Smart Health 30 (2023) 100430

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

DeepaMed: Deep learning-based medication adherence of Parkinson's disease using smartphone gait analysis

Hamza Abujrida *, Emmanuel Agu, Kaveh Pahlavan

Worcester Polytechnic Institute, Worcester, MA, USA

ARTICLE INFO

Keywords:
Parkinson's disease
Gait assessment
Smartphone sensing
Remote monitoring
Machine learning
Deep learning
Conventional neural networks
Medication adherence
Personalized medicine

ABSTRACT

Objectives: Parkinson's disease (PD) is a neurodegenerative chronic disorder with multiple motor and non-motor symptoms. As PD has no ultimate cure, physicians aim to delay PD complications, especially those that degrade the patient's quality of life such as motor symptoms and dyskinesia. Patients' lack of adherence to prescribed medication is a major challenge for physicians, especially for patients suffering from chronic conditions. The Centers for Disease Control and Prevention (CDC) estimates that medication non-adherence causes 30 to 50 percent of chronic disease treatment failures and 125,000 deaths per year in the USA (U.S. Foods and Drugs Administration (FDA) "Why You Need to Take Your Medications as Prescribed or Instructed" https://www.fda.gov/drugs/special-features/why-you-need-take-your-medications-prescribed-or-instructed. June 2021). In PD patients particularly, adherence varies between 10% and 67% (Straka et al., 2019Straka, Igor, et al. "Adherence to pharmacotherapy in patients with Parkinson's disease taking three and more daily doses of medication." Frontiers in neurology 10 (2019): 799).

Objective: The goal of this work is to remotely determine whether PD patients have taken their medication, by analyzing gait data gathered from their smartphone sensors. Using this approach, physicians can track the level of medication adherence of their PD patients.

Methodology: Using data from the mPower study (Bot et al., 2016), we selected 152 PD patients who recorded at least 3 walks before and 3 after taking medications and 304 healthy controls (HC) who recorded 3 walks at minimum. We extracted each subject's gait cycle from their accelerometer and gyroscope sensors data. The sensor data corresponding to gait cycles were fed to DeePaMed; a multilayer Conventional Neural Network (CNN), crafted for patches of gait strides. DeePaMed classified 30 s of a walk as either PD patient "On" vs. "Off" medication, or if the gait data belongs to an HC.

Results: Our DeePaMed model was able to discriminate PD patients on-vs off-medication and baseline HC walk with an accuracy of **98.2%**. The accuracy of our CNN model surpassed that of traditional Machine Learning methods by over **17%**. We also found that our model performed best with inputs containing a minimum of 10 full gait strides.

Conclusion: Medication non-adherence can be accurately predicted using smartphone sensing of the motor symptoms of PD, suggesting that PD patients' medication response and non-adherence can be monitored remotely via smartphone-based measures.

* Corresponding author.
E-mail addresses: hiabujrida@wpi.edu (H. Abujrida), emmanuel@wpi.edu (E. Agu), kaveh@wpi.edu (K. Pahlavan).

https://doi.org/10.1016/j.smhl.2023.100430
Received 1 August 2021; Received in revised form 23 February 2022; Accepted 13 September 2023
Available online 26 September 2023
2352-6483/© 2023 Elsevier Inc. All rights reserved.

1. Introduction and background

PD is the most common neurodegenerative movement disorder. Nearly one million people in North America are living with PD, this is expected to rise to 1.2 Million by 2030 (Marras et al., 2018). Symptoms of PD become noticeable when the brain cells that produce dopamine begin to die off. Dopamine, which works as a neurotransmitter plays an important role in the control and coordination of the human brain. With less dopamine, the human brain loses its ability to control movements, leading to tremor, stiffness, and muscle pain (Parkinsons' Foundation, 2021) as shown in Fig. 1). Today there is no therapeutic treatment for PD. However, there are drugs that mitigate its symptoms. L-dopa drug for instance is commonly used by many PD patients to enhance the brain's supply of dopamine and reduce the severity of PD symptoms when taken regularly.

When PD progresses to a severe state, patients lose their ability to walk, talk and can experience depression, fatigue, and memory loss, resulting in a lack of adherence to prescribed medication. Approximately only 67% of PD patients were found to be adherent to their PD medications (Kulkarni et al., 2008) with the level of non-adherence increasing as the daily dosage increases (Saini et al., 2009). Non-adherence to medication raises the mean annual medical cost to $15,826 compared to $9228 for PD patients who adhere to medication (71% increase). Non-adherence also leads to more hospitalization (2.3 vs. 1.8) and office visits (17.0 vs. 15.9) (Davis et al., 2010), which all increase the burden of PD complications on PD patients. There is a need for remote methods to monitor patients' medication adherence continuously and to remind them to take their medications on time. Sensor-rich smartphones that are now owned by over 85% of the US population (Pew Research Center, 2021) present an opportunity as a viable platform for remote monitoring of medication adherence.

In this work, we introduce **DeepaMed**, a data-driven neural networks methodology, that uses the Smartphone's built-in accelerometer and gyroscope sensors to evaluate PD patients' medication adherence and response to medication based on their gait. DeepaMed consists of a multilayer Conventional Neural Network (CNN), that analyzes smartphone gait data (Bot et al., 2016) collected by patients themselves or their caregivers while performing smartphone-based gait, speech, memory, and tapping tests in an uncontrolled home environment. *DeepaMed* autonomously extracted discriminative features from a raw triaxial accelerometer and gyroscope gait data. One-dimensional (1 d) filters were used to extract gait features from individual signal components, and multi-dimensional filters captured overall signal variations. As DeepaMed runs on patients' smartphones data, its simplicity and low cost facilitate monitoring of medication adherence of PD patients by having them walk at their homes before and after taking medication. While walking, the smartphone in the PD patients' pockets seamlessly records sensor data for offline *DeepaMed* analysis.

Envisioned Usage Scenario for DeePaMed: We envision that DeePaMed will be utilized in an active patient involvement scenario in which the patient's doctor requires them to attach a smartphone to their hip (or front pocket), click a start button, and walk in a straight line before and after taking their medication each day to provide a sample of their walk. We envision that the PD patient's doctor will receive DeePaMed's assessment on whether the patient is adhering to their medication directly and use it remotely to decide when changes in the patients' medication are needed, and intervene early, which would ultimately reduce the frequency of hospital visits. We considered an alternate passive data collection scenario in which DeepaMed opportunistically collects smartphone gait data as the patient moves around their home including when they walk over to get their medication. However, while more convenient, opportunistic, passive data collection encounters several issues including whether the patient happens to be sitting close to their medication on a given day and just reaches out to get it and thus not walk, not walking in a straight line and taking steps that are too few for meaningful assessment (e.g. less than 10-Strides (20 steps)). Consequently, we decided on our active patient involvement approach for this work and hence used the mPower dataset. We will explore a version of DeePaMed that requires minimal patient involvement in future work.

Desirable attributes of the DeepaMed smartphone-based medication adherence system include.

- **Low cost:** the wide adoption of smartphones minimizes the cost of our approach to patients and their caregivers and eliminates the need for dedicated hardware such as infrared cameras, motion sensors, or any purpose-built hardware.

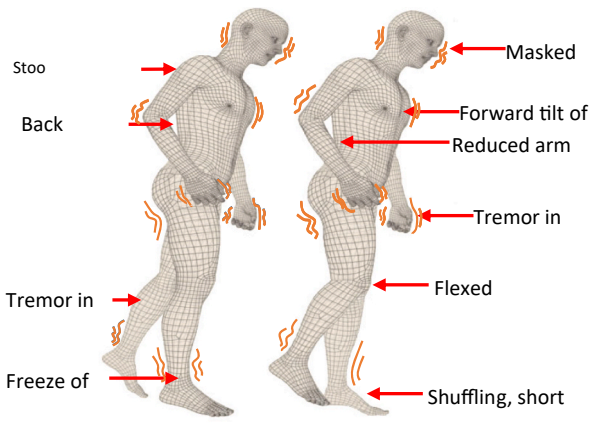

Fig. 1. Symptoms of Parkinson's disease.

- **Continuous Remote Monitoring**: Unlike methods such as MEMS and urine or blood test that monitor discrete medication-related events from drug bottles or blood samples respectively, our DeepaMed approach analyzes data corresponding to continuous windows of time, submitted by PD patients every time they walk before/after taking their medication.
- **Seamless data recording at the home setting:** walking is a simple activity that presents a minimal burden to PD patients for data collection.
- **Anomalies detection:** DeepaMed not only differentiates walks before/after medication but can also discriminate regular walk for HC. That can help isolate any mislabeled data and any walking data recorded by someone other than the patient.

Prior work has explored various methods for measuring medication non-adherence, these include the reliance on physician's judgment (Gilbert et al., 1980; Ose et al., 2012), patient self-reporting and management (Clarkesmith et al., 2017), tablet count (Grosset et al., 2006), tracking drug levels in urine and blood (Dutton et al., 1993), and using electronic devices (Grosset et al., 2005; Leopold et al., 2004; Pearce et al., 1998; Zhang et al., 2019). While tracking the drug level in blood and urine is accurate, it requires visiting a lab, which presents a burden for patients. Moreover, as the majority of the medications are taken outside of the clinic, physicians tend to rely more on self-reporting and self-management. As PD progresses, the loss of memory and depression dramatically increases the patients' non-adherence, which further increases as the patients' daily dose increases (Saini et al., 2009). Passive, continuous, and unobtrusive methods of monitoring medication adherence would increase patients' compliance and reduce missed episodes. Compared to traditional methods. Electronic measures and AI seem to be the most practical approach to track PD patients' medication intake.

A comparison of PD studies and electronic adherence measurement studies are presented in Table 1. A good number of studies (Grosset et al., 2005, 2009; Leopold et al., 2004) rely on Medication Event Monitoring System (MEMS) drug bottles, there are several commercial products, but the main function is to record the date/time of opening the medication bottle. There are multiple issues with this approach, First, sometimes the patient may open the medicine bottle but do not consume the medication. Secondly, MEMS approaches do not monitor the patients' symptoms after taking the medication to also monitor the patient's response. Infrared motion sensing presented in (Tucker et al., 2015) provides a better approach to compare the differences of gait before and after taking medication. But due to the complexity of the system, it is not feasible in the patient's home environment. The Smartphone sensors approach presented in (Zhang et al., 2019) is better suited for PD medication adherence problems, the processing of the walk time

Table 1
Summary of the most recent PD medication adherence studies.

	Study	Monitoring device	Duration of the study	Participants	Pros	Limitation
1	(Koesmahargyo et al., 2020) Koesmahargyo et al.	Machine learning, smartphone front-facing camera	varied	4182	Dosage reminders, uncontrolled environment.	Accuracy is low, can be improved with DL methods. Not focused on PD.
2	(Abujrida et al., 2020) Abujrida et al.	Smartphone sensors, Machine learning	24 weeks	456	Analyzed various aspects of PD gait. Uncontrolled environment.	Accuracy can be improved with DL methods. Not focused on Medication Adherence.
3	(Grosset et al., 2005) Grosset et al.	MEMS drug bottle	12 weeks	68	Dosage reminders, long-life battery, bottle opening sensing and count.	Patient misuse of the device, patients lost device, no guarantee for taking the dose except opening-closing the bottle, limited # participants.
4	(Grosset et al., 2009) Grosset et al.	MEMS drug bottle	4 weeks	112		
5	(Leopold et al., 2004) Leopold et al.	MEMS drug bottle	4 weeks	39		
6	(Tucker et al., 2015) Tucker et al.	line of motion sensing (Infrared Cam)	Duration of data collection session(s).	7	Examine the difference in symptoms between on/off medication states.	Camera angle calibration and sensor constraints make it difficult for self-recording at home, model overfit on individual participants, small #participants.
7	(Zhang et al., 2019) Zhang et al.	Smartphone sensors. (CNN with Spectrogram signal conversion).	12 weeks	247	Continuous monitoring, with smartphones, uncontrolled environment.	Unnecessarily conversion of time series by calculating spectrogram, adding complexity, no controls in the study, limiting the study to binary classification.
8	(Kan et al., 2018) Kan et al.	Smartphone app recording medication intake and ball-game app to measure tremor.	Duration of data collection	10	Combine both inputs from participants and sensors reading.	Small # participants who are not PD patients. ML/DL methods can extract informative features from sensors data.
9	(Lakshminarayana et al., 2017) Lakshminarayana et al.	Smartphone app recording adherence using questionnaires.	16 weeks	158	Uncontrolled environment. Improving adherence by using the PTA app.	depending on PD patients' inputs instead of measuring change by smartphone sensors.
10	(Kalantarian et al., 2015) Kalantarian et al.	Smartwatch detection of taking medication	Duration of data collection.	20	Automatic detection of medication intake.	The use of hand-crafted features can lead to overfitting. ML/DL methods can improve accuracy and reduce overfitting.

series and calculation of spectrograms from the raw sensor data resulted in unnecessary complexity and reduced the overall accuracy of the model. The Smartwatch approach (Kalantarian et al., 2015) seems to be promising. However, it needs to be validated on a larger dataset.

Our main contributions are.

- **Task-specific pipeline of steps for pre-processing of gait data:** Including noise filtering by calculating the moving average for the walking signal, and gravity/breathing effects cancelation by subtracting the signal mean and detecting walking periods from continuous gait data. Pre-processing steps facilitate gait feature extraction using a deep learning approach without any form of data conversion, or handcrafted features. As manual feature extraction is error-prone and tedious.
- **A custom 1D CNN for Parkinson' smartphone gait analyses:**

One dimension CNN has proven to be successful in multiple applications including speech recognition and electrocardiogram monitoring (Sawicki et al., 2021). To the best of our knowledge, this is the first work to innovatively explore 1 d CNN's to infer medication adherence of smartphone gait signals from Parkinson's patients.

- **Directly analyzing the time-series sensor data and eliminating the need for conversion to an image as in prior work**

Other studies on PD gait involve converting the data to images. We explored this method and confirmed that our DeePaMed exceeded the performance of models that work on image representations of the walk signal. Our 6-layers 1 d CNN model operates on raw sensor data that yielded savings on data-conversion processing costs. Each dimension of the signal is processed individually to extract features per dimension, while inter-dimension features are extracted using multi-dimensional filters. In rigorous evaluation, we demonstrate that our DeepaMed DL model consistently outperforms a comprehensive list of benchmark models that included the previous state-of-the-art, discriminating PD patients on-vs off-medication and baseline HC walk with an accuracy of **98.2%.** DeepaMed outperformed traditional Machine Learning methods by at least **17%.**

- **Model simplicity reduces overfitting**

By comparing DeePaMed to complex DL models, measured by the number of trainable parameters. We found that DeePaMed with **0.26** million parameters provides the best testing accuracy of **98%**. Models with more trainable parameters overfitted the training data and performed less well on the testing data, for instance, the CNN-LSTM model is the second-best performer with **2.9** million parameters and testing accuracy of **93%**.

The rest of the paper is as follows, Section 2 describes participants' selection criteria, data preprocessing and our DeepaMed methodology, in Section 3 we present our results for both the machine learning approach and DeepaMed. In section 4 we discuss our findings, and we conclude the paper in Section 4.

2. 2- Methodology

2.1. mPower dataset

To train the DeepaMed CNN model, we utilized data previously gathered in the mPower study (Bot et al., 2016). mPower was a clinical observational study on PD, managed by Sage Bionetworks. Data were collected using the patients' iPhones (Apple Inc., Cuppertino CA, USA). Data gathered via the app's interface included demographic and Unified Parkinson's Disease Rating Scale (UPDRS) surveys in addition to continuous sensor-based recordings from PD patients and HC. The mPower study had a large number of participants (n = 9520), the number of recorded activities including 35,410 walking, 78,887 tapping, and 8569 memory records, as shown in Table 2. All data were recorded in an uncontrolled home environment 3 times per day before, after, and at another time of the day. For HC, data were recorded at any time of the day.

In this work, we utilized the mPower walking activity, in addition to participant surveys and labels of the data. mPower app asked participants to place the smartphone vertically in their front pant pocket and walk for 30 s in a straight line as shown in Fig. 2, participants were then asked to stand for 30 s, turn and walk back for 30 s. We focused only on the first 30 s of the walk. During the walk, raw data gathered included the smartphone's gyroscope and accelerometer data, sampled at 100 Hz. The number of walking records per participant varied from a few to hundreds. 658 PD patients performed 24,001 walks, and 2165 H C performed 10,585 walks

Table 2
MPower Dataset.

Activity	Frequency	Participants	Records
Demographic Survey	Once	6805	6805
Walking	**Daily**	**3101**	**35,410**
Tapping	Daily	8003	78,887
Voice	Daily	5826	65,022
Memory	Daily	968	8569

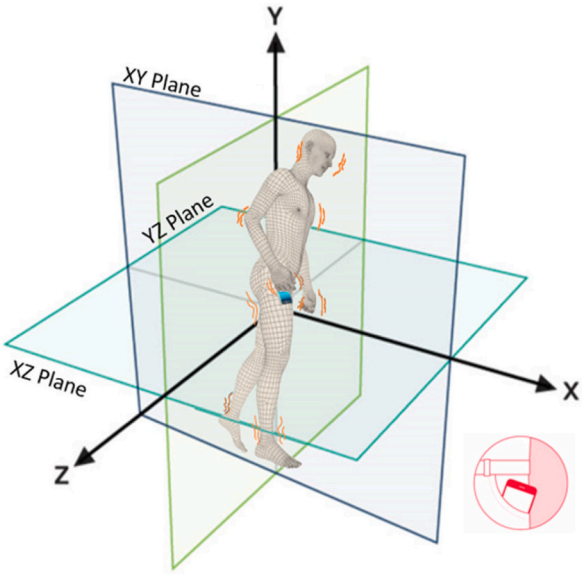

Fig. 2. Participant Walking with Smartphone in pocket.

in the first 6 months of the study. As our selection criterion, we selected only participants who recorded a minimum of 3 walks, resulting in a total number of 815 participants.

2.2. Selection of participants

The number of walks recorded by each participant varied significantly, as shown in Fig. 3. When selecting PD patients, we kept only the participant who has 3 walks minimum before taking medication and 3 walks minimum after taking medication, and dropped the rest of the participants. Since there is only one category of walks for HC, we kept any participant who had 3 walks minimum, Table 3 shows the demographic breakdown of the selected dataset. For any participant who had more than 3 walks of each category, we chose the walks recorded closer to the time of filling the demographic survey. Any walking record that had missing values such as sensor reading or key demographic values have been dropped.

The result is a dataset of 152 PD patients (each had 6 walking records), and 304 H C each had 3 walking records. Participants also filled UPDRS survey describing the severity of gait anomalies, a sample of the UPDRS questions is shown in Table 4, and their corresponding results for the selected sample are shown in Fig. 4.

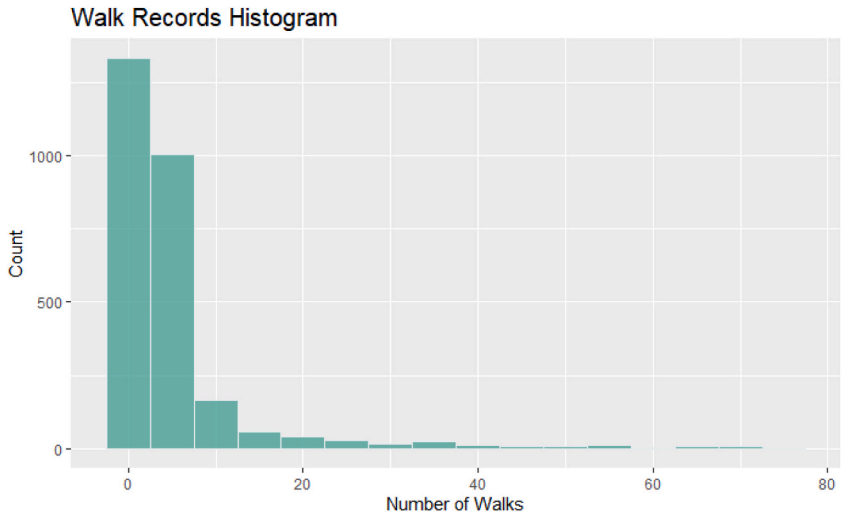

Fig. 3. Number of Walks per participant.

Table 3
Selected sample.

PD Diagnosis	Classes	Age Mean ± std	Male: Female	RACE	Participants	Records	Training Records	Testing Records
PD Patients	Before Medication	63.57 ± 8.09	89:63	"White/Caucasian" 74.5%	152	456	410	46
	After Medication			"Latino/Hispanic" 7.2%		456	410	46
				"Mixed" 2.7%				
				"South Asian" 2.7%				
				"Middle Eastern" 2.3%				
				"Black or African" 2.1%				
				"East Asian" 2.1%				
				"Other" 6.1%				
Healthy Controls	Another Time	40.14 ± 15.45	213:91	"White/Caucasian" 93.1%	304	912	820	92
				"Latino/Hispanic" 1.1%				
				"Mixed" 0.56%				
				"South Asian" 0.56%				
				"Middle Eastern" 0.56%				
				"Black or African" 1.1%				
				"East Asian" 1.1%				
				"Other" 1.7%				

Table 4
Sample of mds-updrs questions.

Question	Variable name	Variable details
Over the past week, have you usually had shaking or tremor?	MDS-UPDRS2.10	one of: {'Normal', 'Slight', 'Mild', 'Moderate', 'Severe'} mapping to {0, 1, 2, 3, 4}
Over the past week, have you usually had problems with balance and walking?	MDS-UPDRS2.12	one of: {'Normal', 'Slight', 'Mild', 'Moderate', 'Severe'} mapping to {0, 1, 2, 3, 4}

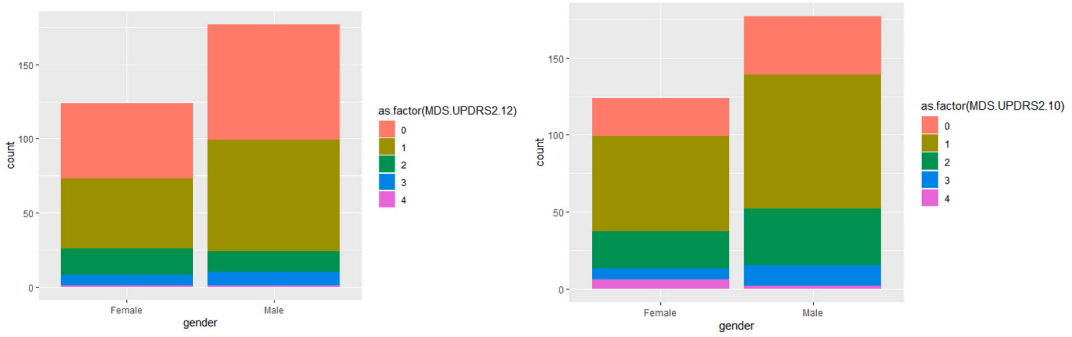

Fig. 4. UPDRS2.10 (Shaking) and UPDRS2.12 (walking imbalance) severity by gender.

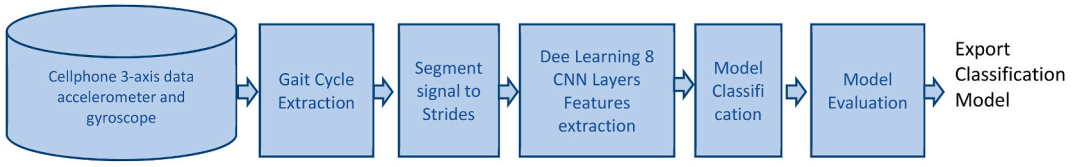

Figure 5. Flow Diagram for data collection, feature extraction, and classification.

2.3. Signal pre-processing

Fig. 5 presents our methodology. The first stage is data mining of the mPower gait records and surveys. We processed the mPower data and extracted the relevant survey information and the 3D gait signals. Data were then arranged in a useable format. The two main signals, shown in Fig. 6, were the smartphone's acceleration (from the accelerometer sensor) and rotation (from the gyroscope sensor). These are given below:

$$\alpha(i) = [\alpha x\ (i), \alpha y\ (i), \alpha z\ (i)]T\ (\text{in m / s2})$$

$$\omega(\text{i}) = [\omega \text{x}\ (\text{i}), \omega \text{y}\ (\text{i}), \omega \text{z}\ (\text{i})]\text{T}\ (\text{in deg / s})$$

where i denotes discrete-time, α indicates acceleration, and ω represents rotation.

To smooth the signal and remove sudden high-frequency variation, the moving average was calculated for the sensors' signals. A moving average is calculated and used to replace each value in the sequence with the average of 5 points around it, given by the following formula:

$$MA_{\alpha x} = \frac{1}{n}\sum_{i=0}^{n-1}(\alpha x_{(-i)})$$

Smartphones capture 3 dimensions of acceleration and rotation, shown in Fig (6). To combine the signals generated on the 3 axes, we calculated the signal magnitudes. The mean of the signal was then subtracted to eliminate gravity or any constant factors such as breathing. The resulting formula is given below:

$$MagNG_{\alpha} = \left(\sum_{i=1}^{n}(||\alpha_{(i)}||) - \overline{Mag_{\alpha}}\right) \tag{1}$$

$$MagNG_{\omega} = \left(\sum_{i=1}^{n}(||\omega_{(i)}||) - \overline{Mag_{\omega}}\right) \tag{2}$$

where *MagNGα* and *MagNGω* are the vector magnitudes of the acceleration and the rotation rates and $\overline{Mag_{\alpha}}$ $\overline{Mag_{\omega}}$ are the means of the

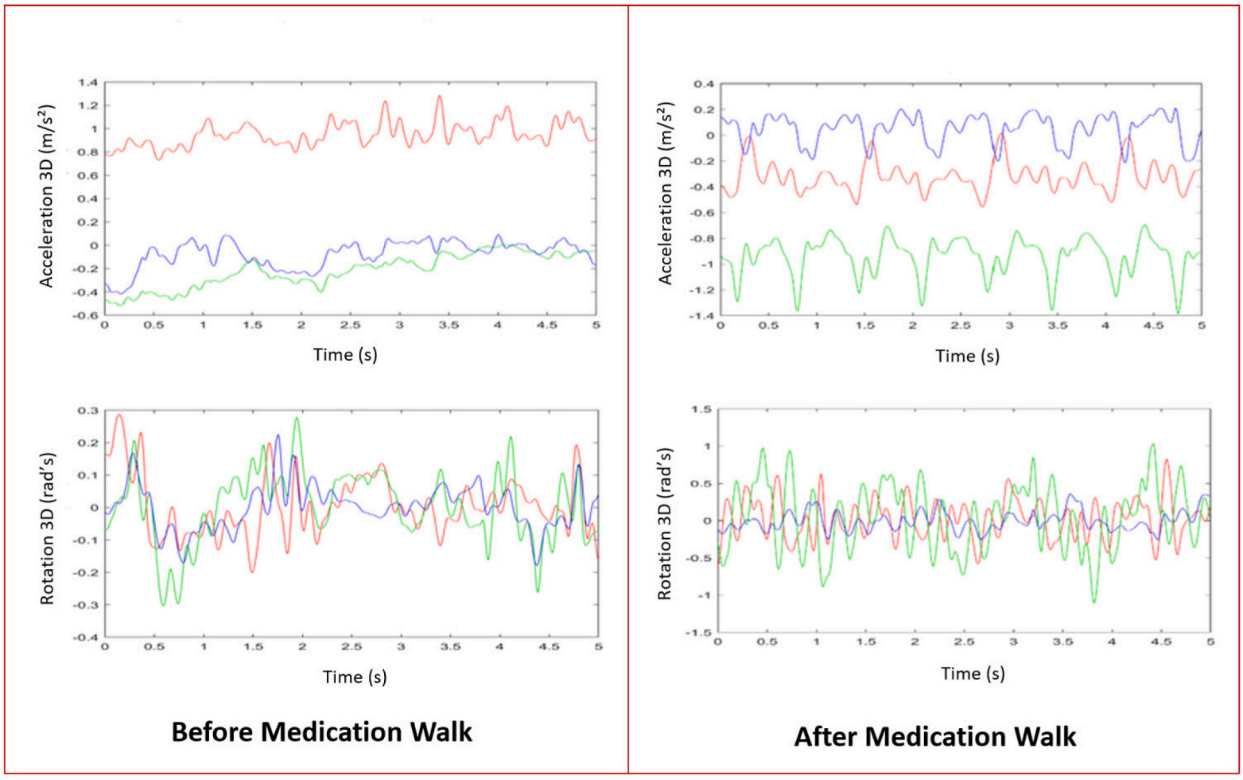

Fig. 6. Acceleration and rotation signals before and after taking medication.

acceleration and rotation.

One of the issues of the mPower dataset is that participant's compliance with the Smartphone app varied from one participant to another. Also, the correct positioning of the Smartphone in the participants' front pockets could not be verified. Consequently, the resulting walking records had gait and non-gait data. Therefore, as our goal was to analyze data for periods when the participants walked, it was essential to identify and extract gait cycles. Inspired by Cyclepro (Ma et al., 2019), we extracted gait segments from the time series and split the accelerometer and gyroscope signals into one-stride segments, according to the following steps; First, we calculated the signal magnitude according to equation (1). Then we generated multiple templates. A template, which represents one walking stride, is selected based on two consecutive local minima's of the signal magnitude. We selected multiple templates then chose the top template by comparing the template std (statistical standard deviation) with the std of the total walking signal magnitude. Cross-correlation is calculated between the top templates and the *MagNGα* signal to identify the repetitive walking pattern. Cross-correlation is calculated based on equation (3) below:

$$C(i) = \frac{\sum_{j=1}^{T} Mag_t(j) MagNG_\alpha(i+j)}{\sum_{j=1}^{T} Mag_t(j)^2} \tag{3}$$

Where Mag_t is the Template Signal Magnitude.

The output of equation (3) is normalized and used for detecting the walking strides. First, We slide a window of regular cadence size and find the local maxima's within this window. Second, within the window, we select only the highest maxima and filter the rest of the maxima's out. Finally, the final peak is selected using Otsu's method, this method finds the similarity of the peaks by minimizing the sum of inner variance within the peaks and separate the invalid peaks.

The resulting maxima's from the final step are used to split Gyroscope and Accelerometer signals into 1-Stride segments. Records of valid gait cycles are then sorted, correctly labeled, and saved for DL processing. The result was a dataset with gait data segmented into single-stride segments, each labeled with the time of the walk, before/after medication for PD patients, and at-another-time for HC.

2.4. Problem formulation

Given the gait signal from the inertia accelerometer and gyroscope sensors, and the signal magnitude calculated earlier, we expressed the input to the CNN network as:

$$X = [\alpha x(i), \alpha y(i), \alpha z(i), MagNG\alpha(i), \omega x(i), \omega y(i), \omega z(i), MagNG\omega(i)]T$$

where i denotes discrete-time, α indicates acceleration, ω represents rotation and T denotes the Stride length. The end goal was to classify the signal into one of the three classes, before medication, after medication, and at another time (HC). These classification classes can be expressed as $Y = (y_1, y_2, y_3)$, where the output $o_{(i)} = P(y_i | X)$, or the probability of $(X \in y_i)$. If the medication state of X is y, then we can express y as:

$$y = \arg \max_{y(i)} \{ o_{(i)} | \, 1 \le i \le 3 \}$$

As a result, we chose the class with the maximum probability from the SoftMax layer as the output of the network.

To infer medication adherence, we experimented with solving the classification problem by traditional Machine Learning (ML) algorithms and deep learning DeePaMed architecture, the two approaches will be discussed in the next two sections.

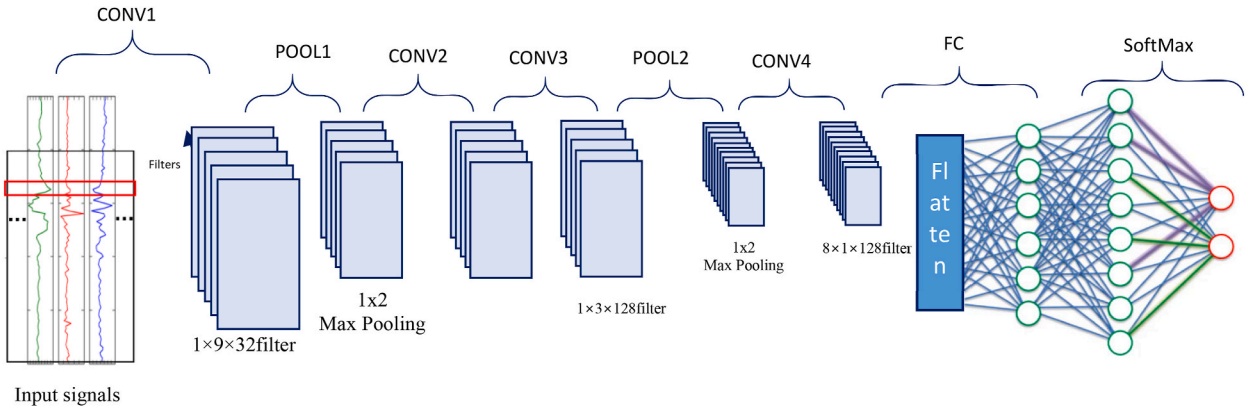

Fig. 7. Flow Diagram for data collection, feature extraction, and classification.

2.5. DeePaMed network architecture

Inspired by prior neural networks based work by Zou et al. (Zou et al., 2020), we created the CNN architecture shown in Fig. 7. The network consists of the feature extracting layers followed by fully-connected and classification layers, as shown in Table 5. Feature extraction consists of four convolutions and two pooling layers. While Convolution layers generate the features, pooling layers are added to downsize the features map.

The mPower study collected data in an uncontrolled environment. The participant had the freedom of positioning the phone in their pocket, which led to an inconsistent axis orientation from one participant to another. To consider this, axis independent features were extracted individually using one-dimensional filters, enabling the capture of the PD signal variation on that particular axis. The overall signal variation is captured by feeding the accelerometer and gyroscope signal magnitudes to the network. Following this mechanism, we ensured that every spike of PD gait fluctuation was captured and analyzed throughout the 30-sec walk, in chunks of *N*-Strides patches. Axis interdependent features were extracted at a later stage in the network, specifically at the Conv4 layer, using 8×1 filters. Convolutional and pooling layers are followed by fully connected and classification layers.

2.6. Machine learning baseline models

Data was prepared for ML by calculating features from the signal magnitudes MagNGα and MagNGω. Features calculated in the time, frequency, statistical, and wavelet domains using Matlab (Mathworks) are shown in Table 6. Our previous work (Abujrida et al., 2017, 2020) has details about our ML methodology and feature selection.

Time and statistical features were calculated directly from the MagNGα and MagNGω. Frequency domain features were calculated from the fast Fourier transform (FFT) and power spectral density (PSD). Frequency domain features were subsequently extracted for each walking segment record. Wavelet domain features were calculated from the discrete wavelet transform (DWT) of the MagNGx signal. After the calculation of features, a dataset is prepared and fed to different ML algorithms for medication adherence classification. The result of classification is discussed in the results section.

2.7. Encoding walking signal as images

The main reason behind the boom of DL models is the success of computer vision and image classification applications. To leverage the development brought by computer vision models, and to compare our DeePaMed model, we encoded the walking signal to an image format. There are multiple ways we considered to covert time-series to an image including; color-coded spectrogram (CCS), Markov Transition Field (MTF), and Gramian Angular Field (GAF). We choose GAF due to its impressive results for different time-series datasets (Wang & Tim, 2015).

Given a Time series $\{x_1, x_2, \ldots, x_n\}$, GAF can be calculated by first normalizing the signal using the equation:

$$\widetilde{x}_i = \frac{(x_i - \min(x_i)) + (x_i - \max(x_i))}{(\max(x_i) - \min(x_i))}$$

From the normalized time series and the time stamp values t_i we can calculate the polar representation using the equation:

$$\begin{cases} \varphi = \arccos(\widetilde{x}_i), -1 \le \widetilde{x}_i \le 1, \widetilde{x}_i \in \widetilde{X} \\ r = \dfrac{t_i}{N}, t_i \in \mathbb{N} \end{cases}$$

Where the time series consists of *N* timestamps t_i. The resulting map on the polar coordinates is unique and reversible for each time series. The polar representation also preserves the absolute temporal relations, so that we will not lose these temporal patterns when converting the walking signal. To capture the correlation between different time intervals the gram matrix is calculated according to the following equation:

$$G = \begin{bmatrix} \cos(\varphi_1 + \varphi_1) & \cdots & \cos(\varphi_1 + \varphi_n) \\ \vdots & \ddots & \vdots \\ \cos(\varphi_n + \varphi_1) & \cdots & \cos(\varphi_n + \varphi_n) \end{bmatrix}$$

Fig. 8 below shows the GAF conversion of the 10-Strides walking signal and its equivalent polar representation. All the selected

Table 5
DeepaMed CNN structure.

Layer Name	Filter size	Filters	Feature Map
Conv1	1×9	32	$8 \times 90 \times 32$
Pool1	1×2	N/A	$8 \times 45 \times 32$
Conv2	1×3	64	$8 \times 45 \times 64$
Conv3	1×3	128	$8 \times 45 \times 128$
Pool2	1×2	N/A	$8 \times 22 \times 128$
Conv4	8×1	128	$1 \times 22 \times 128$

Table 6
Features definitions and descriptions.

S·N	Feature	Feature definition	Description
Time Domain Features and their use cases for Gait analysis			
1	*Number of Steps*	*Local Peaks*	*The number of steps taken in a given time interval*
2	*Average Step Time*	$\frac{time}{\#Steps}$	*The average time elapsed for each step*
3	*Average Cadence*	$\frac{\#steps}{time}$	*The ratio of the total number of steps to the total time*
4	*Skewness*[a]	$\frac{\frac{1}{n}\sum (x_i-\mu_x)^3}{\left[\frac{1}{n}\sum (x_i-\mu_x)^2\right]^{3/2}}$	*Asymmetry of the signal distribution*
5	*Coefficient of Variation of Step Time*	$\frac{\sqrt{\frac{1}{n}\sum (interval_i-\mu_{interval})^2}}{\mu_{interval}}$	*The within-subject standard deviation of the stride interval divided by the mean stride interval*
6	*Average Step Length*	$\frac{0.084}{averageStepTime}+1.89$	*The average distance covered by each step*
7	*Gait Velocity*	$\frac{\left(\frac{0.084}{averageStepTime}+1.89\right)}{averageStepTime}$	*The ratio of the total distance covered by the total time*
8	*Minimum and Maximum Difference*[a]	$\max(x_i)-\min(x_i)$	*Global maximum of one step minus global minimum of one step averaged over all steps of one subject*
9	*Root Mean Square*[a]	$\sqrt{\frac{1}{n}\sum {x_i}^2}$	*Root Mean Square or quadratic mean is a statistical measure*
10	*Entropy Rate*[a]	$-\sum possibility_{unique\ freq} \times log_2(possibility_{unique\ freq})$	*The uncertainty measure of the signal, and the regularity of a signal when its anticipated that consecutive data points are related*
11	*Sway Area*[b] *X.Y, Y.Z, X.Z*	$\pi(AB)$	*Area of an ellipse that encloses the 95 percent confidence interval of all observed gyroscope points in the XY, YZ, and XZ planes. (A and B are the lengths of the semi-major and semi-minor axes of the ellipse)*
Frequency Domain Features and their use cases for Gait analysis			
12	*Harmonic Ratio*[a]	$\frac{\sum_{i=1,3,5,\ldots} V_i}{\sum_{j=2,4,6,\ldots} V_j}$	*Harmonic Ratio quantifies the harmonic composition of the accelerations for a given stride* via *DFT*
13	*Average Power*[a]	$\frac{total\ power\ of\ the\ signal}{bandwidth\ of\ the\ signal}$	*The mean of the total power underneath the curve of the PSD estimate for a signal*
14	*The ratio of Spectral Peak*[a] *(with Welch, FFT, DCT)*	$\frac{\max(power_{freq})}{mean(power_{freq})}$	*The ratio of the energies of low and high-frequency bands*
15	*Signal Noise Ratio*[a]	$\frac{power_{signal}}{power_{noise}}$	*Power of the whole signal over the power of its computed noise*
16	*The energy in Band 0.5 to* 3 Hz[a]	$\int_{0.5}^{3} psd_f df$	*The energy in a frequency band describes components of distinct frequencies in the signal, and the frequency range is recommended as 0.5* Hz *to* 3 Hz
17	*Windowed Energy in Band 0.5 to* 3 Hz[a]	$\int_{0.5}^{3} windowed\ psd_f df$	*The energy in the frequency band of 5-s windows with an overlap of 2.5 s; windows from the complete signal sequence are averaged*
18	*Peak Frequency*[a]	$\max(power_f)$	*The maximum spectral power*
19	*Spectral Centroid*[a]	$\frac{\sum f \times {power_f}^2}{\sum {power_f}^2}$	*The frequency that divides the spectral power distribution into two equal parts*
20	*Bandwidth*[a]	$\frac{\sum (f-spectralCentroid)^2 \times {power_f}^2}{\sum {power_f}^2}$	*The difference between the uppermost and lowermost frequencies/range of frequencies in the signal (Weighted Average)*
Wavelet Domain Features and their use cases for Gait analysis			
21	*Wavelet Bandwidth*[a]	$\frac{cA' * cA}{(cA' * cA + cD' * cD)}$	*The relative energy contribution in a time-frequency band*
22	*Wavelet Entropy Rate*[a]	$-\sum possibility_{unique\ freq} \times log_2(possibility_{unique\ freq})$	*Wavelet entropy represents the signal disorder in the time-frequency domain*
Statistical Features and their use cases for Gait analysis			
23	*Zeroth-Lag Cross-Correlation Coefficient*[a]	$\frac{\sum (x_i-\mu_x)(y_i-\mu_y)}{\sqrt{\sum (x_i-\mu_x)^2 \sum (y_i-\mu_y)^2}}$	*The agreement or similarity between 2 directional acceleration signals*
24	*Kurtosis*[a]	$\frac{\frac{1}{n}\sum (x_i-\mu_x)^4}{\left[\frac{1}{n}\sum (x_i-\mu_x)^2\right]^2}$	*The extent to which the distribution of signal amplitudes lies predominantly on the left of the mean amplitude*
25	*Standard Deviation*[a]	$\sqrt{\frac{1}{n}\sum (x_i-\mu_x)^2}$	*Measure for signal spreading, defined as the square of standard deviation*
Lifestyle Features and their use cases for Gait analysis			
26	*GELTQ.1a*	*The number of times the participant performed strenuous exercise for more than 15 min over the past week.*	
27	*GELTQ.1b*	*The number of times the participant performed moderate exercise for more than 15 min over the past week.*	
28	*GELTQ.1c*	*The number of times the participant performed minimal effort exercise for more than 15 min over the past week.*	
29	*Smoked*	*Ever smoked? (True/false question).*	
30	*Age*	*Participant's age (a number in years).*	

(*continued on next page*)

Table 6 (*continued*)

Time Domain Features and their use cases for Gait analysis			
S·N	Feature	Feature definition	Description
31	*Years. Smoking*	*Number of years participant has smoked (a number)*	
32	*Packs.*per.*day*	*Number of packets smoked* per *day.*	
32	*Gender*	*Female/male*	

Table 6 shows the definition of features, where x_i refers to a data sequence from which the feature is to be calculated; μ_x refers to the average of all x_i; $interval_i$ refers to a sequence of stride intervals; $\mu_{interval}$ refers to the average of all $interval_i$; V_i refers to the amplitude of odd ordered harmonic frequency in the frequency domain, and V_j refers to that of even-ordered harmonic frequency ; *dwt* refers to discrete wavelet transform (approximation coefficients vector *cA* and detail coefficients vector *cD*).

[a] Features that are calculated for both accelerometer and gyroscope.

[b] Features that are calculated for gyroscope only.

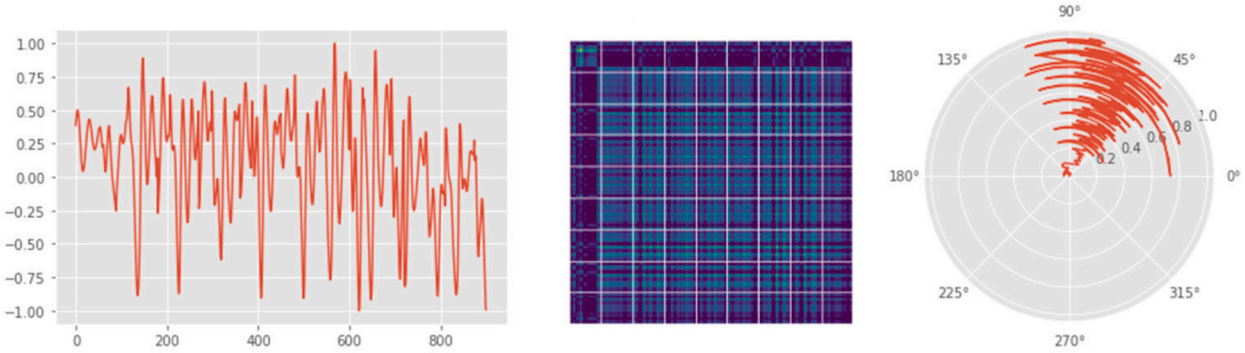

Fig. 8. GAF for a walking signal and the corresponding polar representation.

participants' walks were encoded to images using the above GAF encoder and made available for further processing by DL image-classification models.

2.8. Comparison to top pre-trained models

Pre-trained models have become widely available and utilized in DL applications. Today they are included in libraries such Keras, Tensorflow, Matlab, and PyTorch. Examples of such pre-trained models include ResNet50, Inception, SqueezeNet, MobileNet, and EfficientNet. Those models are trained on huge datasets and can be used to significantly improve the performance of medical imaging diagnosis (Deepak et al., 2019; Deniz et al., 2018; Hussein et al., 2019).

To evaluate our DeePaMed model, we compared DeePaMed performance to the state-of-the-art pre-trained DL models mentioned above, in terms of model architecture (depth and parameters size), and performance (Training, and Testing Accuracy). We adopted the method of transfer learning (TL) shown in Fig. 9 and followed the procedure outlined below.

Step1: We converted all the Accelerometer and Gyroscope signals to an image representation using the GAF method.
Step2: Image data were randomly split to 90% training set and the remaining 10% as a test set.
Step3: The network architecture was adapted to our medication adherence prediction task by replacing the last CONV, FC, and classification layers to match our 3 classes of data.
Step4: The network was trained.
Step5: The network was evaluated on the test set and the performance metrics reported.

3. Results

Based on 760 walking records from 456 participants, we conducted a supervised classification experiment by training the ML

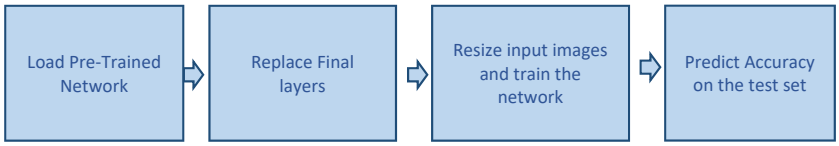

Fig. 9. Illustration of the mechanism for TL.

algorithms and our DeepaMed network on 90% of the data and testing on the remaining 10%. To evaluate modules' performance, the following performance metrics, that are widely used in mobile health applications, have been adopted.

- Accuracy: The percentage of correctly predicted samples. Accuracy formulated is

$$= \frac{\text{Correctly Classified Samples}}{\text{Total Testing Samples}}$$

- Precision: Precision is the fraction of the correctly predicted samples, to the total positive predicted samples.

  Precision measures the robustness of the tested module against false positives.

$$= \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

- Recall: Also called the Sensitivity, is the fraction of the correctly predicted positive samples to the total samples in the classification class. Recall gives an idea of the classification misses.

$$= \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

  F1-score: is the weighted average of Precision and Recall, it takes false positives and false negatives into account

$$= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

3.1. Performance of baseline machine learning models on medication inference

In this work, we experimented with various ML algorithms families, including Decision Trees (DT), Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Logistic Regression (LR), Naïve Bayes, and Ensemble Classifiers (EC) including Random Forest (RF) and bagged trees. The results including performance metrics are presented in Table 7. Random Forest is the best classifier for distinguishing between medication states, with an accuracy of 83.4%, precision of 83.0%, and F1-Score of 85%.

Based on the results seen in Table 7. ML could not achieve an acceptable performance when classifying various walks before/after taking medication and at another time (HC), mainly because the handcrafted features were not able to linearly or non-linearly discriminate between the different classes.

3.2. DeePaMed multiple strides

To overcome the inadequate performance with ML algorithms and due to the demonstrated performance of neural networks on various gait analysis and classification problems, we explored a deep learning approach (DeepaMed) to extract various abstractions of gait features over multiple convolutions and pooling layers. We experimented with varying the input data sizes, consisting of a different number of strides, The classification accuracy with 2 strides segments was 86.6% as presented in Table 8. This result exceeds the performance of traditional ML algorithms with a reasonable margin, However, PD step-to-step signal variations due to the effects of tremor, shuffling made it hard for DeePaMed to distinguish those variations from the signal calm that happens after taking medication, this led us to experiment different segment sizes. The results as can be seen in Table 8 shows performance increase as we fed input containing more strides to the network. With more strides fed into the network, DeepaMed could learn stride-to-stride variations better, an attribute that distinguishes PD gait from regular HC walk. Since taking the medication calms the patient and reduced PD gait anomalies, stride-to-stride variations also reduce, improving our results for discriminating the walk before and after taking medication. This can also be seen in the walk segment presented in Fig. 6.

It can be seen from Table 8 that we got the best results with 10-Strides segments. We computed the ROC curve for this best-performing model. ROC curve is a good measure of how the model distinguishes between classes, by plotting FPR on the x-axis

Table 7
Comparison of ML and DeepaMed algorithms.

Medication Inference:	Precision	Accuracy	Recall	F1-Score
Random Forest	83.0%	83.4%	87.1%	85.0%
Bagged Trees	80.1%	80.2%	78.0%	79.0%
Cubic SVM	74.9%	74.9%	75.1%	75.0%
Weighted KNN	78.5%	78.6%	77.8%	78.1%
Logistic Regression	55.3%	55.6%	54.9%	55.1%
Fine Tree	74.7%	74.9%	73.8%	74.24%
Naïve Bayes	61.75%	52.9%	60.0%	60.9%

Table 8
DeepaMed performance with various numbers of strides.

Stride overlaps	Strides	Accuracy	Precision	Recall	F1 Score
1	2	86.6%	86.3%	85.6%	85.6%
2	3	88.6%	86.6%	88.3%	87.6%
3	4	91.9%	90.4%	91.0%	93.34%
4	5	93.7%	92.4%	93.3%	92.7%
5	6	93.5%	92.0%	92.0%	92.0%
6	7	96.5%	95.6%	95.6%	96.0%
7	8	97.48	96.7%	97.0%	97.0%
8	9	97.1%	96.3%	96.3%	96.3%
9	10	**98.2%**	**97.7%**	**97.7%**	**98.0%**

against TPR on the y-axis. As expected FPR is low for high TPR particularly for this model, which leads to an area under ROC curve of 0.97/1, shown in Fig. 10.

3.3. Impact of network parameters tuning

We explored improving performance further by tuning DeepaMed's parameters. Dropout is a technique to reduce overfitting in Neural Networks. Dropout works by randomly dropping out some neurons' output during the training phase, at a rate specified by the dropout rate. The goal is to generalize the model and preventing complex coadaptation. When studying the impact of stride length, we fixed the dropout rate to 50% But we noticed that training accuracy surpassed the testing accuracy regardless of the number of strides used. Here we fixed the number of strides to 10 and varied the dropout rate. Results are presented in Table 9 below. We notice that the difference between Training and Testing accuracy reached the lowest level when using 80% as the training dropout rate. The dropout rate was always kept at 0% (or no dropout) during the testing phase.

As the dropout rate was increased to 0.8, the testing accuracy got very close to the training accuracy, and overfitting decreased dramatically. Fig. 9 shows how the performance of the network improved while training the network, we measured the network loss and accuracy for both the training and testing sets, as Fig. 11 shows the accuracy surpassed 90% after 20 epochs and very small differences between the training and test plots demonstrates that our models were not overfitting.

3.4. Compare DeePaMed to image-based TL models

Using the method of TL shown in Fig. 12, we investigated different DL models on the challenging mPower dataset. Detailed results are presented in Table 10. The table shows the complexity of the model in terms of parameters size, the memory needed, and the depth of the network. Table 10 also shows the training and test accuracy in comparison to our DeePaMed model. The TL models with the best performance were Inseption-V3, and ResNet-50 which agrees with the results of PDMove (Zhang et al., 2019). Inception-V3 and ResNet-50 achieved a testing accuracy of 87% and 81% respectively. However, by examining the training and test accuracies, we can observe that overfitting occurs in all of the models. All the models performed well on the training set, as shown in Fig(13). But once validated on the unseen testing set TL models misclassify many of the walk images to the wrong class. This observation will be further discussed in the discussion section.

3.5. Comparison of DeePaMed to LSTM parallel LSTM and CNN-LSTM models

Recently deep learning methods such as LSTM and variations of CNN-LSTM have achieved promising results in the problem of gait authentication and human activity tracking (Sawicki et al., 2021; Zeng et al., 2021). We investigated the performance of the Vanilla LSTM, parallel LSTM, and CNN-LSTM models and compared the performance to our DeePaMed model. Results are shown in Table 11.

As shown in Fig (14), the CNN-LSTM model uses CNN layers as automatic feature extractors and the LSTM for time series prediction. CNN-LSTM has previously been used for activity recognition and image/video description (Donahue et al., 2015), as well as gait analysis (Anwary et al., 2021). The model extract features using CNN Convolution layers. The extracted features are then flattened and provided as input to the LSTM network to extract features before the final classification is done using the softmax layer. The results of CNN-LSTM are the best of all the models that we studied as shown in Fig (16), we believe this happens because it combines both the differentiation power of 1D Conv layers and the time series prediction of the LSTM network. However, the CNN-LSTM network also overfits the training data, while the accuracy on the testing set did not exceed 93%.

A parallel LSTM architecture has been explored for Human Activity Recognition (HAR) and led to better performance than shallow ML algorithms with a performance comparable to CNN models (Yu et al., 2018). The parallel LSTM model we utilized is presented in Fig (15). The architecture consists of 6 parallel LSTM nodes, where each part of the 3D Accelerometer and Gyroscope signals is fed to one LSTM layer. The output of all LSTM layers is combined using a concatenation layer, followed by an FC layer and softmax for the overall prediction calculation. One issue with this network is that it learns slowly, and it is performance is lower than DeePaMed, as shown in Fig. 16.

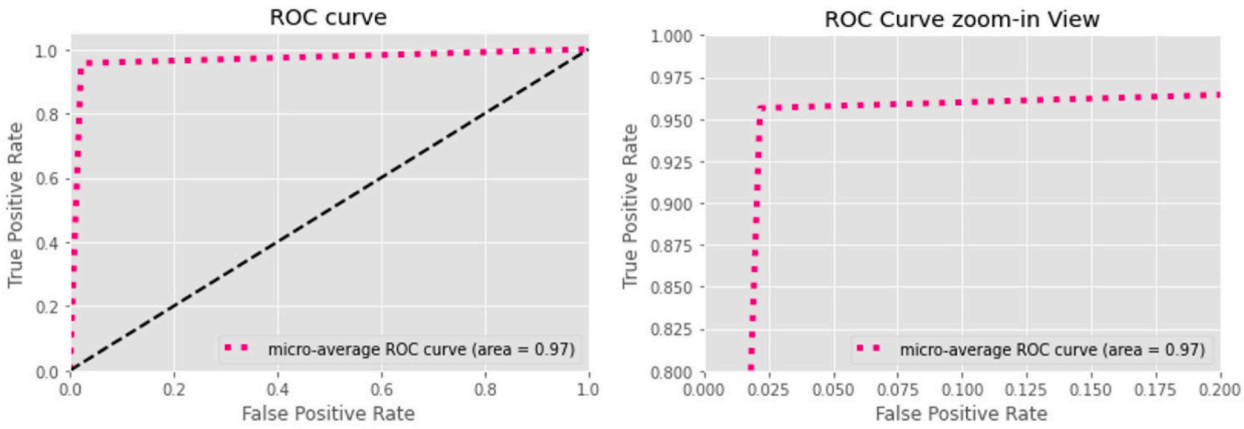

Fig. 10. ROC curve for the best performing 10-Strides model.

Table 9
DeepaMed performance with a variable dropout rate.

	Dropout keep_probe	Dropout rate = 1-keep_prob	Train Accuracy	Test Accuracy	Notes
1	1	0	100%	88.1%	10 strides and 50 epocs
2	0.9	0.1	100%	90.9%	
3	0.8	0.2	100%	95.4%	
4	0.7	0.3	100%	97.7%	
5	0.6	0.4	100%	97.3%	
6	0.5	0.5	100%	98.2%	
7	0.4	0.6	100%	97.9%	
8	0.3	0.7	100%	98.9%	
9	**0.2**	**0.8**	**99.6%**	**99.2%**	
10	0.1	0.9	98.4%	99.1%	

Fig. 11. 10-Strides model Performance.

4. Discussion

As PD progresses to a severe state, patients tend to miss medication doses for various reasons, including depression and memory loss. The level of non-adherence increases as the daily dosage increase (Saini et al., 2009). Patients also start developing resistance to taking medication when they notice that medications are not effective. A study found that 20 to 40 out of 100 people with Parkinson's noticed that the drugs are becoming less effective after five years of treatment (InformedHealth.org, 2019). By comparing walks before and after medication. DeepaMed can continuously detect and alert both patient and caregiver of nonadherence and medication ineffectiveness, resulting in a reduction of clinic visits and providing more data points for the treatment process.

Due to the complexity of PD, no one treatment that fits all patients. Instead of standardizing the treatment, and to embrace the concept of "Personalized Medicine (PM)", treatment needs to be prescribed based on the susceptibility of specific subtypes of PD to side effects with consideration of lifestyle, genetic framework, personality, and pharmacogenetics. (Titova et al., 2017). In PM, Physicians adjust the treatment plan based on occasional hospital visits and input from patients. Alternately, DeepaMed can provide continuous

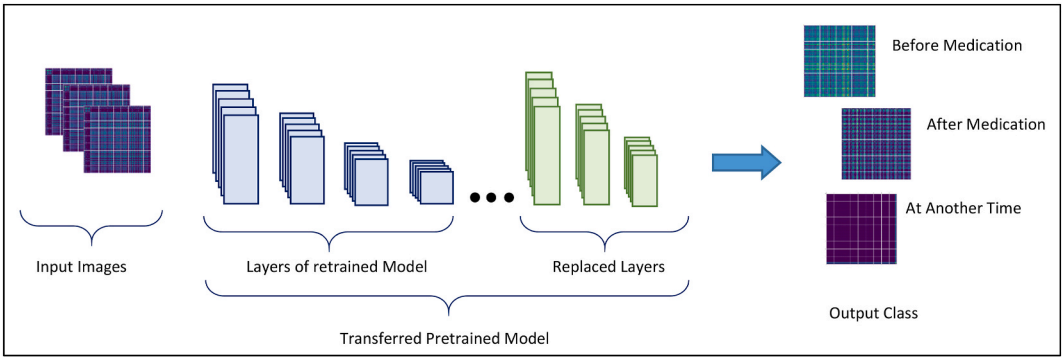

Fig. 12. TL using the Pre-trained model.

Table 10
Comparison Pretrained TL models and DeepaMed.

Model	Depth (layers)	Parameters (millions)	size (MB)	Training Accuracy	Testing Accuracy
DeePaMed	7	0.26	0.98	99%	98%
ResNet-50	50	25.6	96	99%	81%
EfficientNet	82	5.3	20	99%	73%
Inception-V3	48	23.9	89	99%	86%
ShuffleNet	50	1.4	5.4	99%	70%

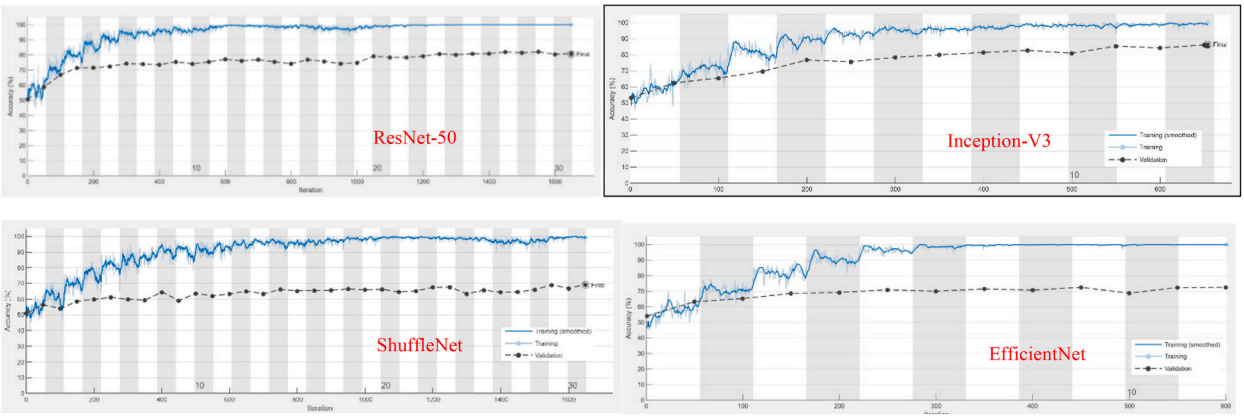

Fig. 13. Results of Pretrained TL models.

Table 11
Comparison of LSTM models and DeepaMed.

Model	Depth (layers)	Parameters (millions)	size (MB)	Training Accuracy	Testing Accuracy
DeePaMed	7	0.26	0.98	99%	98%
LSTM	3	0.08	0.31	61%	56%
Parallel LSTM	5	38.4	151.4	99%	79%
CNN-LSTM	9	2.9	11.35	99%	93%

inputs from the patients, so that doctors can monitor and adjust the individual treatment to fit each patient's unique case.

Some commercial mobile apps compromise security by accessing smartphone sensors. These apps can record the microphone, monitor location, and take photos, all without the user's permission. To address these privacy concerns, the mPower app gives full control of the sensors' recording to the user and limits the gait data collection to 30 s before and 30 s after taking medication. Furthermore, our signal processing includes smoothing of the signal, subtracting signal mean, and calculation of signal magnitude, all will help hide the personal signal variation. Also, our experiments are conducted at the population level using de-identified patient data. The goal is to identify and characterize signal changes that reflect the relatively calm gait that occurs after the medication is

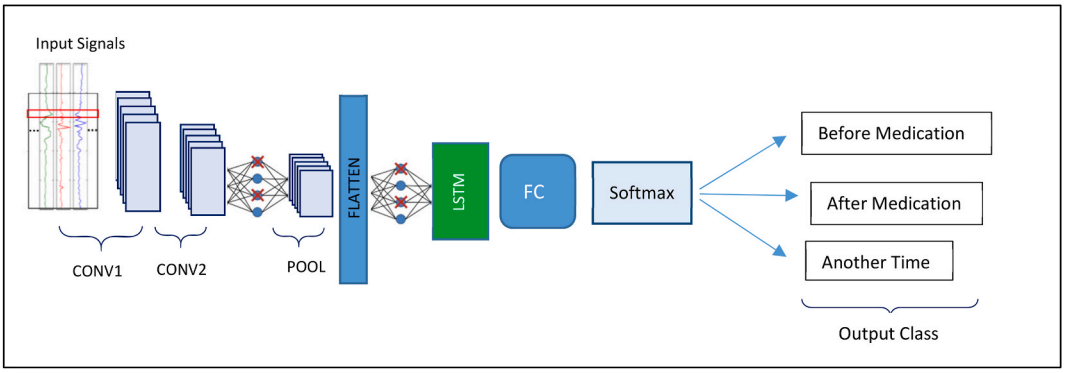

Fig. 14. CNN-LSTM Architecture used.

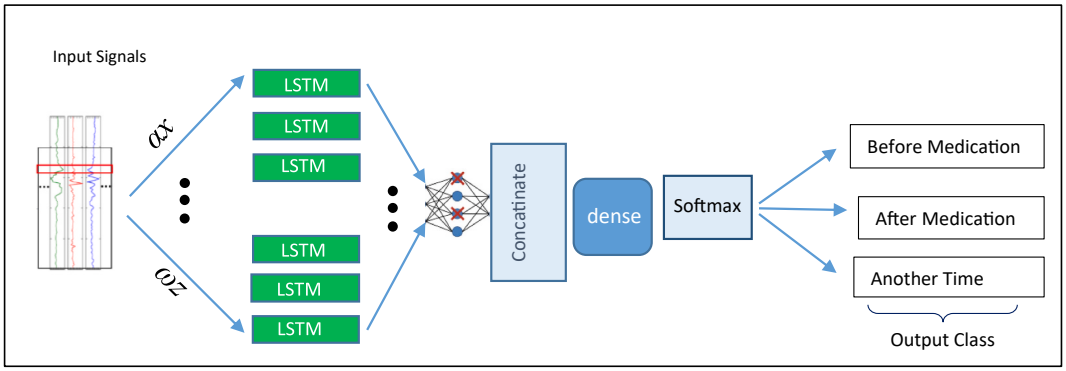

Fig. 15. Parallel LSTM Architecture used.

taken. In comparison to segment-level classification that we explore, subject-level classification will require more data per subject. Consequently, an algorithm with more depth will be needed to classify the subject unique walking signal.

An overfitted model is a statistical model that contains more parameters than can be justified by the data (Everitt & Anders, 2010). When such a model predicts a trend in very noisy data, like in our case the mPower dataset, it tends to overfit the training set, and perform less well on the test set. In Tables 10–11 we showed how DeePaMed overpass the performance of all the pre-trained TL models and LSTM models, that is because DeePaMed has the right complexity and depth to predict the medication adherence trends in the mPower data.

There are some limitations for DeepaMed that we are planning to work on in the future. Based on Table 8, we were able to reach acceptable accuracy with input containing 7 gait strides or more. We will investigate improving the model to provide higher accuracy with a minimal number of walking strides. Also, there are other associated comorbidities that may affect gait, including dementia and Hip and knee osteoarthritis, in the future, we will break the dataset based on the associated comorbidities and study each subset of data individually. And Finally due to the nature of the mPower study, the mPower dataset had unbalanced data bins. We had more HC with acceptable walking records than PD patients, resulting in more observation for HC's (Not-Taking-Medication) class and subsequently adding more weight and increasing the frequency bias for this HC class compared to PD patient classes. In the future, we plan to use a better-balanced dataset.

5. Conclusion

PD Patients' lack of adherence to prescribed medication is a major challenge that PD physicians face. The existing methods of remotely measuring medication adherence are inconvenient, not continuous or passive. In this work, we introduced DeepaMed, a deep learning smartphone approach that can differentiate PD walk before and after medication for PD patients. DeepaMed can also distinguish a non-PD walk by HC's based on the smartphone's acceleration and rotation signals. After experimenting DeepaMed on 452 participants, we prove that medication non-adherence can be accurately predicted using smartphone sensing of the motor symptoms of PD gait. Our DeePaMed model was able to discriminate PD patients on-vs off-medication and baseline HC walk with an accuracy of **98.2%**. The accuracy of our CNN model surpassed that of traditional Machine Learning methods by at least **17%**. We also found that the CNN model performed best with inputs containing a minimum of 10 full gait strides and a 0.8 dropout rate. Our findings suggest that remote monitoring of medication adherence using the smartphone is feasible. In future work, we plan to analyze PD gait using DL methods and compare the findings to the previous work that was based on ML and handcrafted features.

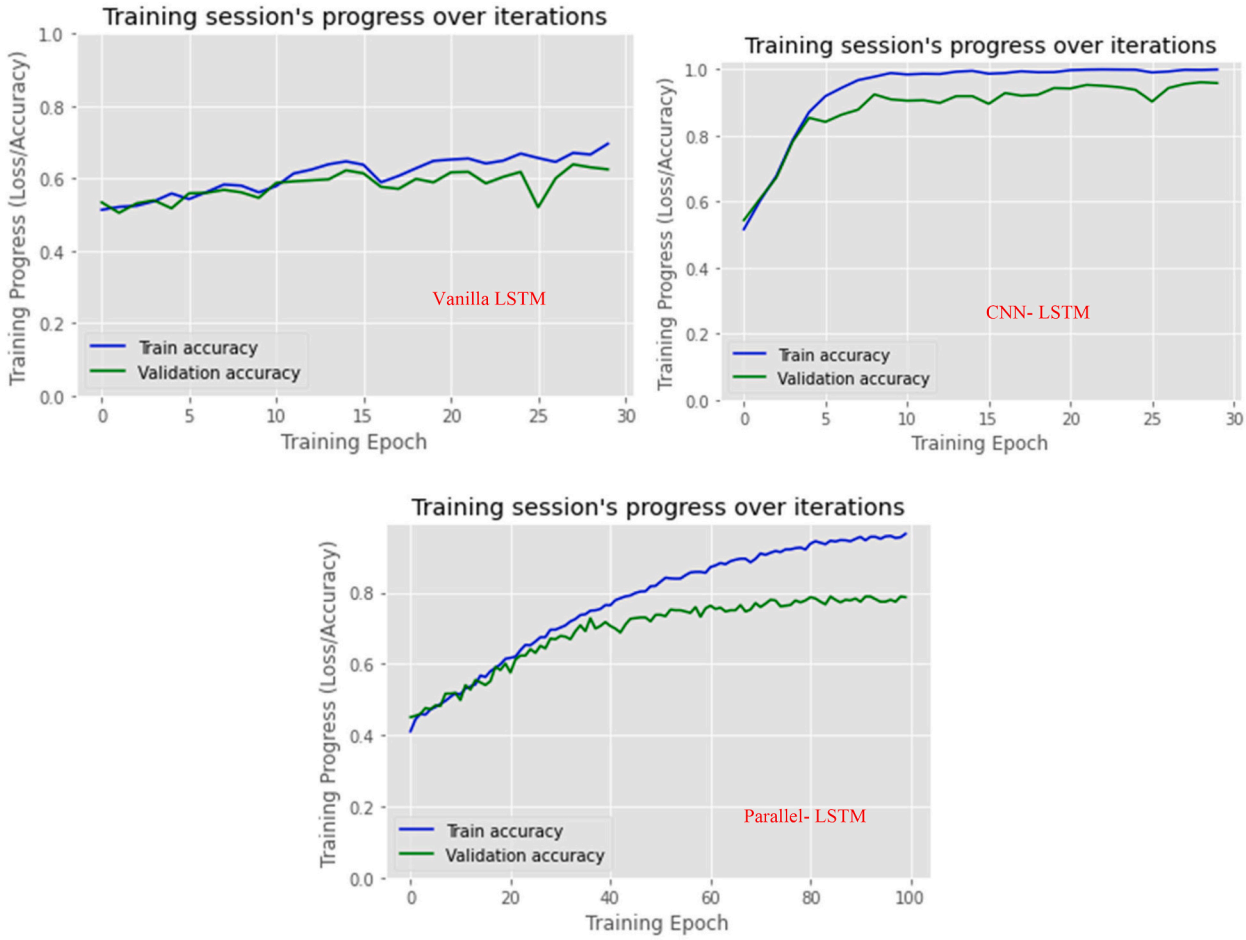

Fig. 16. Results for LSTM, Parallel LSTM and CNN-LSTM models.

CRediT author statement

All authors collaborated on producing this manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

References

Abujrida, H., Agu, E., & Pahlavan, K. (2017). Smartphone-based gait assessment to infer Parkinson's disease severity using crowdsourced data. In *2017 IEEE healthcare innovations and point of care technologies (HI-poct)*. IEEE.

Abujrida, H., Agu, E., & Pahlavan, K. (2020). Machine learning-based motor assessment of Parkinson's disease using postural sway, gait and lifestyle features on crowdsourced smartphone data. *Biomedical Physics & Engineering Express, 6.3*, Article 035005.

Anwary, A. R., Arifoglu, D., Jones, M., Vassallo, M., & Bouchachia, H. (2021). In *Insole-based real-time gait analysis: Feature extraction and classification*. IEEE.

Bot, B. M., Suver, C., Neto, E. C., Kellen, M., Klein, A., Bare, C., … Friend, S. H. (2016). The mPower study, Parkinson disease mobile data collected using ResearchKit. *Scientific Data, 3*(1), 1–9.

Clarkesmith, D. E., Pattison, H. M., Khaing, P. H., & Lane, D. A. (2017). Educational and behavioral interventions for anticoagulant therapy in patients with atrial fibrillation. *Cochrane Database of Systematic Reviews, 4*.

Davis, K. L., Edin, H. M., & Allen, J. K. (2010). Prevalence and cost of medication nonadherence in Parkinson's disease: Evidence from administrative claims data. *Movement Disorders, 25*(4), 474–480. https://doi.org/10.1002/mds.22999.PMID:20131374

Deepak, S., & Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. *Computers in Biology and Medicine, 111*, Article 103345.

Deniz, E., Şengür, A., Kadiroğlu, Z., Guo, Y., Bajaj, V., & Budak, Ü. (2018). Transfer learning based histopathologic image classification for breast cancer detection. *Health Information Science and Systems, 6.1*, 1–7.

Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., & Darrell, T. (2015). Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Dutton, J., Copeland, L. G., Playfer, J. R., & Roberts, N. B. (1993). Measuring L-dopa in plasma and urine to monitor therapy of elderly patients with Parkinson's disease treated with L-dopa and a dopa decarboxylase inhibitor. *Clinical Chemistry, 39.4*, 629–634.

Everitt, B. S., & Anders, S. (2010). *The Cambridge dictionary of statistics*.

Grosset, K. A., et al. (2006). Measuring therapy adherence in Parkinson's disease: A comparison of methods. *Journal of Neurology, Neurosurgery & Psychiatry, 77.2*, 249–251.

Gilbert, J. R., Evans, C. E., Haynes, R., & Tugwell, P. (1980). Predicting compliance with a regimen of digoxin therapy in family practice. *Canadian Medical Association Journal, 123*(2), 119.

Grosset, D., Antonini, A., Canesi, M., Pezzoli, G., Lees, A., Shaw, K., … Senard, A. (2009). Adherence to antiparkinson medication in a multicenter European study. *Movement Disorders: Official Journal of the Movement Disorder Society, 24*(6), 826–832.

Grosset, K. A., Bone, I., & Grosset, D. G. (2005). Suboptimal medication adherence in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society, 20*(11), 1502–1507.

Hussein, S., Kandel, P., Bolan, C. W., Wallace, M. B., & Bagci, U. (2019). Lung and pancreatic tumor characterization in the deep learning era: Novel supervised and unsupervised learning approaches. *IEEE Transactions on Medical Imaging, 38*(8), 1777–1787.

InformedHealth.org. (2019). *Parkinson's disease: Medication for Parkinson's disease*. https://www.ncbi.nlm.nih.gov/books/NBK293715/.

Kalantarian, H., Alshurafa, N., & Sarrafzadeh, M. (2015). Detection of gestures associated with medication adherence using smartwatch-based inertial sensors. *IEEE Sensors Journal, 16.4*, 1054–1061.

Kan, V., Rajanen, D., Opoku Asare, K., & Ferreira, D. (2018). Stop: A smartphone-based game for Parkinson's disease medication adherence. In *Proceedings of the 2018 ACM international joint conference and 2018 international symposium on pervasive and ubiquitous computing and wearable computers*.

Koesmahargyo, V., Abbas, A., Zhang, L., Guan, L., Feng, S., Yadav, V., & Galatzer-Levy, I. R. (2020). Accuracy of machine learning-based prediction of medication adherence in clinical research. *Psychiatry Research, 294*, 113558.

Kulkarni, A. S., Balkrishnan, R., Anderson, R. T., Edin, H. M., Kirsch, J., & Stacy, M. A. (2008). Medication adherence and associated outcomes in medicare health maintenance organization-enrolled older adults with Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society, 23.3*, 359–365.

Lakshminarayana, R., Wang, D., Burn, D., Chaudhuri, K. R., Galtrey, C., Guzman, N. V., Hellman, B., James, B., Pal, S., Stamford, J., & Steiger, M. (2017). Using a smartphone-based self-management platform to support medication adherence and clinical consultation in Parkinson's disease. *NPJ Parkinson's disease, 3.1*, 1–10.

Leopold, N. A., Polansky, M., & Hurka, M. R. (2004). Drug adherence in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society, 19.5*, 513–517.

Ma, Y., Ashari, Z. E., Pedram, M., Amini, N., Tarquinio, D., Nouri-Mahdavi, K., Pourhomayoun, M., Catena, R. D., & Ghasemzadeh, H. (2019). CyclePro: A robust framework for domain-agnostic gait cycle detection. *IEEE Sensors Journal, 19.10*, 3751–3762.

Marras, C., Beck, J. C., Bower, J. H., Roberts, E., Ritz, B., Ross, G. W., Abbott, R. D., Savica, R., Van Den Eeden, S. K., Willis, A. W., & Tanner, C. M. (2018). Prevalence of Parkinson's disease across North America. *NPJ Parkinson's disease, 4.1*, 1–7.

Ose, D., Mahler, C., Vogel, I., Ludt, S., Szecsenyi, J., & Freund, T. (2012). Let's talk about medication: Concordance in rating medication adherence among multimorbid patients and their general practitioners. *Patient Preference and Adherence, 6*, 839.

Postural Instability, (June 2021). https://www.parkinson.org/Understanding-Parkinsons/Symptoms/Movement-Symptoms/Postural-Instability.

Pearce, R. K., Banerji, T., Jenner, P., & David Marsden, C. (1998). De novo administration of ropinirole and bromocriptine induces less dyskinesia than L-dopa in the MPTP-treated marmoset. *Movement Disorders: Official Journal of the Movement Disorder Society, 13.2*, 234–241.

Pew Research Center. (2021). *Mobile Mobile Fact Sheet*. https://www.pewresearch.org/internet/fact-sheet/mobile/.

Sawicki, A., & Saeed, K. (2021). Application of LSTM networks for human gait-based identification. In *International conference on dependability and complex systems*. Cham: Springer.

Titova, N., & Ray Chaudhuri, K. (2017). Personalized medicine in Parkinson's disease: Time to be precise. *Movement Disorders, 32.8*, 1147–1154.

Tucker, C. S., Behoora, I., Nembhard, H. B., Lewis, M., Sterling, N. W., & Huang, X. (2015). Machine learning classification of medication adherence in patients with movement disorders using non-wearable sensors. *Computers in Biology and Medicine, 66*, 120–134.

Wang, Zhiguang, & Tim. (2015). Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. *Oates*, 1.

Yu, T., Chen, J., Yan, N., & Liu, X. (2018). A multi-layer parallel lstm network for human activity recognition with smartphone sensors. In *2018 10th International conference on wireless communications and signal processing (WCSP)*. IEEE.

Zeng, X., Zhang, X., Yang, S., Shi, Z., & Chi, C. (2021). Gait-based implicit authentication using edge computing and deep learning for mobile devices. *Sensors, 21.13*, 4592.

Zhang, H., Xu, C., Li, H., Rathore, A. S., Song, C., Yan, Z., … Xu, W. (2019). Pdmove: Towards passive medication adherence monitoring of Parkinson's disease using smartphone-based gait assessment. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 3*(3), 1–23.

Zou, Q., Wang, Y., Wang, Q., Zhao, Y., & Li, Q. (2020). Deep learning-based gait recognition using smartphones in the wild. *IEEE Transactions on Information Forensics and Security, 15*, 3197–3212.

Saini, S. D., Schoenfeld, P., Kaulback, K., Dubinsky, M. C., & Care, A. J. M. (2009). Effect of medication dosing frequency on adherence in chronic diseases. *American Journal of Managed Care, 15*(6), e22–e33, 1 Jun. 2009.