PDLens: Smartphone Knows Drug Effectiveness among Parkinson's via Daily-Life Activity Fusion

Hanbin Zhang[1], Gabriel Guo[1], Chen Song[1], Chenhan Xu[1], Kevin Cheung[1], Jasleen Alexis[1], Huining Li[1], Dongmei Li[2], Kun Wang[3], Wenyao Xu[1]
[1]University at Buffalo, Buffalo, New York, USA
[2]University of Rochester Medical Center, Rochester, New York
[3]University of California, Los Angeles, California, USA
{hanbinzh,gguo2,csong5,chenhanx,kcheung8,jasleena,wenyaoxu}@buffalo.edu
huinli@outlook.com,dongmei_li@urmc.rochester.edu,wangk@ucla.edu

ABSTRACT

Drug effectiveness management is a complicated and challenging task in chronic diseases, like Parkinson's Disease (PD). Drug effectiveness control is not only linked to personal out-of-pocket cost but also affecting the quality of life among patients with chronic symptoms. In the current practice, although that health and medical professionals still play a key role in the personalized treatment plan, the critical decision on drug selection falls upon the individual report when patients call in or visit the clinics. Unfortunately, most of the patients with chronic diseases either fail to report their day-to-day symptoms or have a limited access to medical resources due to economic constraints. In this paper, we present *PDLens*, a first smartphone-based system to detect drug effectiveness among Parkinson's in daily life. Specifically, *PDLens* can extract digital behavioral markers related to PD drug responses from everyday activities, including phone calls, standing, and walking. *PDLens* models the PD symptom severity on drug treatment and detects the change of severity scores before and after drug intake. A ranking-based multi-view deep neural network is developed to decide the drug effectiveness upon the symptom severity changes. To validate the performance of *PDLens*, we conduct a pilot study with 81 PD patients and monitor their smartphone activities and severity changes over 33693 drug intake events across six (6) months. Compared with the standard clinical drug effectiveness test developed by Motor Disorder Society, results reveal that *PDLens* is a promising tool to facilitate drug effectiveness detection among PD patients in their daily lives.

CCS CONCEPTS

• **Human-centered computing** → *Ubiquitous and mobile computing*.

Corresponding Contact: wenyaoxu@buffalo.edu; wangk@ucla.edu.

MobiCom '20, September 21–25, 2020, London, United Kingdom

ACM ISBN 978-1-4503-7085-1/20/09…$15.00
https://doi.org/10.1145/3372224.3380889

KEYWORDS

Mobile Health, Parkinson's Disease, Drug Effectiveness.

ACM Reference Format:
Hanbin Zhang[1], Gabriel Guo[1], Chen Song[1], Chenhan Xu[1], Kevin Cheung[1], Jasleen Alexis[1], Huining Li[1], Dongmei Li[2], Kun Wang[3], Wenyao Xu[1]. 2020. PDLens: Smartphone Knows Drug Effectiveness among Parkinson's via Daily-Life Activity Fusion. In *The 26th Annual International Conference on Mobile Computing and Networking (MobiCom '20), September 21–25, 2020, London, United Kingdom.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3372224.3380889

1 INTRODUCTION

Nearly 10 million people worldwide have Parkinson's disease (PD) [1]. PD is one of the most common neurodegenerative diseases among the elderly, and often leads to pain, immobility and disabilities. Once PD is onset, the brain gradually stops making dopamine, a chemical that helps send signals in human brain [2, 3]. Currently, the most prescribed PD drug is levodopa (also called L-Dopa), a natural chemical that passes into the Parkinson's brain and converts to dopamine [4]. Nowadays, there are tens of PD drugs in the L-Dopa class, such as Sinemet, Rytary, and Apomorphine [5]. Thanks to the development of PD medicine science, medical professionals can usually select the most proper drugs for PD patients based on the clinical diagnosis and patient-reported symptoms.

Nevertheless, drug resistance [6] is very common during PD treatment. Existing studies [7] show that nearly 50% PD patients' symptoms do not respond well to the prescribed dopaminergic drugs, and become resistant to L-Dopa treatment with disease progression and longer disease duration (*e.g.,* PD onset after five years). Besides the less effectiveness in PD treatment, side effects to drug resistance, such as nausea [8], hallucinations [9] and dyskinesia [10], seriously affect the quality of life and even become a grave threat to life [11]. If drug resistance can be identified in time, PD professionals can adjust drug doses or forms, and effects of drug resistance can be significantly relieved.

In recent years, mobile health technologies enable ubiquitous diagnosis and screening in many medical applications [12, 13]. Therefore, we ask a question: *is it possible to have a highly-accessible and usable solution to drug effectiveness detection by leveraging ubiquitous technologies, such as smartphones*? We know that, once PD is onset, the nerve cells responsible for producing dopamine progressively die, which further causes a deficit in motor and non-motor features [2, 14–16]. When dopamine stays at a low level, PD patients will develop related symptoms, such as vocal impairment, tremors, and

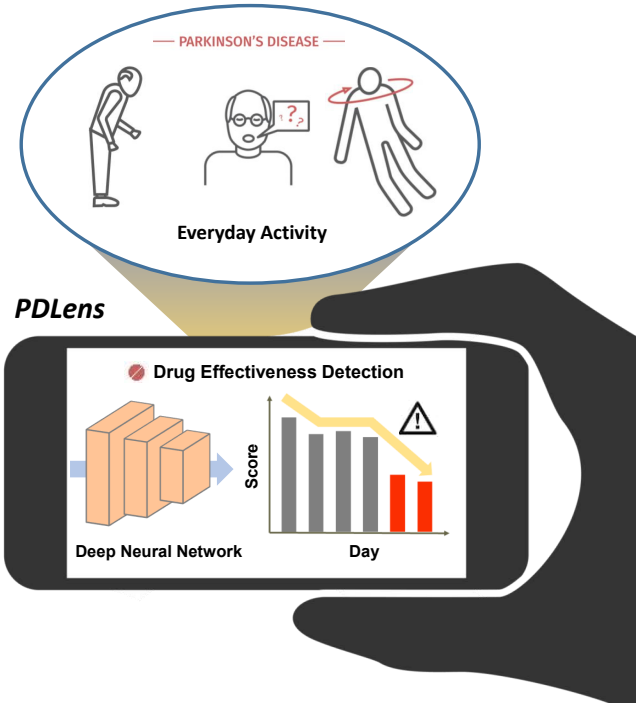


Figure 1: Smartphone can work as a drug lens to augment drug effectiveness detection among PD patients through monitoring daily-life activities.

difficulty walking. For example, a patient can feel better ("on time") as a new dose of drug starts to take effect, while feeling worse ("off time") before he is due for a different drug dose or form [6, 15]. This fact motivates us to explore a smart health solution to detect drug effectiveness by exploiting symptom fluctuations in daily life.

The goal of our work is to compare symptom severity before and after drug intake to qualify drug effectiveness. For instance, when the drug works well, the symptom gets relieved. As the disease progresses, the substantia nigra cells degenerate, and their capability to store dopamine becomes significantly affected. In this case, the drug effectiveness is weakened, and the symptomatic relief is obstructed. Since the dopamine level is highly correlated to PD symptoms, a smartphone can collect PD biomarkers, assess symptom severity, and remind users occurrence of drug resistance in daily life. The benefits of this smartphone-based solution are two folds.

- Burden-free Sensing: A smartphone is packed with a rich set of built-in sensors. These sensors allow a smartphone to collect multiple PD biomarkers from the daily-life activity without extra sensors.
- Continuous Detection: Compared with clinic-based solutions, a smartphone-based approach is daily accessible. Therefore, it allows continuous detecting and reminds the emergence of drug resistance in good time.

However, there still exists challenges. Although a rich set of built-in sensors in the smartphone can detect PD biomarkers, traditional supervised machine learning process requires us to label the symptom severity for the model training before it can detect symptom severity in routine usage. This process even costs much in clinics to fully specify the symptom severity, not to mention in daily environment. To address this problem, we propose a ranking-based solution which utilizes the following key observation. Even measuring symptom severity at a given time is challenging, acquiring comparison of symptom severity at two different times is easy. Considering drug intake happens between two timestamps (called i and j), we acquire that symptom severity at time i is more severe than that at time j due to that drug intake relieves PD symptom.

To this end, we present *PDLens*, a smartphone-based end-to-end system to facilitate self-detection of PD drug effectiveness (see Fig. 1). It leverages the built-in sensors (*i.e.,* accelerometer, gyroscope, and microphone) to collect digital PD biomarkers, *i.e.,* gait, balance, and voice from user daily activities. In particular, gait refers to the walking patterns, and balance measures the falling risk when standing. These digital PD biomarkers are sensitive to dopamine levels in brains, thus respond to drug intake [17]. Afterward, we design and implement a drug effectiveness detector that consists of a ranking-based siamese neural network. First of all, we adopt a CNN-RNN architecture as the backbone for feature extraction, where our convolutional neural network (CNN) is a residual convolutional network and our recurrent neural network (RNN) is a standard variety long short-term memory (LSTM). Then, a motor-symptom scorer that consists of a RankNet scores symptom severity. This motor-symptom scorer builds on that the comparison of symptom severity that can be easily accessed in daily life. Thus, it can achieve symptom severity detection without the need for labeling the severity level. Moreover, our proposed ranking-based multi-view neural network has different architectures in training and inference phase. The architecture in training phase accepts two samples from different timestamps (*e.g., i* and *j*), and calculates the probability that symptom at time i is more severe than that at time j. While in inference phase, the architecture accepts one sample and infers symptoms severity. We finally acquire the drug effectiveness by calculating the difference of symptom severity between before and after drug intake.

We evaluate our system on a collected dataset, including 81 participants enrolled in a six-month-long cohort study. As the first smartphone-based system that augments drug effectiveness detection in daily life, *PDLens* achieves 91.2% accuracy to detect drug response and 74.4% accuracy to screen patients with poor drug effectiveness. Moreover, the score predicted in *PDLens* is correlated to that in standard clinical drug effectiveness test with a Pearson correlation coefficient of -0.36 for enrolled participants. Our results show that *PDLens* can work as the first line of defense for detecting drug resistance.

We conclude our contributions as follows.

- To the best of our knowledge, we perform the first study to investigate that everyday activity fusion can augment drug effectiveness detection of PD. We assess three PD biomarkers, *i.e.,* voice, gait, and balance through the on-board sensors of a commercial off-the-shelf smartphone. Their variability between before and after drug intake provides valuable insights to measure drug effectiveness.

- We design and implement *PDLens*, a smartphone-based system for detecting drug effectiveness in daily life. We hypothesize that, for a PD patient, the symptom severity before and after drug intake can be assessed by smartphone. To validate this hypothesis, we design a ranking-based multi-view deep neural network. It fuses three smartphone activities, automatically extracts high-level features, and detects symptom severity. After comparing the symptom severity before and after drug intake, we infer the drug effectiveness.
- We evaluate our proposed *PDLens* on a dataset collected from a daily-life scenario. Our results reveal that a smartphone can assess this difference presented in symptom severity before and after drug intake. Moreover, this difference is an effective indicator of drug effectiveness. Our discovery will not only pave the way for augmenting drug effectiveness of PD patients in daily life but also pave the way for drug effectiveness management in other related chronic diseases.

2 BACKGROUND

In this section, we will briefly introduce the Parkinson's disease (PD).

2.1 PD Rationale and Symptoms

Parkinson's disease (PD) is a long-term degenerative disorder of the central nervous system (CNS) [18, 19]. At the very beginning, the death of cells happens in the substantia nigra due to unclear factors. Then, it destroys the dopamine pathway and results in insufficient dopamine in these areas. As dopamine takes an essential role in transmitting signals which help people move their bodies, insufficient dopamine thereby induces PD symptoms [2, 3].

PD symptoms can be categorized as motor symptoms and non-motor symptoms, where the most apparent motor symptoms can include tremor, bradykinesia (slowness of movement), and postural instability [20, 21], while non-motor symptoms include organs disorders (*i.e.,* vocal impairment and stomachic dysfunction) and mood disorders [20, 21]. When the disease progresses, these symptoms progressively make patients lose their mobility. In our study, we focus on both motor and non-motor symptoms, where impairment movement and balance belong to motor symptoms, while voice disorder belongs to non-motor symptoms.

2.2 PD Treatment

Although PD cannot be cured, PD patients greatly benefit from drug intake, which slows down the disease progression, relieves symptoms, and improves the quality of life [20]. Once initializing the treatment, patients are motivated to monitor their health conditions. They usually have significant symptoms relieved after drug intake. Over time, however, drug effectiveness frequently diminishes or becomes less consistent due to disease progression and other comprehensive reasons [6]. We call this phenomenon as drug resistance, which affects dopamine absorption. Fig. 2 is an example comparing the good and poor drug effectiveness. The dashed line marks the threshold of on-off time. In "on time", patients feel symptoms relieved; while in "off time", PD symptoms become significant again. We can observe that drug resistance significantly narrows down the length of "on time" .

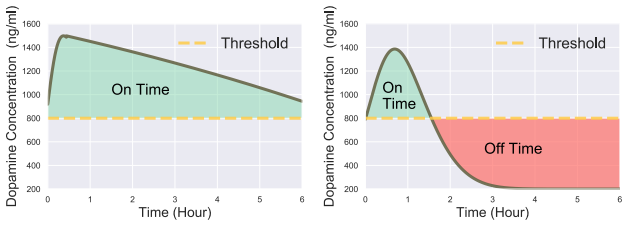


Figure 2: An example of good (left) and poor (right) drug effectiveness. Dash line marks the on-off time threshold. Drug resistance impairs the dopamine absorption.

Since everyone's symptoms and constitution are different, the progression of drug effectiveness is usually unpredictable. Self-management is unreliable because PD progresses very slowly, and patients usually do not realize drug resistance until late time. Thus, to help relieve symptoms and improve the quality of life, patients are encouraged to visit clinics and upgrade their drug plan, including prescription (*e.g.*, levodopa, dopamine agonists, and MAO B inhibitors [22]), daily dosage amount and dosage time, in good time.

2.3 PD Biomarkers and Smartphone Activities

In current clinical trials, disease disability and impairment are traditionally measured by PD professionals using the Unified Parkinson's Disease Rating Scale (UPDRS). However, the resources such as time and economic burdens limit these visits. With the proliferation of smartphones, researchers nowadays expect to utilize the built-in sensors to enable continuous measurement of symptom fluctuation in daily life [23–25].

In this paper, we explore the feasibility of detecting drug effectiveness using a smartphone. We collect digital PD biomarkers from three everyday activity, *i.e.,* walking, standing, and talking. By measuring the symptom severity before and after drug intake, we detect drug effectiveness.

3 PDLENS OVERVIEW

In this section, we present an overview of *PDLens*, including the application scenario and system protocol.

3.1 Application Scenario

Our system, *PDLens*, builds on the fact that drug-intake induces symptom fluctuation in daily life. After drug intake, symptoms get relieved; while symptoms return after drug effectiveness ends. *PDLens* leverages built-in sensors to collect digital PD biomarkers from everyday activities, including talking, walking, and standing at a non-clinical environment. Since these activities are prevalent, data collection can be burden-free or even passive. *PDLens* then uses a score to quantify symptom severity. Further, it compares this score between before and after drug intake to detect drug effectiveness.

3.2 Protocol

PDLens consists of three parts, *i.e.,* a data collector, a data preprocessor, and a drug effectiveness detector, respectively, as shown in Fig. 3.

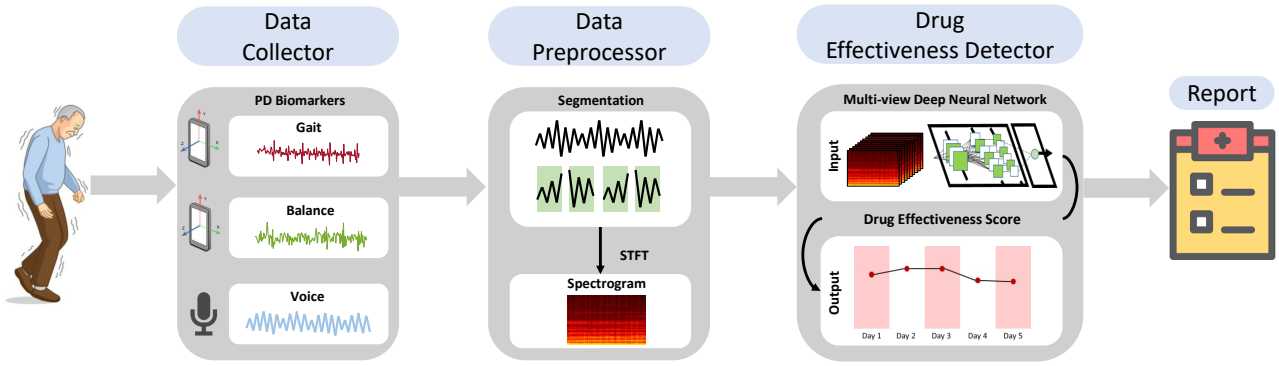


Figure 3: The proposed *PDLens* consists of a data collector, a data preprocessor, and a drug effectiveness detector. It leverages the built-in sensors to extract digital biomarkers (*i.e.,* gait, balance, and voice) from daily-life activities (*i.e.,* walking, standing, and speaking), and utilizes a customized deep neural network to achieve the detection of PD drug effectiveness.

Data Collector: Data collection happens in a nonclinical environment (*e.g.,* at home or office). Our system utilizes the built-in sensors (*i.e.,* microphone, accelerometer, and gyroscope) to collect three different activities that highly reveal the symptom severity. In this study, we design three experiments.

1) Voice Activity: Vocal impairment is one of the most common and core PD symptoms. Surveys have unearthed that as many as 90% PD patients report onset of speech impairment after their PD onset [26, 27]. PD vocal pathology stems from impaired function of laryngeal, articulatory, and respiratory muscles, which is induced by insufficient dopamine.

We utilize a microphone at the bottom of the smartphone to record voice data. Data collection happens at home. In the experiment, each participant is required to perform a 10-second-long sustaining vowel sample. A built-in microphone records each sample with a sample rate of 44.1 kHz.

2) Walking Activity: Gait impairment is one of the most common and disabling PD symptom [28–30]. In early stages, the gait slows and step length shortens. When the disease progresses, shuffling steps and freezing of gait (FOG) can happen as well. Since most gait symptoms respond to dopaminergic medication, we can leverage this gait variability as an indicator of drug effectiveness.

We utilize the built-in accelerometer and gyroscope in a smartphone to record walking data. Each participant is required to put his/her smartphone in the pocket and walk unassisted for about 20 steps in a straight line. Each sample is recorded with a sample rate of 100 Hz.

3) Standing Activity: Balance impairment is a common symptom in mid-stage and advanced stage of PD. During disease progression, PD affects the basal ganglia, a brain region to keep balance [31, 32]. Poor balance can make PD patients more likely to fall.

We utilize the built-in accelerometer and gyroscope in a smartphone to assess the balance. Each participant is required to put their smartphone in the pocket and stand still for about 30 seconds. Each sample is recorded with a sample rate of 100 Hz.

Data Preprocessor: As different sensing modalities achieve data collection, the data structure is different. Considering the input structure of a CNN requires unified data format (*e.g,* a 2-D tensor), We are motivated to preprocess this temporal domain data first before feeding them into a classifier. For this purpose, we adopt spectrogram representation due to the following two benefits: 1) Spectrogram can transform the temporal data with different length into a 2-D figure, which fits the input structure of a deep neural network; 2) Spectrogram augments features of both the time domain and the frequency domain, which can be helpful to improve the performance of drug effectiveness detection.

Drug Effectiveness Detector: The drug effectiveness detector is responsible for extracting features from represented data and detecting drug effectiveness. It consists of a multi-view deep neural network that contains three parts: 1) **Multi-modal Sensor Fusion** addresses multiple smartphone activities fusion; 2) **Feature Extractor** extracts features from sensor data; 3) **Motor-symptom Scorer** scores PD symptom severity from extracted features.

4 DRUG EFFECTIVENESS DETECTOR

4.1 Design Consideration

Challenges: In Section 2 and Section 3, we describ that a smartphone can collect daily-life activities that contain PD biomarkers. However, *PDLens* is expected to address the following three design challenges for meeting the daily-life usage.

The first challenge is how to fuse heterogeneous data. Different sensing modalities carry the information from various views, which complement each other and allow for useful information gain. However, leveraging data from heterogeneous sensors is challenging. This challenge mainly comes from distinctive properties and structures. In our study, *PDLens* adopts the built-in sensors of a smartphone to collect three PD biomarkers and the challenging can be seen as two-fold. First, sample rates can be entirely different among different sensors. Inertial sensors (*i.e.,* accelerometer and gyroscope) have a sample rate of 100 Hz, but a microphone has a sample rate of 44.1 kHz. Second, data channels are different. Audio data is one-dimensional, while inertial sensors generate three-dimensional data.

The second challenge is how to quantify drug effectiveness. To measure the drug effectiveness, we intend to calculate the difference of symptom severity before and after drug intake. It can be

challenging for a traditional approach. For example, considering a method using Euclidean distance (L2 norm), L2 distance can be large between two samples that both are collected before drug intake. Another approach is to use some traditional machine learning algorithms. However, this approach requires hand-crafted feature engineering and the prior knowledge about drug effectiveness in smartphone data is far from being enough today. We can extract some common features (*e.g.,* average value, variance value, and RMS value) from the data but we probably lose a lot of key information, which suffers the accuracy loss. Moreover, a traditional model can be incapable of learning key correlations among different sensing modalities.

The third challenge is how to get the symptom severity label. As described in Section 3, drug-intake induces symptom fluctuation in daily life. Thus, we expect to measure the symptom severity before and after drug intake to detect the drug effectiveness. Intuitively, we can leverage a traditional regression model to achieve this goal. However, such a supervised learning method requires an expert to help quantify and label the symptom severity. In fact, an accurate score cannot be elicited without costing massive labors and resources in clinics, not to mention in a daily-life environment.

Our Solution: To address these challenges, we design and implement the drug effectiveness detector, a ranking-based siamese neural network consisting of a data representation module and a multi-view deep learning model (as shown in Fig. 4). We claim that our drug effectiveness detector can address three aforementioned challenges.

First, PDLens leverages the time-frequency representation to project the heterogeneous data into the same dimension (see Section 3.2). After transforming data from the temporal domain to the time-frequency domain, we design and implement a ranking-based multi-view deep neural network to combine information from multiple sensors. Each image is passed through the first part of our network separately, then aggregated at a view-pooling layer. In particular, we adopt element-wise maximum operation across the multi-view to aggregate the information [33].

Second, our feature extractor that includes a convolutional neural network (CNN) and a long short-term memory (LSTM) is able to automatically extract features thereby avoid hand-crafted feature engineering. Moreover, this deep learning-based solution performs better than traditional approaches in the following three aspects: 1) back propagation allows to adjust weights for error correction; 2) multi-layer neurons allow to capture both low-level and high-level features; 3). activation functions (*i.e., ReLU* layer) provide the ability to understand non-linear relationship between the data.

Third, this ranking-based solution utilizes the fact that drug intake relieves the symptom severity. Therefore, *PDLens* only expects to know the timestamp (*i.e.,* before and after drug intake) without the need to fully examine the symptom severity. Moreover, *PDLens* is a personalized model, which will not require labels to measure the severity among patients [34].

4.2 Feature Extractor

The feature extractor contains a convolutional neural network (CNN) and a long short-term memory (LSTM) architecture.

CNN: We adopt a residual network as our backbone for its ability to learn identified mapping well [37]. Specially, we employ 4 residual blocks, each with 4 convolutional layers. There are 64 filters in the first residual block, and the number of filters doubles for each following residual blocks. We use a 7×7 frequency-time filter for the first convolutional layer, followed by 3×3 filters for the remaining convolutional layers.

The dimension of the last convolutional layer will be in [Batch×Channel×Height×Width] format. The first element is the batch size; the second element is the dimension of feature maps; the third and the fourth elements are frequency and time context in our system, respectively. According to our description mentioned above, our output follows [Batch×512×7×7] format.

LSTM: After performing frequency modeling, we then unroll and pass the output of the last convolutional layer to LSTM [38] for modeling the time-domain signal. According to the strategy introduced in previous work [39, 40], we especially implement 2 LSTM layers for dimensionality reduction. Since the time dimensionality is 7, each LSTM layer is unrolled for 7 time steps for training. The first LSTM layer has an input of 512×7 and an output of 64; while the second LSTM layer has an input of 64 and an output of 64. Finally, a view-pooling layer achieves the feature vectors concatenations.

4.3 Motor-symptom Scorer

After extracting and fusing features from daily-life activities, original raw data are processed and transformed into tuples $[x_i^p, ..., x_n^p]$, where $x \in R^n$ is the collected data associated with patient $p \in P$, and $i \in I$ is the timestamp. Then, we intend to identify a function that can correctly score PD symptom severity. For this purpose, we utilize a ranking-based idea, which leverages a Sigmoid function [41] to achieve this scoring for a twofold reason. First, the Sigmoid function can be implemented through a Sigmoid layer that has been shown to lead to a good probability estimation. Second, the Sigmoid function is differentiable, which enables the backpropagation to update weights. More details about theorem and proof can refer to [36].

Problem Formulation: Given a pair of feature vectors with different timestamps, we adopt a function, called $g(\cdot)$, to map them to a numeric value:

$$\begin{cases} s_i = g(x_i) \\ s_j = g(x_j) \end{cases}. \tag{1}$$

We define $X_i \succ X_j$ as the drug intake happens at a time between i and j. Thus, the PD symptom at time i is more severe than that at time j. These two outputs are then mapped to a learned probability that x_i should be ranked higher than x_j via a Sigmoid function:

$$P_{ij} \equiv P(X_i \succ X_j) \equiv \frac{1}{1 + e^{-\sigma(s_i - s_j)}}, \tag{2}$$

where σ determines the shape of this Sigmoid function.

Loss Function: With knowing the probability function, we adopt the cross-entropy as a cost function to penalize the deviation of model output probabilities from desired probabilities. let $\overline{P_{ij}}$ be a known probability, we can formulate this cost function as:

$$L = -\overline{P_{ij}} log P_{ij} - (1 - \overline{P_{ij}}) log(1 - P_{ij}). \tag{3}$$

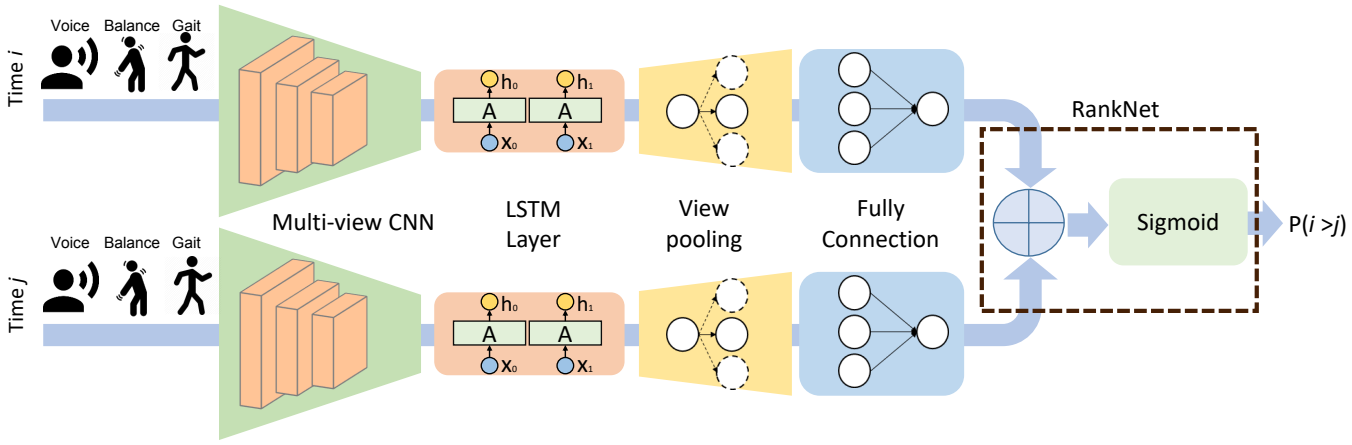


Figure 4: The drug effectiveness detection is achieved by our proposed ranking-based multi-view siamese neural network [35]. Convolutional layers and LSTM layers are adopted for feature extraction. Then, RankNet [36] is adopted to compare the symptom severity at two different time. Weights in neural network are not shared among each view; while weights are shared between two branches i and j.

Substituting Eq. (2) to Eq. (3) and calculating the gradient, we can obtain:

$$\frac{\partial C}{\partial s_i} = -\frac{1}{1+e^{s_i - s_j}}. \tag{4}$$

This gradient is then employed for weights updating during the backpropagation.

4.4 Training and Inference

Our ranking-based multi-view siamese neural network has a different architecture in training and inference phase, respectively (see Fig. 5).

Training: The training architecture accepts two inputs. Given a pair of inputs (*i.e.,* x_i and x_j), the output is a probability that symptom at time i is more severe than that at time j. We label the inputs with $S_{ij} \in \{0, +1, -1\}$. Let $S_{ij} = 1$, if symptom at time i is more severe than at time j; while $S_{ij} = -1$, if symptom at time j is more severe than at time i. Moreover, $S_{ij} = 0$ means symptom at time i and j has a same level of severity. Referring to Section 2, the severity relationship is deterministically known, where the "before drug intake" is the most severe and the "after drug intake" is the least severe. Let $\overline{P_{ij}} = \frac{1}{2}(1 + S_{ij})$. Combining the above formula with Eq. (3), we have:

$$L = \frac{1}{2}(1 - S_{ij})\sigma(s_i - s_j) + log(1 + e^{-\sigma(s_i - s_j)}). \tag{5}$$

Note that the cost is $log2$ when $s_i = s_j$, which means the model tries to push it away from each other where data at two inputs are from a different time. For patients with good drug effectiveness, this pushing works well. However, this pushing is obstructed for patients with poor drug effectiveness. The reason is that the difference of symptom severity between before and after drug intake is small when a patient does not respond to drug well. In this case, s_i is close to s_j.

Inference: After pre-training our network on existing collected data, the architecture accepts one input for inference. The output is the score measuring symptom severity. Through calculating the difference of symptom severity before and after drug intake, we acquire the drug effectiveness.

When a drug loses its effectiveness, data collected from before and after drug intake will have the same data representation. The difference of symptom severity before and after drug intake, the drug effectiveness, will be small.

Case Study: We use an example to better explain this procedure. When x_i is "before drug intake" and x_j is "after drug intake", we let $S_{ij} = 1$ as we have no prior knowledge if drug works. In this case, we have:

$$L_1 = log(1 + e^{-(s_i - s_j)}). \tag{6}$$

We set σ as 1 since it only controls the shape of a Sigmoid function. The backpropogation allows the feature extractor to push away s_i and s_j from each other in order to minimize this loss function L_1.

When x_i and x_j are both from "before drug intake", we let $S_{ij} = 0$ and the loss function will be:

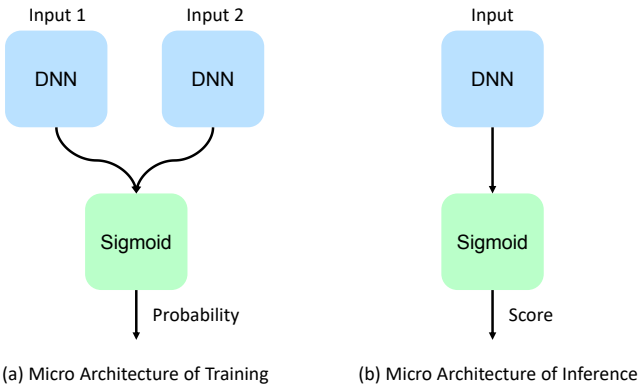
$$L_2 = \frac{1}{2}(s_i - s_j) + log(1 + e^{-(s_i - s_j)}). \tag{7}$$

To minimize this loss function L_2, the backpropogation will try to make s_i and s_j equal.

When drug works well, the data collected from before and after drug intake will have different data representation. $min\ L_1$ and $min\ L_2$ work well with this situation. However, $min\ L_1$ and $min\ L_2$ compete with each other if drug loses its effectiveness. In this case, the data collected from either before or after drug intake will have the same data representation. Since weights in our proposed network are shared between two branches, weights optimized for L_1 will not be suitable for L_2.

5 BENCHMARK PREPARATION

In this section, we introduce the participant enrollment and system implementation.



(a) Micro Architecture of Training (b) Micro Architecture of Inference

Figure 5: Our proposed architecture accepts two and one input in the training and inference phase, respectively.

5.1 Participant Enrollment

Our research is approved by the Western Institutional Review Board (IRB). During six months, more than hundreds of PD patients join in this cohort study, and 81 of all participants (see Table 1) who complete the collection of three activities are qualified in *PDLens*. From these participants, we collect 11668 voice activity samples, 11014 walking activity samples, and 11014 standing activity samples, respectively.

Moreover, each subject owns a healthcare provider (*e.g.,* Neurologist and PD professionals) who advises the drug intake. Even so, drug effectiveness varies from person to person. To measure this drug effectiveness, we request each participant to take a standard clinical drug effectiveness test to comprehensively evaluate the "off time" in daily life (see Table 2). These questions belong to a part of MDS-UPDRS [42], a standard clinical used rating tool to gauge PD. We further employ the score to define good and poor drug effectiveness.

Characteristic	Number
Age (Years), M (SD)	65.3 (6.0)
Time Since PD Onset, M (SD)	7.8 (4.5)
Time Since Medication Onset, M (SD)	4.0 (3.7)
Male, n (%)	52 (64.2)
Healthcare Provider, n (%)	81 (100)
High School Degree or Above, n (%)	81 (100)

Table 1: Demographic characteristics of enrolled PD participants (N = 81).

5.2 Deep Neural Network Implementation

We implement our deep neural network in PyTorch. Throughout these experiments (see Section 6), we adopt stochastic gradient descent with momentum (SGDM) as the optimizer with an initial learning rate of 0.01. During training, data augmentation methods are adopted, including random resize and crop, random horizontal lip, and color jitter.

5.3 Smartphone Implementation

We convert our trained PyTorch model to an intermediate ONNX (Open Neural Network Exchange) representation. The ONNX file is loaded into Caffe2, a scalable lightweight deep learning framework [43]. From Caffe2, we export two protobuf (Google Protocol Buffer [44]) models. One file contains the input and parameters to the model. The other file guides run-time execution [45]. The advantage of protobuf is that it is serialized and platform-neutral, allowing for execution of the same model on both desktop and smartphone platforms [44]. We then adopt a script from [46] to execute the protobuf models on smartphone through Android Debug Bridge.

6 EVALUATION

In this section, we evaluate our system on the collected smartphone dataset.

6.1 Overall Performance

We present the overall performance study. We will first investigate whether our system can differentiate the conditions before and after drug intake according to symptom severity. On this basis, we further investigate whether our system can screen participants who develop drug resistance.

Stage 1: Evaluation of Drug Response Detection

Experimental Goals: We expect to investigate the ability of *PDLens* to differentiate the conditions before and after drug intake.

Evaluation Metrics: We adopt a confusion matrix to describe the classification performance comprehensively. The first row from left to right is True Positive (TP) and False Positive (FP), separately; the second row from left to right is False Negative (FN) and True Negative (TN), separately. With this confusion matrix, we calculate metrics such as $accuracy = \frac{TP+TN}{TP+TN+FP+FN}$, $precision = \frac{TP}{TP+FP}$, and $recall = \frac{TP}{TP+FN}$.

Ground Truth: We instruct our participants to take the experiments and report the timestamps. These timestamps correspond to two labels, which are "before drug intake" and "after drug intake", respectively.

- "before drug intake" describes the timestamp when a participant does not take drug in a day, and the motor-symptom becomes most severe in a day.
- "after drug intake" describes the status when a participant feels relieved after drug intake.

We adopt these reported labels as ground truth.

Person-dependent Results: We first report our person-centered results. For each user, we adopt 80% of data for training, and the remaining 20% of data is used for testing. Recall that our model is not a classification one, we adopt a least-square criterion to generate a threshold in the testing set. We predict a sample as "before drug intake" if the output is higher than this threshold, whereas predicting a sample as "after drug intake" if the output is lower than this threshold.

Fig. 6 shows the classification performance on all enrolled participants. The positive class is defined as "after drug intake". Our system achieves 91.0% accuracy on average, suggesting that we can well differentiate stages between before and after drug intake. The

Table 2: The standard clinical drug effectiveness test for "off time" rating. All questions come from MDS-UPDRS.

Questions	Clinical Rating Scale				
	0	1	2	3	4
Over the past week, have you usually had problems with your speech?					
Over the past week, have you usually had troubles handling your food and using eating utensils? For example, do you have trouble handling finger foods or using forks, knives, spoons, chopsticks?					
Over the past week, have you usually had problems dressing? For example, are you slow or do you need help with buttoning, using zippers, putting on or taking off your clothes or jewelry?					
Over the past week, have you usually been slow or do you need help with washing, bathing, shaving, brushing teeth, combing your hair, or with other personal hygiene?					
Over the past week, have people usually had trouble reading your handwriting?					
Over the past week, have you usually had trouble doing your hobbies or other things that you like to do?					
Over the past week, do you usually have trouble turning over in bed?					
Over the past week, have you usually had shaking or tremor?					
Over the past week, have you usually had problems with balance and walking?					
Over the past week, do you suddenly stop or freeze as if your feet are stuck to the floor?					

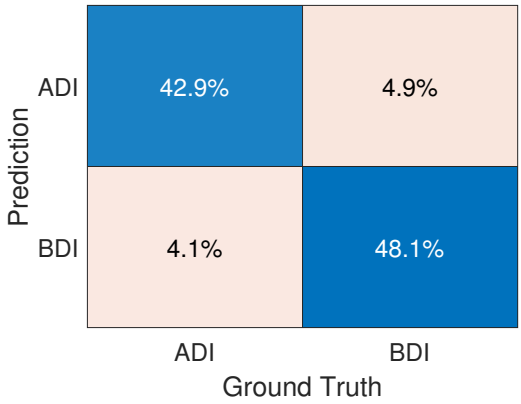


Figure 6: The average normalized confusion matrix of person-centered result. ADI is short for after drug intake, and BDI is short for before drug intake.

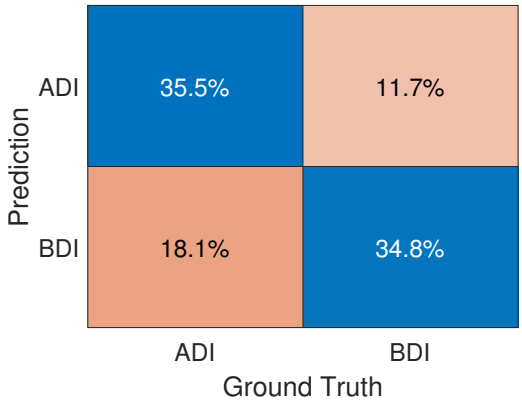


Figure 7: The average normalized confusion matrix of person-independent result. ADI is short for after drug intake, and BDI is short for before drug intake.

recall is close to precision, suggesting that we can differentiate two stages without obvious bias.

Person-independent Results: We further report the person-independent results. We adopt a 20-fold leave-one-out validation. In each experiment, we randomly select one participant as testing and train our model on the remaining participants.

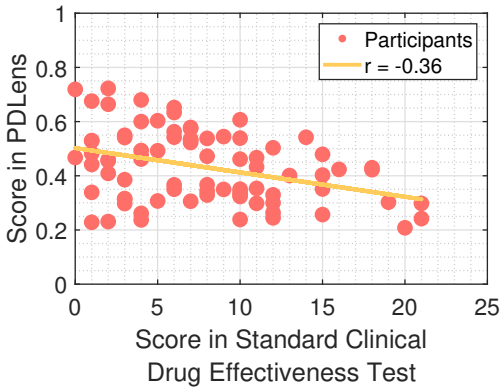
Fig. 7 shows that our system is capable of predicting the conditions before and after drug intake with 70.0% accuracy. FP is lower than FN, suggesting that our model more likely predict a sample as "before drug intake". This performance, however, is lower than the person-centered model. Person-independent drug intake detection is essentially more challenging since drug effectiveness can evolve different among different subjects due to various disease progressions and medical prescriptions.

Stage 2: Evaluation of Drug Effectiveness Detection

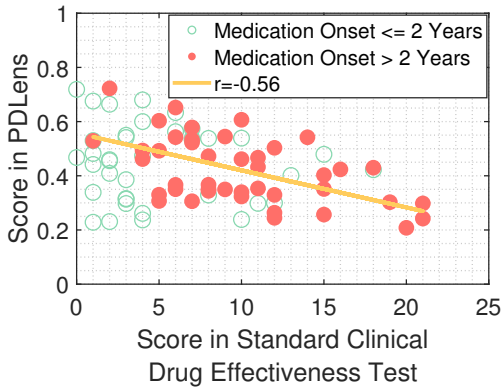
Experimental Goals: On the basis of stage 1, we further investigate the ability of *PDLens* to detect drug effectiveness. In particular, our system should be able to screen patients who are becoming resistant to current drug.

Evaluation Metrics: We compare the difference of symptom severity before and after drug intake to measure drug effectiveness. As expected, a participant who responses to drug well exhibits a noticeable difference in symptom severity between before and after drug intake; while a participant who is resistant to drug exhibits a non-obvious or even no difference.

Ground Truth: As shown in Table 2, we use a questionnaire to test each participant the drug effectiveness. There are totally ten questions, and each question is anchored with five responses that link to commonly accepted clinical terms : 0 = normal, 1 = slight, 2 = mild, 3 = moderate, and 4 = severe. These questions belong to the MDS-UPDRS [42], which is a standard test to gauge PD conditions in clinical medicine. Fig. 9 shows the CDF graph of score in standard clinical drug effectiveness test on a total of 81 participants, where the highest score is 21, the lowest score is 0, and the median value is



(a) $r = -0.36$ for all participants.



(b) $r = -0.56$ for participants medication onset more than two years.

Figure 8: A scatter plot of data on variable "score in standard clinical drug effectiveness test" and "score in *PDLens*". Our result reveals that the prediction of *PDLens* is correlated with the ground truth from the standard clinical drug effectiveness test.

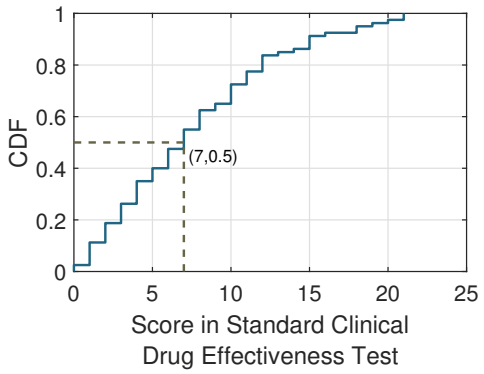


Figure 9: The CDF graph of score in standard clinical drug effectiveness test. The highest score is 21 and lowest score is 0. The median value is 7.

7. Specifically, the higher score in this standard drug effectiveness test indicates the poorer drug effectiveness.

Results: Fig. 8 identifies the correlation between the score in standard drug effectiveness test and that in *PDLens*. Although all the patients respond to drug, the drug effectiveness varies from person to person, where the highest score is 0.72, and the lowest score is 0.2. Moreover, the score predicted by *PDLens* is correlated to that in standard drug effectiveness test with a Pearson correlation coefficient of -0.36. This result meets our expectation that a participant with poorer drug will correspondingly receive a lower score in *PDLens*.

We further observe that participants receiving a lower score in the clinical test have a broader variation in drug effectiveness detection. After checking into demographics, we find out that most of those participants are medication onset less than two years. For those participants, the detection of *PDLens* can exist bias. The reason is that when first initializing medication, "on time" can last until the due for the next dose of drug. Therefore, the difference of symptom severity between before and after drug intake is negligible. We further investigate that the Pearson correlation coefficient is -0.56 for participants who are medication onset for more than two years. Since drug resistance usually appears later after medication onset (*e.g.,* after two years), *PDLens* can detect the occurrence of drug resistance.

Considering that medication onset time does affect our detection, we then investigate if prior knowledge of medication onset time can assist *PDLens* to screen poor effectiveness participants. In particular, we define "good drug effectiveness" if a participant averagely feels no (0 = no) or very slight (1 = slight) uncomfortable, and define the remaining participants as "poor drug effectiveness". In Fig. 10, the red dots mark the "good drug effectiveness" and the green dots mark the "poor drug effectiveness", respectively. This experiment is conducted with a linear support vector machine (SVM) classifier. We adopt a 10-fold hold-out validation. In each fold, we randomly hold 15% data as validation. We achieve 74.4% accuracy on average. The non-linear kernel is tried as well, but the improvement is not obvious.

Our results (Fig. 8 and Fig. 10) reveal that *PDLens* can detect the drug effectiveness of PD. That is, participants with poor drug effectiveness receive a corresponding low score from *PDLens*. Although detection can exist bias for PD patients who recently start drug intake, our system can work as the first line of defense to remind them a clinic visit.

6.2 Understanding Roles of drug Effectiveness Detector

In this section, we explore the roles of our model in detecting drug effectiveness.

Roles of Ranking

Experimental Setup: We evaluate the role of ranking in *PDLens*. As a baseline, we train a classifier with our CNN-RNN model as the backbone for feature extraction and formulate the output as a

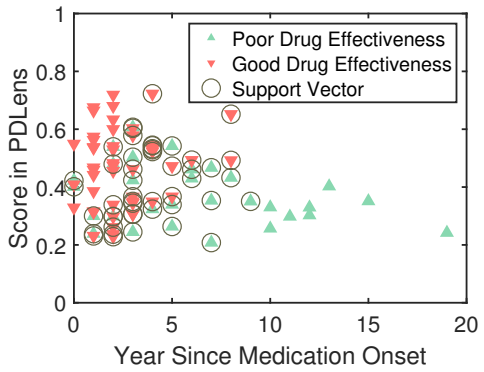
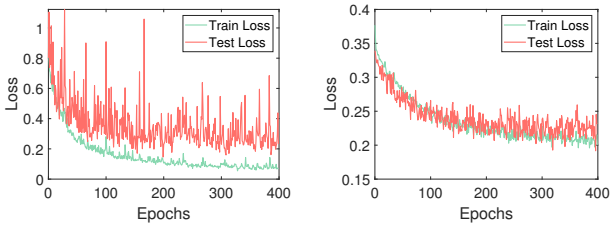
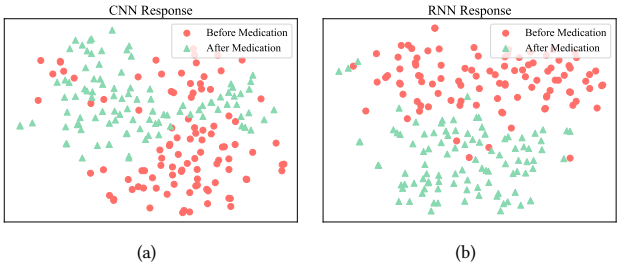


Figure 10: This figure visualizes data distribution on variable "year since medication onset" and "drug effectiveness score".



(a) Training and testing loss comparison in the prepared classification model. (b) Training and testing loss comparison in our ranking model.

Figure 11: A comparison of loss between different ranking and classification model.



(a) (b)

Figure 12: A comparison of CNN and RNN response. The model will more likely predict a sample as "after drug intake" without LSTM layers.

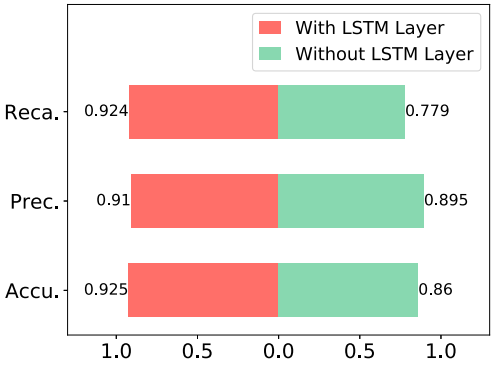


Figure 13: The comparison of performance (*i.e.,* accuracy, precision, and recall) between with and without LSTM layers.

binary classification problem. Both ranker and classifier adopt the same optimizer and learning rate (see Section 5.2).

Results: We find that ranker achieves an average accuracy of 2.5% higher than that of classifier. This result shows that the ranking model converges better. For further exploration, we visualize the training and testing loss between ranker and classifier, as shown in Fig. 11. For the classification model, the testing loss has a large variation, and the overfitting occurs at about the 100th epochs. Instead, the ranking model has a small variation in testing loss, and the overfitting is not obvious. Note that when $S_i = S_j$, the cost is $log2$, so the model incorporates a margin. That is, collected samples with different labels, but receiving the same scores, are still pushed away from each other. This mechanism, to some extent, prevents overfitting, thereby helps convergence better.

Roles of RNN

Experimental Setup: We adopt the t-SNE embedding [47] to visualize the network response. In addition, we train a model by removing LSTM layers as a baseline.

Results: We find that overall accuracy decreases about 4.7% by removing the LSTM layers. Take one participant for example. Fig. 12 visualizes the CNN and RNN response, and Fig. 13 compares its performance. It shows that RNN augments the differentiation of two stages. The recall is 12% higher than the precision, suggesting that the model more likely predict a collected sample as "after drug intake" without LSTM layers. This result also suggests that LSTM layers do learn some temporal domain features that benefit drug effectiveness detection.

6.3 System Overhead Analysis

In this section, we evaluate the footprint of our model on mobile devices. Since this model is appropriate for daily-life usage, it should own low resource cost and high efficiency. We implement our model on two different types of smartphones employing Google's Protocol Buffers [44].

Run Time Measurement: We set up the batch size as 1. To measure the run time, we continuously run our model on a smartphone for 100 times and calculate the average value.

Battery Usage Measurement: We adopt *AccuBattery* [48], a tool for real-time battery usage monitoring. We measure the rate of battery usage before and after we start our model on a smartphone, respectively, and calculate its difference.

Results: Table 3 shows the footprint of our model on two different smartphones. Our model's run time elapsing per iteration can be lowered by using more capable smartphones. The run time per iteration for Google Pixel 2 (2017) is 1770.88 *ms*, while the run time

Phone Brand	Average Run Time (ms)	Average CPU (%)	Battery Usage (mA)
Pixel 2	1770.88	15.0	358
Galaxy S9	1200.87	14.0	488

Table 3: System overhead on different smartphones.

per iteration for Samsung Galaxy S9 (2018) is 1200.87 ms. However, it seems that the superior performance of Galaxy S9 requires a greater energy cost, as the rate of battery usage for Galaxy S9 is 488 mA, compared to 358 mA for Pixel 2. Yet, in terms of practical battery cost in daily usage, this difference is not significant, as explained below.

Assuming that drug intake happens three times a day, six iterations of our model will run per day. The actual battery cost of our model is negligible in daily usage. Pixel 2 is projected to use 1.06 mAh per day, less than 0.1% of its 2700 mAh battery. Similar to Pixel 2, Galaxy S9 is projected to use 0.98 mAh per day, less than 0.1% of its 3000 mAh battery. Furthermore, both smartphones take approximately one second to make an inference with the model, which is a sufficient speed to deliver real-time results to users.

7 PDLENS STABILITY STUDY

Although talking, walking, and standing are three everyday activities that can be easily accessed, a smartphone may still miss some activities in daily life, and we are responsible for exploring the system performance in this situation. Despite requiring retraining model, this phase can be simultaneously finished during data collection. For example, we can train multiple models in advance and choose appropriate pre-trained models to deal with different cases of missing data.

Same as described in Section 6.1, we evaluate both drug response detection (stage 1) and drug effectiveness detection (stage 2) when there exists only biomarker of voice, gait, and balance, respectively.

Experimental Setup: We separately train a model for each smartphone activity per participant. The hyper-parameter for training is the same as we train a multi-view deep neural network. The ground truth, as we described above, is the collected time (*i.e.,* "before drug intake" and "after drug intake") and the standard clinical drug effectiveness test for "off time" rating.

Results: Fig. 14 compares the performance of detecting drug response between each single smartphone activity and the multi-view one. As expected, the ranking-based multi-view model achieves the highest accuracy. In fact, we observe that our ranking-based multi-view model can always achieve about $1 \sim 2\%$ accuracy than the best case of single activity. Moreover, voice can achieve a higher accuracy than gait and balance due to that vocal impairment is the most widely existing PD symptom in all stages, and drug intake can relieve this vocal impairment.

Table 4 reports the performance of drug effectiveness detection. The first column is the result on all participants, and the second column describes the result on participants who start medication

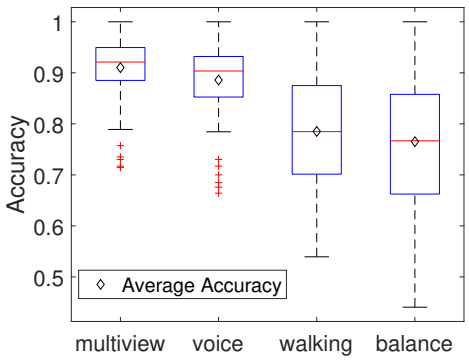


Figure 14: A comparison of accuracy between the activity fusion and single activity. The central, the bottom and top edges of the box marks the median, the 25th and 75th percentiles, respectively. Symbol '+' marks the outliers.

	Pearson Correlation on All PD Patients
Multi-view	-0.36
Voice	-0.35
Gait	-0.28
Balance	-0.04

Table 4: Pearson correlation coefficient between score in *PDLens* and score in standard clinical drug effectiveness test.

longer than two years. The balance, however, achieves almost no correlation. One reason is that balance impairment is not obvious for most PD patients until late stages. Thus, depending on balance only to achieve drug effectiveness detection is not accurate enough.

8 IMPLICATION

In this section, we discuss the extension of *PDLens*, including passive PD severity sensing and precise medicine.

Passive PD Severity Sensing: As a progressive neurodegenerative disorder, PD causes significant physical impairment and declined quality of life. Nowadays, PD management requests patients to accept regular assessment and close monitoring of symptoms for adjusting drug usage such as dosage and frequency. However, this assessment in a clinic is subjective and clinical professionals cannot continuously monitor the trajectory of symptom severity between two clinic visits. As a smartphone-based sensing system, *PDLens* shows the feasibility to monitor symptom severity via fusing daily-life activities, such as putting a smartphone in the pocket or recording the phone talk. We believe that an extension of *PDLens* can work as a passive sensing system to continuously monitor the symptom severity without extra physical and economic burden.

Precision Medicine: Current clinical professionals usually choose the most prescribed PD drug. However, individual variability in genes, symptom severity and physical condition result in different therapeutic effect. This problem called precision medicine is a

relatively new but significant topic in current healthcare. Clinical-based approaches cannot track drug-induced symptom variation in daily life. As a mobile system, *PDLens* shows the feasibility to achieve this goal. Besides monitoring the symptom severity before and after drug intake, *PDLens* can expand to monitor intra-day drug metabolism. The collected data can assist clinical professionals to optimize the prescription such as drug type, dosage and frequency.

9 DISCUSSION

We identify that smartphone can know drug effectiveness of Parkinson's disease patients. Our system, *PDLens*, builds on the fact that drug intake can relief motor symptoms. Through collecting daily-life activities, such as voice, walking, and standing, before and after drug intake, *PDLens* can quantify the difference of symptom severity, thereby infer the drug effectiveness.

The observed phenomenon is consistent with previous studies that smartphone can measure the symptom severity [49]. One difference is that most of existing work adopts a supervised learning-based method. That is, the relationship of symptom severity among patients should be known in advance. This method can require large labour to label these relationships. Instead, our proposed method might be more like an unsupervised learning method. The training phase will not require the severity information from others.

Strengths of our study include the long follow-up period, both before and after the drug intake across a total time of 6 months, as well as large sample of PD patients and collected daily-life activities. In addition, we leverage an AI-based solution in detecting symptom severity, which allows us to examine the non-linear relationship among features. This non-linear relationship might be missed in conventional models (*e.g.,* SVM algorithm). Moreover, the AI-based solution enables automatic feature selection that rarely has been done in conventional models.

PDLens marks a closer step towards drug effectiveness detection of PD in daily life. However, it exhibits some limitations. First, the correlation between system output and score in drug effective test is not very high. One reason comes from the difference between subjective and objective evaluation. The other reason is that wearing-off phenomenon not always happens on every subject. In this case, gap between before and after drug intake becomes insignificant.

Second, *PDLens* leverages a multi-model deep learning architecture to fuse multiple activities. However, it is challenging to quantify how much one activity contributes. The reason is that our deep learning architecture will nonlinearly map the input to a latent space. That is different from conventional linear models where the weights can be considered as the contribution. We leave such considerations as extensions to our system.

10 RELATED WORK

Related work falls within two areas, including mobile health systems and Parkinson's disease (PD) related research.

10.1 Mobile Health Systems

Mobile health (mHealth) is an emerging area of interest for researchers in recent years [50–61]. Stresssense [12] recognized stress from the human voice using smartphones. Pho_2 [62] measured the blood oxygen level of a person with a built-in camera. iSleep [63] leveraged embedded microphone to monitor events related to sleep quality, such as snoring and body movement. Farhan *et al.* [64] applied data from GPS and accelerometer of the smartphone to perform depression screening. Healthaware [65] utilized the embedded accelerometer to monitor daily physical activities and the built-in camera to analyze food items to control obesity. SymDetector [66] employed the built-in microphone to detect respiratory symptoms, such as coughs, sniffles, and sneezes. Some other work focuses on medication management. Kalantarian *et al.* [67] utilized the accelerometers and gyroscopes of a smartwatch to detect gestures that can predict medication adherence, including opening pill bottles. Bae *et al.* [68] adopted the built-in step counter, distance tracker, and motion sensor of a Fitbit to predict hospital readmission for cancer patients post-surgery.

Our work utilizes the built-in sensors of a smartphone to collect digital PD biomarkers from a user's daily activity to augment drug effectiveness detection of PD.

10.2 PD Related Research

In the past twenty years, understanding PD symptoms using mobile sensors attracts tremendous attentions from researchers. These studies include but not limit to gait assessment [69–73], finger dexterity [74–79], and vocal impairment [56, 80–83]. Sharma *et al.* [84] developed SPARK, which utilizes a smartphone and a smartwatch to monitor dysfunctional speech, gait abnormalities, limb dyskinesia, and voice disorder. Arora *et al.* [23] developed a smartphone application to detect and monitor multiple PD biomarkers. Zhan *et al.* [49] proposed an objective measurement of PD severity and tested construct validity. *PDMove* [85] achieved the medication adherence monitoring through gait assessment using a smartphone. PDLens is totally different from *PDVocal* [13] in following three aspects. 1) The application is different. *PDVocal* expects to address PD detection whereas *PDLens* aims to detect drug effectiveness for patients with PD; 2) The data collection is different; *PDVocal* collects data from both PD and non-PD people; while *PDLens* collects data from only PD patients but data needs to be collected before and after drug intake; 3) The sensing modality is different. *PDVocal* adopts the built-in microphone to collect audio data only whereas *PDLens* collects 3 types of activities using the built-in microphone, accelerometer, and gyroscope.

Most existing work target PD diagnosis and screening. However, *PDLens* is the first solution that augments drug effectiveness detection using a smartphone.

11 CONCLUSIONS

In this paper, we present *PDLens*, the first mobile health system that utilizes everyday activity to augment drug effectiveness detection in daily life. It works by collection of digital PD biomarkers through the built-in sensors, detection of symptom severity with a customized deep neural network, and detection of drug effectiveness. We believe that *PDLens* demonstrates a promising step in the real-world deployment of a drug management mobile system in the future.

REFERENCES

[1] M. Parish, "The pathophysiology and nursing care for parkinson's disease," 2019.

[2] H.-C. Cheng, C. M. Ulane, and R. E. Burke, "Clinical progression in parkinson disease and the neurobiology of axons," *Annals of neurology*, vol. 67, no. 6, pp. 715–725, 2010.

[3] T. C. Thannickal, Y.-Y. Lai, and J. M. Siegel, "Hypocretin (orexin) cell loss in parkinson's disease," *Brain*, vol. 130, no. 6, pp. 1586–1595, 2007.

[4] M. D. Yahr, R. C. Duvoisin, M. J. Schear, R. E. Barrett, and M. M. Hoehn, "Treatment of parkinsonism with levodopa," *Archives of neurology*, vol. 21, no. 4, pp. 343–354, 1969.

[5] R. Dhall and D. L. Kreitzman, "Advances in levodopa therapy for parkinson disease: review of rytary (carbidopa and levodopa) clinical efficacy and safety," *Neurology*, vol. 86, no. 14 Supplement 1, pp. S13–S24, 2016.

[6] A. Haycox, C. Armand, S. Murteira, J. Cochran, and C. Francois, "Cost effectiveness of rasagiline and pramipexole as treatment strategies in early parkinson's disease in the uk setting," *Drugs & aging*, vol. 26, no. 9, pp. 791–801, 2009.

[7] W. J. Mutch, I. Dingwall-Fordyce, A. W. Downie, J. G. Paterson, and S. K. Roy, "Parkinson's disease in a scottish city." *Br Med J (Clin Res Ed)*, vol. 292, no. 6519, pp. 534–536, 1986.

[8] A. Barbeau, "L-dopa therapy in parkinson's disease: a critical review of nine years' experience," *Canadian Medical Association Journal*, vol. 101, no. 13, p. 59, 1969.

[9] R. Inzelberg, S. Kipervasser, and A. D. Korczyn, "Auditory hallucinations in parkinson's disease," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 64, no. 4, pp. 533–535, 1998.

[10] H. Narabayashi, F. Yokochi, and Y. Nakajima, "Levodopa-induced dyskinesia and thalamotomy." *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 47, no. 8, pp. 831–839, 1984.

[11] W. M. McDonald, I. H. Richard, and M. R. DeLong, "Prevalence, etiology, and treatment of depression in parkinson's disease," *Biological psychiatry*, vol. 54, no. 3, pp. 363–375, 2003.

[12] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury, "Stresssense: Detecting stress in unconstrained acoustic environments using smartphones," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 351–360.

[13] H. Zhang, C. Song, A. Wang, C. Xu, D. Li, and W. Xu, "Pdvocal: Towards privacy-preserving parkinson's disease detection using non-speech body sounds," 2019.

[14] H. Bernheimer, W. Birkmayer, O. Hornykiewicz, K. Jellinger, and F. Seitelberger, "Brain dopamine and the syndromes of parkinson and huntington clinical, morphological and neurochemical correlations," *Journal of the neurological sciences*, vol. 20, no. 4, pp. 415–455, 1973.

[15] C. D. Marsden and J. Parkes, ""On-off" effects in patients with parkinson's disease on chronic levodopa therapy," *The Lancet*, vol. 307, no. 7954, pp. 292–296, 1976.

[16] G. U. Höglinger, P. Rizk, M. P. Muriel, C. Duyckaerts, W. H. Oertel, I. Caille, and E. C. Hirsch, "Dopamine depletion impairs precursor cell proliferation in parkinson disease," *Nature neuroscience*, vol. 7, no. 7, p. 726, 2004.

[17] P. Svenningsson, E. Westman, C. Ballard, and D. Aarsland, "Cognitive impairment in patients with parkinson's disease: diagnosis, biomarkers, and treatment," *The Lancet Neurology*, vol. 11, no. 8, pp. 697–707, 2012.

[18] H. Zhao, C. Wang, N. Zhao, W. Li, Z. Yang, X. Liu, W. Le, and X. Zhang, "Potential biomarkers of parkinson's disease revealed by plasma metabolic profiling," *Journal of Chromatography B*, vol. 1081, pp. 101–108, 2018.

[19] J. Wang, J. G. Hoekstra, C. Zuo, T. J. Cook, and J. Zhang, "Biomarkers of parkinson's disease: current status and future perspectives," *Drug discovery today*, vol. 18, no. 3-4, pp. 155–162, 2013.

[20] T. M. Cutson, K. C. Laub, and M. Schenkman, "Pharmacological and nonpharmacological interventions in the treatment of parkinson's disease," *Physical therapy*, vol. 75, no. 5, pp. 363–373, 1995.

[21] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[22] P. M. C. Group *et al.*, "Long-term effectiveness of dopamine agonists and monoamine oxidase b inhibitors compared with levodopa as initial treatment for parkinson's disease (pd med): a large, open-label, pragmatic randomised trial," *The Lancet*, vol. 384, no. 9949, pp. 1196–1205, 2014.

[23] S. Arora, V. Venkataraman, A. Zhan, S. Donohue, K. Biglan, E. Dorsey, and M. Little, "Detecting and monitoring the symptoms of parkinson's disease using smartphones: A pilot study," *Parkinsonism & related disorders*, vol. 21, no. 6, pp. 650–653, 2015.

[24] W. Maetzler, J. Domingos, K. Srulijes, J. J. Ferreira, and B. R. Bloem, "Quantitative wearable sensors for objective assessment of parkinson's disease," *Movement Disorders*, vol. 28, no. 12, pp. 1628–1637, 2013.

[25] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average parkinson's disease symptom severity," *Journal of the royal society interface*, vol. 8, no. 59, pp. 842–855, 2010.

[26] J. R. Duffy, *Motor Speech Disorders-E-Book: Substrates, Differential Diagnosis, and Management*. Elsevier Health Sciences, 2013.

[27] J. A. Whitfield, C. Delong, A. M. Goberman, and M. Blomgren, "Fluency adaptation in speakers with parkinson disease: a motor learning perspective," *International journal of speech-language pathology*, vol. 20, no. 7, pp. 699–707, 2018.

[28] G. Yogev, N. Giladi, C. Peretz, S. Springer, E. S. Simon, and J. M. Hausdorff, "Dual tasking, gait rhythmicity, and parkinson's disease: which aspects of gait are attention demanding?" *European journal of neuroscience*, vol. 22, no. 5, pp. 1248–1256, 2005.

[29] B. R. Bloem, J. M. Hausdorff, J. E. Visser, and N. Giladi, "Falls and freezing of gait in parkinson's disease: a review of two interconnected, episodic phenomena," *Movement disorders: official journal of the Movement Disorder Society*, vol. 19, no. 8, pp. 871–884, 2004.

[30] A. Schrag, M. Jahanshahi, and N. Quinn, "What contributes to quality of life in patients with parkinson's disease?" *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 69, no. 3, pp. 308–312, 2000.

[31] J. Sian, D. T. Dexter, A. J. Lees, S. Daniel, Y. Agid, F. Javoy-Agid, P. Jenner, and C. D. Marsden, "Alterations in glutathione levels in parkinson's disease and other neurodegenerative disorders affecting basal ganglia," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 36, no. 3, pp. 348–355, 1994.

[32] F. Blandini, G. Nappi, C. Tassorelli, and E. Martignoni, "Functional changes of the basal ganglia circuitry in parkinson's disease," *Progress in neurobiology*, vol. 62, no. 1, pp. 63–88, 2000.

[33] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3d shape recognition," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 945–953.

[34] Q. Suo, F. Ma, Y. Yuan, M. Huai, W. Zhong, J. Gao, and A. Zhang, "Deep patient similarity learning for personalized healthcare," *IEEE transactions on nanobioscience*, vol. 17, no. 3, pp. 219–227, 2018.

[35] S. Chopra, R. Hadsell, Y. LeCun, *et al.*, "Learning a similarity metric discriminatively, with application to face verification," in *CVPR (1)*, 2005, pp. 539–546.

[36] C. J. Burges, "From ranknet to lambdarank to lambdamart: An overview," *Learning*, vol. 11, no. 23-581, p. 81, 2010.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[38] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2016.

[39] M. Zhao, S. Yue, D. Katabi, T. S. Jaakkola, and M. T. Bianchi, "Learning sleep stages from radio signals: A conditional adversarial architecture," in *International Conference on Machine Learning*, 2017, pp. 4100–4109.

[40] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W. Geng, "A novel attention-based hybrid cnn-rnn architecture for semg-based gesture recognition," *PloS one*, vol. 13, no. 10, p. e0206049, 2018.

[41] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in *International Workshop on Artificial Neural Networks*. Springer, 1995, pp. 195–201.

[42] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel, *et al.*, "Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): Scale presentation and clinimetric testing results," *Movement disorders*, vol. 23, no. 15, pp. 2129–2170, 2008.

[43] "A new lightweight, modular, and scalable deep learning framework," https://research.fb.com/downloads/caffe2/, accessed 07/29/19.

[44] "Protocol buffers," https://developers.google.com/protocol-buffers/, accessed 08/06/19.

[45] "Transfering a model from pytorch to caffe2 and mobile using onnx," https://pytorch.org/tutorials/advanced/super_resolution_with_caffe2.html, accessed 08/06/19.

[46] "Transfering a model from pytorch to caffe2 and mobile using onnx," https://github.com/pytorch/pytorch/blob/master/binaries/speed_benchmark.cc, accessed 07/25/19.

[47] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[48] "Accubattery," https://www.accubatteryapp.com/, accessed 08/01/19.

[49] A. Zhan, S. Mohan, C. Tarolli, R. B. Schneider, J. L. Adams, S. Sharma, M. J. Elson, K. L. Spear, A. M. Glidden, M. A. Little, *et al.*, "Using smartphones and machine learning to quantify parkinson disease severity: The mobile parkinson disease score," *JAMA neurology*, 2018.

[50] K. Sha, G. Zhan, W. Shi, M. Lumley, C. Wiholm, and B. Arnetz, "Spa: a smart phone assisted chronic illness self-management system with participatory sensing," in *Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments*. ACM, 2008, p. 5.

[51] T. Denning, A. Andrew, R. Chaudhri, C. Hartung, J. Lester, G. Borriello, and G. Duncan, "Balance: towards a usable pervasive wellness application with accurate activity inference," in *Proceedings of the 10th workshop on Mobile Computing Systems and Applications*. ACM, 2009, p. 5.

[52] N. Oliver and F. Flores-Mangas, "Healthgear: Automatic sleep apnea detection and monitoring with a mobile phone." *JCM*, vol. 2, no. 2, pp. 1–9, 2007.

[53] Z. Jin, J. Oresko, S. Huang, and A. C. Cheng, "Hearttogo: a personalized medicine technology for cardiovascular disease prevention and detection," in *Life Science Systems and Applications Workshop, 2009. LiSSA 2009. IEEE/NIH.* IEEE, 2009, pp. 80–83.

[54] O. Akinbode, O. Longe, and B. Amosa, "Mobile-phone based patient compliance system for chronic illness care in nigeria," *Journal of Computer Science & Technology*, vol. 12, 2012.

[55] F. Lin, J. Ajay, J. Langan, L. Cavuoto, I. Nwogu, H. Subryan, and W. Xu, "Qmforms: A portable and cost-effective upper extremity rehabilitation system," *Smart Health*, vol. 14, p. 100080, 2019.

[56] H. Zhang, A. Wang, D. Li, and W. Xu, "Deepvoice: A voiceprint-based mobile health framework for parkinson's disease identification," in *Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on.* IEEE, 2018, pp. 214–217.

[57] C. Song, A. Wang, F. Lin, M. Asmani, R. Zhao, Z. Jin, J. Xiao, and W. Xu, "Tempospatial compressed sensing of organ-on-a-chip for pervasive health," *IEEE journal of biomedical and health informatics*, vol. 22, no. 2, pp. 325–334, 2017.

[58] F. Lin, C. Song, X. Xu, L. Cavuoto, and W. Xu, "Patient handling activity recognition through pressure-map manifold learning using a footwear sensor," *Smart Health*, vol. 1, pp. 77–92, 2017.

[59] F. Lin, A. Wang, L. Cavuoto, and W. Xu, "Toward unobtrusive patient handling activity recognition for injury reduction among at-risk caregivers," *IEEE Journal of biomedical and health informatics*, vol. 21, no. 3, pp. 682–695, 2016.

[60] J. Castner, K. Klingman, S. Sullivan, W. Xu, and A. Titus, "Hitting home with technology development for asthma," *The Lancet Respiratory Medicine*, vol. 4, no. 2, pp. 102–103, 2016.

[61] J. Ajay, C. Song, A. Wang, J. Langan, Z. Li, and W. Xu, "A pervasive and sensor-free deep learning system for parkinsonian gait analysis," in *Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on.* IEEE, 2018, pp. 108–111.

[62] N. Bui, A. Nguyen, P. Nguyen, H. Truong, A. Ashok, T. Dinh, R. Deterding, and T. Vu, "Pho2: Smartphone based blood oxygen level measurement systems using near-ir and red wave-guided light," in *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems.* ACM, 2017, p. 26.

[63] T. Hao, G. Xing, and G. Zhou, "isleep: unobtrusive sleep quality monitoring using smartphones," in *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems.* ACM, 2013, p. 4.

[64] A. A. Farhan, C. Yue, R. Morillo, S. Ware, J. Lu, J. Bi, J. Kamath, A. Russell, A. Bamis, and B. Wang, "Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data." in *Wireless Health*, 2016, pp. 30–37.

[65] C. Gao, F. Kong, and J. Tan, "Healthaware: Tackling obesity with health aware smart phone systems," in *Robotics and Biomimetics (ROBIO), 2009 IEEE International Conference on.* Ieee, 2009, pp. 1549–1554.

[66] X. Sun, Z. Lu, W. Hu, and G. Cao, "Symdetector: Detecting sound-related respiratory symptoms using smartphones," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, pp. 97–108.

[67] H. Kalantarian, N. Alshurafa, and M. Sarrafzadeh, "Detection of gestures associated with medication adherence using smartwatch-based inertial sensors," *IEEE Sensors Journal*, vol. 16, no. 4, pp. 1054–1061, 2016.

[68] S. Bae, A. K. Dey, and C. A. Low, "Using passively collected sedentary behavior to predict hospital readmission," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 616–621.

[69] S. T. Moore, H. G. MacDougall, and W. G. Ondo, "Ambulatory monitoring of freezing of gait in parkinson's disease," *Journal of neuroscience methods*, vol. 167, no. 2, pp. 340–348, 2008.

[70] S. Mazilu, M. Hardegger, Z. Zhu, D. Roggen, G. Troster, M. Plotnik, and J. M. Hausdorff, "Online detection of freezing of gait with smartphones and machine learning techniques," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on.* IEEE, 2012, pp. 123–130.

[71] K. Wang, B. Li, D. Gu, K. Dai, and L. Zhou, "A smartphone based system for freezing of gait monitoring for parkinson's disease patients," in *Communication Software and Networks (ICCSN), 2017 IEEE 9th International Conference on.* IEEE, 2017, pp. 1529–1533.

[72] L. Pepa, F. Verdini, M. Capecci, and M. Ceravolo, "Smartphone based freezing of gait detection for parkinsonian patients," in *Consumer Electronics (ICCE), 2015 IEEE International Conference on.* IEEE, 2015, pp. 212–215.

[73] H. Kim, H. J. Lee, W. Lee, S. Kwon, S. K. Kim, H. S. Jeon, H. Park, C. W. Shin, W. J. Yi, B. S. Jeon, *et al.*, "Unconstrained detection of freezing of gait in parkinson's disease patients using smartphone," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE.* IEEE, 2015, pp. 3751–3754.

[74] C. D. Wurster, H. Graf, H. Ackermann, K. Groth, J. Kassubek, and A. Riecker, "Neural correlates of rate-dependent finger-tapping in parkinson's disease," *Brain Structure and Function*, vol. 220, no. 3, pp. 1637–1648, 2015.

[75] A. H. Butt, E. Rovini, D. Esposito, G. Rossi, C. Maremmani, and F. Cavallo, "Biomechanical parameter assessment for classification of parkinson's disease on clinical scale," *International Journal of Distributed Sensor Networks*, vol. 13, no. 5, p. 1550147717707417, 2017.

[76] B. M. Eskofier, S. I. Lee, J.-F. Daneault, F. N. Golabchi, G. Ferreira-Carvalho, G. Vergara-Diaz, S. Sapienza, G. Costante, J. Klucken, T. Kautz, *et al.*, "Recent machine learning advancements in sensor-based mobility analysis: Deep learning for parkinson's disease assessment," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the.* IEEE, 2016, pp. 655–658.

[77] A. Bermeo, M. Bravo, M. Huerta, and A. Soto, "A system to monitor tremors in patients with parkinson's disease," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the.* IEEE, 2016, pp. 5007–5010.

[78] B. P. Printy, L. M. Renken, J. P. Herrmann, I. Lee, B. Johnson, E. Knight, G. Varga, and D. Whitmer, "Smartphone application for classification of motor impairment severity in parkinson's disease," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE.* IEEE, 2014, pp. 2686–2689.

[79] G. Figueras, V. Parra, M. Huerta, A. Marzinotto, R. Clotet, R. González, A. Moreno, K. Pinto, D. Rivas, R. Alvizu, *et al.*, "Smartphone application for quantitative measurement of parkinson tremors," in *VI Latin American Congress on Biomedical Engineering CLAIB 2014, Paraná, Argentina 29, 30 & 31 October 2014.* Springer, 2015, pp. 785–788.

[80] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, L. O. Ramig, *et al.*, "Suitability of dysphonia measurements for telemonitoring of parkinson's disease," *IEEE transactions on biomedical engineering*, vol. 56, no. 4, pp. 1015–1022, 2009.

[81] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and analysis of a parkinson speech dataset with multiple types of sound recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828–834, 2013.

[82] L. Jeancolas, H. Benali, B.-E. Benkelfat, G. Mangone, J.-C. Corvol, M. Vidailhet, S. Lehericy, and D. Petrovska-Delacrétaz, "Automatic detection of early stages of parkinson's disease through acoustic voice analysis with mel-frequency cepstral coefficients," in *Advanced Technologies for Signal and Image Processing (ATSIP), 2017 International Conference on.* IEEE, 2017, pp. 1–6.

[83] A. H. P. Alavijeh, Y. P. Raykov, R. Badawy, J. Jensen, M. G. Christensen, and M. A. Little, "Quality control of voice recordings in remote parkinson's disease monitoring using the infinite hidden markov model," in *Ieee International Conference on Acoustics, Speech and Signal Processing (icassp).* IEEE, 2019.

[84] V. Sharma, K. Mankodiya, F. De La Torre, A. Zhang, N. Ryan, T. G. Ton, R. Gandhi, and S. Jain, "Spark: personalized parkinson disease interventions through synergy between a smartphone and a smartwatch," in *International Conference of Design, User Experience, and Usability.* Springer, 2014, pp. 103–114.

[85] H. Zhang, C. Xu, H. Li, S. A. Rathore, Z. Yan, D. Li, F. Lin, K. Wang, and W. Xu, "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*, vol. 3, no. 3, 2019.