolutional neural network and 77.6% for the best support vector machine. While the classification network demonstrated marginally superior results, the support vector machine required significantly less training time, suggesting its potential suitability for practical applications.

Keywords: human activity recognition · machine learning · feature extraction · feature learning · convolutional neural network · support vector machine · activities of daily living

1 Introduction

As a consequence of demographic change, the number of older people is rising rapidly. In old age, individuals experience a decline in cognitive abilities and, most notably, in motor functions. To assess physical decline, the activities of daily living (ADL) are frequently evaluated through the administration of a

⋆ Both authors contributed equally to this research.

questionnaire. However, this approach often results in a considerable degree of measurement noise, as self-perception frequently differs from the actual measured reality. It would be more prudent to assess the ADL over an extended period to identify changes and implement preventative measures to delay dependence. Technical assistance would be beneficial, as the scarcity of medical professionals necessitates that such assessments be conducted independently to alleviate the burden on them [12]. Additionally, individuals exhibit disparate behaviors in their own homes than under the supervision of a medical professional. Therefore, it may be advantageous to implement home access systems or to use wearables to facilitate continuous monitoring [7,22].

This study examines the potential for measuring ADL using wearable technology. For this purpose, we self-designed and implemented all the steps of a Human Activity Recognition (HAR) system. The study design is depicted in Figure 1. Once the research question was defined, eight dynamic ADLs were selected for analysis. Dynamic movements, such as walking or jogging, exhibit variations over time, while static activities like sitting or lying down produce relatively constant readings. We decided to incorporate dynamic activities - even activities that require higher levels of physical engagement, like jogging or jumping - into our analysis, in order to enable the identification of subtle behavioral changes at an early stage. Detecting these changes early, while the individual is still capable of performing such activities, allows for timely interventions. Interventions, such as targeted exercise and training, can help to extend the period during which the person remains physically active and dynamic, thereby maintaining their independence for a longer time. Studies have shown that especially jumping, which is a rebound exercise, improved the Timed-Up and Go (TUG) test results in the rebound group twice compared to the control group [21]. This significant reduction in TUG test times not only indicates an improved ability to stand up from a chair more quickly, walk faster with better balance, turn more efficiently, and sit down with greater control without assistance, but also underscores the importance of jumping in assessing functionality in older adults. Jumping challenges and enhances coordination, strength, and balance, which are critical components for maintaining independence and reducing the risk of falls in this population [20]. Our approach aims to recognize eight dynamic activities in order to address emerging issues before they significantly impact the person's daily life.

The activities were performed and recorded in a separate data set of eight test subjects. A sensor belt was employed to quantify acceleration and rotation. Following the preprocessing of the recorded data, two types of classifiers were analyzed: a deep learning approach (DL) and a machine-learning (ML) approach. In addition, two data splits were analyzed: a person-specific approach, in which the data from all subjects is used for training, and a LOSO-CV to ascertain whether the data from an unknown subject can be robustly classified. In order to facilitate the testing phase, the data pertaining to a single subject was excluded from the training set. These approaches are then compared with each other and the reasons for any misclassifications are explained.

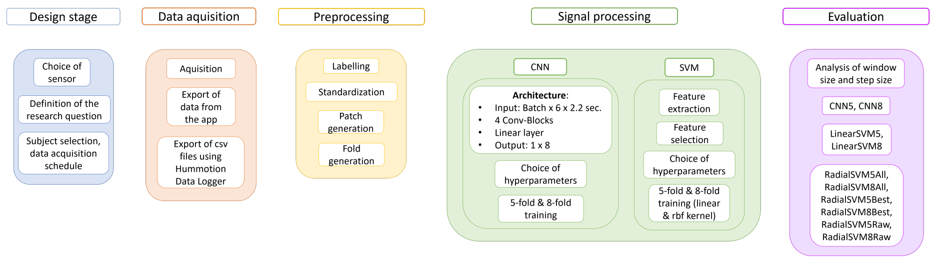

Fig. 1. Study design. The single steps of the study are presented, starting with the design stage, followed by the data acquisition, preprocessing and signal processing of the data. In the end, the evaluation steps are presented.

2 Related Work

In this section, we discuss the most common design choices of HAR system components for healthcare in existing work. HAR systems can be categorized based on the data acquisition tools used, classification methods employed and the complexity of the activities being analyzed [25]. We will briefly cover these aspects in the following paragraphs.

In terms of the acquisition tools, especially for healthcare applications such as activity monitoring of elderly people, wearable devices are preferred over cameras because of privacy concerns and because they satisfy the long-term usability of a monitored environment [25,16].

Inertial sensor-based systems that employ accelerometers, gyroscopes and magnetometers can be easily integrated into wearable devices, which can be used to record movements on various body sites at the same time. However, the inconvenience and intrusiveness of wearing numerous devices can impose additional burdens on users, thus many studies prefer to use only one sensor, such as smartphones, which have become indispensable in people's everyday lives [8]. The authors in [15] proposed using one Inertial Measurement Unit (IMU) integrated into a belt to ensure effortless use and discreet sensor placement, particularly to allow an easy process of future self-assessments at home. The employed sensor must have the ability to model the motion information of the activities of interest so that it will be able to classify it [25].

Defining activities of interest depends on the target application or desired health-related outcome [6]. The authors in [13] proposed six activities - sitting, standing, walking, laying, going up and going downstairs, for elderly patients' monitoring of their general health status and dynamism. In contrast, we decided to use only dynamic activities, and not consider pose-related activities in our models.

Depending on the nature of the activity being analyzed - e.g. simple or complex actions, poses, locomotion patterns or ADLs - as well as the amount of training data available, various machine-learning models can be used for accurate activity recognition [18].

Traditional ML approaches like SVMs, decision trees and KNNs performed well across many HAR application areas [18] but require feature engineering to execute optimally [25].

Especially SVMs have been shown in [3,18] to perform well with data of unknown distribution and once its decision boundary is established, SVM proves to be a robust classifier and scales effectively for high-dimensional data. These facts motivated our decision to use the SVM in our research as the traditional ML approach of choice.

Traditional ML methods are increasingly outperformed by deep neural networks (DNNs), particularly in their ability to learn complex representations from multidimensional data [9].

Deep learning is the second big approach for HAR after traditional ML using handcrafted features but requires a substantial amount of training data to learn intricate features effectively. DL is an evolution of ML in that it uses deeper networks to autonomously learn and recognize more complex and hierarchical patters in large amounts of data without the need for manual feature extraction. Deep Artificial Neural Networks (ANNs) emulate the human neural system, with the primary goal of extracting non-linear relationships from data for classification, thus automatically extracting and learning the features instead of manually engineering and selecting them [10].

Convolutional Neural Networks (CNN) are powerful neural networks in image processing tasks using convolutions to learn representative spatial features [4]. CNNs perform well in recognizing locomotion activities like walking, cycling, sitting, or standing, but more complex activities made up of multiple simpler activities would require more specialized models like recurrent neural networks (RNNs) or hybrid architectures [2].

Since the collected dynamic activities are not very complex, and because CNNs have been shown to work well and converge faster on multidimensional time-series data when compared with RNNs [3], a CNN is selected as the deep learning model for the subsequent study.

3 Methods

In the following, the underlying methods of this study will be outlined. First, the recorded dataset will be presented. Subsequently, the deep learning and machine-learning approaches employed are outlined, along with a description of the experiments conducted.

3.1 Dataset

This study aims to examine the potential of deep learning and machine learning classifiers for the activities of daily living. To attain this objective, this eight dynamic activities were selected:

- Bend

- Walk
- Jump
- Jog
- Rotate left
- Rotate right
- Walking up the stairs
- Walking down the stairs

In this research, a self-acquired data set is employed. To attain this objective, we employed the SmarTracks Sensor DX5.0 Timing in conjunction with the TB40 belt. The data is recorded via an Android smartphone using the Smart Run app [17]. The Smart Run app paired with a SmarTracks DX5.0 sensor provides real-time timing results, which are streamed directly to your smartphone. The app also provides a way of exporting the raw data recorded with the SmarTracks Sensor, which we have used for our study. All subjects were introduced to the functionality of the device and were permitted to initiate and terminate the measurements independently. The smartphone was retained on the body throughout the recording period to ensure continuous connectivity with the sensor. The DX5.0 Timing comprises three sensors: a three-axis accelerometer, a three-axis gyroscope and a three-axis magnetometer. Each of the sensors was initiated with a sampling rate of 250 Hz. Eight subjects, designated S_i with $i \in \{1, 2, 3, 4, 5, 6, 7, 8\}$, were included in the study. In order to obtain standardized measurements, it is essential that the placement of the sensor belt and the position of the sensor unit remain consistent across all subjects. The belt was secured around the waist, between the L3 and L5 lumbar vertebral bodies, in accordance with the methodology described by Hellmers et al. [15]. Four of the test subjects were female, and four were male. The mean age of the subjects was 24.5 years, with a standard deviation of 3.2 years. The mean height was 1.71 meters, with a standard deviation of 0.07 meters. The cohort thus represents a targeted area of the population from 21 to 27 years.

Each subject was required to conduct each of the eight activities continuously for approximately two minutes, with the aim of achieving a balanced distribution of the activities. In order to facilitate subsequent data labelling, only the described dynamic activities were recorded, without transitions between different exercises. This approach enabled the assignment of one label to each whole measurement, alleviating the subsequent analysis.

As the accelerometer and gyroscope are capable of displaying movements with a high degree of precision, the magnetometer measurements were excluded from the subsequent discussion to avoid calibration and interference problems and to reduce the dimensionality of the data as it is also done in the experiments of Pesenti et al. [23]. As the test subjects conducted the measurements independently and there was a brief interval between the start of the measurement and the start of the activity, the initial and last five seconds of the individual measurements were excluded as the initial step in the preprocessing procedure. This was due to the fact that the subjects frequently placed the smartphone in their trouser pockets after initiating the recording process, allowing them to perform the desired movements with their hands free. For the two stair-climbing activities, the window was set to 2.5 seconds. This ensured that only the movement itself was presented in the data, as it was observed during the data recording that the subjects started to climb the stairs at a considerably faster pace than the planned five seconds.

The subsequent step involved standardizing the data across the three axes of the accelerometer and gyroscope, by using z-normalization, which is a standard preprocessing technique that can significantly enhance the learning capabilities of deep neural networks and also removes certain types of bias effects [1,5]. Consequently, the mean and standard deviation were calculated for all subjects. Afterwards, a windowing operation was performed in order to augment the data. After performing a search to find the optimal window and stride sizes, the width of the window was set to 2.2 seconds, while the stride was set to 0.1 seconds. The process of finding the optimal window and stride sizes is described in Chapter 3.4. The data after each preprocessing step is shown in Figure 2.

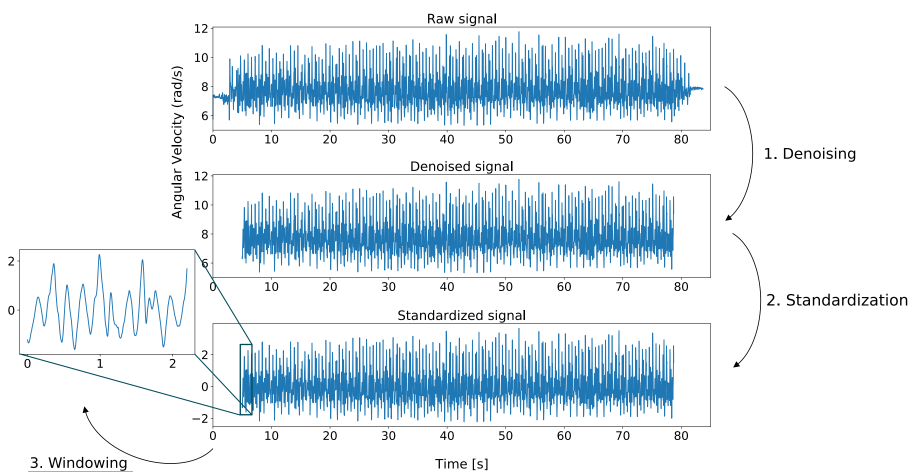

Fig. 2. Preprocessing pipeline of the data. A measurement of the gyroscope of the x-axis of subject S_1 of the activity *Walk* is shown. The data is visualized after each preprocessing step: after the denoising and the standardization (z-normalization).

Given the specified window length of 2.2 seconds and stride of 0.1 seconds, the resulting number of segments over all subjects amounts to 64164. The exact number of data, dependent on subject and activity, can be found in Table 1. The table shows that the segments of the classes *Walking up the stairs* and *Walking down the stairs* are significantly fewer than those of the other classes. This results in class imbalance due to the fact that the stair which was used is not particularly long, with each recording therefore being approximately ten to fifteen seconds in length. Furthermore, the first and last seconds of each recording have been omitted, which has also resulted in a reduction in the number of segments.

Table 1. A summary of the segments recorded per subject and activity, with a window length of 2.2 seconds and a stride of 0.1 seconds.

Subject	**Activity**							
	Bend	**Walk**	**Jump**	**Jog**	**Rotate left**	**Rotate right**	**Walking up the stairs**	**Walking down the stairs**
S_1	1173	1090	1119	920	946	1173	364	284
S_2	1119	1441	1115	898	1062	1124	847	781
S_3	1222	1427	1211	1020	1179	1173	989	751
S_4	1312	1086	1165	960	1116	1169	921	891
S_5	1169	1020	1234	1039	1211	1192	871	770
S_6	1180	1312	1169	1079	1188	1161	860	767
S_7	1103	1035	1161	1073	1138	1146	444	305
S_8	1222	822	1119	701	1010	756	476	383
Total	9500	9233	9293	7690	8850	8894	5772	4932

3.2 Classification network

The deep learning approach utilizes a similar architecture to the one proposed in [24], which is illustrated in Figure 3.

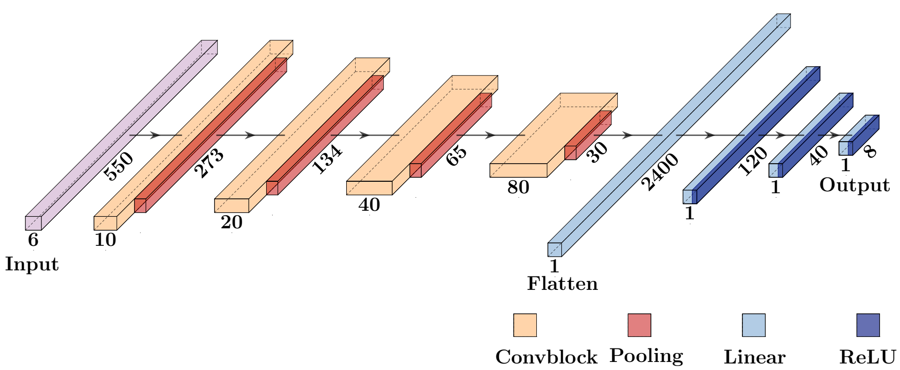

Fig. 3. Architecture of the proposed classification network. Four Conv-Blocks followed by max-pooling layers are used for feature extraction. Three fully-connected layers are used to obtain the output size of eight in order to represent the class probabilities of the activities.

The classification network receives a tensor of the shape $(B \times C \times N)$ as input, where B represents the batch size, C the number of channels and N the length of the sequence. Given that the input comprises patches of 2.2s from three axes of the accelerometer and the gyroscope, the dimensions of the input tensor are $C = 6$ and $N = 550$. The classification network consists of four

Conv-Blocks. Each Conv-Block comprises a convolutional layer, which doubles the dimensions, followed by a batch normalization and a rectified linear unit (ReLU) activation. A second convolutional block, comprising a convolutional layer, batch normalization and ReLU activation, is then applied, maintaining the number of channels. The Conv-Block is shown in Figure 4.

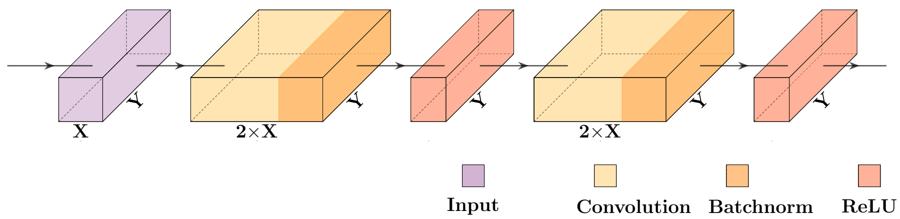

Fig. 4. Architecture of the proposed Conv-Block. The feature size is first doubled by a convolution, followed by a batch normalization and an activation layer. A second block of convolution, normalization and activation is applied, keeping the feature size.

Following each convolutional block, the dimensions are reduced by a factor of 50% through a max-pooling operation. Subsequently, the feature maps resulting from the final Conv-Block are flattened and fed through three linear layers, with the objective of achieving an output size of eight, which defines the probability for each activity class.

3.3 Support Vector Machine

In the ML approach, a support vector machine (SVM) was employed. As no features are learned automatically in this context, we utilize handcrafted features from the preprocessed segments that are suitable for IMU data [26]. Therefore, the following time-domain features are included:

- Minimum (Min)
- Maximum (Max)
- Arithmetic Mean (Mean)
- Standard deviation (SD)
- Variance (Var)
- Zero crossing rate (ZCR)
- Mean crossing rate (MCR)
- Positive peak count (PPK)
- Negative peak count (NPK)

In addition to a linear kernel, a Gaussian radial basis function (RBF) kernel is also analyzed. Given that this is a multi-class problem, the 'one versus rest' separation is employed.

3.4 Experiments

We describe our experiments in the following section as follows:

First, we present the search for the best stride and window length, which was performed to prepare the data for training. Subsequently, we describe how we split the data into three distinct sets, namely training, validation and test set, via a K-fold cross-validation approach. Finally, we provide the hyperparameters of the two approaches and the evaluation metrics for the validation of our models.

Since it is important for the classification of dynamic movements based on sensor data that the movements are completely contained in a window, the analysis of the window size is of great importance. Hellmers et al. [15] achieved the best results with a window length of 1.853 and a stride of 0.249 for dynamic activities. Therefore, we chose a range around these values to find out the best window length and step size for our recorded activities. For this purpose, we trained the CNN on a fold with 20 different window sizes ranging from 0.4 to 2.3 seconds with a fixed stride of 0.25 seconds and calculated the accuracy in each case to get the range of the sufficient window length. In the second step, a cross-analysis of window length between 1.8 and 2.2 and strides between 0.1 and 0.3 is conducted in order to identify the optimal pair of values. These are then employed in the subsequent analysis of all approaches, since related work has shown that the window and step sizes found for specific activities are optimal irrespective of the classification method [11].

Both the CNN and the SVM approach are validated through a K-fold cross-validation. First, a 5-fold validation is performed. For this purpose, the data of the individual activities is mixed across all subjects and equally distributed so that each task is presented with the same frequency in each fold. For the validation set, 10% of the training data was used at random. This resulted in an average training set of 46,200 segments, a validation set of 5,134 segments, and a test set of 12,830 segments. This approach may be beneficial in the context of subject-specific systems, which involve the adaptation of a classifier to a known set of subjects. In order to assess the generalizability of the classifiers on data from a subject, of which there is no previous knowledge in the training, an 8-fold cross-validation was also analyzed. In this case, one subject was excluded from the training set and used for testing across all activities. This resulted in the use of seven subjects for training purposes and one subject for testing. Additionally, 10% of the training data was randomly selected for validation, resulting in the following average quantities: A total of 51385 segments were used for training, 5710 for validation and 7069 for testing.

The classification network was trained for 200 epochs, with an early stopping mechanism built in to terminate the training if the validation loss did not improve for 10 epochs. Additionally, a batch size of 16 was selected. The Adam optimizer was used to optimize the weights with an initial learning rate of 0.001. A learning rate scheduler was applied, which reduces the learning rate by 0.1 every 20 epochs. The cross-entropy loss function was employed to quantify the error. Given that the classes are not balanced, it is necessary to address this issue in the classifiers. Therefore, the weights were assigned to the classes according to their probability of occurrence. The weights of the optimal network were utilized for the analysis of the test data set. In the following, the classification network with the 5-fold cross-validation will be referred to as *CNN5* whereas the one with the 8-fold cross-validation will be referred to as *CNN8*.

In order to identify the optimal features for the SVM, a random feature elimination (RFE) was employed. This involves training the SVM on the initial feature set and recursively considering an ever smaller feature set until only one feature remains, resulting in a ranking of the features. The lower the ranking, the more important the feature is, so the features with the lowest rankings are the ones that are selected. It is also necessary to address the issue of class imbalance in the SVMs. So the SVM with linear kernel was trained with weights automatically set inversely proportional to the class frequencies, thus compensating for any imbalance in the classes. The maximum number of iterations was set to 50,000. The dual parameter was set to false, as a greater number of segments were designated as features. A grid search was conducted to identify the optimal values for parameters C and *tol*. This resulted in a C of 10 and a *tol* of 1e-05 for the linear kernel. For the SVM with the RBF kernel, the automatic balancing of the classes was also applied, and a maximum number of iterations of 50,000 was selected. The gamma parameter was set to scale, and the C and *tol* parameters were also determined here using the grid search method. In the following, the SVM trained with the linear kernel on the best features resulting from the RFE with a 5-fold or an 8-fold cross-validation is labelled with *LinearSVM5* and *LinearSVM8* respectively. The SVM with the RBF kernel is trained with the raw data (*RadialSVM5Raw* and *RadialSVM8Raw*), on all hand-crafted features (*RadialSVM5All* and *RadialSVM8All*) and also on the best features (*RadialSVM5Best* and *RadialSVM8Best*) for the 5-fold and 8-fold cross-validation.

To validate the test dataset for both the CNN and the SVM, a confusion matrix was constructed based on the predictions and the ground truth. The following metrics are employed to assess the performance of the classifiers: accuracy, macro F1-score, precision and recall.

4 Results

This section provides the analysis results of the optimal window and stride selection and of the optimal feature selection using RFE. In addition, the results for the metrics accuracy, macro F1-score, precision and recall of the approaches *CNN5*, *CNN8*, *LinearSVM5*, *LinearSVM8*, *RadialSVM5All*, *RadialSVM8All*, *RadialSVM5Best*, *RadialSVM8Best*, *RadialSVM5Raw* and *RadialSVM8Raw* are shown.

4.1 Window and stride

The results of the window length and stride analysis are presented in Figure 5.

It can be observed in Figure 5 that the accuracy improves as the window length increases. When recording the data for the activity *Jump*, there were

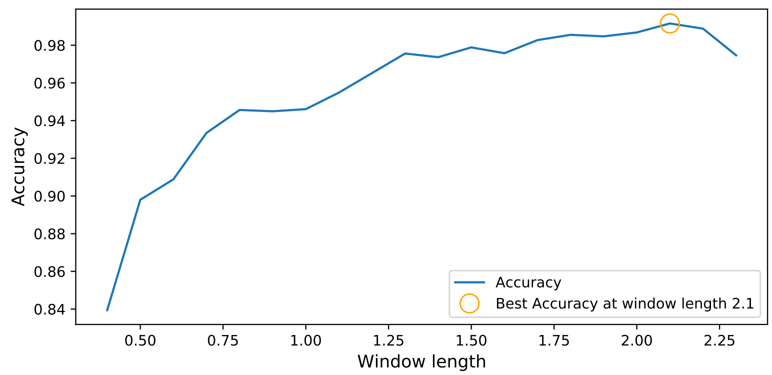

Fig. 5. Analysis of the optimal window length. The best accuracy was reached with a window length of 2.1 seconds (yellow circle).

often small pauses between the jumps, which means that the required window must be large enough to cover the entire motion. The best accuracy values are between 1.8 and 2.2 seconds, so these values were used in the next step for a cross-analysis with the stride.

The results of the cross-analysis are presented in Figure 6. The x-axis initially displays a fixed window length, after which the three strides between 0.1 and 0.3 are presented. It can be observed that the accuracy is highest for the smallest stride of 0.1, which yields the greatest amount of data, and then declines as the stride increases. The highest accuracy is achieved with a window length of 2.2 seconds and a stride of 0.1 seconds. These values will be used for all the approaches in the following.

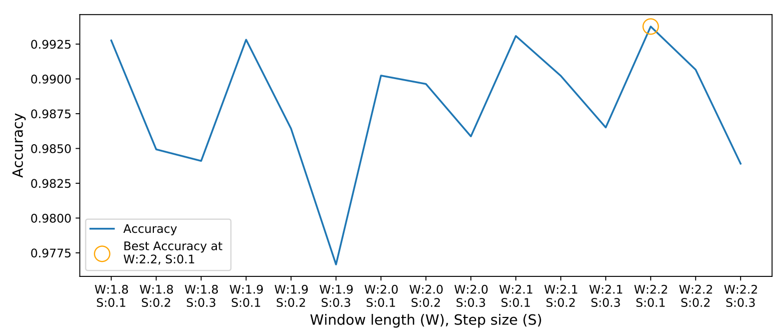

Fig. 6. Cross analysis of window length and stride. The x-axis first displays a fixed window length W, followed by three strides S from 0.1 to 0.3. The highest accuracy is achieved with $W = 2.2$ and $S = 0.1$.

4.2 CNN

The CNN was trained using K-fold cross-validation with K=5 (CNN5) and K=8 (CNN8). Table 2 shows the results of the metrics of CNN5 per fold, as well as the results averaged over all folds.

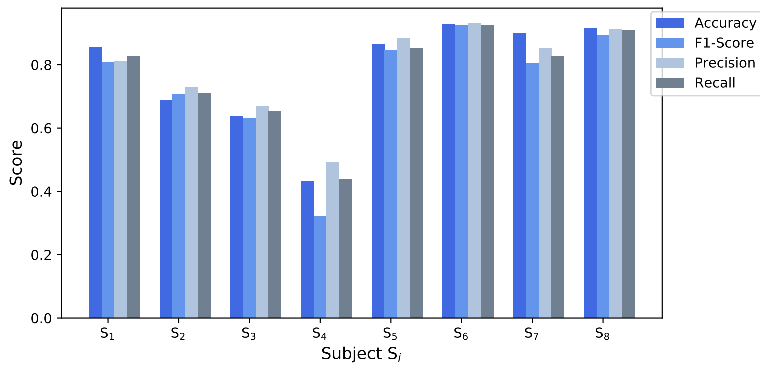

Fig. 7. Subject-wise scores for accuracy, macro F1-score, precision and recall of CNN8. Subject S_4 achieved notably lower scores.

Table 2. CNN 5-Fold comparison in terms of different metrics.

Fold	Accuracy	F1-Score	Precision	Recall
1	**0.996**	**0.996**	**0.996**	**0.996**
2	**0.996**	**0.996**	0.995	**0.996**
3	0.995	0.994	0.994	0.994
4	0.995	0.993	0.994	0.993
5	0.994	0.994	0.994	0.993
Average	0.995	0.994	0.994	0.994

Figure 8 shows that the tasks can be separated very well and there are only few misclassifications.

To show the extent to which the CNN generalizes even with a LOSO-CV approach of a subject, Figure 7 displays the scores of the 8-fold for accuracy, macro F1-score, Review 1: Use consistent capitalization and recall. It can be observed that subject S_4 achieved notably lower scores compared to the other subjects. The scores of subjects S_2 and S_3 are also well below average.

As illustrated in Figure 9, a diagonal can still be discerned in the confusion matrix, although there is a significant increase in misclassifications. It is noteworthy that a considerable number of tasks have been identified as *Walk*. Other

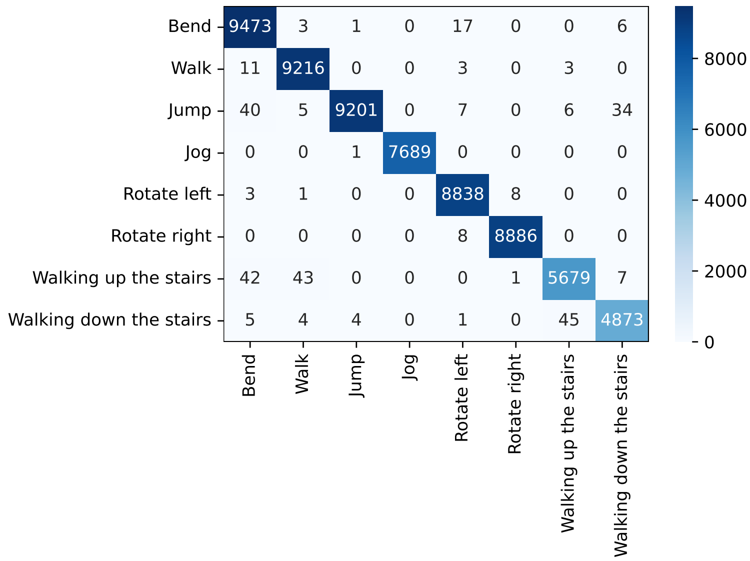

Fig. 8. Confusion matrix of CNN5. Only a few misclassifications have occurred.

common errors in classification included the misclassification of *Walk* as *Walking up the stairs* and *Jump* as *Walking down the stairs*. Furthermore, the activity *Jog* was often misclassified as *Jump*, and the activity *Rotate* was frequently misclassified as *Bend*.

In order to ascertain which activities can be classified as particularly well or poorly, Figure 10 illustrates the accuracy for each activity. It can be observed that specific movements, such as *Bend* or *Rotate*, which differ significantly from the others in terms of execution, can be classified with the greatest accuracy. The activity *Walk* is the least well-recognized.

Table 3. Comparison of the average values of CNN5 and CNN8 in terms of different metrics

Model	**Accuracy**	**macro score**	**F1- Precision**	**Recall**
CNN5	0.995	0.995	0.995	0.994
CNN8	0.778	0.743	0.786	0.786

Table 3 shows the comparison of the average values of *CNN5* and *CNN8* in terms of different metrics. *CNN5* demonstrates superior performance, as the data exhibits greater variability across the folds.

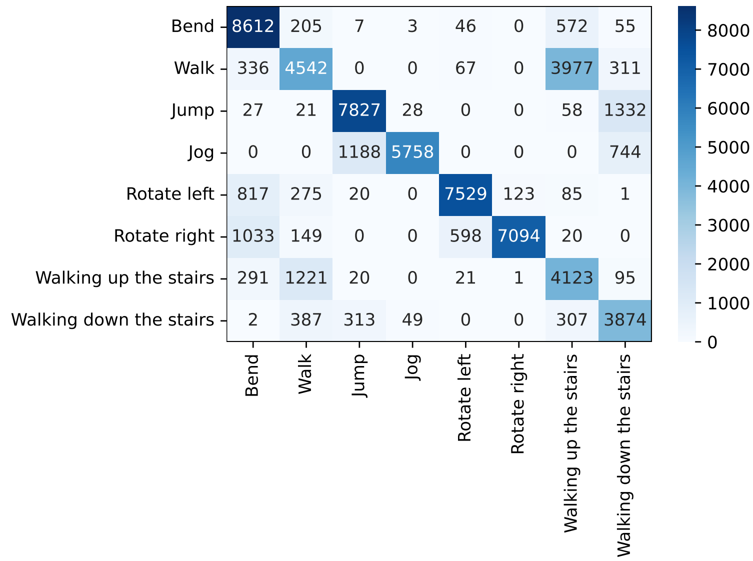

Fig. 9. Confusion matrix of CNN8.

4.3 SVM

The results of the optimal feature selection using RFE are shown in Figure 11. The features comprise the nine statistical values for the six channels, resulting in a total of 54 features. The figure displays the summed ranks of the six channels for the respective statistical values. The lower the rank, the more important the feature is. Therefore, the mean crossing rate was identified as the most important feature, and the top five features were selected for further experimentation: mean crossing rate, zero crossing rate, standard deviation, mean and variance.

Table 4 presents the averaged results of the SVM models. In a manner analogous to that observed in the case of *CNN5* and *CNN8*, the outcomes of the 5-fold models are significantly better than those of the 8-fold models. In the 5-fold cross-validation, *RadialSVM5All* achieved the most favorable results, closely followed by *RadialSVM5Best*. In the LOSO-CV, *RadialSVM8Best* achieved the highest scores.

Figure 12 illustrates the accuracy of *RadialSVM8Best* task-wise. This analysis allows us to identify which activities may be classified more or less effectively. In comparison to *CNN8* (see Figure 10), the results are notably more balanced; however, *Walk* is also the activity that is classified with the greatest difficulty. This phenomenon can also be observed in the confusion matrix of *RadialSVM8Best* in Figure 13. Here, the activity *Walk* was most frequently classified as *Walking down the stairs* too; other misclassifications were similar to those observed in the *CNN8*.

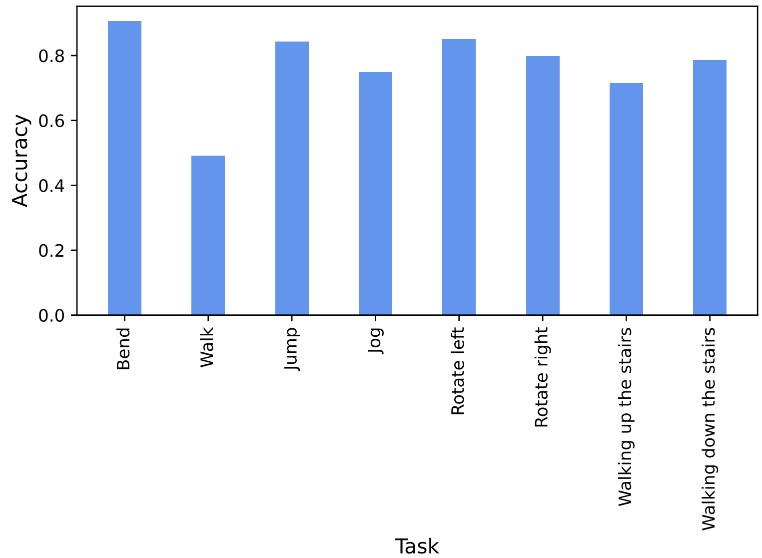

Fig. 10. Task-wise accuracy of CNN8.

5 Discussion

The objective of this study was to investigate the classification of activities of daily living based on self-recorded sensor data from a non-intrusive and easy-to-use sensor belt using a deep learning approach and a machine learning approach. Similarly to [16], we observed that our approach to recognize activities via a single inertial sensor worn at the waist, is both easily applicable and well-suited for recordings in a home environment.

One limitation in our study is related to the generalizability of our models, which were trained predominantly on data from younger individuals. Consequently, these models may not perform optimally when applied to elderly populations, as highlighted in [14,19]. To address this limitation we plan to include

Table 4. SVM metrics

Model	Accuracy	F1-Score	Precision	Recall
LinearSVM5	0.983	0.980	0.980	0.979
RadialSVM5All	0.992	**0.991**	**0.991**	**0.992**
RadialSVM5Best	**0.998**	0.975	0.976	0.975
RadialSVM5Raw	0.985	0.983	0.983	0.983
LinearSVM8	0.771	0.728	0.769	0.759
RadialSVM8All	0.763	0.728	**0.788**	0.755
RadialSVM8Best	**0.776**	**0.731**	0.771	**0.766**
RadialSVM8Raw	0.699	0.653	0.695	0.700

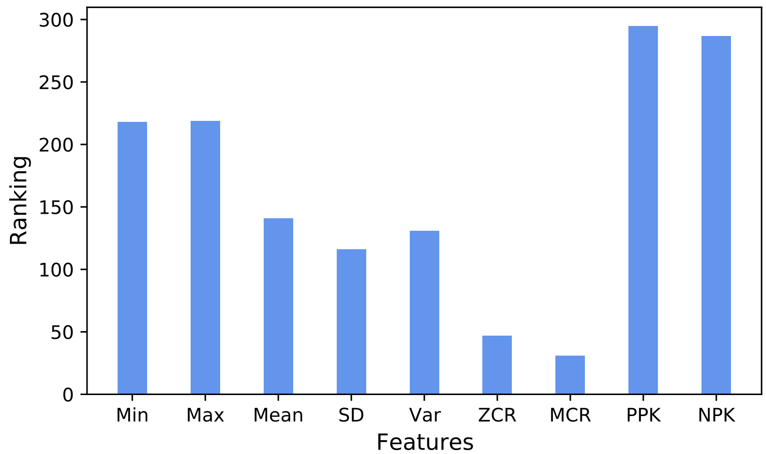

Fig. 11. Ranking of relevance of the handcrafted statistical features. The ranking was calculated using RFE. Lower rankings indicate a higher relevance.

more diverse training data, particularly varying in age and mobility in the future. Still, our study serves as a foundational investigation, demonstrating the effectiveness of using a single sensor belt for activity recognition. This approach is promising and warrants further exploration with datasets specifically collected from elderly participants.

Both a person-specific system analysis and a LOSO-CV were conducted. In the person-specific system analysis, data from all test subjects was used for testing, while in the LOSO-CV, data from one test subject was used for testing. The following section will present a comparison between the deep learning approach and the machine learning approach. Subsequently, the misclassifications are analyzed.

5.1 Comparison of SVM and CNN

Table 5. CNN vs SVM

Model	**Accuracy**	**F1-Score**	**Precision**	**Recall**
CNN8	**0.778**	**0.743**	**0.786**	**0.768**
RadialSVM8Best	0.776	0.731	0.771	0.766

The results show that the 5-fold approaches, where all subjects were included in the training- and test data, yielded highly accurate results, with an average accuracy of 99.5% for the classification network and 98.95% for all SVM approaches. Only a few misclassifications were observed.

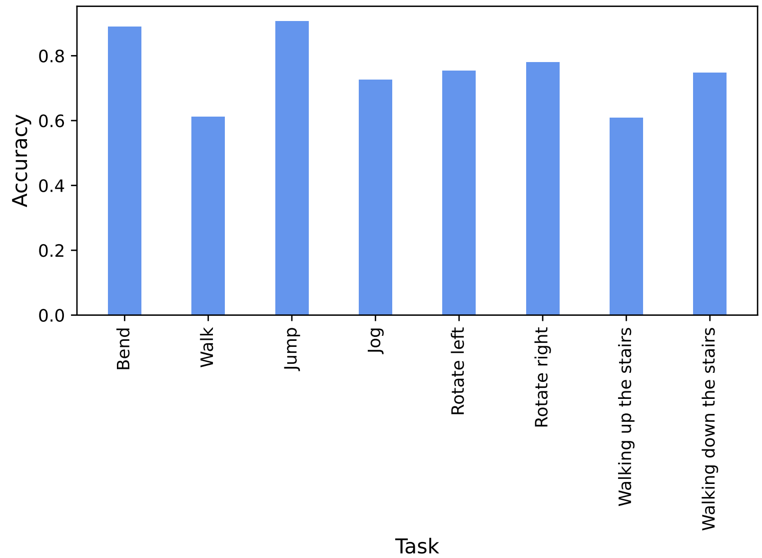

Fig. 12. Task-wise accuracy of RadialSVM8Best.

Table 5 presents a comparative analysis of the DL approach and the ML approach, demonstrating the performance of *CNN8* and *RadialSVM8Best* in terms of accuracy, macro F1-score, precision and recall. In contrast, *CNN8* demonstrates slightly superior performance, yet the SVM necessitates considerably less training time. Specifically, a single fold requires only a few minutes with the SVM, whereas the CNN requires approximately two hours for the same task.

The results of the LOSO-CV are presented in Figure 15. A comparable trend can be observed in the case of subjects presenting a greater difficulty level (outliers), resulting in similar levels of inaccuracy across the board (subjects S_2, S_3 and S_4). Similarly, subjects S_6, S_7 and S_8 demonstrate the most favorable outcomes with both approaches.

Figure 14 illustrates the receiver operating characteristic (ROC) curve for both approaches in the optimal fold. The ROC curves of the individual activities for the *RadialSVM8Best* are displayed in the upper diagram, and those of the *CNN8* are displayed on the bottom. It is evident that the SVM curves are of a high quality, whereas the CNN curves demonstrate lower performance, particularly for the classes *Walk* (AUC of 0.79) and *Walking down the stairs* (AUC of 0.55). This is attributed to the presence of issues between the classes. Furthermore, the CNN exhibits superior results for the remaining folds in comparison to the SVM, suggesting that it may also be an outlier for this fold.

Nevertheless, the difference between the CNN and the SVM in the third decimal place is marginal, which is why they can be regarded as equally good. This can be attributed to the fact that the training set may not yet be sufficiently

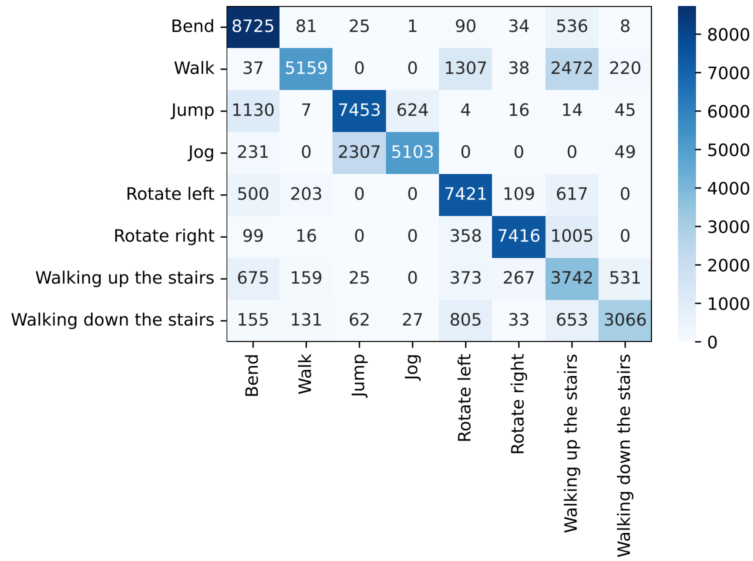

Fig. 13. Confusion Matrix of RadialSVM8Best.

comprehensive for a DL approach, resulting in comparable performance between traditional and DL methods, which may be preferable in certain circumstances. These findings are consistent with the results of related studies that have employed a similar analytical approach.

This study did not examine more sophisticated DL architectures that are well-suited for temporal data. This may also be a reason why a traditional DL network is inadequate for processing this data.

5.2 Misclassifications

The two approaches (CNN and SVM) were trained with K-fold cross-validation, with K=5 and K=8, respectively. At K=5, all subjects were distributed evenly across the folds, resulting in the absence of outliers. This explains the high scores and the low rate of misclassification.

In the LOSO-CV, one subject is omitted from training in each repetition, and the classifier is then tested on this subject. Consequently, it is possible to identify significantly larger differences between the folds. In particular, subject S_4 exhibits a notably low level of accuracy when classified with the CNN (see Figure 7. It was apparent from the outset of the data acquisition process that subject three displayed a markedly unusual range of movements. These included a tendency to walk or run upstairs with a pronounced strolling gait, as well as a pronounced lateral movement when rotating. Consequently, his data exhibited a notable divergence from the mean, rendering him an outlier within the data set. This could be a potential explanation for the observed lower scores.

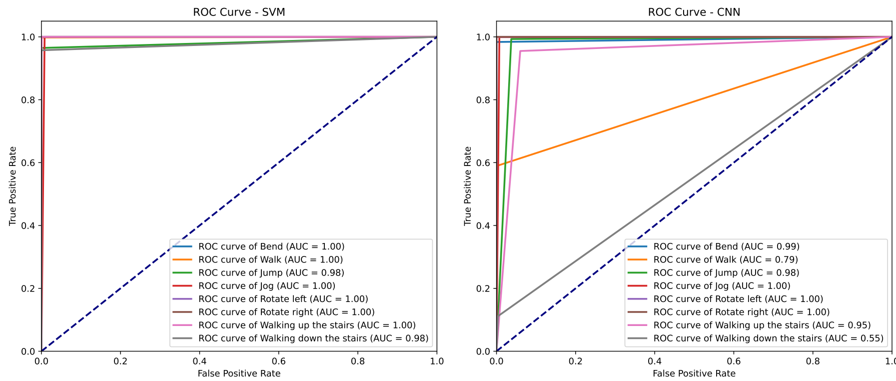

Fig. 14. ROC curves for the *RadialSVM8Best* (left) and the *CNN8* (right) of Subject S_7.

All experiments employing the LOSO-CV methodology demonstrate comparable misclassifications of movements. The frequent misclassification of the activity *Walk* as a separate movement is likely due to the fact that this movement is often part of other activities, such as *Walking down the stairs* or *Walking up the stairs*. As a consequence of its inclusion in almost all other movements, it is arguably the most challenging to distinguish from the others.

The classification of *Walk* as *Walking up the stairs* was likely due to the presence of a plateau on the stairs and subjects' slow pace of ascent. Furthermore, the classification of *Jump* as *Walking down the stairs* was based on the observation of a similar acceleration on the z-axis and a gait that resembled jumping. Similarly, the activity of *Jog* was misclassified as *Jump* due to a similar acceleration observed on the z-axis. The classification of *Rotate* as *Bend* was likely due to the observation that a significant proportion of subjects exhibited a slight hip rotation when bending, particularly when only one arm was moved towards the floor.

6 Conclusion

In order to maintain the independence of older people for as long as possible, the activities of daily living should be tracked regularly so that changes in behavior can be detected as early as possible. In this study, the recognition of eight different activities was investigated using deep learning and machine learning approaches. For this purpose, a self-recorded data set of eight subjects was applied, whereby accelerometer and gyroscope data were recorded. A classification network and various SVM approaches were analyzed. Two data splits were examined: a subject-specific one, in which the training data consisted of all subjects, and a LOSO-CV, in which one subject was excluded from the training and was used for testing. The person-specific approach achieved accuracies

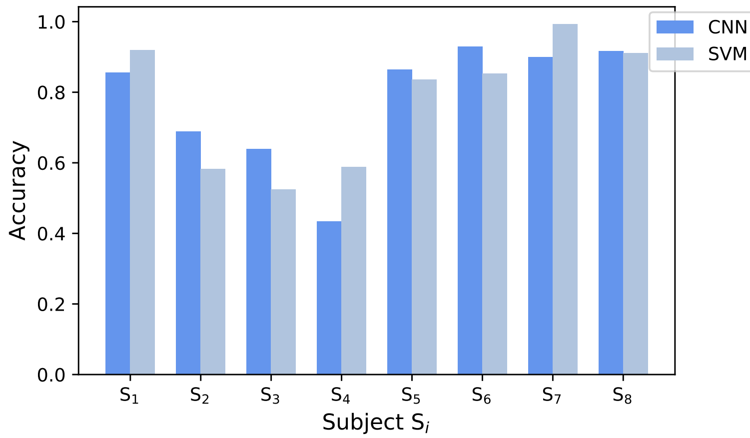

Fig. 15. Subject-wise accuracy of CNN8 and RadialSVM8Best. Subjects S_2 and S_3 and S_3 achieve the lowest accuracy.

of 99.5% and 99.8% for the CNN and for the best SVMs. Also, the LOSO-CV results in appropriate accuracies of 77.8% for the CNN and 77.6% for the best SVM. A number of misclassifications have been identified, some of which are attributable to data recording errors. In summary, the SVM and the CNN yield comparable outcomes, with the CNN demonstrating marginally superior results but also significantly longer training times. In conclusion, the preferred choice of the classifier might depend on the circumstances. On the one hand, the SVM might be the right choice if low computational complexity is preferred, e.g. for real-time systems. On the other hand, if classifier performance is of more importance and the long training time can be disregarded, then CNN might be the better choice.

As was recognized in the LOSO-CV variant, a high data variance is important. It could therefore be useful to include additional test subjects in order to better compensate for outliers such as test subject S_3. In addition, a similar group of people is currently covered, in which the activities do not differ greatly. Therefore, subjects from different age groups could also be included.

For future work, other ML or DL methods could also be taken into consideration. Given that the data are time series, network architectures that contain temporal components, such as recurrent neural networks (RNNs) or long short-term memory (LSTM) networks, could be tested. With regard to the SVM, further features could be investigated that originate from areas other than the time domain, such as Fourier coefficients. A combination of both approaches could also be investigated. In order to avoid the use of hand-crafted features, a CNN could be employed to identify relevant features. Subsequently, a SVM could be utilized to estimate the correct activity class based on the selected features. We can conclude that using one easy-to-use, non-intrusive IMU sensor is able to model dynamic movements and can be successfully used for the recognition of

ADL using traditional ML methods such as the SVM, as well as DL methods such as CNNs.

Acknowledgments. The study is funded by the German Federal Ministry of Education and Research (Project No. 01ZZ2007).

Disclosure of Interests. The authors have no conflict of interest.

References

1. Aggarwal, C.C.: Neural Networks and Deep Learning - A Textbook. Springer (2018)
2. Augustinov, G., Nisar, M.A., Li, F., Tabatabaei, A., Grzegorzek, M., Sohrabi, K., Fudickar, S.: Transformer-based recognition of activities of daily living from wearable sensor data. In: Proceedings of the 7th International Workshop on Sensor-Based Activity Recognition and Artificial Intelligence. iWOAR '22, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3558884.3558895, https://doi.org/10.1145/3558884.3558895
3. Bento, N., Rebelo, J., Barandas, M., Carreiro, A.V., Campagner, A., Cabitza, F., Gamboa, H.: Comparing handcrafted features and deep neural representations for domain generalization in human activity recognition. Sensors (Basel) **22**(19), 7324 (September 2022)
4. Bevilacqua, A., MacDonald, K., Rangarej, A., Widjaya, V., Caulfield, B., Kechadi, T.: Human activity recognition with convolutional neural networks. In: Machine Learning and Knowledge Discovery in Databases, pp. 541–552. Springer International Publishing (2019). https://doi.org/10.1007/978-3-030-10997-4_33, https://doi.org/10.1007/978-3-030-10997-4_33
5. Ciortuz, G., Grzegorzek, M., Fudickar, S.: Effects of time-series data pre-processing on the transformer-based classification of activities from smart glasses. In: Proceedings of the 8th International Workshop on Sensor-Based Activity Recognition and Artificial Intelligence. iWOAR '23, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3615834.3615858, https://doi.org/10.1145/3615834.3615858
6. Ciortuz, G., Pour, H.H., Fudickar, S.: Evaluating movement and device-specific deepconvlstm performance in wearable-based human activity recognition. In: Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2: HEALTHINF. pp. 746–753. INSTICC, SciTePress (2024). https://doi.org/10.5220/0012471300003657
7. Edemekong, P.F., Bomgaars, D., Sukumaran, S., Levy, S.B.: Activities of daily living (2019), https://digitalcollections.dordt.edu/faculty_work/1222
8. Fan, C., Gao, F.: Enhanced human activity recognition using wearable sensors via a hybrid feature selection method. Sensors **21**(19), 6434 (September 2021). https://doi.org/10.3390/s21196434, http://dx.doi.org/10.3390/s21196434
9. Ferrari, A., Micucci, D., Mobilio, M., Napoletano, P.: Hand-crafted features vs residual networks for human activities recognition using accelerometer. pp. 153–156 (June 2019). https://doi.org/10.1109/ISCE.2019.8901021
10. Ferrari, A., Micucci, D., Mobilio, M., Napoletano, P.: Trends in human activity recognition using smartphones. Journal of Reliable Intelligent Environments **7**(3), 189–213 (July 2021). https://doi.org/10.1007/s40860-021-00147-0, http://dx.doi.org/10.1007/s40860-021-00147-0

11. Fida, B., Bernabucci, I., Bibbo, D., Conforto, S., Schmid, M.: Varying behavior of different window sizes on the classification of static and dynamic physical activities from a single accelerometer. Medical Engineering & Physics **37**(7), 705–711 (2015). https://doi.org/https://doi.org/10.1016/j.medengphy.2015.04.005, https://www.sciencedirect.com/science/article/pii/S1350453315001009
12. Fudickar, S., Kiselev, J., Frenken, T., Wegel, S., Dimitrowska, S., Steinhagen-Thiessen, E., Hein, A.: Validation of the ambient tug chair with light barriers and force sensors in a clinical trial. Assistive Technology **32**(1), 1–8 (May 2018). https://doi.org/10.1080/10400435.2018.1446195, http://dx.doi.org/10.1080/10400435.2018.1446195
13. Hayat, A., Morgado-Dias, F., Bhuyan, B.P., Tomar, R.: Human activity recognition for elderly people using machine and deep learning approaches. Information **13**(6) (2022). https://doi.org/10.3390/info13060275, https://www.mdpi.com/2078-2489/13/6/275
14. Hayat, A., Morgado-Dias, F., Bhuyan, B.P., Tomar, R.: Human activity recognition for elderly people using machine and deep learning approaches. Information **13**(6) (2022). https://doi.org/10.3390/info13060275, https://www.mdpi.com/2078-2489/13/6/275
15. Hellmers, S., Izadpanah, B., Dasenbrock, L., Diekmann, R., Bauer, J.M., Hein, A., Fudickar, S.: Towards an automated unsupervised mobility assessment for older people based on inertial tug measurements. Sensors **18**(10) (2018). https://doi.org/10.3390/s18103310, https://www.mdpi.com/1424-8220/18/10/3310
16. Hellmers, S., Kromke, T., Dasenbrock, L., Heinks, A., Bauer, J.M., Hein, A., Fudickar, S.: Stair climb power measurements via inertial measurement units - towards an unsupervised assessment of strength in domestic environments. In: Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies. SCITEPRESS - Science and Technology Publications (2018). https://doi.org/10.5220/0006543900390047, http://dx.doi.org/10.5220/0006543900390047
17. Humotion: Smart run app - reliable running data with the push of a button (2018), https://smartracks.run/de/smartracks-run-app/
18. Kulsoom, F., Narejo, S., Mehmood, Z., Chaudhry, H., Butt, A., Bashir, A.: A review of machine learning-based human activity recognition for diverse applications. Neural Computing and Applications **34** (August 2022). https://doi.org/10.1007/s00521-022-07665-9
19. Matthies, D.J., Haescher, M., Nanayakkara, S., Bieber, G.: Step detection for rollator users with smartwatches. In: Proceedings of the 2018 ACM Symposium on Spatial User Interaction. p. 163–167. SUI '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3267782.3267784, https://doi.org/10.1145/3267782.3267784
20. Moran, J., Ramirez-Campillo, R., Granacher, U.: Effects of jumping exercise on muscular power in older adults: A meta-analysis. Sports Medicine **48**(12), 2843–2857 (Oct 2018). https://doi.org/10.1007/s40279-018-1002-5, http://dx.doi.org/10.1007/s40279-018-1002-5
21. Okemuo, A.J., Gallagher, D., Dairo, Y.M.: Effects of rebound exercises on balance and mobility of people with neurological disorders: A systematic review. PLOS ONE **18**(10), e0292312 (Oct 2023). https://doi.org/10.1371/journal.pone.0292312, http://dx.doi.org/10.1371/journal.pone.0292312
22. Osakwe, Z.T., Larson, E., Agrawal, M., Shang, J.: Assessment of activity of daily living among older adult patients in home healthcare and skilled

nursing facilities: An integrative review. Home Healthcare Now **35**(5) (2017), https://journals.lww.com/homehealthcarenurseonline/fulltext/2017/05000/assessment_of_activity_of_daily_living_among_older.4.aspx
23. Pesenti, M., Invernizzi, G., Mazzella, J., Bocciolone, M., Pedrocchi, A., Gandolla, M.: Imu-based human activity recognition and payload classification for low-back exoskeletons. Scientific Reports **13**(1), 1184 (2023). https://doi.org/10.1038/s41598-023-28195-x, https://doi.org/10.1038/s41598-023-28195-x
24. Ronao, C.A., Cho, S.B.: Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications **59**, 235–244 (2016). https://doi.org/https://doi.org/10.1016/j.eswa.2016.04.032, https://www.sciencedirect.com/science/article/pii/S0957417416302056
25. Serpush, F., Menhaj, M.B., Masoumi, B., Karasfi, B.: Wearable sensor-based human activity recognition in the smart healthcare system. Computational Intelligence and Neuroscience **2022** (February 2022). https://doi.org/10.1155/2022/1391906, http://dx.doi.org/10.1155/2022/1391906
26. Steven Eyobu, O., Han, D.S.: Feature representation and data augmentation for human activity classification based on wearable imu sensor data using a deep lstm neural network. Sensors **18**(9) (2018). https://doi.org/10.3390/s18092892, https://www.mdpi.com/1424-8220/18/9/2892