Detecting Medication Usage in Parkinson's Disease Through Multi-modal Indoor Positioning: A Pilot Study in a Naturalistic Environment

Abstract

Parkinson's disease (PD) is a progressive neurodegenerative disorder that leads to motor symptoms, including gait impairment. The effectiveness of levodopa therapy, a common treatment for PD, can fluctuate, causing periods of improved mobility ("on" state) and periods where symptoms re-emerge ("off" state). These fluctuations impact gait speed and increase in severity as the disease progresses. This paper proposes a transformer-based method that uses both Received Signal Strength Indicator (RSSI) and accelerometer data from wearable devices to enhance indoor localization accuracy. A secondary goal is to determine if indoor localization, particularly in-home gait speed features (like the time to walk between rooms), can be used to identify motor fluctuations by detecting if a person with PD is taking their levodopa medication or not. The method is evaluated using a real-world dataset collected in a free-living setting, where movements are varied and unstructured. Twenty-four participants, living in pairs (one with PD and one control), resided in a sensor-equipped smart home for five days. The results show that the proposed network surpasses other methods for indoor localization. The evaluation of the secondary goal reveals that accurate room-level localization, when converted into in-home gait speed features, can accurately predict whether a PD participant is taking their medication or not.

1 Introduction

Parkinson's disease (PD) is a debilitating neurodegenerative condition that affects approximately 6 million individuals globally. It manifests through various motor symptoms, including bradykinesia (slowness of movement), rigidity, and gait impairment. A common complication associated with levodopa, the primary medication for PD, is the emergence of motor fluctuations that are linked to medication timing. Initially, patients experience a consistent and extended therapeutic effect when starting levodopa. However, as the disease advances, a significant portion of patients begin to experience "wearing off" of their medication before the next scheduled dose, resulting in the reappearance of parkinsonian symptoms, such as slowed gait. These fluctuations in symptoms negatively impact patients' quality of life and often necessitate adjustments to their medication regimen. The severity of motor symptoms can escalate to the point where they impede an individual's ability to walk and move within their own home. Consequently, individuals may be inclined to remain confined to a single room, and when they do move, they may require more time to transition between rooms. These observations could potentially be used to identify periods when PD patients are experiencing motor fluctuations related to their medication being in an ON or OFF state, thereby providing valuable information to both clinicians and patients.

A sensitive and accurate ecologically-validated biomarker for PD progression is currently unavailable, which has contributed to failures in clinical trials for neuroprotective therapies in PD. Gait parameters are sensitive to disease progression in unmedicated early-stage PD and show promise as markers of disease progression, making measuring gait parameters potentially useful in clinical trials of disease-modifying interventions. Clinical evaluations of PD are typically conducted in artificial clinic or laboratory settings, which only capture a limited view of an individual's motor function. Continuous monitoring could capture symptom progression, including motor fluctuations, and sensitively quantify them over time.

While PD symptoms, including gait and balance parameters, can be measured continuously at home using wearable devices containing inertial motor units (IMUs) or smartphones, this data does not show the context in which the measurements are taken. Determining a person's location within a home (indoor localization) could provide valuable contextual information for interpreting PD symptoms. For instance, symptoms like freezing of gait and turning in gait vary depending on the environment, so knowing a person's location could help predict such symptoms or interpret their severity. Additionally, understanding how much time someone spends alone or with others in a room is a step towards understanding their social participation, which impacts quality of life in PD. Localization could also provide valuable information in the measurement of other behaviors such as non-motor symptoms like urinary function (e.g., how many times someone visits the toilet room overnight).

IoT-based platforms with sensors capturing various modalities of data, combined with machine learning, can be used for unobtrusive and continuous indoor localization in home environments. Many of these techniques utilize radio-frequency signals, specifically the Received Signal Strength Indication (RSSI), emitted by wearables and measured at access points (AP) throughout a home. These signals estimate the user's position based on perceived signal strength, creating radio-map features for each room. To improve localization accuracy, accelerometer data from wearable devices, along with RSSI, can be used to distinguish different activities (e.g., walking vs. standing). Since some activities are associated with specific rooms (e.g., stirring a pan on the stove is likely to occur in a kitchen), accelerometer data can enhance RSSI's ability to differentiate between adjacent rooms, an area where RSSI alone may be insufficient.

The heterogeneity of PD, where symptoms and their severity vary between patients, poses a challenge for generalizing accelerometer data across different individuals. Severe symptoms, such as tremors, can introduce bias and accumulated errors in accelerometer data, particularly when collected from wrist-worn devices, which are a common and well-accepted placement location. Naively combining accelerometer data with RSSI may degrade indoor localization performance due to varying tremor levels in the acceleration signal. This work makes two primary contributions to address these challenges.

(1) We detail the use of RSSI, augmented by accelerometer data, to achieve room-level localization. Our proposed network intelligently selects accelerometer features that can enhance RSSI performance in indoor localization. To rigorously assess our method, we utilize a free-living dataset (where individuals live without external intervention) developed by our group, encompassing diverse and unstructured movements as expected in real-world scenarios. Evaluation on this dataset, including individuals with and without PD, demonstrates that our network outperforms other methods across all cross-validation categories.

(2) We demonstrate how accurate room-level localization predictions can be transformed into in-home gait speed biomarkers (e.g., number of room-to-room transitions, room-to-room transition duration). These biomarkers can effectively classify the OFF or ON medication state of a PD patient from this pilot study data.

2 Related Work

Extensive research has utilized home-based passive sensing systems to evaluate how the activities and behavior of individuals with neurological conditions, primarily cognitive dysfunction, change over time. However, there is limited work assessing room use in the home setting in people with Parkinson's.

Gait quantification using wearables or smartphones is an area where a significant amount of work has been done. Cameras can also detect parkinsonian gait and some gait features, including step length and average walking speed. Time-of-flight devices, which measure distances between the subject and the camera, have been used to assess medication adherence through gait analysis. 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. 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 behavior, and there have been some privacy concerns around the use of video data at home.

RSSI data from wearable devices is a type of data with fewer privacy concerns; it can be measured continuously and unobtrusively over long periods to capture real-world function and behavior in a privacy-friendly way. In indoor localization, 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 from the wearable to the access point. 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 localization accuracy. Some works utilize 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 location estimate. Other works try to utilize 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 localization models to further leverage the temporal dependencies across time-series readings.

It has been suggested that we cannot rely on RSSI alone for indoor localization in home environments for PD subjects due to shadowing rooms with tight separation. Some researchers 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. Classic machine learning approaches such as Random Forest (RF), Artificial Neural Network (ANN), and k-Nearest Neighbor (k-NN) are tested, and the result shows that the RF outperforms other methods in tracking a person in indoor environments. Others 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. The proposed sensor fusion framework uses location fingerprinting in combination with a pedestrian dead reckoning (PDR) algorithm to reduce positioning errors.

Looking at this multi-modality classification/regression problem from a time series perspective, there has been a lot of exploration in tackling a problem where each modality can be categorized as multivariate time series data. LSTM and attention layers are often used in parallel to directly transform raw multivariate time series data into a low-dimensional feature representation for each modality. Later, various processes are done to further extract correlations across modalities through the use of various layers (e.g., concatenation, CNN layer, transformer, self-attention). Our work is inspired by prior research where we only utilize accelerometer data to enrich the RSSI, instead of utilizing all IMU sensors, in order to reduce battery consumption. In addition, unlike previous work that stops at predicting room locations, we go a step further and use room-to-room transition behaviors 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 and 10 Access Points (APs) placed throughout the residential home, each measuring the RSSI. 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 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 healthy control volunteer (HC). This pairing was chosen to enable PD vs. HC comparison, for safety reasons, and also to increase the naturalistic social behavior (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 on 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 localization labels. Multiple human labelers used software called ELAN to watch up to 4 simultaneously-captured video files at a time. The resulting labeled data recorded the kitchen, hallway, dining room, living room, stairs, and porch. The duration of labeled 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. 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 localization: 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 synchronize the data with RSSI values. With a 5-second time window and a 5Hz sampling rate, each RSSI data sample has an input of size (25 x 20), and accelerometer data has an input of size (25 x 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 normalized.

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 hypothesize 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, the hallway is used as a hub connecting all other rooms labeled, 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-localization 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 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 that is used to train a different binary classifier to predict whether a person with PD is ON or OFF medications.

Ethical approval: Full approval from the NHS Wales Research Ethics Committee was granted on December 17, 2019, and Health Research Authority and Health and Care Research Wales approval was confirmed on January 14, 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.

4 Methodologies and Framework

We introduce Multihead Dual Convolutional Self Attention (MDCSA), a deep neural network that utilizes dual modalities for indoor localization in home environments. The network addresses 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 localization, 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) Modeling 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 MDCSA architecture 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 $x^r = [x^r_1, x^r_2, ..., x^r_T] \in \mathbb{R}^{T\times r}$ and accelerometer data $x^a = [x^a_1, x^a_2, ..., x^a_T] \in \mathbb{R}^{T\times a}$ within T time units, where $x^r_t = [x^r_{t1}, x^r_{t2}, ..., x^r_{tr}]$ represents RSSI signals from r access points, and $x^a_t = [x^a_{t1}, x^a_{t2}, ..., x^a_{ta}]$ represents accelerometer data from a spatial directions at time t with $t < T$. Given feature vectors $x_t = [x^r_t, x^a_t]$ with $u \in \{r, a\}$ representing RSSI or accelerometer data at time t, and $t < T$ representing the time index, a positional embedding h^u_t for RSSI or accelerometer can be obtained by:

$$h^u_t = (W_u x^u_t + b_u) + \tau_t \tag{1}$$

where $W_u \in \mathbb{R}^{u\times d}$ and $b_u \in \mathbb{R}^d$ are the weight and bias to learn, d is the embedding dimension, and $\tau_t \in \mathbb{R}^d$ is the corresponding position encoding at time t.

4.2 Locality Enhancement with Self-Attention

Since 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. Utilizing 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 utilize a recurrent neural network such as a long-short term memory (LSTM) approach. However, in LSTM layers, the local context is summarized based on the previous context and the current input. Two similar patterns separated by a long period of time might have different contexts if they are processed by the LSTM layers. We utilize 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 $\hat{x}_1 \in \mathbb{R}^{N\times d}$ and a secondary input $\hat{x}_2 \in \mathbb{R}^{N\times d}$ and yields:

$$DCSA(\hat{x}_1, \hat{x}_2) = GRN(Norm(\phi(\hat{x}_1) + \hat{x}_1), Norm(\phi(\hat{x}_2) + \hat{x}_2)) \tag{2}$$

with

$$\phi(\hat{x}) = SA(\Phi_k(\hat{x})W_Q, \Phi_k(\hat{x})W_K, \Phi_k(\hat{x})W_V) \tag{3}$$

where $GRN(.)$ is the Gated Residual Network to integrate dual inputs into one integrated embedding, $Norm(.)$ is a standard layer normalization, $SA(.)$ is a scaled dot-product self-attention, $\Phi_k(.)$ is a 1D-convolutional layer with a kernel size $\{1, k\}$ and a stride of 1, $W_K \in \mathbb{R}^{d\times d}, W_Q \in \mathbb{R}^{d\times d}, W_V \in \mathbb{R}^{d\times d}$ 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 to capture global dependencies across time steps. It is embedded as part of the DCSA architecture. Inspired by utilizing multihead self-attention, we utilize our DCSA with various kernel lengths with the same aim: allowing asymmetric long-term learning. The multihead DCSA takes in two inputs $\hat{x}_1, \hat{x}_2 \in \mathbb{R}^{N \times d}$ and yields:

$$MDCSA_{k_1,...,k_n}(\hat{x}_1, \hat{x}_2) = \Xi_n(\phi_{k_1,...,k_n}(\hat{x}_1, \hat{x}_2)) \tag{4}$$

with

$$\phi_{k_i}(\hat{x}_1, \hat{x}_2) = SA(\Phi_{k_i}(\hat{x}_1)W_Q, \Phi_{k_i}(\hat{x}_2)W_K, \Phi_{k_i}(\hat{x}_1, \hat{x}_2)W_V) \tag{5}$$

where $\Phi_{k_i}(.)$ is a 1D-convolutional layer with a kernel size $\{1, k_i\}$ and a stride k_i, $W_K \in \mathbb{R}^{d \times d}$, $W_Q \in \mathbb{R}^{d \times d}$, $W_V \in \mathbb{R}^{d \times d}$ are weights for keys, queries, and values of the self-attention layer, and $\Xi_n(.)$ concatenates the output of each $DCSA_{k_i}(.)$ in temporal order. For regularization, a normalization 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 $h^r = [h_1^r, ..., h_T^r]$ and accelerometer $h^a = [h_1^a, ..., h_T^a]$, produced by Eq. 1, are then fed to an MDCSA layer with various kernel sizes $[k_1, ..., k_n]$:

$$h = MDCSA_{k_1,...,k_n}(h^r, h^a) \tag{6}$$

to yield $h = [h_1, ..., h_T]$ with $h_t \in \mathbb{R}^d$ and $t < T$.

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:

$$\hat{y}_t = CRF(\phi(h_t)) \tag{7}$$

$$q'(h_t) = W_p h_t + b_p \tag{8}$$

where $W_p \in \mathbb{R}^{d \times m}$ and $b_p \in \mathbb{R}^m$ are the weight and bias to learn, m is the number of room locations, and $h = [h_1, ..., h_T] \in \mathbb{R}^{T \times d}$ is the refined embedding produced by Eq. 7. Even though the transformer can take into account neighbor 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 maximize 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., minimizing 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 h_t as:

$$\hat{f}_t = W_f h_t + b_f \tag{9}$$

where $W_f \in \mathbb{R}^{d \times 1}$ and $b_f \in \mathbb{R}$ are the weight and bias to learn, and $\hat{f} = [\hat{f}_1, ..., \hat{f}_T] \in \mathbb{R}^T$ 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) $\hat{e} = [\phi(h_1), ..., \phi(h_T)]$ are trained to generate the likelihood estimate of room predictions, while the binary classification output $\hat{f} = [\hat{f}_1, ..., \hat{f}_T]$ 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 functions described as:

$$L(\hat{e}, y, \hat{f}, f) = L_{LL}(\hat{e}, y) + \sum_{t=1}^{T} L_{BCE}(\hat{f}_t, f_t) \tag{10}$$

$$L_{LL}(\hat{e}, y) = \sum_{i=0}^{T} P(\phi(h_i))q_i^T(y_i|y_{i-1}) - \sum_{i=0}^{T} P(\phi(h_i))[q_i^T(y_i|y_{i-1})] \tag{11}$$

$$L_{BCE}(\hat{f}, f) = -\frac{1}{T} \sum_{t=0}^{T} f_t \log(\hat{f}_t) + (1 - f_t) \log(1 - \hat{f}_t) \tag{12}$$

where $L_{LL}(.)$ represents the negative log-likelihood and $L_{BCE}(.)$ denotes the binary cross-entropy, $y = [y_1, ..., y_T] \in \mathbb{R}^T$ is the actual room locations, and $f = [f_1, ..., f_T] \in \mathbb{R}^T$ is the binary value whether at time t the room is the referenced room or not. $P(y_i|y_{i-1})$ denotes the conditional probability, and $P(y_t|y_{t-1})$ denotes the transition matrix cost of having transitioned from y_{t-1} to y_t.

5 Experiments and Results

We compare our proposed network, MDCSA1,4,7 (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 localization. - A modified transformer encoder in combination with a CRF layer representing a model with the capability to capture global dependency and enforce dependencies in temporal aspects. - A state-of-the-art model for multimodal and multivariate time series with a transformer encoder to learn asymmetric correlations across modalities. - An alternative to the previous model, representing it with a GRN layer replacing the context aggregation layer and a CRF layer added as the last layer. - MDCSA1,4,7 4APS, as an ablation study, with our proposed network (i.e., MDCSA1,4,7) using 4 access points for the RSSI (instead of 10 access points) and accelerometer data (ACCL) as its input features. - MDCSA1,4,7 RSSI, as an ablation study, with our proposed network using only RSSI, without ACCL, as its input features. - MDCSA1,4,7 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 localization. For the modified transformer encoder, at each time step t, RSSI x_t^r and accelerometer 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 a specific optimizer in combination with a Look-Ahead algorithm 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. 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 localization 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 hypothesize 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 (see Section 3), and 2) the normalized 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 the MDCSA network and other approaches for room-level classification. For room-level classification, the MDCSA network outperforms other networks and RF with a minimum improvement of 1.3% for the F1-score over the second-best network in each cross-validation type, with the exception of the ALL-HC validation. The improvement is more significant in the 4m-HC and 4m-PD validations, when the training data are limited, with an average improvement of almost 9% for the F1-score over the alternative to the state-of-the-art model.

The LOO-HC and LOO-PD validations show that a model that has the ability to capture the temporal dynamics across time steps will perform better than a standard baseline technique such as a Random Forest. The modified transformer encoder and the state-of-the-art model perform better in those two validations due to their ability to capture asynchronous relations 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 a considerable amount of training data, the modified transformer encoder performs worst in these two validations. The state-of-the-art model performs quite well due to its ability to capture local context via LSTM for each modality. However, in general, its 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. The MDCSA network has all the capabilities that the state-of-the-art model has, with an improvement in suppressing the accelerometer modality when needed via the GRN layer embedded in DCSA. Suppressing the noisy modality seems to have a strong impact on maintaining the performance of the network when the training data is limited. This is validated by how the alternative to the state-of-the-art model (i.e., the state-of-the-art model with added GRN and CRF layers) outperforms the standard state-of-the-art model by an average of 2.2% for the F1-score in the 4m-HC and 4m-PD validations. It is further confirmed by MDCSA1,4,7 4APS against MDCSA1,4,7 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 MDCSA1,4,7 4APS RSSI model performed the best in the 4m-PD validation. However, the omission of accelerometer data affects the model's ability to differentiate 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 MDCSA1,4,7 4APS RSSI model has low performance in predicting the hallway compared to the full model of MDCSA1,4,7. As a consequence, the MDCSA1,4,7 4APS RSSI model cannot produce in-home gait speed features as

accurately, as shown in Table 3.

Room-to-room Transition and Medication Accuracy: We hypothesize 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 localization, focusing on the time a person is predicted to spend in a hallway between rooms. We show, in Table 3, that the mean transition duration for all transitions studied produced by the MDCSA1,4,7 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 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 the MDCSA network, a minimum of 15% improvement over the baseline demographic features can be obtained, with the biggest gain obtained in the 4m-PD validation data. In the 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 TENER failed to capture any transitions between the dining room and living room across all periods that have ground truths. MDCSA networks can provide medication state prediction and maintain their performance across all cross-validations thanks to the addition of Eq. 13 in the loss function.

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 generalizability 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 the transfer of this deep learning model to other varied home settings may introduce variations in localization 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 labeled) 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.

6 Conclusion

We have presented the MDCSA model, a new deep learning approach for indoor localization utilizing 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 state-of-the-art accuracy for indoor localization. The availability of accelerometer data does indeed enrich the RSSI features, which, in turn, improves the accuracy of indoor localization.

Accurate room localization using these data modalities has a wide range of potential applications within healthcare. This could include tracking of gait speed during rehabilitation from orthopedic surgery, monitoring wandering behavior 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.

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 MDCSA1,4,7

Training	Model	Room-Level Localisation		Medication State	
		Precision	F1-Score	F1-Score	AUROC
ALL-HC	RF	95.00	95.20	56.67 (17.32)	84.55 (12.06)
	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)
	MDCSA1,4,7 4APS	92.22	92.22	53.47 (12.63)	73.48 (6.18)
	MDCSA1,4,7 RSSI	94.70	94.90	51.14 (11.95)	68.33 (18.49)
	MDCSA1,4,7 4APS RSSI	93.30	93.10	*64.52* (11.44)	*81.84* (6.30)
	MDCSA1,4,7	**94.90**	**95.10**	**64.13** (6.05)	**80.95** (10.71)
	Demographic Features			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)	*89.82* (2.60)	55.34 (13.67)	73.77 (9.84)
	Alt DTML	90.52 (2.17)	89.71 (2.83)	49.56 (17.26)	73.26 (10.65)
	MDCSA1,4,7 4APS	88.01 (6.92)	88.08 (5.73)	59.52 (20.62)	74.35 (16.78)
	MDCSA1,4,7 RSSI	90.26 (2.43)	89.48 (3.47)	58.84 (23.08)	76.10 (10.84)
	MDCSA1,4,7 4APS RSSI	88.55 (6.67)	88.75 (5.50)	42.34 (13.11)	72.58 (6.77)
	MDCSA1,4,7	*91.39* (2.13)	**91.06** (2.62)	*55.50* (15.78)	**83.98** (13.45)
	Demographic Features			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)
	MDCSA1,4,7 4APS	86.44 (6.96)	85.93 (6.05)	47.26 (14.47)	72.62 (11.16)
	MDCSA1,4,7 RSSI	87.61 (6.64)	87.21 (5.44)	45.71 (17.85)	67.76 (10.73)
	MDCSA1,4,7 4APS RSSI	87.20 (7.17)	87.00 (6.12)	41.33 (17.72)	66.26 (12.11)
	MDCSA1,4,7	**88.04** (6.94)	**87.82** (6.01)	*49.99* (13.18)	*81.08* (8.46)
	Demographic Features			43.89 (14.43)	60.95 (25.16)
4m-HC	RF	74.27 (8.99)	69.87 (7.21)	50.47 (12.63)	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)
	MDCSA1,4,7 4APS	81.42 (6.95)	78.65 (7.59)	42.87 (17.34)	67.09 (7.42)
	MDCSA1,4,7 RSSI	81.69 (6.85)	77.12 (8.46)	49.95 (17.35)	69.71 (11.55)
	MDCSA1,4,7 4APS RSSI	82.80 (7.82)	79.37 (8.98)	43.57 (23.87)	65.46 (15.78)
	MDCSA1,4,7	**83.32** (6.65)	**80.24** (6.85)	*55.43* (10.48)	*78.24* (6.67)
	Demographic Features			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
	MDCSA1,4,7 4APS	81.02 (8.48)	*76.85* (10.94)	49.97 (7.80)	69.10 (7.64)
	MDCSA1,4,7 RSSI	77.47 (12.54)	73.99 (13.00)	41.79 (16.82)	67.37 (16.86)
	MDCSA1,4,7 4APS RSSI	*83.01* (6.42)	**79.77** (7.05)	41.18 (12.43)	63.16 (11.06)
	MDCSA1,4,7	**83.30** (6.73)	76.77 (13.19)	*48.61* (12.03)	*76.39* (12.23)
	Demographic Features			36.69 (18.15)	50.53 (15.60)

In naturalistic settings, in-home mobility can be measured through the use of indoor localization models. We have shown, using room transition duration results, that our PD cohort takes 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 localization outputs to detect the dopamine-related gait fluctuations in PD that impact patients' quality of life and are important in clinical decision-making. We have also demonstrated that our indoor localization system provides precise in-home gait speed features in PD with a minimal average offset to the ground

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
Ground Truth		18.71 (18.52)	14.65 (6.03)	10.64 (11.99)
ALL-HC	RF	16.18 (12.08)	14.58 (10.22)	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)	*10.84* (10.81)
	MDCSA	**17.70** (16.17)	**14.94** (9.71)	**10.76** (9.59)
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)	14.01 (8.08)	10.37 (12.44)
	MDCSA	**17.70** (17.42)	**14.34** (9.48)	**11.07** (13.60)
! 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	16.98 (15.15)	15.26 (8.85)	9.99 (13.03)
	MDCSA	**16.42** (14.04)	**14.48** (9.81)	**10.77** (14.18)
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	16.89 (18.07)	*14.68* (13.57)	*9.31* (15.70)
	MDCSA	**18.15** (19.12)	**15.32** (14.93)	**11.89** (17.55)
4m-PD	RF	11.52 (16.07)	8.73 (12.90)	N/A
	TENER	8.75 (14.89)	N/A	N/A
	Alt DTML	*14.75* (13.79)	**13.47** (17.66)	N/A
	MDCSA	**17.96** (19.17)	*14.74* (10.83)	**10.16** (14.03)

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.

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 various grants and institutions.

Statistical Significance Test

It could be argued that all the localization models compared in Table 1 might not be statistically different due to the fairly high standard deviation across all types of cross-validations, which is caused by the relatively small number of participants. In order to compare multiple models over cross-validation sets and show the statistical significance of our proposed model, we perform the Friedman test to first reject the null hypothesis. We then performed a pairwise statistical comparison: the Wilcoxon signed-rank test with Holm's alpha correction.

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
Kitchen-Living OFF	17.2 sec	Kitchen-Living ON	14.0 sec	75.0	2.824
Dining-Kitchen OFF	12.9 sec	Dining-Kitchen ON	9.2 sec	76.0	2.903
Dining-Living OFF	10.4 sec	Dining-Living ON	9.0 sec	64.0	1.961