# Miss-placement Prediction of Multiple On-body Devices for Human Activity Recognition

Robin Dönnebrink
robin.doennebrink@motionminers.com
MotionMiners GmbH
Dortmund, NRW, Germany

Fernando Moya Rueda
fernando.moya@motionminers.com
MotionMiners GmbH
Dortmund, NRW, Germany

Maximilian Stach
maximilian.stach@motionminers.com
MotionMiners GmbH
Dortmund, NRW, Germany

René Grzeszick
rene.grzeszick@motionminers.com
MotionMiners GmbH
Dortmund, NRW, Germany

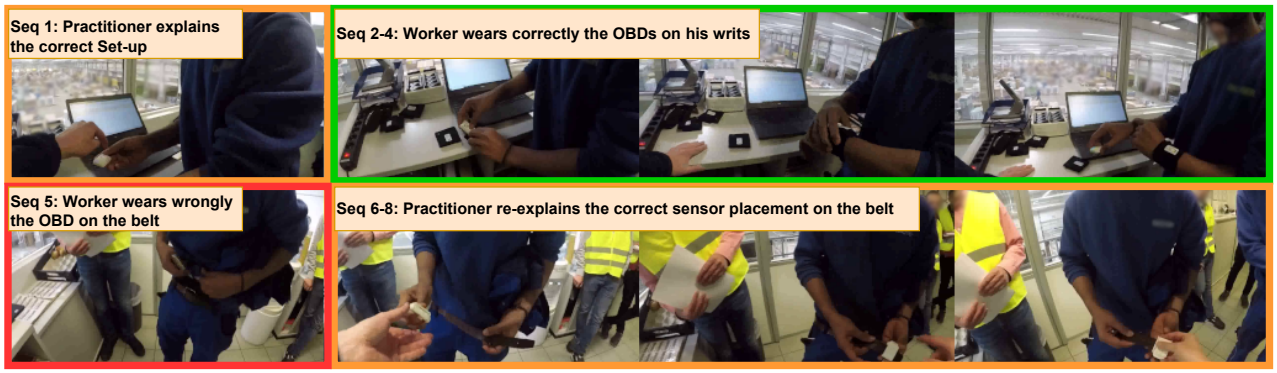


**Figure 1: Worker setting up an OBDs on his wrists and torso under the supervision of a practitioner.**

## ABSTRACT

Nowadays, in industrial applications, automatic human activity recognition plays a central role. Especially human-centered activity recognition methods using on-body devices (OBDs) address situations where the identity has to be protected. However, practitioners strongly assume that end-users use OBDs correctly at deployment. In reality, this is hardly the case. Thus, there is a need for a robust activity-recognition system, either at the recording stage or at recognition stage. This contribution addresses a combination of both stages. It proposes a miss-placement recognition of OBDs on the human body when performing an activity. We deploy a limb-oriented temporal convolutional neural network to either recognize a miss-placement occurring or the type of miss-placement. Primarily results on a proposed dataset suggest that miss-placement classification is possible, which can be used for end-user feedback during recording or leveraged in data post-processing.


*iWOAR 2023, September 21–22, 2023, Lübeck, Germany*

ACM ISBN 979-8-4007-0816-9/23/09…$15.00
https://doi.org/10.1145/3615834.3615838

## KEYWORDS

Multi-channel time-series, On-body devices, Deep Learning, Human Activity Recognition, Dataset

**ACM Reference Format:**
Robin Dönnebrink, Fernando Moya Rueda, Maximilian Stach, and René Grzeszick. 2023. Miss-placement Prediction of Multiple On-body Devices for Human Activity Recognition. In *8th international Workshop on Sensor-Based Activity Recognition and Artificial Intelligence (iWOAR 2023), September 21–22, 2023, Lübeck, Germany*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3615834.3615838

## 1 INTRODUCTION

Human activity recognition (HAR) seeks to predict human actions from multi-channel time-series recordings, such as videos or on-body devices (OBDs). OBDs are helpful for many situations where videos are restricted, for example, in industrial processes [4, 13]. Besides, they extend the usability of HAR out of laboratory settings or constrained environments; they protect a user's identity to a certain extent [12], as sequences are not easily interpretable as in the case of videos. A reliable analysis of a process conformed by particular tasks depends strongly on a robust HAR system. Unfortunately, a robust HAR system is hard to obtain as humans perform activities differently, or distinctive activities can be carried out similarly—there is a vast variability in human activities. Furthermore, there is a strong component of difficulty regarding the sensor settings, environmental noise, unclear recording protocols or guidelines, and end-user deployment, being the latter very prominent in real

scenarios. For example, in logistic scenarios, workers minimize the importance of keeping with recording settings, as they are focused on carrying out their daily job. Fig. 1 shows a sequence where a worker gets corrected after miss-placing an OBD on the torso wrongly rotated. These phenomena also happen in medical applications, such as in HAR at nursery [6].

There is a relation between the location of OBDs with the recognition of a specific activity [8–10]. For example, locating devices on the legs might be more suitable for recognising walking than placing them on the neck; or locating them on the wrists to recognise picking activities. Besides, the sensor localisation might provide context information for enhancing the recognition [10]. The sensor configuration affects HAR performance, as trained classifiers rely on a specific sensor set-up. A model expecting recordings from a sensor located on a left arm performs poorly when receiving a right arm—even a loosely attached OBD affects performance as discussed in [10]. Localisation of devices on the human body has been considered as a classification problem [8–10]. The authors of these papers empirically show that recognition drops when miss-placement occurs. Alternatively, collecting recording of various sensor set-ups might be a solution, but the time expenditure for such recordings and annotations increases with each set-up, facing limitations of resources. Attempts have been carried out to transfer learning under different locations. However, the transfer learning performance is limited [11, 14]. Addressing HAR from the practitioner perspective, questions remain: How to create HAR datasets and a HAR model? How robust is the HAR system with respect to the recording sensor set-up at deployment? Can miss-placements be recognised from the sensor recordings themselves? How does the change of the miss-placement set-up affect HAR? What happens when the users deploy the sensors incorrectly? What happens when users are not supervised at recording?

This work proposes a method for recognising miss-placement of OBDs during recording, considering this as a classification problem as in [8–10, 19]. It uses a human limb-oriented temporal deep architecture for such classification. Furthermore, it combines a binary miss-placement classification with a multi-class miss-placement classification. For that, this paper also introduces a novel dataset with two subsets with different miss-placement situations, one with the recording of the expected recording set-up. This contribution is the first regarding miss-placement of a HAR system already deployed in the industry. It offers a solution to a problem by predicting when any miss-placement occurs for end-user feedback or, ideally, to be leveraged in the post-processing of the data.

This paper is divided as follows: First, Sec. 2 presents the related work on multi-channel time-series HAR and OBD localisation as a classification problem. Second, the proposed miss-placement classification method is presented in Sec. 3. Then, Sec. 4 introduces the MotionMiners Miss-placement dataset along with the experimental set-up. Sec. 5 displays the classification results considering two tasks, whether a miss-placement occurs or the type of miss-placement. Finally, a conclusion is drawn in Sec. 6.

## 2 RELATED WORK

Multi-channel time-series HAR is strongly addressed using deep architectures. Different deep architectures have been proposed for solving HAR, namely, temporal CNNs, CNNs using spectrogram images, recurrent neural networks (RNNs), CNN in combination with RNNs, transformers, and conformers [3, 7, 22]. These architectures generally comprise a feature extractor, a feature aggregator, and a classifier, commonly an multilayer perceptron (MLP) with a softmax layer. All bring different capabilities; for example, CNNs are excellent feature extractors [5, 24], RNNs such as long short-term memory (LSTM) learn temporal relations, and transformers can process contextual information efficiently. Moreover, these architectures process multi-channel time-series data from OBD located on the human body.

Recent surveys reveal a variety of sensor technologies and configurations for multi-channel time-series HAR [1, 13, 15]. The authors in [15] carried out an extended revision of many publications for HAR in Production and Logistic scenarios. They found no clear picture emerges of the placement. Dataset authors have placed OBDs on the arms, hands, legs, feet, torso, head, and smartphone in the pocket. To notice the most prominent location is the torso. However, the authors did not find a direct relationship between the placement position and the activities. There is hardly any connection between the activity categories to recognize and the sensor set-up. They are particular for individual datasets and applications of multi-channel time-series HAR, and they hardly follow any community-agreed standard. The reviewers could not establish a link between the sensor placement and the recognised activities. While some researchers attach the sensors to the arm performing a motion, e.g., drilling or opening a window, other approaches try to recognise Activities of Daily Living and locomotion activities with sensors placed on the waist; similarly, when researchers aim to recognise locomotion with on-body devices placed on the hand. Placing sensors on the head or a helmet is proposed relatively rarely, e.g., in [21].

The authors in [4] propose using only three OBDs for Production and Logistic scenarios. They argue that these set-ups do not represent a considerable impairment of employees when performing manual work. Beyond that, researchers and practitioners may face the problem that subjects cannot wear the sensors as specified due to physical constraints.

Ongoing work optimizing the number and position of sensors is limited to virtual sensors [2, 11, 23]. The bottleneck of this research branch is the vast effort for recording and labeling data for each sensor configuration and position to derive best practices. Consequently, the multi-channel time-series HAR research community needs a validated method to determine the best sensor configuration and position based on a given application. Furthermore, deviating parameters of the sensor set-up between individual recording sessions and datasets further impede generalized transfer learning human activity recognition.

Previous works on the localisation of OBDs on the human body considered the problem as a classification task [8–10]. However, they carried out experiments assuming a single sensor recording. They concentrated on finding which sensor positioning is best for a particular activity class.

The authors in [8, 9] proposed a statistical pattern recognition method for HAR, predicting *Working on desk*, *Walking*, *Making coffee*, *Giving a presentation*, *Walking staircases* of three recordings of $\approx$ 15 min from six subjects. Besides, they considered a HAR method for only *Walking* activity along with localisation of OBDs

within a body part; that is, a sort of miss-placement of a single OBD as a dislocation within a body part when walking. They exclude orientation in their experiments.

In [10], a combination of HAR and sensor localisation of a single OBD on seven locations on the human body is proposed using CNNs and spectrograms. They aimed to find a relation between a device's specific location on the human body and the performed activity. This problem is solved by an evaluation sensor by sensor. The used method consists of two parts: an image generation using Short-time Fourier Transform (STFT), and a classification by means of a CNN. During the recording the ODBs are positioned in seven different location on the human body. They proposed a network for HAR following a two-stream simplified VGG architecture [16]. Each stream handles the accelerometer and gyroscope feature representation, predicting activity and the sensor location. They carried out HAR using all seven devices combined, setting up a sort of baseline for their work. The sensor recordings are classified into a set of activities comparing to the baseline, which uses the combination of all seven devices is set. Further, the best position of a sensor for a specific activity is determined. The authors explained the low performance in some results as the sensor tagged as *thigh* is, in fact, loosely placed in the front pocket of the participants' trousers, adding noise to the recordings and deviating from the expected signals by the classifier.

The sensor localisation is considered as a classification problem by the authors in [19]. They considered the localization classification using each device's recordings individually. The authors sought to recognise locomotion activities, e.g., *Walking*, *Standing*, *Climbing Stairs*, and *Lying*. *Chest*, *Forearm*, *Head*, *Shin*, *Thigh*, *Upper Arm*, and *Waist* are considered as locations on the human body. Different from [8–10], the signals have been recorded simultaneously. However, the setting is kept fixed during all experiments. A two-step method is used to first differentiate between static and dynamic activities and secondly to predict the device position per subject.

Methods for OBD localisation on the human body assume that the end-user follows a correct recording guideline for wearing the devices. Besides, most of the works have considered single-sensor position classification. However, a sensor set-up at deployment will involve all the expected sensors in use and classification of whether corrections at recording by the end-user can be done prior to a large recording or a post-processing step for reliable processing of the data is lacking. Additionally, as reported by [10], experimental set-ups are affected by miss-placements detected only after manually revising their results.

## 3 MISS-PLACEMENT RECOGNITION

The miss-placement of OBDs is defined as a classification problem where the different miss-placements configurations are considered as classes. These miss-placements might occur following the design of the recording devices. They consider the degree of freedom of the user when putting on the devices.

With this, a temporal CNN is deployed for carrying out the classification. A temporal CNN maps multichannel time-series recordings from sensors to a set of activities; in this case, a set of miss-placements or wrong placements of the sensors. Besides, correctly wearing the sensors is considered an additional class. Moreover, we propose a binary classification that considers all miss-placements types vs correct wearing of the OBDs. This is similar to the problem considered by [10, 19]. However, the classification is carried out for all the OBDs in the recording setup and not only for a single one.

The proposed architecture is shown in Figure 2. Similar to the IMU-CNN architecture proposed in [4, 11], the data from each IMU or OBD is processed separately. Four convolution operations process the normalised OBD data—a channel-wise zero mean and unit variance is deployed as normalisation step. Each layer applies 64 2D convolving kernels $[5 \times 1]$ to the input signal. The calculated feature map of each OBD is mapped to an OBD-wise representation with 256 features, by a fully connected layer. Afterward, all three representations are concatenated by another fully connected layer with 256 neurons. For the two classification tasks, two separate, fully connected layer follows. A pseudo probability $\tilde{y^s}$ of the sensor miss-placement is calculated with the help of a fully connected layer of size C activated by the softmax function. There C is the number of different miss-placements to consider. The other output layer containing two neurons represents $\tilde{y^m}$ and is activated by the sigmoid function to indicate if the sample was recorded by a miss-placed OBD.

The architecture is trained using a combination of the categorical- and the binary-cross-entropy loss given by:

$$\begin{aligned}\mathbf{E}(y, \tilde{y}) &= \rho_s * \mathbf{E}\left(y^s, \tilde{y^s}\right) + \rho_m * \mathbf{E}\left(y^m, \tilde{y^m}\right) \\ &= \rho_s \left(-\frac{1}{n}\sum_{i=1}^{n}\sum_{c=1}^{C} log\left(y_{i,c}^{\tilde{s}}\right) * y_{i,c}^{s}\right) \\ &+ \rho_m \left(-\frac{1}{n}\sum_{i=1}^{n}\left(y_i^m log\left(\tilde{y_i^m}\right) + \left(1 - y_i^m\right) log\left(1 - \tilde{y_i^m}\right)\right)\right),\end{aligned} \tag{1}$$

with $\rho_m$ the proportion for the binary and $\rho_s$ for the multi-class classification result should feed into the total loss. Thereby both proportions sum to one. Batch SGD is used to minimize the loss with a Nesterov momentum of 0.9, based on the formula of Sutskever et al. [18]. Since different learning rates are analysed, the learning rate is experiment dependent. It will be described in detail in Sec. 5.

After each epoch, the model performance on a holdout validation dataset is determined by averaging the performance for both tasks. To avoid overfitting, the previous trained model is restored once the score has decreased. Moreover, the regularization technique dropout [17] with the probability of 50 % is used in the fully connected layer, except the two classification layers.

## 4 MOTIONMINERS MISSPLACEMENT DATASET

The MotionMiners Miss-placement Dataset $MP$[1] is composed of recordings of seven subjects carrying out different activities in the intralogistics, using a sensor set-up of OBDs as in [4, 13] for industrial applications. Here, the position and orientation of the OBD change with respect to the recording-and-usage guidelines. The OBDs are labeled with respect to their expected location on the human body, namely, $OBD_R$, $OBD_L$ and $OBD_T$ on the right

[1]Access to Dataset and its preprocessing scripts via Zenodo with DOI: 10.5281/zenodo.8272091.

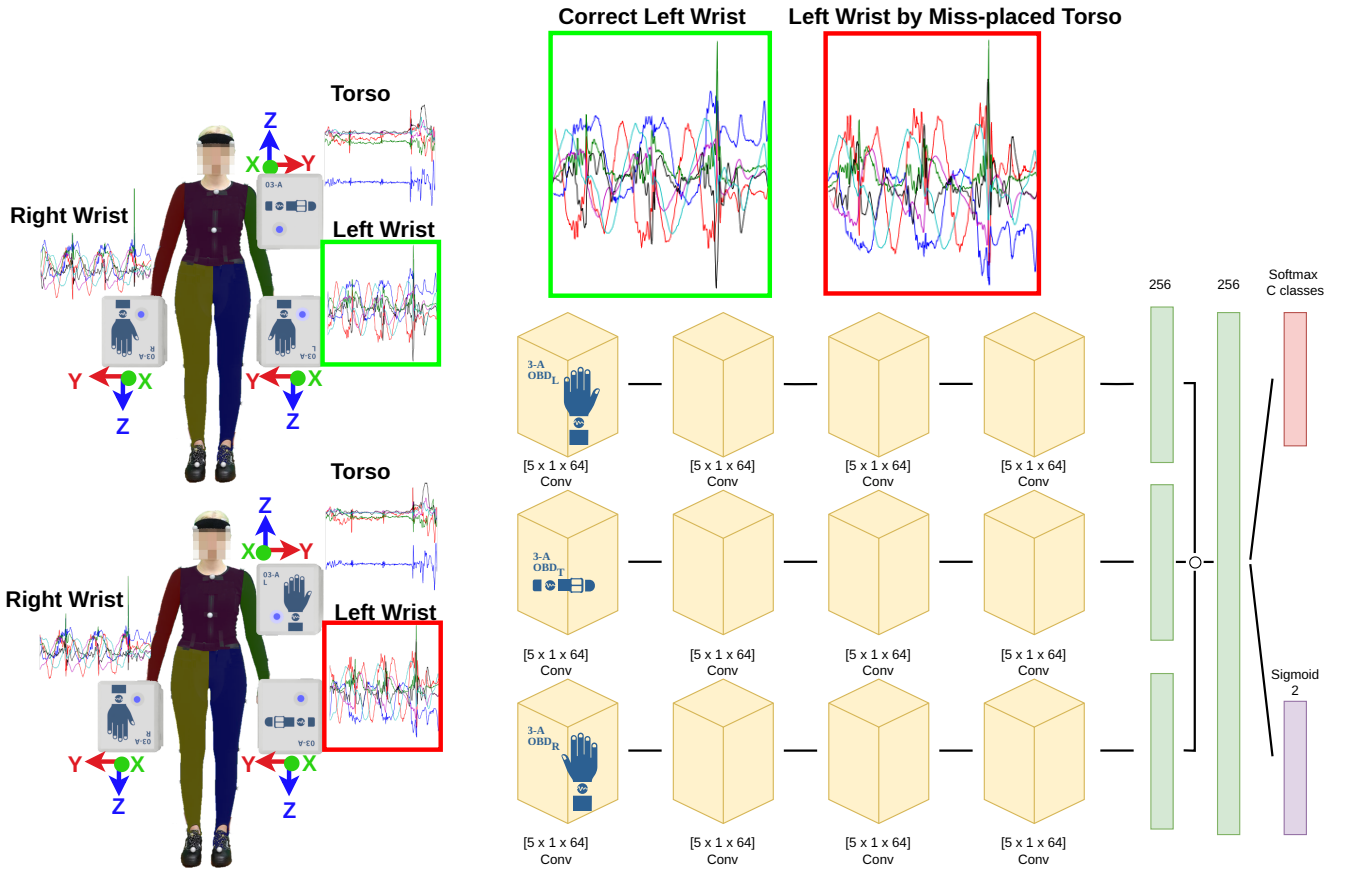


**Figure 2: Human limb-oriented temporal CNN for miss-placement classification. Subject on the lower part toggles the OBDs between the left wrist and the torso. The recordings on the OBD tagged for torso miss-placed on the left wrist shows differences with respect to the expected OBD tagged left on correct location (upper part of the image in green and red boxes), where signals are inter-exchanged and rotated, shown with the change of axes colours.**

arm, left arm, and frontal torso. Tab. 1 presents the different miss-placement classes of the dataset. This dataset considers the miss-placement as a classification problem, similar to [9] considering localisation; however, differently, the *MP* dataset considers rotations miss-placements—commonly appear on deployment from practitioners experience.

The *MP* dataset contains recordings of seven subjects performing six activities: *Standing*, *Walking*, *Handling Centred*, *Handling Upwards*, *Handling Downwards*, and an additional *Sync.*. Each subject carried out each activity under the case of up to 15 different miss-placement situations, including a correct set-up of the devices. The *MP* dataset is divided in two subsets, $MP_A$ and $MP_B$.

The $MP_A$ constitutes a first attempt to address the problem with regular results for the correct wearing of OBDs; Tab. 7-Column: $MP_A$ shows these results. The amount of recordings where the subject wears the OBDs correctly is proportional in the dataset; however, this is not the case in reality. The $MP_B$ is created considering two sets of OBDs synchronously recording, with one OBD always correctly worn. This recording process reduced recording time and provided miss-placement recordings and their ground truth. Besides, $MP_B$ collects a more variate type of miss-placements. In general, the $MP_A$ comprises recordings from three subjects, whereas $MP_B$ of four subjects. For $MP_A$, each subject carried out 40 recordings with a duration of 2 min per activity. The $MP_B$ is more extensive, containing recordings of four subjects and two sets of synchronous OBDs, where one set is worn correctly, while the second is worn following the classes in Table 1. Each subject carries out a sequence of 5 min, following activities of picking boxes at different heights and carrying out the activities with no script. In total, each subject performed 15 of such recordings.

This dataset considers for this preliminary study seven male subjects of $\approx 28 \pm 2.5$ years with height $\approx 177.9 \pm 8.1$cm. Besides, they are all right handed. However, this dataset is envisioned to be extended considering female workers, left handed, and a more variate height depending on availability of volunteers. New versions of the *MP* dataset will be uploaded via its official entry in Zenodo.

**Table 1: Miss-placement $MP$ classes and their proportion in the sets $MP_A$% and $MP_B$%. Rotations are considered with respect to the local frame of the OBD as shown in Fig. 2. These are the most common miss-placement situations reported in practice.**

| Class | Miss-Placement | $MP_A$% | $MP_B$% | Meaning |
|---|---|---|---|---|
| $K_0$ | *Correct* | 14.88% | 50.00% | Correct worn on-body devices |
| $K_1$ | *Toggling Arms* | 12.48% | 2.55% | Subject switches the $OBD_R$ and $OBD_L$ located on the wrists. |
| $K_2$ | *Rotate Torso 180 z-axis* | 11.54% | 2.81% | Subject wears $OBD_T$ rotating 180° with respect to its z-axis. |
| $K_3$ | *Rotate Torso 180 x-axis* | | 1.71% | Subject wears $OBD_T$ rotating 180° with respect to its x-axis. |
| $K_4$ | *Torso moving to left side* | 12.42% | 3.20% | Subject moves the $OBD_T$ to their left side of the hip. |
| $K_5$ | *Torso moving to right side* | 12.49% | 3.35% | Subject moves the $OBD_T$ to their right side of the hip. |
| $K_6$ | *Rotate Left 90* | 11.53% | 4.42% | Subject wears $OBD_L$ rotating 90° with respect to the x-axis. |
| $K_7$ | *Rotate Right 90* | 12.48% | 3.18% | Subject wears $OBD_R$ rotating 90° with respect to the x-axis. |
| $K_8$ | *Rotate Left 180* | | 3.73% | Subject wears $OBD_L$ rotating 180° with respect to the x-axis. |
| $K_9$ | *Rotate Right 180* | | 3.55% | Subject wears $OBD_R$ rotating 180° with respect to the x-axis. |
| $K_{10}$ | *Rotate Both arm 180* | 12.18% | 3.64% | Subject wears $OBD_R$ and $OBD_L$ rotating 180° with respect to the x-axis. |
| $K_{11}$ | *Toggling Torso to Left* | | 3.77% | Subject switches the $OBD_T$ and $OBD_L$ located on the wrists. |
| $K_{12}$ | *Toggling Torso to Right* | | 3.62% | Subject switches the $OBD_T$ and $OBD_R$ located on the wrists. |
| $K_{13}$ | *Clock-wise rotation of all OBDs* | | 3.51% | Subject moves the $OBD_T \rightarrow OBD_R$ and $OBD_R \rightarrow OBD_L$. |
| $K_{14}$ | *Counter clock-wise rotation of all OBDs* | | 3.49% | Subject moves the $OBD_T \rightarrow OBD_L$ and $OBD_L \rightarrow OBD_R$. |
| $K_{15}$ | *Loose belt* | | 3.48% | Subject looses the belt where the $OBD_T$ is placed. |

As preliminary experiments on the $MP$, we deployed an out-of-the-shelf temporal CNN for HAR [11] on the dataset measuring HAR performance. The activity classification performance decreases dramatically under miss-placement situations—being the *Walking* activity the only one robust enough. These results are, somehow, expected following the experiments of HAR and location in the literature, presented in [9, 10]. However, handling activities predictions deprecate drastically.

## 5 EXPERIMENTS AND RESULTS

We evaluate the miss-placement (MP) recognition on the two MotionMiners $MP$ datasets, measuring the classification performance following a cross-fold validation over the subjects in each set, where each subject becomes the testing, and the others are used for training and validation. Considering a practitioner perspective, two scenarios are evaluated: a binary miss-placement occurring and a miss-placement classification. These two scenarios are handy when feeding back to the end-user possible errors in a recording, at least in the sensor of the sensor set-up on the human body. The miss-placement classification considers the seven miss-placement classes occurring in the $MP_A$ and the 16 in the $MP_B$. In addition, we consider a combination of the binary and miss-placement classification controlled by the $\rho_m$ and $\rho_s$ variables, see Eq. 1.

Several hyperparameters influence whether the training of a temporal CNN leads to a satisfied prediction on the test set or not. Since no other contribution has yet tried to solve the miss-placement problem, a grid search is performed by analysing the set of hyperparameters on the small dataset $MP_A$. As shown in Tab. 2, four hyperparameters are analysed. The standard machine learning parameters amount of training epochs,$Ep$, and the used learning rate, $\gamma$, are evaluated. The other two hyperparameters are problem specific, namely the window width and the proportions $\rho_m$ and $\rho_s$. The window width $W$ is used in the sliding-window approach to create the data. The four widths correspond to $[150, 200, 400, 800]$ samples or $[1.5, 2, 4, 8]$s respectively. This parameter is evaluated since it is not known how much context data is needed to detect a miss-placed OBD. Moreover, the proportions $\rho_m$ and $\rho_s$ are evaluated as a tuple which sums to one. They impact the calculated loss and may solve the problem that both datasets are skewed regarding one task. On the one hand, $MP_A$ contains more samples of miss-placed OBDs in terms of the binary classification. On the other hand, $MP_B$ is strongly unbalanced towards the correct set-up ($K_0$) for multi-class classification. To consider the unbalanced proportion of classes, the performance is measured using the well-known weighted F1 ($F1_w$).

**Table 2: Analysed hyperparameters on $MP_A$**

| Hyperparameter | Values |
|---|---|
| Learning rate $\gamma$ | 0.005, 0.01, 0.1 |
| Epochs $Ep$ | 10, 15, 20, 30, 40 |
| Window width $W$ | 150, 200, 400, 800 |
| Proportions ($\rho_s$, $\rho_m$) | (0, 1), (0.25, 0.75), (0.5, 0.5), (0.75, 0.25), (1, 0) |

During all the experiments, the batch size is fixed to 100 samples. Furthermore, time-warping is used as augmentation. As proposed by Um et al. random sinusoidal curves are generated using a cubic spline to smoothly distort the time intervals between samples [20]. Similar to [4], a Gaussian noise, drawn from $\mathcal{N}(0, 10^{-3})$, has been added to simulate inaccuracies during the sampling process. In addition, values have been scaled with a random scalar sampled from the Gaussian distribution $\mathcal{N}(1, 10^{-3})$.

### 5.1 Results on $MP_A$

Tab. 3 and Tab. 4 show the $F1_w$[%] over the hyperparameter search. Best configurations are chosen for evaluating the combination of both tasks, combination of the binary and the categorical cross-entropy loss given by Eq. 1. The results of the grid search over all hyperparameters are summarised in Tab. 3 and Tab. 4 where for each possible window width the best $\gamma$ and $Ep$ is shown. Considering the binary task of a recording under miss-placement set-ups from Tab. 3, the temporal CNN shows a performance $\approx 77$ % for all of the window sizes. Even with more context data, if at least one OBD was miss-placed, the recognition did not improve significantly. The temporal CNN shows a classification performance of around $\approx 36$ % in Tab. 4. The classification performance of the miss-placement type is not as promising as for the binary task.

Tab. 5 shows the $F1_w$ when solving the binary miss-placement task after hyperparameter grid search and combining with a miss-classification loss under different proportions. All in all, the networks perform relatively well with an $F1_w$ score $\approx 82.66$ % for the

**Table 3: $MP_A$: $F1_w$[%] Learning rate $\gamma$ vs Width $W$ and Epochs $Ep$ vs Width $W$ for binary classification.**

| Width $W$ | Learning rate $\gamma$ | | | Epochs $Ep$ | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0.005 | 0.01 | 0.1 | 10 | 15 | 20 | 30 | 40 |
| 150 | **76.11** | 76.08 | 75.15 | **77.11** | 74.01 | 74.68 | 76.37 | 76.74 |
| 200 | 76.99 | 76.54 | **78.54** | 77.23 | 77.80 | 76.24 | 77.69 | **77.81** |
| 400 | **77.93** | 76.73 | 76.52 | 75.85 | **79.19** | 78.01 | 75.98 | 76.27 |
| 800 | 74.63 | **76.23** | 75.42 | 74.87 | **77.13** | 75.37 | 76.08 | 73.69 |

**Table 4: $MP_A$: $F1_w$[%] metric for Learning rate $\gamma$ vs Width $W$ and Epochs $Ep$ vs Width $W$ for multi-class classification.**

| Width $W$ | Learning rate $\gamma$ | | | Epochs $Ep$ | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0.005 | 0.01 | 0.1 | 10 | 15 | 20 | 30 | 40 |
| 150 | **36.34** | 35.77 | 34.05 | 34.76 | 32.32 | **37.42** | 35.08 | 37.37 |
| 200 | **35.56** | 34.86 | 34.76 | 34.71 | 33.02 | 35.42 | **36.44** | 35.70 |
| 400 | 34.11 | **35.96** | 33.54 | 33.93 | **35.64** | 33.05 | 34.65 | 35.41 |
| 800 | **36.54** | 35.92 | 34.28 | **37.16** | 36.70 | 35.37 | 34.03 | 34.64 |

binary classification. Regarding determining the miss-placement, Tab. 6 displays the $F1_w$ score for the proportions tuples and the best hyperparameters. If the network is trained without respect to the cross-entropy loss, $\rho_s = 0$, the predictions are worse regarding the miss-placement class. Once this loss is also considered during training, the performance significantly increases. Then for each window width, a configuration exists with a $F1_w$ greater than 49 %.

A combination of both tasks, binary and miss-placement classification, improves the performance significantly from ≈79.19 % up to ≈82.66 % for the first one (Tab. 3 vs. Tab. 5), and from ≈37.42 % up to ≈ 58.15 % for the latter one (Tab. 4 vs. Tab. 6). Generally, it has been shown that the classifier performance improves considerably when combining the binary with the multi-class classification.

**Table 5: $MP_A$: Grid search result in terms of $F1_w$[%] for binary MP task**

| Best configuration | $(\rho_s, \rho_m)$ | | | | |
|---|---|---|---|---|---|
| | (1, 0) | (0.75, 0.25) | (0.5, 0.5) | (0.25, 0.75) | (0, 1) |
| $W = 150$, $\gamma = 0.005$, $Ep = 10$ | 64.01 | 79.94 | 76.36 | 80.26 | **81.22** |
| $W = 200$, $\gamma = 0.1$, $Ep = 40$ | 78.55 | 77.82 | 80.15 | **82.66†** | 78.78 |
| $W = 400$, $\gamma = 0.005$, $Ep = 15$ | 71.87 | 78.64 | 79.20 | **81.50** | 80.87 |
| $W = 800$, $\gamma = 0.01$, $Ep = 15$ | 65.20 | 77.65 | 77.90 | **81.39** | 80.31 |

**Table 6: $MP_A$: Grid search result in terms of $F1_w$[%] for multi-class classification**

| Best configuration | $(\rho_s, \rho_m)$ | | | | |
|---|---|---|---|---|---|
| | (0, 1) | (0.25, 0.75) | (0.5, 0.5), | (0.75, 0.25) | (1, 0) |
| $W = 150$, $\gamma = 0.005$, $Ep = 20$ | 7.18 | **51.06** | 48.65 | 39.64 | 50.98 |
| $W = 200$, $\gamma = 0.005$, $Ep = 30$ | 13.87 | 48.02 | 49.20 | 34.44 | **49.49** |
| $W = 400$, $\gamma = 0.01$, $Ep = 15$ | 10.53 | **50.47** | 44.54 | 43.57 | 35.93 |
| $W = 800$, $\gamma = 0.005$, $Ep = 10$ | 9.56 | 51.49 | 48.31 | **58.15**$^*$ | 35.06 |

Tab. 7 and 8 show the confusion matrix for the best classifier regarding the binary or multi-class classification. The classifier performs relatively well for miss-placements considering rotations on the wrists ($K_6$, $K_7$, $K_{10}$) or toggling the devices located on the torso to the wrists ($K_4$, $K_5$). However, recognising when the devices are toggled between the wrists ($K_1$) or rotations on the torso is challenging ($K_2$). Interestingly, what is considered correct recorded shows poor performance when considering ($K_0$) all the classes simultaneously. This problem can be addressed considering the negative of the binary prediction. However, this also shows that the network cannot generalise what is expected as a correct recording.

**Table 7: Confusion matrix in terms of $F1_w$[%] for binary classification for best network from Tab. 5 (†) and 9 on the $MP_A$ and the $MP_B$**

| | $MP_A$ | | $MP_B$ | |
|---|---|---|---|---|
| | MP | Correct | MP | Correct |
| MP | 91.88 | 8.12 | 88.27 | 11.73 |
| Correct | 81.99 | 18.01 | 8.98 | 91.02 |

**Table 8: $MP_A$: Confusion matrix for the best classifier ($^*$)**

| | $K_0$ | $K_1$ | $K_2$ | $K_4$ | $K_5$ | $K_6$ | $K_7$ | $K_{10}$ |
|---|---|---|---|---|---|---|---|---|
| $K_0$ | 20.67 | 4.69 | 20.06 | 0.07 | 23.89 | 6.37 | 23.74 | 0.52 |
| $K_1$ | 8.05 | 48.60 | 5.18 | 3.50 | 1.40 | 23.19 | 7.60 | 2.48 |
| $K_2$ | 37.50 | 5.23 | 19.78 | 0.14 | 9.18 | 0.41 | 27.42 | 0.34 |
| $K_4$ | 1.92 | 0.65 | 4.78 | 64.72 | 2.98 | 0.63 | 6.21 | 18.11 |
| $K_5$ | 17.21 | 5.52 | 11.22 | 0.90 | 59.24 | 0.17 | 5.74 | 0.00 |
| $K_6$ | 1.44 | 20.57 | 5.26 | 1.36 | 3.11 | 67.90 | 0.30 | 0.07 |
| $K_7$ | 1.84 | 0.00 | 0.24 | 0.00 | 0.00 | 3.29 | 94.63 | 0.00 |
| $K_{10}$ | 0.00 | 0.23 | 0.00 | 0.05 | 0.00 | 4.71 | 0.00 | 95.01 |

## 5.2 Results on $MP_B$

Tab. 9 shows the binary classification performance on the Set B, following the best configuration on Tab. 5 for each window width. There is an approximated increase in binary classification performance up to ≈ 8% between $MP_A$ and $MP_B$. Tab. 7 highlights that the binary prediction performance of the correct recordings on the $MP_B$ is significantly higher than on the $MP_A$. This outcome is explained as the $MP_A$ is composed of a balanced amount of recordings for all the miss-placements, including correct ones; half of the $MP_B$ recordings belong to the correct set-up, as recordings were recorded synchronously. Tab. 9 presents the performance on $MP_B$, using the best configuration for the binary classification determined for each window width in Tab. 5. For all $W$ the binary classification task can be solved very accurately.

**Table 9: $MP_B$: Binary classifier**

| Configuration: $(W, \gamma, Ep, (\rho_s, \rho_m))$ | $F1_w$[%] |
|---|---|
| $(150,\ 0.005,\ 10,\ (0, 1))$ | 89.74 |
| $(200,\ 0.1,\ 40,\ (0.25, 0.75))$ | 86.56 |
| $(400,\ 0.005,\ 15,\ (0.25, 0.75))$ | 88.83 |
| $(800,\ 0.01,\ 15,\ (0.25, 0.75))$ | 89.19 |

For each window width the $F1_w$ score using the configurations from Tab. 6 is presented in Tab. 11. Except for the window width of 200, the results could be improved compared to $MP_A$. Since $MP_B$ is skewed towards $K_0$ and the hyperparameter $\rho_s = 1$ has been chosen for $W = 200$, the trained network learned to classify every input as $K_0$. All other classifiers achieved a score greater than 61%.

The confusion matrix of the best classifier(‡) is shown in Tab. 10. The classification of correct samples works very reliably. Moreover, the detection if $OBD_T$ is rotated with respect to the z- or x-axis ($K_2$, $K_3$) cannot be detect at all. Interestingly, the classifier can more

**Table 10: $MP_B$: Confusion matrix for best classifier (‡) in Tab. 11**

| | $K_0$ | $K_1$ | $K_2$ | $K_3$ | $K_4$ | $K_5$ | $K_6$ | $K_7$ | $K_8$ | $K_9$ | $K_{10}$ | $K_{11}$ | $K_{12}$ | $K_{13}$ | $K_{14}$ | $K_{15}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $K_0$ | 92.53 | 0.00 | 0.00 | 0.00 | 1.89 | 2.70 | 0.70 | 0.21 | 0.01 | 1.23 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.71 |
| $K_1$ | 5.69 | 21.53 | 0.00 | 0.00 | 0.27 | 0.00 | 3.20 | 0.58 | 2.79 | 0.37 | 57.90 | 0.00 | 0.00 | 0.00 | 2.69 | 4.97 |
| $K_2$ | 11.25 | 1.63 | 0.00 | 0.00 | 22.71 | 9.68 | 17.07 | 0.00 | 0.00 | 0.00 | 4.28 | 3.20 | 0.00 | 0.00 | 0.00 | 30.17 |
| $K_3$ | 50.05 | 0.00 | 0.00 | 0.00 | 1.02 | 0.00 | 11.84 | 0.20 | 0.00 | 0.05 | 19.31 | 0.00 | 0.00 | 10.82 | 6.66 | 0.05 |
| $K_4$ | 13.15 | 0.00 | 9.60 | 0.00 | 13.82 | 15.04 | 0.98 | 0.08 | 0.00 | 0.00 | 0.08 | 0.00 | 28.08 | 0.22 | 0.46 | 18.49 |
| $K_5$ | 23.67 | 0.00 | 5.52 | 0.00 | 8.53 | 12.03 | 14.41 | 1.19 | 0.00 | 0.00 | 0.00 | 0.18 | 0.00 | 25.56 | 0.13 | 8.79 |
| $K_6$ | 16.12 | 0.18 | 0.00 | 0.00 | 2.16 | 12.27 | 21.97 | 2.24 | 11.39 | 1.45 | 12.49 | 0.22 | 2.69 | 0.16 | 4.16 | 12.51 |
| $K_7$ | 5.59 | 0.52 | 0.00 | 0.00 | 4.15 | 2.70 | 0.66 | 54.50 | 9.01 | 0.85 | 1.50 | 6.00 | 0.00 | 0.63 | 0.00 | 13.89 |
| $K_8$ | 22.04 | 0.63 | 0.00 | 0.00 | 0.28 | 0.65 | 51.37 | 1.02 | 16.09 | 0.00 | 1.35 | 0.00 | 0.07 | 0.00 | 3.23 | 3.27 |
| $K_9$ | 24.90 | 0.17 | 0.00 | 0.00 | 0.66 | 0.12 | 1.73 | 13.62 | 0.00 | 49.23 | 5.47 | 3.49 | 0.00 | 0.00 | 0.61 | 0.00 |
| $K_{10}$ | 0.36 | 9.17 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.36 | 1.52 | 1.24 | 86.66 | 0.00 | 0.00 | 0.00 | 0.52 | 0.14 |
| $K_{11}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 2.00 | 0.00 | 0.00 | 0.00 | 52.72 | 0.00 | 45.27 | 0.00 | 0.00 |
| $K_{12}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 13.61 | 1.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 50.80 | 4.48 | 29.98 | 0.02 |
| $K_{13}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 27.68 | 0.00 | 72.32 | 0.00 | 0.00 |
| $K_{14}$ | 0.00 | 0.94 | 0.00 | 0.00 | 0.00 | 3.65 | 11.88 | 0.00 | 0.15 | 0.00 | 0.75 | 0.05 | 6.93 | 0.67 | 71.47 | 3.50 |
| $K_{15}$ | 15.02 | 0.00 | 0.00 | 0.00 | 25.94 | 0.85 | 16.96 | 1.50 | 0.00 | 0.07 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 | 38.81 |

easily distinguish if $OBD_R$ has been rotated with respect to the x-axis ($K_7$, $K_9$) than $OBD_L$ ($K_6$, $K_8$). Toggling of the OBDs ($K_{11}$-$K_{14}$) can be detected almost reliably, except for switching $OBD_R$ and $OBD_L$ ($K_1$).

**Table 11: $MP_B$: Multi-class classifier**

| Configuration: $(W, \gamma, Ep, (\rho_s, \rho_m))$ | $F1_w$[%] |
|---|---|
| $(150,\ 0.005,\ 20,\ (0.25, 0.75))$ | 61.69 |
| $(200,\ 0.005,\ 30,\ (1, 0)$ | 33.34 |
| $(400,\ 0.01,\ 15,\ (0.25, 0.75)$ ‡ | 61.70 |
| $(800,\ 0.01,\ 10,\ (0.75, 0.25)$ | 61.48 |

## 6 CONCLUSIONS

This manuscript addresses the problem of determining miss-placement of OBD recordings intended for HAR from a practitioner perspective. This work uses a temporal CNN conceived for processing OBD recordings from human limbs to classification. Furthermore, a dataset is presented, with recordings of subjects performing different activities in the intralogistics with a variety of different OBD set-ups differing from the expected at recording is introduced.

In general, predicting whether a miss-placement occurs in a recording is possible. Determining the specific miss-placement is possible for some miss-placement types but challenging for others. The classification performance for miss-placed OBDs on the wrist is relatively high. However, this outcome strongly differs for rotational miss-placements on the torso since the detection was very poor. Moreover, the recognition of the miss-placements with respect to toggling devices is contradictory. Toggling between the wrists could not be reliably detected on both given data sets. Recording of several OBD sets capturing synchronous data leverages the data creation effort. This recording process allows to capture of correct worn OBDs along with miss-placements data.

Despite the presented results, much research can be conducted, considering other types of ODBs set-ups, a larger number of OBDs—here, the sensors set-up was based on industrial settings. An extended *MP* dataset is planned. Moreover, the proposed dataset $MP_B$ can be used to train an autoencoder, a generative adversarial network (GAN), or a transformer intended to map the probably miss-placed input signals to the signals recorded by the correct worn OBDs.

## REFERENCES


[1] Sizhen Bian, Mengxi Liu, Bo Zhou, and Paul Lukowicz. 2022. The State-of-the-Art Sensing Techniques in Human Activity Recognition: A Survey. *Sensors* 22, 12 (Jan. 2022), 4596. https://doi.org/10.3390/s22124596 Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

[2] Anindya Das Antar, Masud Ahmed, and Md Atiqur Rahman Ahad. 2019. Challenges in Sensor-based Human Activity Recognition and a Comparative Analysis of Benchmark Datasets: A Review. In *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*. 134–139. https://doi.org/10.1109/ICIEV.2019.8858508

[3] Florenc Demrozi, Graziano Pravadelli, Azra Bihorac, and Parisa Rashidi. 2020. Human Activity Recognition Using Inertial, Physiological and Environmental Sensors: A Comprehensive Survey. *IEEE Access* 8 (2020), 210816–210836. https://doi.org/10.1109/ACCESS.2020.3037715

[4] Rene Grzeszick, Jan Marius Lenk, Fernando Moya Rueda, Gernot A. Fink, Sascha Feldhorst, and Michael ten Hompel. 2017. Deep Neural Network Based Human Activity Recognition for the Order Picking Process. In *Proceedings of the 4th International Workshop on Sensor-Based Activity Recognition and Interaction* (Rostock, Germany) *(iWOAR '17)*. Association for Computing Machinery, New York, NY, USA, Article 14, 6 pages. https://doi.org/10.1145/3134230.3134231

[5] Nils Hammerla, Shane Halloran, and Thomas Plötz. 2016. Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16, Vol. 8)*. AAAI Press, New York, New York, USA, 1533–1540.

[6] Sylvia Kaczmarek, Martin Fiedler, Andreas Bongers, Sebastian Wibbeling, and Rene Grzeszick. 2023. Dataset and Methods for Recognizing Care Activities. In *Proceedings of the 7th International Workshop on Sensor-Based Activity Recognition and Artificial Intelligence* (Rostock, Germany) *(iWOAR '22)*. Association for Computing Machinery, New York, NY, USA, Article 14, 8 pages. https://doi.org/10.1145/3558884.3558891

[7] Yeon-Wook Kim, Woo-Hyeong Cho, Kyu-Sung Kim, and Sangmin Lee. 2022. Inertial-Measurement-Unit-Based Novel Human Activity Recognition Algorithm Using Conformer. *Sensors* 22, 10 (2022). https://doi.org/10.3390/s22103932

[8] Kai Kunze and Paul Lukowicz. 2008. Dealing with Sensor Displacement in Motion-Based Onbody Activity Recognition Systems. In *Proceedings of the 10th International Conference on Ubiquitous Computing* (Seoul, Korea) *(UbiComp '08)*. Association for Computing Machinery, New York, NY, USA, 20–29. https://doi.org/10.1145/1409635.1409639

[9] Kai Kunze, Paul Lukowicz, Holger Junker, and Gerhard Tröster. 2005. Where am I: Recognizing On-body Positions of Wearable Sensors. In *Location- and Context-Awareness*, Thomas Strang and Claudia Linnhoff-Popien (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 264–275.

[10] Isah Lawal and Sophia Bano. 2020. Deep Human Activity Recognition With Localisation of Wearable Sensors. *IEEE Access* PP (08 2020), 1–1. https://doi.org/

10.1109/ACCESS.2020.3017681

[11] Fernando Moya Rueda and Gernot A. Fink. 2021. From Human Pose to On-Body Devices for Human-Activity Recognition. In *2020 25th International Conference on Pattern Recognition (ICPR)*. https://doi.org/10.1109/ICPR48806.2021.9412283 tex.ids= rueda_human_2021 ISSN: 1051-4651.

[12] Nilah Ravi Nair, Fernando Moya Rueda, Christopher Reining, and Gernot A. Fink. 2023. Multi-Channel Time-Series Person and Soft-Biometric Identification. arXiv:2304.01585 [cs.CV]

[13] Friedrich Niemann, Christopher Reining, Fernando Moya Rueda, Nilah Ravi Nair, Janine Anika Steffens, Gernot A. Fink, and Michael ten Hompel. 2020. LARa: Creating a Dataset for Human Activity Recognition in Logistics Using Semantic Attributes. *Sensors* (2020). https://doi.org/10.3390/s20154083

[14] Francisco Javier Ordóñez Morales and Daniel Roggen. 2016. Deep Convolutional Feature Transfer across Mobile Activity Recognition Domains, Sensor Modalities and Locations. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers*. Association for Computing Machinery, Heidelberg, Germany, 92–99. https://doi.org/10.1145/2971763.2971764

[15] Christopher Reining, Friedrich Niemann, Fernando Moya Rueda, Gernot A. Fink, and Michael ten Hompel. 2019. Human Activity Recognition for Production and Logistics—A Systematic Literature Review. *Information* 10, 8 (2019). https://doi.org/10.3390/info10080245

[16] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs.CV]

[17] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* 15, 56 (2014), 1929–1958. http://jmlr.org/papers/v15/srivastava14a.html

[18] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. 2013. On the Importance of Initialization and Momentum in Deep Learning. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28* (Atlanta, GA, USA) *(ICML'13)*. JMLR.org, III–1139–III–1147.

[19] Timo Sztyler, Heiner Stuckenschmidt, and Wolfgang Petrich. 2017. Position-aware activity recognition with wearable devices. *Pervasive and Mobile Computing* 38 (2017), 281–295. https://doi.org/10.1016/j.pmcj.2017.01.008 Special Issue IEEE International Conference on Pervasive Computing and Communications (PerCom) 2016.

[20] Terry T. Um, Franz M. J. Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić. 2017. Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring Using Convolutional Neural Networks. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (Glasgow, UK) *(ICMI '17)*. Association for Computing Machinery, New York, NY, USA, 216–220. https://doi.org/10.1145/3136755.3136817

[21] Johann P. Wolff, Florian Grützmacher, Arne Wellnitz, and Christian Haubelt. [n. d.]. Activity Recognition using Head Worn Inertial Sensors. In *Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction - iWOAR '18* (Berlin, Germany, 2018). ACM Press, 1–7. https://doi.org/10.1145/3266157.3266218

[22] Rui Xi, Mengshu Hou, Mingsheng Fu, Hong Qu, and Daibo Liu. 2018. Deep Dilated Convolution on Multimodality Time Series for Human Activity Recognition. In *International Joint Conference on Neural Networks (IJCNN)*. IEEE, Rio de Janeiro, Brazil, 1–8. https://doi.org/10.1109/IJCNN.2018.8489540

[23] Chengshuo Xia and Yuta Sugiura. 2021. Optimizing Sensor Position with Virtual Sensors in Human Activity Recognition System Design. *Sensors* 21, 20 (Jan. 2021), 6893. https://doi.org/10.3390/s21206893 Number: 20 Publisher: Multidisciplinary Digital Publishing Institute.

[24] Rui Yao, Guosheng Lin, Qinfeng Shi, and Damith C. Ranasinghe. 2018. Efficient Dense Labelling of Human Activity Sequences from Wearables using Fully Convolutional Networks. *Pattern Recognition* 78 (2018), 252–266. https://doi.org/10.1016/j.patcog.2017.12.024