



# Multimodal Indoor Localisation in Parkinson’s Disease for Detecting Medication Use: Observational Pilot Study in a Free-Living Setting

Ferdian Jovan  
Faculty of Engineering, University of  
Bristol  
Bristol, UK  
ferdian.jovan@bristol.ac.uk

Catherine Morgan  
Translational Health Sciences, Bristol  
Medical School  
Bristol, UK  
catherine.morgan@bristol.ac.uk

Ryan McConville  
Faculty of Engineering, University of  
Bristol  
Bristol, UK  
ryan.mcconville@bristol.ac.uk

Emma L. Tonkin  
Faculty of Engineering, University of  
Bristol  
Bristol, UK  
e.l.tonkin@bristol.ac.uk

Ian Craddock  
Faculty of Engineering, University of  
Bristol  
Bristol, UK  
ian.craddock@bristol.ac.uk

Alan Whone  
North Bristol NHS Trust, Southmead  
Hospital  
Bristol, UK  
alan.whone@bristol.ac.uk

## ABSTRACT

Parkinson’s disease (PD) is a slowly progressive, debilitating neurodegenerative disease which causes motor symptoms including gait dysfunction. Motor fluctuations are alterations between periods with a positive response to levodopa therapy (“on”) and periods marked by re-emergence of PD symptoms (“off”) as the response to medication wears off. These fluctuations often affect gait speed and they increase in their disabling impact as PD progresses. To improve the effectiveness of current indoor localisation methods, a transformer-based approach utilising dual modalities which provide complementary views of movement, Received Signal Strength Indicator (RSSI) and accelerometer data from wearable devices, is proposed. A sub-objective aims to evaluate whether indoor localisation, including its in-home gait speed features (i.e. the time taken to walk between rooms), could be used to evaluate motor fluctuations by detecting whether the person with PD is taking levodopa medications or withholding them. To properly evaluate our proposed method, we use a free-living dataset where the movements and mobility are greatly varied and unstructured as expected in real-world conditions. 24 participants lived in pairs (consisting of one person with PD, one control) for five days in a smart home with various sensors. Our evaluation on the resulting dataset demonstrates that our proposed network outperforms other methods for indoor localisation. The sub-objective evaluation shows that precise room-level localisation predictions, transformed into in-home gait speed features, produce accurate predictions on whether the PD participant is taking or withholding their medications.

## CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; • **Information systems** → *Sensor networks*; • **Applied computing** → **Health care information systems**.

## KEYWORDS

indoor localisation, parkinson disease, multimodal learning, time-series data, transformer networks

## ACM Reference Format:

Ferdian Jovan, Catherine Morgan, Ryan McConville, Emma L. Tonkin, Ian Craddock, and Alan Whone. 2023. Multimodal Indoor Localisation in Parkinson’s Disease for Detecting Medication Use: Observational Pilot Study in a Free-Living Setting. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD ’23)*, August 6–10, 2023, Long Beach, CA, USA. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3580305.3599872>

## 1 INTRODUCTION

Parkinson’s disease (PD) is a debilitating neurodegenerative disease affecting around 6 million people worldwide. It is characterised by a variety of movement-related (motor) symptoms, such as slowness of movement, rigidity and gait dysfunction [23]. A complication of the mainstay medication used to treat PD, levodopa, is that patients start to experience motor symptom fluctuations related to medication timings. When levodopa is first started, patients experience a smooth and prolonged therapeutic response. As disease progresses (and in a substantial proportion of patients within the first five years), patients start to “wear off” from their medications before the next dose, causing a reemergence of parkinsonian symptoms including slowness of gait. These symptom fluctuations impair patients’ quality of life and often necessitate changes in medication regime. Motor symptoms can become severe enough to hinder the subject’s gait and movement around their own house [49]. As a result, the subject may be more likely to stay in one room; once they move, they may typically need more time to transition between rooms. Such outcomes could be used to detect ON and OFF medication motor fluctuations in PD and to inform clinicians and patients of such symptoms.



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

KDD ’23, August 6–10, 2023, Long Beach, CA, USA  
© 2023 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-0103-0/23/08.  
<https://doi.org/10.1145/3580305.3599872>

Furthermore, a sensitive and accurate ecologically-validated biomarker of PD progression is currently lacking [15], resulting in multiple failures of clinical trials testing putative neuroprotective therapies in PD [11, 21, 27]. Gait parameters are sensitive to disease progression symptom change in unmedicated early-stage PD [52] and show promise as markers of disease progression [44], making measuring gait parameters potentially of use in clinical trials of disease-modifying interventions which typically recruit recently diagnosed patients [42]. Clinical evaluation of PD is normally undertaken in an artificial clinic or laboratory environment where only a snapshot view of the individual’s motor function can be captured. Constant monitoring could capture symptom progression, including motor fluctuations, and sensitively quantify them over time [36].

Although PD symptoms including gait and balance parameters can be measured continuously at home (with varying degrees of reliability and accuracy) by wearable devices containing inertial motor units (IMUs) or smartphones [12, 14, 37, 38], this data does not show the context in which the measurements are taken (for example where someone is at the time of the symptom). Knowing which room someone is in (indoor localisation) could add valuable holistic information to the interpretation of symptoms of PD. For example, symptoms like freezing of gait [2] and turning in gait [18] vary according to the nature of the setting the person is in, so knowing where someone is could help predict such symptoms or interpret their severity. Furthermore, knowing how much time someone spends alone or with others in a room is a step towards understanding their societal participation [25], which affects quality of life in PD [13]. Localisation could also add valuable information in the measurement of other behaviors such as non-motor symptoms such as urinary function [16, 20] (e.g. how many times someone visits the toilet room overnight).

To perform indoor localisation in home environments, IoT-based platforms with sensors capturing various modalities of data combined with machine learning can be used to provide an unobtrusive and continuous localisation [39]. Typically, many of these techniques take advantage of the radio-frequency signals, the Received Signal Strength Indication (RSSI), emitted by wearables and measured at access points (AP) throughout a home. These signals to estimate the user’s position from the perceived signal strength, thereby creating radio-map features for each room [22]. To provide more accurate localisation, accelerometer data measured by wearable devices, equipped with RSSI measured at receivers, can also be used as it provides a means to distinguish different activities (e.g., walking vs standing). Furthermore, as some activities are tied to particular rooms (e.g. stirring a pan on the hob is very likely to be in a kitchen), accelerometer data may enrich RSSI in differentiating adjacent rooms, which RSSI alone may struggle with [35].

If accelerometer data are to provide extra features for separating adjacent rooms, greater consideration must be given to data generalisation across different PD patients. As PD is a heterogeneous disease, the symptoms experienced and their severity may vary from one patient to another [19]. These severe symptoms, such as tremor, may affect the generalisation of accelerometer data which are prone to bias and accumulated errors [34], especially those worn on the patient’s wrists, which is a common and well accepted placement location [17]. Naively combining the accelerometer data

with the RSSI may impair the performance of indoor localisation due to differing levels of tremor manifesting in the acceleration signal. In this work, we make two main contributions.

- (1) We describe the utilisation of RSSI enriched by the accelerometer data to perform room-level localisation. Our proposed network<sup>1</sup> intelligently chooses accelerometer features which may improve the RSSI performance in performing indoor localisation. To properly evaluate our proposed method, we use a free-living (a person living their life freely, without external intervention) dataset created by our group, where the movements and mobility are greatly varied and unstructured as expected in real-world conditions. Our evaluation on such a unique dataset, which includes subjects with and without PD, demonstrates that our proposed network outperforms other approaches in all cross-validation categories.
- (2) We also demonstrate how the accurate room-level localisation predictions can be transformed into in-home gait speed biomarkers (e.g. number of room-to-room transition, room-to-room transition duration) which can be used to effectively classify the OFF or ON medication state of a PD patient from this pilot study data.

## 2 RELATED WORK

There has been substantial work using home-based passive sensing systems to assess how the activities and behaviour of people with neurological disease (mainly cognitive dysfunction) change over time [33, 55]. There is very limited work assessing room use in the home setting in people with Parkinson’s.

However, gait quantification using wearables or smartphones is an area where a significant amount of work has been done (with several systematic reviews such as these[4, 7]). Cameras can detect also Parkinsonian gait and some gait features including step length and average walking speed [46]. Time of flight devices (which measure distances between the subject and the camera [24]) have been used to assess medication adherence through gait analysis [45]. From free-living data, one approach to gait and room use evaluation in home settings is by emitting and detecting radio waves to non-invasively track movement. Gait analysis using radio wave technology shows promise to track disease progression, severity and medication response [30]. However, this approach cannot identify who is doing the movement and also suffers from technical issues when the radio waves are occluded by another object. Much of the work done so far using video to track PD symptoms has focused on the performance of structured clinical rating scales during telemedicine consultations as opposed to naturalistic behaviour [41], and there have been some privacy concerns around the use of video data at home [48].

RSSI data produced from wearable devices is a type of data with fewer privacy concerns; it can be measured continuously and unobtrusively over long periods of time to capture real-world function and behavior in a privacy-friendly way. In indoor localisation, fingerprinting using RSSI is the typical technique used to estimate the wearable (user) location by using signal strength data representing a coarse and noisy estimate of the distance access point

<sup>1</sup>Code available at <https://github.com/ferdianjovan/Multihead-Dual-Convolutional-Self-Attention>

from the wearable [5, 40]. RSSI signals are not stable, they fluctuate randomly due to shadowing, fading and multi-path effects. However, many techniques have been proposed in recent years to tackle these fluctuations, and, indirectly, improve the localisation accuracy. Some of the works [54] utilise deep neural networks (DNN) to generate coarse positioning estimates from RSSI signals, which are then refined by a hidden Markov model (HMM) to produce a final estimate location. Other works, [22], try to utilise a time-series of RSSI data and exploit the temporal connections within each access point to estimate room-level position. A CNN is used to build localisation models to further leverage the temporal dependencies across time-series readings.

It has been suggested that we cannot rely on RSSI alone for indoor localisation in home environments for PD subjects due to shadowing rooms with tight separation [32, 35, 39]. Sansano et al. combine RSSI signals and inertial measurement unit (IMU) data to test the viability of leveraging other sensors in aiding the positioning system to produce a more accurate location estimate [39]. Classic machine learning approaches such as Random Forest (RF), Artificial Neural Network (ANN), k-Nearest Neighbour (k-NN) are tested, and the result shows that the RF outperforms other methods in tracking a person in indoor environments. Poulouse et al. combine smartphone IMU sensor data and Wi-Fi received signal strength indication (RSSI) measurements to estimate the exact location (in Euclidean position X, Y) of a person in indoor environments [35]. The proposed sensor fusion framework uses location fingerprinting in combination with a pedestrian dead reckoning (PDR) algorithm to reduce the positioning errors.

Looking at this multi-modality classification / regression problem from a timeseries perspective, there has been a lot of explorations in tackling a problem where each modality can be categorised as multivariate timeseries data [8, 28, 51]. LSTM and attention layers are often used in parallel to directly transform raw multivariate time series data into low-dimensional feature representation for each modality. Later, various processed is done to further extract correlations across modalities through the use of various layers (e.g. concatenation, CNN layer, transformer, self-attention) [28, 51]. Our work is inspired by Sansano-Sansano et al. [39] where we only utilise accelerometer data to enrich the RSSI, instead of utilising all IMU sensors, in order to reduce battery consumption. In addition, unlike Sansano-Sansano et al. who stop at predicting room locations, we go a step further and use room-to-room transition behaviours, as features for a binary classifier predicting whether people with PD are taking their medications or withholding them.

### 3 COHORT AND DATASET

**Dataset.** This dataset was collected using wristband wearable sensors, one on each wrist of all participants, containing tri-axial accelerometers<sup>2</sup> and 10 Access Points (APs) placed through the residential home (see Fig. 1 for house layout and AP location), each measuring the RSSI [26]. The wearable devices wirelessly transmit data using the Bluetooth Low Energy (BLE) standard which can be received by the 10 APs. Each AP records the transmitted packets

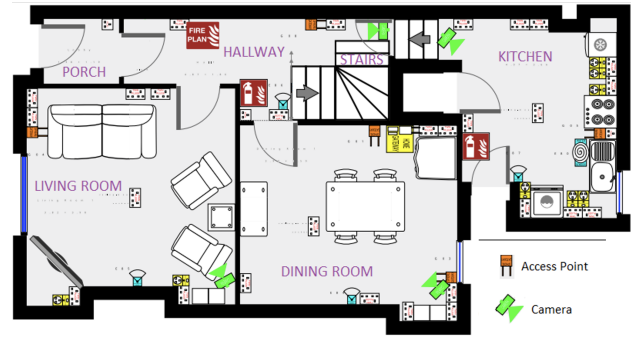


Figure 1: Layout of the residential home setting.

from the wearable sensor which contains the accelerometer readings sampled at 30Hz, with each AP recording RSSI values sampled at 5 Hz.

The dataset contains 12 spousal/parent-child/friend-friend pairs (24 participants in total) living freely in a smart home for five days. Each pair consists of one person with PD and one person as the healthy control volunteer (HC). This pairing was chosen to enable PD vs HC comparison, for safety reasons and also to increase the naturalistic social behaviour (particularly amongst the spousal pairs who already lived together). From the 24 participants, five females and seven males have PD. The average age of the participants is 60.25 (PD 61.25, Control 59.25) and the average time since PD diagnosis for the person with PD is 11.3 years (range 0.5-19).

To measure the accuracy of the machine learning models, wall-mounted cameras are installed in the ground floor of the house which capture red-green-blue (RGB) and depth data 2-3 hours daily (during daylight hours at times when participants were at home). The videos were then manually annotated to the nearest millisecond to provide localisation labels. Multiple human labellers used a widely available software called ELAN [1] to watch up to 4 simultaneously-captured video files at a time. The resulting labelled data recorded the kitchen, hallway, dining room, living room, stairs, and porch. The duration of labelled data recorded by the cameras for PD and HC is 72.84 and 75.31 hours, respectively, which provides a relatively balanced label set for our room-level classification<sup>3</sup>. Finally, to evaluate the ON/OFF medication state, participants with PD were asked to withhold their dopaminergic medications so that they were in the practically-defined OFF medications state for a temporary period of several hours during the study. Withholding medications removes their mitigation on symptoms, leading to mobility deterioration which can include slowing of gait.

**Data pre-processing for indoor localisation.** The data from the two wearable sensors worn by each participant were combined at each time point, based on their modality, i.e. twenty RSSI values (corresponding to 10 APs for each of the two wearable sensors), and accelerometry traces in six spatial directions (corresponding to the three spatial directions (x, y, z) for each wearable) were recorded at each time point. The accelerometer data is resampled to 5Hz to synchronise the data with RSSI values. With a 5-second

<sup>2</sup>The wearables are custom-designed and purposefully do not utilise gyroscope and magnetometer sensors to increase battery life.

<sup>3</sup>Approaches generalise to all rooms over the entire period, however we limit the time frame and rooms to only those that were manually annotated (on ground floor).

time window and 5Hz sampling rate, each RSSI data sample has an input of size  $(25 \times 20)$  and accelerometer data has an input of size  $(25 \times 6)$ . Imputation for missing values, specifically for RSSI data, is applied by replacing the missing values with a value that is not possible normally (i.e., -120dB). Missing values exist in RSSI data whenever the wearable is out of range of an AP. Finally, all time-series measurements by the modalities are normalised.

**Data pre-processing for medication state.** Our main focus is for our neural network to continuously produce room predictions which are then transformed into in-home gait speed features, particularly for persons with PD. We hypothesise that during their OFF medication state, the deterioration in mobility of a person with PD is exhibited by how they transition between rooms. These features include 'Room-to-room Transition Duration', and the 'Number of Transitions' between two rooms. 'Number of Transitions' represents how active PD subjects are within a certain period of time, while 'Room-to-room Transition Duration' may provide insight into how severe their disease is by the speed with which they navigate their home environment. With the layout of the house where participants stayed (see Fig. 1), the hallway is used as a hub connecting all other rooms labelled, and 'Room-to-room Transition' shows the transition duration (in seconds) between two rooms connected by the hallway. The transition between (1) kitchen and living room, (2) kitchen and dining room, and (3) dining room and living room are chosen as the features due to their commonality across all participants. For these features, we limit the transition time duration (i.e. the time spent in the hallway) to 60 seconds to exclude transitions likely to be prolonged and thus may not be representative of the person's mobility.

These in-home gait speed features are produced by an indoor-localisation model by feeding RSSI signals and accelerometer data from 12 PD participants from 6 a.m. to 10 p.m. daily which are aggregated into 4 hour windows. From this, each PD participant will have 20 data samples (four data samples for each of the five days), each of which contains six features (three for the mean of room-to-room transition duration, and three for the number of room-to-room transitions). There is only one 4-hour window during which the person with PD is OFF medications. These samples are then used to train a binary classifier<sup>4</sup> determining whether a person with PD is ON or OFF their medications.

For a baseline comparison to the in-home gait speed features, demographic features which include age, gender, years of PD, and MDS-UPDRS III score (the gold-standard clinical rating scale score used in clinical trials to measure motor disease severity in PD) are chosen. Two MDS-UPDRS III scores are assigned for each PD participant; one is assigned when a person with PD is ON medications, and the other one is assigned when a person with PD is OFF medications. For each in-home gait speed feature data sample, there will be a corresponding demographic feature data sample which are used to train a different binary classifier to predict whether a person with PD is ON or OFF medications.

**Ethical approval.** Full approval from NHS Wales Research Ethics Committee 6 was granted on 17<sup>th</sup> December 2019, and Health Research Authority and Health and Care Research Wales approval

<sup>4</sup>The Random Forest is chosen as a binary classifier; It is not the one used for indoor localisation.

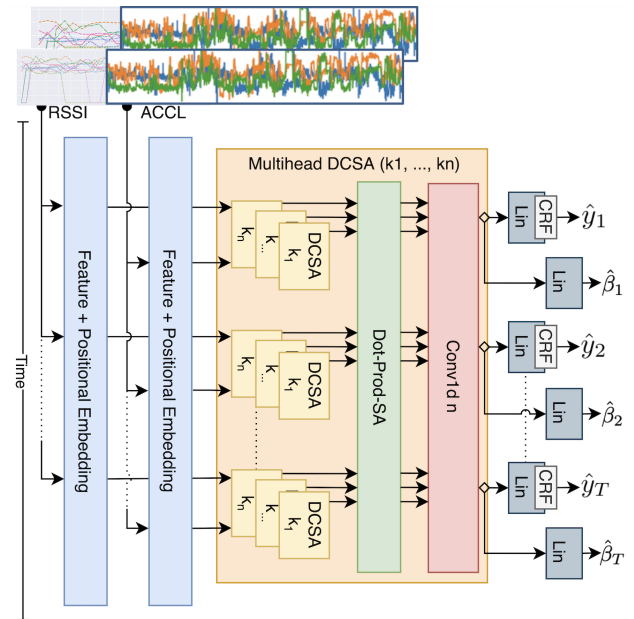


Figure 2: MDCSA architecture.

confirmed on 14<sup>th</sup> January 2020; the research was conducted in accord with the Helsinki Declaration of 1975; written informed consent was gained from all study participants. In order to protect participant privacy supporting data is not shared openly. It will be made available to bona fide researchers subject to a data access agreement. If you wish to apply to access this data, please email [data-bris@bristol.ac.uk](mailto:data-bris@bristol.ac.uk).

## 4 METHODOLOGIES AND FRAMEWORK

We introduce Multihead Dual Convolutional Self Attention (MDCSA), a deep neural network that utilises dual modalities for indoor localisation in home environments. The network tackles two challenges that arise from multimodality and time-series data:

- (1) **Capturing multivariate features and filtering multimodal noises.** RSSI signals, which are measured at multiple access points within a home received from wearable communication, have been widely used for indoor localisation [22], typically using a fingerprinting technique that produces a ground truth radio map of a home. Naturally, the wearable also produces acceleration measurements which can be used to identify typical activities performed in a specific room, and thus we can explore if accelerometer data will enrich the RSSI signals, in particular to help distinguish adjacent rooms, which RSSI-only systems typically struggle with. If it will, how can we incorporate these extra features (and modalities) into the existing features for accurate room predictions, particularly in the context of PD where the acceleration signal may be significantly impacted by the disease itself?
- (2) **Modelling local and global temporal dynamics.** The true correlations between inputs both intra-modality (i.e. RSSI signal among access points) and inter-modality (i.e. RSSI signal against accelerometer fluctuation) are dynamic.

These dynamics can affect one another within a local context (e.g. cyclical patterns) or across long-term relationships. Can we capture local and global relationships across different modalities?

The MDSCSA architecture, shown in Figure 2, addresses the aforementioned challenges through a series of neural network layers which are described in the following sections.

#### 4.1 Modality Positional Embedding

Due to different data dimensionality between RSSI and accelerometer, coupled with the missing temporal information, a linear layer with a positional encoding is added to transform both RSSI and accelerometer data into their respective embeddings. Suppose we have a collection of RSSI signals  $\mathbf{x}^r = [\mathbf{x}_1^r, \dots, \mathbf{x}_T^r] \in \mathbb{R}^{T \times r}$ , and accelerometer data  $\mathbf{x}^a = [\mathbf{x}_1^a, \dots, \mathbf{x}_T^a] \in \mathbb{R}^{T \times a}$  within  $T$  time unit, where  $\mathbf{x}_t^r = [x_t^1, \dots, x_t^r]$  represents RSSI signals from  $r$  access points, and  $\mathbf{x}_t^a = [x_t^1, \dots, x_t^a]$  represents accelerometer data from  $a$  spatial directions at time  $t$  with  $t \leq T$ . Given feature vectors  $\mathbf{x}_t^u = [x_t^1, \dots, x_t^u]$  with  $u \in \{r, a\}$  representing RSSI or accelerometer data at time  $t$ , and  $t \leq T$  representing time index, a positional embedding  $\mathbf{h}_t^u$  for RSSI or accelerometer can be obtained by:

$$\mathbf{h}_t^u = (\mathbf{W}_u \mathbf{x}_t^u + \mathbf{b}_u) + \vec{p}_t \quad (1)$$

where  $\mathbf{W}_u \in \mathbb{R}^{u \times d}$ , and  $\mathbf{b}_u \in \mathbb{R}^d$  are weight and bias to learn,  $d$  is the embedding dimension, and  $\vec{p}_t \in \mathbb{R}^d$  is the corresponding position encoding at time  $t$ .

#### 4.2 Locality Enhancement with Self-Attention

As it is time series data, the importance of an RSSI or accelerometer value at each point in time can be identified in relation to its surrounding values - such as cyclical patterns, trends, or fluctuations. Utilising historical context that can capture local patterns on top of point-wise values, performance improvements in attention-based architectures can be achieved. One straightforward option is to utilise a recurrent neural network such as a long-short term memory (LSTM) approach. However, in LSTM layers, the local context is summarised based on the previous context and the current input. Two similar patterns separated by a long period of time might have different context if they are processed by the LSTM layers [3]. We utilise a combination of *causal convolution* layers and self-attention layers which we name Dual Convolutional Self-Attention (DCSA). The DCSA takes in a primary input  $\hat{\mathbf{x}}_1 \in \mathbb{R}^{T \times d}$  and a secondary input  $\hat{\mathbf{x}}_2 \in \mathbb{R}^{T \times d}$  and yields:

$$DCSA_k(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2) = GRN(Norm(\Phi(\hat{\mathbf{x}}_1) + \hat{\mathbf{x}}_1), Norm(\Phi(\hat{\mathbf{x}}_2) + \hat{\mathbf{x}}_2)) \quad (2)$$

with

$$\Phi(\hat{\mathbf{x}}) = SA(\Phi_k(\hat{\mathbf{x}}) \mathbf{W}_Q, \Phi_k(\hat{\mathbf{x}}) \mathbf{W}_K, \hat{\mathbf{x}} \mathbf{W}_V) \quad (3)$$

where  $GRN(\cdot)$  is Gated Residual Network (GRN), introduced in [28], to integrate dual inputs into one integrated embedding,  $Norm(\cdot)$  is a standard layer normalisation,  $SA(\cdot)$  is a scaled dot-product self-attention introduced in [47],  $\Phi_k(\cdot)$  is a 1D-convolutional layer with a kernel size  $\{1, k\}$  and a stride 1,  $\mathbf{W}_K \in \mathbb{R}^{d \times d}$ ,  $\mathbf{W}_Q \in \mathbb{R}^{d \times d}$

,  $\mathbf{W}_V \in \mathbb{R}^{d \times d}$  are weights for keys, queries and values of the self-attention layer, and  $d$  is the embedding dimension. Note that all weights for GRN are shared across each time step  $t$ .

#### 4.3 Multihead Dual Convolutional Self-Attention

Our approach employs a self-attention mechanism introduced in [47] to capture global dependencies across time steps. It is embedded as part of the DCSA architecture. Inspired by Vaswani et al. [47] in utilising multihead self-attention, we utilise our DCSA with various kernel lengths with the same aim: allowing asymmetric long-term learning. The multihead DCSA, shown as part in Figure 2, takes in two inputs  $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 \in \mathbb{R}^{T \times d}$  and yields:

$$MDCSA_{k_1, \dots, k_n}(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2) = \Phi_n(\Phi_{k_1, \dots, k_n}(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2)) \quad (4)$$

with

$$\begin{aligned} \Phi_{k_1, \dots, k_n}(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2) &= SA(\Xi_{k_1, \dots, k_n}(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2) \mathbf{W}_Q, \\ &\Xi_{k_1, \dots, k_n}(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2) \mathbf{W}_K, \\ &\Xi_{k_1, \dots, k_n}(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2) \mathbf{W}_V) \end{aligned} \quad (5)$$

$\Xi_{k_1, \dots, k_n}(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2) = [DCSA_{k_1}(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2), \dots, DCSA_{k_n}(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2)]$  (6) where  $\Phi_n(\cdot)$  is a 1D-convolutional layer with a kernel size  $\{1, n\}$  and a stride  $n$ ,  $\mathbf{W}_K \in \mathbb{R}^{d \times d}$ ,  $\mathbf{W}_Q \in \mathbb{R}^{d \times d}$ ,  $\mathbf{W}_V \in \mathbb{R}^{d \times d}$  are weights for keys, queries and values of the self-attention layer, and  $\Xi_{k_1, \dots, k_n}(\cdot)$  concatenates the output of each  $DCSA_{k_c}(\cdot)$  in temporal order. For regularisation, a normalisation layer followed by a dropout layer is added after Equation 4.

Following the modality positional embedding layer in subsection 4.1, the positional embeddings of RSSI  $\mathbf{h}^r = [\mathbf{h}_1^r, \dots, \mathbf{h}_T^r]$  and accelerometer  $\mathbf{h}^a = [\mathbf{h}_1^a, \dots, \mathbf{h}_T^a]$ , produced by Eq. 1, are then fed to an MDSCSA layer with various kernel sizes  $[k_1, \dots, k_n]$ :

$$\mathbf{h} = MDCSA_{k_1, \dots, k_n}(\mathbf{h}^r, \mathbf{h}^a) \quad (7)$$

to yield  $\mathbf{h} = [\mathbf{h}_1, \dots, \mathbf{h}_T]$  with  $\mathbf{h}_t \in \mathbb{R}^d$  and  $t \leq T$ .

#### 4.4 Final Layer and Loss Calculation

We apply two different layers to produce two different outputs during training. The room-level predictions are produced via a single conditional random field (CRF) layer in combination with a linear layer applied to the output of Eq. 7 to produce the final predictions as

$$\hat{y}_t = CRF(\Phi(\mathbf{h}_t)) \quad (8)$$

$$\Phi(\mathbf{h}_t) = \mathbf{W}_p \mathbf{h}_t + \mathbf{b}_p \quad (9)$$

where  $\mathbf{W}_p \in \mathbb{R}^{d \times m}$ , and  $\mathbf{b}_p \in \mathbb{R}^m$  are weight and bias to learn,  $m$  is the number of room locations, and  $\mathbf{h} = [\mathbf{h}_1, \dots, \mathbf{h}_T] \in \mathbb{R}^{T \times d}$  is the refined embedding produced by Eq. 7. Even though the transformer can take into account neighbour information before generating the refined embedding at time step  $t$ , its decision is independent; it does not take into account the actual decision made by other refined embeddings  $t$ . We use a CRF layer to cover just that, i.e. to maximise the probability of the refined embeddings of all time steps, so it can better model cases where refined embeddings closest to one another must be compatible (i.e. minimising the possibility for impossible room transitions). When finding the best sequence of room location  $\hat{y}_t$ , the Viterbi Algorithm is used as a standard for the CRF layer.



For the second layer, we choose a particular room as a reference and perform a binary classification at each time step  $t$ . The binary classification is produced via a linear layer applied to the refined embedding  $\mathbf{h}$  as

$$\hat{\beta}_t = \mathbf{W}_\beta \mathbf{h}_t + \mathbf{b}_\beta \quad (10)$$

where  $\mathbf{W}_\beta \in \mathbb{R}^{d \times 1}$ , and  $\mathbf{b}_\beta \in \mathbb{R}$  are weight and bias to learn, and  $\hat{\beta} = [\hat{\beta}_1, \dots, \hat{\beta}_T] \in \mathbb{R}^T$  is the target probabilities for the referenced room within time window  $T$ . The reason to perform a binary classification against a particular room is because of our interest in improving the accuracy in predicting that room. In our application, the room of our choice is the hallway where it will be used as a hub connecting any other room.

**Loss Functions.** During the training process, the MDCSA network produces two kinds of outputs. Emission outputs (outputs produced by Equation 9 prior to prediction outputs)  $\hat{\mathbf{e}} = [\Phi(\mathbf{h}_1), \dots, \Phi(\mathbf{h}_T)]$  are trained to generate the likelihood estimate of room predictions, while the binary classification output  $\hat{\beta} = [\hat{\beta}_1, \dots, \hat{\beta}_T]$  is used to train the probability estimate of a particular room. The final loss function can be formulated as a combination of both likelihood and binary cross entropy loss function described as:

$$\mathcal{L}(\hat{\mathbf{e}}, \mathbf{y}, \hat{\beta}, \beta) = \mathcal{L}_{NLL}(\hat{\mathbf{e}}, \mathbf{y}) + \sum_{i=1}^T \mathcal{L}_{BCE}(\hat{\beta}_i, \beta_i) \quad (11)$$

$$\mathcal{L}_{NLL}(\hat{\mathbf{e}}, \mathbf{y}) = \sum_{\hat{y}} \sum_{i=0}^T P(\Phi(\mathbf{h}_i) | \hat{y}_i) T(\hat{y}_i | \hat{y}_{i-1}) - \sum_{i=0}^T P(\Phi(\mathbf{h}_i) | y_i) T(y_i | y_{i-1}) \quad (12)$$

$$\mathcal{L}_{BCE}(\hat{\beta}, \beta) = -\frac{1}{T} \sum_{i=0}^T \beta_i \log(\hat{\beta}_i) + (1 - \beta_i) \log(1 - \hat{\beta}_i) \quad (13)$$

where  $\mathcal{L}_{NLL}(\cdot)$  represents the negative log-likelihood and  $\mathcal{L}_{BCE}(\cdot)$  denotes the binary cross entropy,  $\mathbf{y} = [y_1, \dots, y_T] \in \mathbb{R}^T$  is the actual room locations, and  $\beta = [\beta_1, \dots, \beta_T] \in \mathbb{R}^T$  is the binary value whether at time  $t$  the room is the referenced room or not.  $P(x | y)$  denotes the conditional probability, and  $T(y_i | y_{i-1})$  denotes the transition matrix cost of having transitioned from  $y_{i-1}$  to  $y_i$ .

## 5 EXPERIMENTS AND RESULTS

We compare our proposed network, MDCSA<sub>1,4,7</sub><sup>5</sup> (MDCSA with 3 kernels of size 1, 4, and 7), with:

- Random Forest (RF) as a baseline technique which has been shown to work well for indoor localisation [43],
- TENER [50] which is a modified transformer encoder in combination with a CRF layer representing a model with capability to capture global dependency and enforce dependencies in temporal aspects,
- DTML [51] represents the state-of-the-art model for multimodal and multivariate time series with a transformer encoder to learn asymmetric correlations across modalities,

- Alt DTML<sup>6</sup> representing DTML with a GRN layer replacing the context aggregation layer and CRF layer added as the last layer,
- MDCSA<sub>1,4,7</sub> 4APS, as an ablation study, with our proposed network (i.e. MDCSA<sub>1,4,7</sub>) using 4 access points for the RSSI (instead of 10 access points) and accelerometer data (ACCL) as its input features,
- MDCSA<sub>1,4,7</sub> RSSI, as an ablation study, with our proposed network using only RSSI, without ACCL, as its input features, and
- MDCSA<sub>1,4,7</sub> 4APS RSSI, as an ablation study, with our proposed network using only 4 access points for the RSSI as its input features.

For RF, all the time series features of RSSI and accelerometry are flattened and merged into one feature vector for room-level localisation. For TENER, at each time step  $t$ , RSSI  $\mathbf{x}_t^r$  and accelerometer  $\mathbf{x}_t^a$  features are combined via a linear layer before they are processed by the networks. A grid search on the parameters of each network is performed to find the best parameter for each model. The parameters to tune are: the embedding dimension  $d$  in {128, 256}, the number of epochs in {200, 300}, and the learning rate in {0.01, 0.0001}. The dropout rate is set to 0.15, and the RAdam optimiser [29] in combination with Look-Ahead algorithm [53] is used for the training with early stopping using the validation performance. For the RF, we perform a cross-validated parameter search for the number of trees ({200, 250}), the minimum number of samples in a leaf node ({1, 5}), and whether a warm start is needed ({True, False}). The Gini impurity is used to measure splits.

**Evaluation Metrics.** We are interested in developing a system to monitor PD motor symptoms in home environments. For example, we will consider if there is any significant difference in the performance of the system when it is trained with PD data compared to being trained with healthy control (HC) data. We tailored our training procedure to test our hypothesis by performing variations of cross-validation. Apart from training our models on all HC subjects (ALL-HC), we also perform four different kinds of cross-validation: 1) We train our models on one PD subject (LOO-PD), 2) We train our models on one HC subject (LOO-HC), 3) We take one HC subject and use only roughly four minutes worth of data to train our models (4m-HC), 4) We take one PD subject and use only roughly four minutes worth of data to train our models (4m-PD). For all of our experiments, we test our trained models on all PD subjects (excluding the one used as training data for LOO-PD and 4m-PD). For room-level localisation accuracy, we use precision and weighted F1-score, all averaged and standard deviated across the test folds.

To showcase the importance of in-home gait speed features in differentiating the medication state of a person with PD, we first compare how accurate the ‘Room-to-room Transition’ duration produced by each network is to the ground truth (i.e. annotated location). We hypothesise that the more accurate the transition is compared to the ground truth, the better mobility features are for medication state classification. For the medication state classification, we then compare two different groups of features with two simple binary classifiers: 1) the baseline demographic features

<sup>5</sup>We drop the ‘<sub>1,4,7</sub>’ part when the context is clear.

<sup>6</sup>Our attempt to see the effect of GRN and CRF layer on a SOTA model.

**Table 1: Room-level and medication state accuracy of all models. Standard deviation is shown in (.), the best performer is bold, while the second best is italicized. Note that our proposed model is the one named MDCSA<sub>1,4,7</sub>**

Training	Model	Room-Level Localisation		Medication State	
		Precision	F1-Score	F1-Score	AUROC
ALL-HC	RF	<b>95.00</b>	<b>95.20</b>	56.67 (17.32)	<b>84.55 (12.06)</b>
	TENER	94.60	94.80	47.08 (16.35)	67.74 (10.82)
	DTML	94.80	94.90	50.33 (13.06)	75.97 (9.12)
	Alt DTML	94.80	95.00	47.25 (5.50)	75.63 (4.49)
	MDCSA <sub>1,4,7</sub> 4APS	92.22	92.22	53.47 (12.63)	73.48 (6.18)
	MDCSA <sub>1,4,7</sub> RSSI	94.70	94.90	51.14 (11.95)	68.33 (18.49)
	MDCSA <sub>1,4,7</sub> 4APS RSSI	93.30	93.10	<b>64.52 (11.44)</b>	<i>81.84 (6.30)</i>
	MDCSA <sub>1,4,7</sub>	<i>94.90</i>	<i>95.10</i>	<i>64.13 (6.05)</i>	80.95 (10.71)
<b>Demographic Features</b>			49.74 (15.60)	65.66 (18.54)	
LOO-HC	RF	89.67 (1.85)	88.95 (2.61)	54.74 (11.46)	69.24 (17.77)
	TENER	90.35 (1.87)	89.75 (2.24)	51.76 (14.37)	70.80 (9.78)
	DTML	90.51 (1.95)	<i>89.82 (2.60)</i>	55.34 (13.67)	73.77 (9.84)
	Alt DTML	<i>90.52 (2.17)</i>	89.71 (2.83)	49.56 (17.26)	73.26 (10.65)
	MDCSA <sub>1,4,7</sub> 4APS	88.01 (6.92)	88.08 (5.73)	<b>59.52 (20.62)</b>	74.35 (16.78)
	MDCSA <sub>1,4,7</sub> RSSI	90.26 (2.43)	89.48 (3.47)	<i>58.84 (23.08)</i>	<i>76.10 (10.84)</i>
	MDCSA <sub>1,4,7</sub> 4APS RSSI	88.55 (6.67)	88.75 (5.50)	42.34 (13.11)	72.58 (6.77)
	MDCSA <sub>1,4,7</sub>	<b>91.39 (2.13)</b>	<b>91.06 (2.62)</b>	55.50 (15.78)	<b>83.98 (13.45)</b>
<b>Demographic Features</b>			51.79 (15.40)	68.33 (18.43)	
LOO-PD	RF	86.89 (7.14)	84.71 (7.33)	43.28 (14.02)	62.63 (20.63)
	TENER	86.91 (6.76)	86.18 (6.01)	36.04 (9.99)	60.03 (10.52)
	DTML	87.13 (6.53)	86.31 (6.32)	43.98 (14.06)	66.93 (11.07)
	Alt DTML	87.36 (6.30)	86.44 (6.63)	44.02 (16.89)	69.70 (12.04)
	MDCSA <sub>1,4,7</sub> 4APS	86.44 (6.96)	85.93 (6.05)	<i>47.26 (14.47)</i>	<i>72.62 (11.16)</i>
	MDCSA <sub>1,4,7</sub> RSSI	<i>87.61 (6.64)</i>	<i>87.21 (5.44)</i>	45.71 (17.85)	67.76 (10.73)
	MDCSA <sub>1,4,7</sub> 4APS RSSI	87.20 (7.17)	87.00 (6.12)	41.33 (17.72)	66.26 (12.11)
	MDCSA <sub>1,4,7</sub>	<b>88.04 (6.94)</b>	<b>87.82 (6.01)</b>	<b>49.99 (13.18)</b>	<b>81.08 (8.46)</b>
<b>Demographic Features</b>			43.89 (14.43)	60.95 (25.16)	
4m-HC	RF	74.27 (8.99)	69.87 (7.21)	<i>50.47 (12.63)</i>	59.55 (12.38)
	TENER	69.86 (18.68)	60.71 (24.94)	N/A	N/A
	DTML	77.10 (9.89)	70.12 (14.26)	43.89 (11.60)	64.67 (12.88)
	Alt DTML	78.79 (3.95)	71.44 (9.82)	47.49 (14.64)	65.16 (12.56)
	MDCSA <sub>1,4,7</sub> 4APS	81.42 (6.95)	78.65 (7.59)	42.87 (17.34)	67.09 (7.42)
	MDCSA <sub>1,4,7</sub> RSSI	81.69 (6.85)	77.12 (8.46)	49.95 (17.35)	<i>69.71 (11.55)</i>
	MDCSA <sub>1,4,7</sub> 4APS RSSI	<i>82.80 (7.82)</i>	<i>79.37 (8.98)</i>	43.57 (23.87)	65.46 (15.78)
	MDCSA <sub>1,4,7</sub>	<b>83.32 (6.65)</b>	<b>80.24 (6.85)</b>	<b>55.43 (10.48)</b>	<b>78.24 (6.67)</b>
<b>Demographic Features</b>			32.87 (13.81)	53.68 (13.86)	
4m-PD	RF	71.00 (9.67)	65.89 (11.96)	N/A	N/A
	TENER	65.30 (23.25)	58.57 (27.19)	N/A	N/A
	DTML	70.35 (14.17)	64.00 (17.88)	N/A	N/A
	Alt DTML	74.43 (9.59)	67.55 (14.50)	N/A	N/A
	MDCSA <sub>1,4,7</sub> 4APS	81.02 (8.48)	<i>76.85 (10.94)</i>	<b>49.97 (7.80)</b>	<i>69.10 (7.64)</i>
	MDCSA <sub>1,4,7</sub> RSSI	77.47 (12.54)	73.99 (13.00)	41.79 (16.82)	67.37 (16.86)
	MDCSA <sub>1,4,7</sub> 4APS RSSI	<i>83.01 (6.42)</i>	<b>79.77 (7.05)</b>	41.18 (12.43)	63.16 (11.06)
	MDCSA <sub>1,4,7</sub>	<b>83.30 (6.73)</b>	76.77 (13.19)	<i>48.61 (12.03)</i>	<b>76.39 (12.23)</b>
<b>Demographic Features</b>			36.69 (18.15)	50.53 (15.60)	

**Table 2: Hallway prediction on limited training data.**

Training	Model	Precision	F1-Score
4m-HC	MDCSA 4APS RSSI	62.32 (19.72)	58.99 (23.87)
	MDCSA 4APS	68.07 (23.22)	60.01 (26.24)
	MDCSA	71.25 (21.92)	68.95 (17.89)
4m-PD	MDCSA 4APS RSSI	58.59 (23.60)	57.68 (24.27)
	MDCSA 4APS	62.36 (18.98)	57.76 (20.07)
	MDCSA	70.47 (14.10)	64.64 (21.38)

**Table 3: Room-to-room transition accuracy (in seconds) of all models compared to the ground truth. Standard deviation is shown in (.), the best performer is bold, while the second best is italicized. A model that fails to capture a transition between particular rooms within a period that has the ground truth is assigned 'N/A' score.**

Data	Models	Kitch-Livin	Kitch-Dinin	Dinin-Livin
<b>Ground Truth</b>				
ALL-HC	RF	16.18(12.08)	<b>14.58(10.22)</b>	10.19(9.46)
	TENER	15.58(8.75)	16.30(12.94)	12.01(13.01)
	Alt DTML	15.27(7.51)	13.40(6.43)	<i>10.84(10.81)</i>
	MDCSA	<b>17.70(16.17)</b>	<i>14.94(9.71)</i>	<b>10.76(9.59)</b>
	MDCSA	<b>17.70(17.42)</b>	<b>14.34(9.48)</b>	<i>11.07(13.60)</i>
LOO-HC	RF	17.52(16.97)	11.93(10.08)	9.23(13.69)
	TENER	14.62(16.37)	9.58(9.16)	7.21(10.61)
	Alt DTML	16.30(17.78)	<i>14.01(8.08)</i>	<b>10.37(12.44)</b>
	MDCSA	<b>17.70(17.42)</b>	<b>14.34(9.48)</b>	<i>11.07(13.60)</i>
	MDCSA	<b>17.70(17.42)</b>	<b>14.34(9.48)</b>	<i>11.07(13.60)</i>
LOO-PD	RF	14.49(15.28)	11.67(11.68)	8.65(13.06)
	TENER	13.42(14.88)	10.87(10.37)	6.95(10.28)
	Alt DTML	<b>16.98(15.15)</b>	<i>15.26(8.85)</i>	<i>9.99(13.03)</i>
	MDCSA	<i>16.42(14.04)</i>	<b>14.48(9.81)</b>	<b>10.77(14.18)</b>
	MDCSA	<i>16.42(14.04)</i>	<b>14.48(9.81)</b>	<b>10.77(14.18)</b>
4m-HC	RF	14.22(18.03)	11.38(15.46)	13.43(18.87)
	TENER	10.75(15.67)	8.59(14.39)	N/A
	Alt DTML	<i>16.89(18.07)</i>	<b>14.68(13.57)</b>	<i>9.31(15.70)</i>
	MDCSA	<b>18.15(19.12)</b>	<i>15.32(14.93)</i>	<b>11.89(17.55)</b>
	MDCSA	<b>18.15(19.12)</b>	<i>15.32(14.93)</i>	<b>11.89(17.55)</b>
4m-PD	RF	11.52(16.07)	8.73(12.90)	N/A
	TENER	8.75(14.89)	N/A	N/A
	Alt DTML	<i>14.75(13.79)</i>	<i>13.47(17.66)</i>	N/A
	MDCSA	<b>17.96(19.17)</b>	<b>14.74(10.83)</b>	<b>10.16(14.03)</b>
	MDCSA	<b>17.96(19.17)</b>	<b>14.74(10.83)</b>	<b>10.16(14.03)</b>

(see Section 3), and 2) the normalised in-home gait speed features. The metric we use for ON / OFF medication state evaluation is the weighted F1-Score and AUROC which are averaged and standard deviated across the test folds.

## 5.1 Experimental Results

**Room-level Accuracy.** The first part of Table 1 compares the performance of MDCSA network and other approaches for room-level classification. For the room-level classification, MDCSA network outperforms other networks and RF with a minimum improvement of 1.3% for the F1-score over the second-best network (i.e.

Alt DTML) in each cross-validation type with the exception of the ALL-HC validation. The improvement is more significant on the 4m-HC and 4m-PD validations, when the training data are limited, with an average improvement at almost 9% for the F1-score over the Alt DTML.

The LOO-HC and LOO-PD validation show that a model that has the ability to capture the temporal dynamics across time steps (e.g. TENER and DTML) will perform better than a standard baseline technique such as a Random Forest. TENER and DTML perform better in those two validations due to their ability to capture asynchronous relation across modalities. However, when the training data becomes limited as in 4m-HC and 4m-PD validations, having extra capabilities is necessary to further extract temporal information and correlations. Due to being a vanilla transformer requiring considerable amount of training data, TENER performs worst in these two validations. DTML performs quite well due to its ability to capture local context via LSTM for each modality. However, in general, DTML's performance suffers in both the LOO-PD and 4m-PD validations as the accelerometer data (and modality) may be erratic due to PD and should be excluded at times from contributing to room classification. MDCSA network has all the capabilities that DTML has with an improvement in suppressing accelerometer modality when needed via GRN layer embedded in DCSA. Suppressing the noisy modality seems to have a strong impact in maintaining the performance of the network when the training data is limited. This is validated by how Alt DTML (i.e. DTML added with GRN and CRF layers) outperforms the standard DTML by an average of 2.2% for the F1-score in in 4m-HC and 4m-PD validations. It is further confirmed by MDCSA<sub>1,4,7</sub> 4APS against MDCSA<sub>1,4,7</sub> 4APS RSSI with the latter model, which does not include the accelerometer data, outperforming the former for the F1-score by an average of 1.6% in the last three cross validations. It is worth pointing out that the MDCSA<sub>1,4,7</sub> 4APS RSSI model performed the best in the 4m-PD validation. However, the omission of accelerometer data affects the model in differentiating rooms that are more likely to have active movement (i.e. hall) than the rooms that are not (i.e. living room). It can be seen from Table 2 that the MDCSA<sub>1,4,7</sub> 4APS RSSI model has low performance in predicting hallway compared to the full model of MDCSA<sub>1,4,7</sub>. As a consequence, the MDCSA<sub>1,4,7</sub> 4APS RSSI model cannot produce in-home gait speed features as competent as the ones produced by the MDCSA<sub>1,4,7</sub>.

**Room-to-room Transition and Medication Accuracy.** We hypothesise that during their OFF medication state, the deterioration in mobility of a person with PD is exhibited by how they transition between rooms. To test this hypothesis, a Wilcoxon signed rank test was used on the annotated data from PD participants undertaking each of the three individual transitions between rooms whilst ON (taking) and OFF (withholding) medications to assess whether the mean transition duration ON medications was statistically significantly shorter than the mean transition duration for the same transition OFF medications for all transitions studied (see Table 4). From this result, we argue that the mean transition duration obtained by each model from Table 1 that is close to the ground truth can capture what the ground truth captures. As mentioned in Section 3, this transition duration for each model is generated by the model continuously performing room-level localisation focusing on the time a person is predicted to spend in a hallway between



**Table 4: PD participant room transition duration with ON and OFF medications comparison using Wilcoxon signed rank tests.**

OFF transitions	Mean transition duration	ON transitions	Mean transition duration	W	z	p
Kitchen-Living OFF	17.2 sec	Kitchen-Living ON	14.0 sec	75.0	2.824	0.001
Dining-Kitchen OFF	12.9 sec	Dining-Kitchen ON	9.2 sec	76.0	2.903	< .001
Dining-Living OFF	10.4 sec	Dining-Living ON	9.0 sec	64.0	1.961	0.026

rooms. We show, in Table 3, that the mean transition duration for all transitions studied produced by MDCSA<sub>1,4,7</sub> model is the closest to the ground truth improving over the second best by around 1.25 seconds across all hall transitions and validations.

The second part of the Table 1 shows the performance of all our networks for medication state classification. The demographic features can be used as a baseline for each type of validation. The MDCSA network, with the exception of the ALL-HC validation, outperforms any other network by a significant margin for the AUROC score. By using in-home gait speed features produced by MDCSA network, a minimum of 15% improvement over the baseline demographic features can be obtained with the biggest gained obtained in the 4m-PD validation data. In 4m-PD validation data, RF, TENER, and DTML could not manage to provide any prediction due to their inability to capture (partly) hall transitions. Furthermore, TENER has shown its inability to provide any medication state prediction from the 4m-HC data validations. It can be validated by Table 3 when the TENER failed to capture any transitions between dining room and living room across all periods that have ground truths. MDCSA networks are able to provide medication state prediction and maintain its performance across all cross-validations thanks to the addition of Eq. 13 in the loss function.

## 6 CONCLUSION

We have presented the MDCSA model, a new deep learning approach for indoor localisation utilising RSSI and wrist-worn accelerometer data. The evaluation on our unique real-world free-living pilot dataset, which includes subjects with and without PD, shows that MDCSA achieves the state-of-the-art accuracy for indoor localisation. The availability of accelerometer data does indeed enrich the RSSI features which, in turn, improves the accuracy of the indoor localisation.

In naturalistic settings, in-home mobility can be measured through the use of indoor localisation models. We have shown, using room transition duration results, that our PD cohort take longer on average to perform a room transition when they withhold medications. With accurate in-home gait speed features, a classifier model can then differentiate accurately if a person with PD is in an ON or OFF medication state. Such changes show the promise of these localisation outputs to detect the dopamine-related gait fluctuations in PD that impact on patients’ quality of life [31] and are important in clinical decision making [9]. We have also demonstrated that our indoor localisation system provides precise in-home gait speed features in PD with a minimal average offset to the ground truth. The network also outperforms other models in the production of in-home gait speed features which is used to differentiate the medication state of a person with PD.

**Limitations and future research.** One limitation of this study is the relatively small sample size (which was planned as this is an exploratory pilot study). We believe our sample size is ample to show proof of concept. This is also the first such work with unobtrusive ground truth validation from embedded cameras. Future work should validate our approach further on a large cohort of people with PD and consider stratifying for sub-groups within PD (e.g. akinetic-rigid or tremor-dominant phenotypes), which would also increase the generalisability of the results to the wider population. Future work in this matter could also include the construction of a semi-synthetic dataset based on collected data to facilitate a parallel and large-scale evaluation.

This smart home’s layout and parameters remain constant for all the participants, and we acknowledge that transfer of this deep learning model to other varied home settings may introduce variations in localisation accuracy. For future ecological validation and based on our current results, we anticipate the need for pre-training (e.g. a brief walkaround which is labelled) for each home, and also suggest that some small amount of ground-truth data will need to be collected (e.g. researcher prompting of study participants to undertake scripted activities such as moving from room to room) to fully validate the performance of our approach in other settings.

Accurate room localisation using these data modalities has a wide range of potential applications within healthcare. This could include tracking of gait speed during rehabilitation from orthopaedic surgery, monitoring wandering behaviour in dementia or triggering an alert for a possible fall (and long lie on the floor) if someone is in one room for an unusual length of time. Furthermore, accurate room use and room-to-room transfer statistics could be used in occupational settings, e.g. to check factory worker location.

## ACKNOWLEDGMENTS

We are very grateful to the study participants for giving so much time and effort to this research. We acknowledge the local Movement Disorders Health Integration Team (Patient and Public Involvement Group) for their assistance at each study design step. This work was supported by the SPHERE Next Steps Project funded by the UK Engineering and Physical Sciences Research Council (EPSRC), [Grant EP/R005273/1]; the Elizabeth Blackwell Institute for Health Research, University of Bristol and the Wellcome Trust Institutional Strategic Support Fund [grant code: 204813/Z/16/Z]; by Cure Parkinson’s Trust [grant code AW021]; and by IXICO [grant code R101507-101]. Dr and Mrs de Pass made a charitable donation to the University of Bristol through the Development and Alumni Relations Office to pay the salary of CM, the PhD student and Clinical Research Fellow.

## REFERENCES

- [1] Pierre-Emmanuel Aguera, Karim Jerbi, Anne Caclin, and Olivier Bertrand. 2011. ELAN: a software package for analysis and visualization of MEG, EEG, and LFP signals. *Computational intelligence and neuroscience* 2011 (2011).
- [2] Marc Bachlin, Daniel Roggen, Gerhard Troster, Meir Plotnik, Noit Inbar, Inbal Meidan, Talia Herman, Marina Brozgol, Eliya Shaviv, Nir Giladi, and Jeffrey M. Hausdorff. 2009. Potentials of Enhanced Context Awareness in Wearable Assistants for Parkinson's Disease Patients with the Freezing of Gait Syndrome. In *2009 International Symposium on Wearable Computers*. 123–130. <https://doi.org/10.1109/ISWC.2009.14>
- [3] Dzmityr Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural Machine Translation by Jointly Learning to Align and Translate. <https://doi.org/10.48550/ARXIV.1409.0473>
- [4] Aishwarya Balakrishnan, Jeevan Medikonda, Pramod Kesavan Nambhoorthi, Manik Manik, and an Natarajan. 2022. Role of Wearable Sensors with Machine Learning Approaches in Gait Analysis for Parkinson's Disease Assessment: A Review. *Engineered Science* 19 (2022), 5–19. <https://doi.org/10.30919/es8e622>
- [5] Óscar Belmonte-Fernández, Adrian Puertas-Cabedo, Joaquín Torres-Sospedra, Raúl Montoliu-Colás, and Sergi Trilles-Oliver. 2017. An Indoor Positioning System Based on Wearables for Ambient-Assisted Living. *Sensors* 17, 1 (2017). <https://doi.org/10.3390/s17010036>
- [6] Alessio Benavoli, Giorgio Corani, and Francesca Mangili. 2016. Should we really use post-hoc tests based on mean-ranks? *The Journal of Machine Learning Research* 17, 1 (2016), 152–161.
- [7] Lorenzo Brognara, Pierpaolo Palumbo, Bernd Grimm, and Luca Palmerini. 2019. Assessing Gait in Parkinson's Disease Using Wearable Motion Sensors: A Systematic Review. *Diseases* 7, 1 (2019). <https://doi.org/10.3390/diseases7010018>
- [8] Defu Cao, Yujing Wang, Juanyong Duan, Ce Zhang, Xia Zhu, Congrui Huang, Yunhai Tong, Bixiong Xu, Jing Bai, Jie Tong, and Qi Zhang. 2020. Spectral Temporal Graph Neural Network for Multivariate Time-series Forecasting. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 17766–17778.
- [9] Carl E. Clarke, Paul Worth, Donald Grosset, and David Stewart. 2009. Systematic review of apomorphine infusion, levodopa infusion and deep brain stimulation in advanced Parkinson's disease. *Parkinsonism Related Disorders* 15 (12 2009), 728–741. Issue 10. <https://doi.org/10.1016/j.parkreldis.2009.09.005>
- [10] Janez Demšar. 2006. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research* 7, 1 (2006), 1–30.
- [11] David Devos, Etienne Hirsch, and Richard Wyse. 2021. Seven Solutions for Neuroprotection in Parkinson's Disease. *Movement Disorders* 36, 2 (2021), 306–316. <https://doi.org/10.1002/mds.28379>
- [12] Giulia Di Lazzaro, Mariachiara Ricci, Mohammad Al-Wardat, Tommaso Schirinzi, Simona Scalise, Franco Giannini, Nicola B Mercuri, Giovanni Saggio, and Antonio Pisani. 2020. Technology-based objective measures detect subclinical axial signs in untreated, de novo Parkinson's disease. *Journal of Parkinson's Disease* 10, 1 (2020), 113–122.
- [13] Ryan P. Duncan and Gammon M. Earhart. 2011. Measuring participation in individuals with Parkinson disease: relationships with disease severity, quality of life, and mobility. *Disability and Rehabilitation* 33, 15–16 (2011), 1440–1446. <https://doi.org/10.3109/09638288.2010.533245> PMID: 21091047.
- [14] Alberto J. Espay, Paolo Bonato, Fatta B. Nahab, Walter Maetzler, John M. Dean, Jochen Klucken, Bjoern M. Eskofier, Aristide Merola, Fay Horak, Anthony E. Lang, Ralf Reilmann, Joe Giuffrida, Alice Nieuwboer, Malcolm Horne, Max A. Little, Irene Litvan, Tanya Simuni, E. Ray Dorsey, Michelle A. Burack, Ken Kubota, Anita Kamondi, Catarina Godinho, Jean-Francois Daneault, Georgia Mitsi, Lothar Krinke, Jeffery M. Hausdorff, Bastiaan R. Bloem, Spyros Papapetropoulos, and on behalf of the Movement Disorders Society Task Force on Technology. 2016. Technology in Parkinson's disease: Challenges and opportunities. *Movement Disorders* 31, 9 (2016), 1272–1282. <https://doi.org/10.1002/mds.26642>
- [15] Alberto J. Espay, Jeffrey M. Hausdorff, Álvaro Sánchez-Ferro, Jochen Klucken, Aristide Merola, Paolo Bonato, Serene S. Paul, Fay B. Horak, Joaquin A. Vizcarra, Tiago A. Mestre, Ralf Reilmann, Alice Nieuwboer, E. Ray Dorsey, Lynn Rochester, Bastiaan R. Bloem, Walter Maetzler, and on behalf of the Movement Disorder Society Task Force on Technology. 2019. A roadmap for implementation of patient-centered digital outcome measures in Parkinson's disease obtained using mobile health technologies. *Movement Disorders* 34, 5 (2019), 657–663. <https://doi.org/10.1002/mds.27671>
- [16] Luc J.W. Evers, Jesse H. Krijthe, Marjan J. Meinders, Bastiaan R. Bloem, and Tom M. Heskes. 2019. Measuring Parkinson's disease over time: The real-world within-subject reliability of the MDS-UPDRS. *Movement Disorders* 34, 10 (2019), 1480–1487. <https://doi.org/10.1002/mds.27790>
- [17] James M. Fisher, Nils Y. Hammerla, Thomas Ploetz, Peter Andras, Lynn Rochester, and Richard W. Walker. 2016. Unsupervised home monitoring of Parkinson's disease motor symptoms using body-worn accelerometers. *Parkinsonism Related Disorders* 33 (2016), 44–50. <https://doi.org/10.1016/j.parkreldis.2016.09.009>
- [18] N Giladi, TA Treves, ES Simon, H Shabtai, Y Orlov, B Kandinov, D Paleacu, and AD Korczyn. 2001. Freezing of gait in patients with advanced Parkinson's disease. *Journal of neural transmission* 108, 1 (2001), 53–61.
- [19] Julia C. Greenland, Caroline H. Williams-Gray, and Roger A. Barker. 2019. The clinical heterogeneity of Parkinson's disease and its therapeutic implications. *European Journal of Neuroscience* 49, 3 (2019), 328–338.
- [20] Neal Hermanowicz, Sarah A Jones, and Robert A Hauser. 2019. Impact of non-motor symptoms in Parkinson's disease: a PMDAAlliance survey. *Neuropsychiatric disease and treatment* 15 (2019), 2205.
- [21] Etienne C. Hirsch and David G. Standaert. 2021. Ten Unsolved Questions About Neuroinflammation in Parkinson's Disease. *Movement Disorders* 36, 1 (2021), 16–24. <https://doi.org/10.1002/mds.28075>
- [22] Mai Ibrahim, Marwan Torki, and Mustafa ElNainay. 2018. CNN based Indoor Localization using RSS Time-Series. In *2018 IEEE Symposium on Computers and Communications (ISCC)*. 01044–01049. <https://doi.org/10.1109/ISCC.2018.8538530>
- [23] J Jankovic. 2008. Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry* 79, 4 (2008), 368–376.
- [24] A. Kolb, E. Barth, R. Koch, and R. Larsen. 2010. Time-of-Flight Cameras in Computer Graphics. *Computer Graphics Forum* 29, 1 (2010), 141–159. <https://doi.org/10.1111/j.1467-8659.2009.01583.x>
- [25] Nenad Kostanjsek. 2011. Use of The International Classification of Functioning, Disability and Health (ICF) as a conceptual framework and common language for disability statistics and health information systems. In *BMC public health*, Vol. 11. Springer, 1–6.
- [26] Michał Kozłowski, Dallan Byrne, Raul Santos-Rodriguez, and Robert Piechocki. 2018. Data fusion for robust indoor localisation in digital health. In *2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. 302–307.
- [27] Anthony E. Lang and Alberto J. Espay. 2018. Disease Modification in Parkinson's Disease: Current Approaches, Challenges, and Future Considerations. *Movement Disorders* 33, 5 (2018), 660–677. <https://doi.org/10.1002/mds.27360>
- [28] Bryan Lim, Sercan Ö. Arık, Nicolas Loeff, and Tomas Pfister. 2021. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting* 37, 4 (2021), 1748–1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- [29] Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. 2019. On the Variance of the Adaptive Learning Rate and Beyond. *CoRR abs/1908.03265* (2019). <https://doi.org/10.48550/ARXIV.1908.03265> arXiv:1908.03265
- [30] Ruichen Liu, Zhelong Wang, Sen Qiu, Hongyu Zhao, Cui Wang, Xin Shi, and Fang Lin. 2022. A Wearable Gait Analysis and Recognition Method for Parkinson's Disease Based on Error State Kalman Filter. *IEEE Journal of Biomedical and Health Informatics* 26, 8 (2022), 4165–4175. <https://doi.org/10.1109/JBHI.2022.3174249>
- [31] Connie Marras, Anthony Lang, Murray Krahn, George Tomlinson, and Gary Naglie. 2004. Quality of life in early Parkinson's disease: Impact of dyskinesia and motor fluctuations. *Movement Disorders* 19 (1 2004), 22–28. Issue 1. <https://doi.org/10.1002/MDS.10642>
- [32] Ankur Pandey, Ryan Sequeira, and Sudhir Kumar. 2021. Joint Localization and Radio Map Generation using Transformer Networks with Limited RSS Samples. In *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*. 1–6.
- [33] Antoine Piau, Katherine Wild, Nora Mattek, and Jeffrey Kaye. 2019. Current State of Digital Biomarker Technologies for Real-Life, Home-Based Monitoring of Cognitive Function for Mild Cognitive Impairment to Mild Alzheimer Disease and Implications for Clinical Care: Systematic Review. *J Med Internet Res* 21, 8 (30 Aug 2019), e12785. <https://doi.org/10.2196/12785>
- [34] Alwin Poulouse, Odongo Steven Eyobu, and Dong Seog Han. 2019. An Indoor Position-Estimation Algorithm Using Smartphone IMU Sensor Data. *IEEE Access* 7 (2019), 11165–11177. <https://doi.org/10.1109/ACCESS.2019.2891942>
- [35] Alwin Poulouse, Jihun Kim, and Dong Seog Han. 2019. A Sensor Fusion Framework for Indoor Localization Using Smartphone Sensors and Wi-Fi RSSI Measurements. *Applied Sciences* 9, 20 (2019). <https://doi.org/10.3390/app9204379>
- [36] Antoine Regnault, Babak Boroojerdi, Juliette Meunier, Massimo Bani, Thomas Morel, and Stefan Cano. 2019. Does the MDS-UPDRS provide the precision to assess progression in early Parkinson's disease? Learnings from the Parkinson's progression marker initiative cohort. *Journal of neurology* 266, 8 (2019), 1927–1936.
- [37] Mariachiara Ricci, Giulia Di Lazzaro, Antonio Pisani, Simona Scalise, Mohammad Alwardat, Chiara Salimei, Franco Giannini, and Giovanni Saggio. 2019. Wearable Electronics Assess the Effectiveness of Transcranial Direct Current Stimulation on Balance and Gait in Parkinson's Disease Patients. *Sensors* 19, 24 (2019). <https://doi.org/10.3390/s19245465>
- [38] Erika Rovini, Carlo Maremmi, and Filippo Cavallo. 2017. How Wearable Sensors Can Support Parkinson's Disease Diagnosis and Treatment: A Systematic Review. *Frontiers in Neuroscience* 11 (2017).

[39] Emilio Sansano-Sansano, Óscar Belmonte-Fernández, Raúl Montoliu, Arturo Gascó-Compte, and Antonio Caballer-Miedes. 2020. Multimodal sensor data integration for indoor positioning in ambient-assisted living environments. *Mobile Information Systems* 2020 (2020).

[40] Corina K. Schindhelm and Asa MacWilliams. 2011. *Overview of Indoor Positioning Technologies for Context Aware AAL Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, 273–291. [https://doi.org/10.1007/978-3-642-18167-2\\_19](https://doi.org/10.1007/978-3-642-18167-2_19)

[41] Krista G Sibley, Christine Girges, Ehsan Hoque, and Thomas Foltynie. 2021. Video-based analyses of Parkinson’s disease severity: A brief review. *Journal of Parkinson’s Disease* 11, s1 (2021), S83–S93.

[42] Diane Stephenson, Reham Badawy, Soania Mathur, Maria Tome, and Lynn Rochester. 2021. Digital progression biomarkers as novel endpoints in clinical trials: A multistakeholder perspective. *Journal of Parkinson’s Disease* 11, s1 (2021), S103–S109.

[43] Thanaphon Suwannaphong, Ryan McConville, and Ian Craddock. 2022. Radio Signal Strength Indication Augmentation for One-Shot Learning in Indoor Localisation. In *Proceedings of the 1st ACM Workshop on Smart Wearable Systems and Applications* (Sydney, NSW, Australia) (*SmartWear ’22*). Association for Computing Machinery, New York, NY, USA, 7–12. <https://doi.org/10.1145/3556560.3560714>

[44] Nima Toosizadeh, Jane Mohler, Hong Lei, Saman Parvaneh, Scott Sherman, and Bijan Najafi. 2015. Motor performance assessment in Parkinson’s disease: association between objective in-clinic, objective in-home, and subjective/semi-objective measures. *PLoS one* 10, 4 (2015), e0124763.

[45] Conrad S. Tucker, Ishan Behoora, Harriet Black Nembhard, Mechelle Lewis, Nicholas W. Sterling, and Xuemei Huang. 2015. Machine learning classification of medication adherence in patients with movement disorders using non-wearable sensors. *Computers in Biology and Medicine* 66 (2015), 120–134. <https://doi.org/10.1016/j.compbiomed.2015.08.012>

[46] Jannis van Kersbergen, Karen Otte, Nienke M de Vries, Bastiaan R Bloem, Hanna M Röhlhling, Sebastian Mansow-Model, Nicolien M van der Kolk, Sebastiaan Overeem, Svitlana Zinger, and Merel M van Gilst. 2021. Camera-based objective measures of Parkinson’s disease gait features. *BMC Research Notes* 14, 1 (2021), 1–6.

[47] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc.

[48] N. M. De Vries, K. Smilowska, J. Hummelink, B. Abramiuc, M. M. Van Gilst, B. R. Bloem, P. H.N. De With, and S. Overeem. 2019. Exploring the Parkinson patients’ perspective on home-based video recording for movement analysis: A qualitative study. *BMC Neurology* 19 (4 2019), Issue 1. <https://doi.org/10.1186/S12883-019-1301-Y>

[49] Max Wuehr, A Huppert, F Schenkel, J Decker, K Jahn, and Roman Schniepp. 2020. Independent domains of daily mobility in patients with neurological gait disorders. *Journal of neurology* 267, 1 (2020), 292–300.

[50] Hang Yan, Bocao Deng, Xiaonan Li, and Xipeng Qiu. 2019. TENER: Adapting Transformer Encoder for Named Entity Recognition. *CoRR* abs/1911.04474 (2019). <https://doi.org/10.48550/ARXIV.1911.04474>

[51] Jaemin Yoo, Yejun Soun, Yong-chan Park, and U Kang. 2021. Accurate Multivariate Stock Movement Prediction via Data-Axis Transformer with Multi-Level Contexts (*KDD ’21*). Association for Computing Machinery, New York, NY, USA, 2037–2045.

[52] Cris Zampieri, Arash Salarian, Patricia Carlson-Kuhta, Kamiar Aminian, John G Nutt, and Fay B Horak. 2010. The instrumented timed up and go test: potential outcome measure for disease modifying therapies in Parkinson’s disease. *Journal of Neurology, Neurosurgery & Psychiatry* 81, 2 (2010), 171–176. <https://doi.org/10.1136/jnnp.2009.173740>

[53] Michael Zhang, James Lucas, Jimmy Ba, and Geoffrey E Hinton. 2019. Lookahead Optimizer: k steps forward, 1 step back. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc.

[54] Wei Zhang, Kan Liu, Weidong Zhang, Youmei Zhang, and Jason Gu. 2016. Deep Neural Networks for wireless localization in indoor and outdoor environments. *Neurocomputing* 194 (2016), 279–287.

[55] Yuri Álvarez López, Jacqueline Franssen, Guillermo Álvarez Narcandi, Janet Pagnozzi, Ignacio González-Pinto Arrillaga, and Fernando Las-Heras Andrés. 2018. RFID Technology for Management and Tracking: e-Health Applications. *Sensors* 18, 8 (2018). <https://doi.org/10.3390/s18082663>

## A STATISTICAL SIGNIFICANCE TEST

It could be argued that all the localisation models compared in Table 1 might not be statistically different due to fairly high standard deviation across all types of cross-validations which is caused by relatively small number of participants. In order to compare multiple models over cross-validation sets and show statistical significance of our proposed model, we perform the Friedman test to first reject the null hypothesis [10]. We then performed a pairwise statistical comparison: the Wilcoxon signed-rank test with Holm’s alpha correction ( $\alpha = 5\%$ ). Finally, we used a critical difference diagram [6] to visualize the results of these statistical tests projected onto the average rank axis, with a thick horizontal line showing a clique of localisation models that are not significantly different (see Figure 3 and 4).

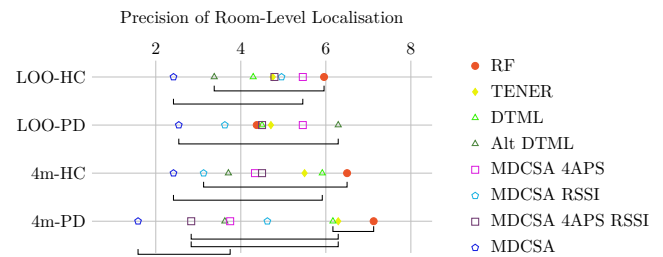


Figure 3: Critical difference diagram for the precision of room-level localisation showing the pairwise statistical comparison of all localisation models across different cross-validation techniques.

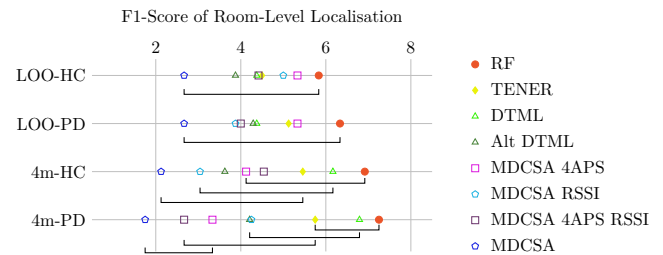


Figure 4: Critical difference diagram for the F1-score of room-level localisation showing the pairwise statistical comparison of all localisation models across different cross-validation techniques.