**MSc Dissertation**

**Literature Review Report**

**MSc in Integrated Machine Learning Systems**

**Student:** Svetlana Grant

**Student number:** 19132626

**Project Title:**

**Machine Learning Approach for Parkinson Disease Monitoring Using Wearable Technologies**

University College London

Dept. of Electrical and Electronic Engineering

2020/2021

**Table of Contents**

Contents

[1. Introduction 3](#_Toc71962832)

[2. Background Theory and Literature Review 5](#_Toc71962833)

[2.1. Approaches to the analysis of wearable sensor data 6](#_Toc71962834)

[2.2. Supervised ML for PD wireless sensor data 7](#_Toc71962835)

[2.3. Unsupervised machine learning and deep learning for PD 8](#_Toc71962836)

[2.4. Transparency and interpretability of machine learning models for PD 9](#_Toc71962837)

[3. Summary 10](#_Toc71962838)

[4. List of References 12](#_Toc71962839)

# Introduction

Parkinson’s Disease is a chronic neurological disease, which affects as many as 5 million people worldwide, with more than double of this number expected by 2030 [1]. In the last ten years, advances in wearable sensor technologies and growing use of machine learning (ML) for medical purposes have opened new possibilities for tracking and monitoring the progress of Parkinson’s Disease (PD). Historically, clinical neurology data for diagnosing and tracking the progress of Parkinson’s Disease has been collected on a small scale in a clinical environment, with long time gaps between patient evaluation sessions. Recently, there has been growing interest in evaluating the use of medical sensors for continuous remote monitoring, and several studies, for example, Tzallas, 2014 [2], Sigcha, 2020 [3] and Mancini, 2021 [4] have tested their use in the home environment.

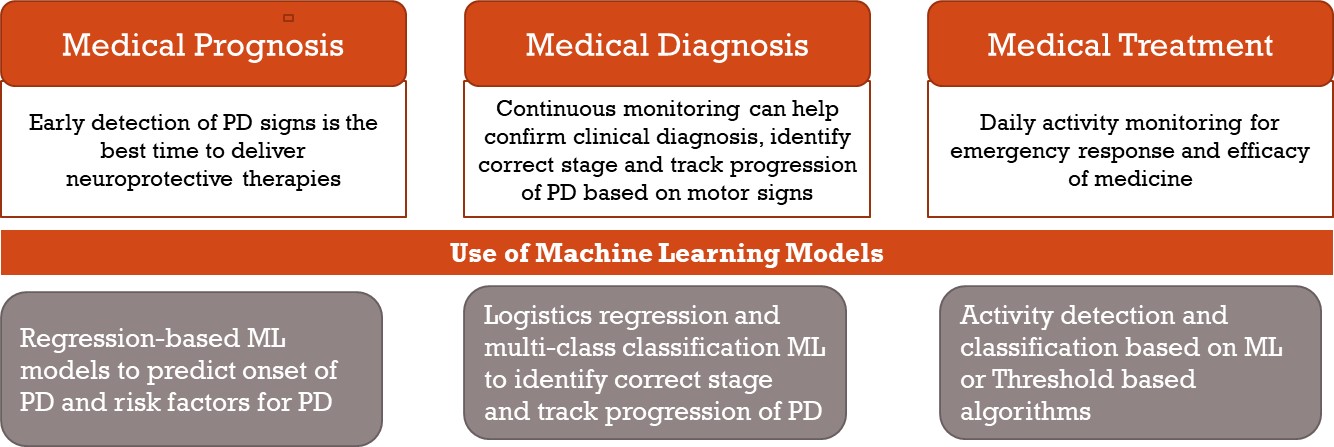
Home monitoring using wearable systems can be particularly beneficial for PD patients who need to be observed and evaluated on a regular basis, who are prone to injuries due to problems with balance and suffer from such debilitating symptoms as dyskinesia (involuntary erratic movements) , bradykinesia (slowness of movement) and freezing of gait (FoG). Wearable sensors can capture continuous changes in patients’ motor and non-motor symptoms, improving accuracy and frequency of observation. Camps et al pointed out that “accurate and automatic detection of FOG in PD patients has the potential of providing neurologists with relevant indicators about the condition status and its evolution, while enabling to design useful cueing wearable devices since these systems become more effective when applied only during the symptom’s episodes.” [5]

Wearable Inertial Measurement Unit (IMU) sensors with accelerometers, gyroscopes and magnetometers, as well as pressure sensors, can be used to measure various symptoms of PD, including tremors, bradykinesia and gait freezing, as well as effects of PD medicines, for example, dyskinesia [6]. In the aftermath of COVID-19, the need for remote monitoring and enablement of medical decision support systems outside of hospitals is expected to increase even further.

In parallel, use of Machine Learning algorithms for the analysis of wearable sensor data has grown in popularity. Earlier wearable sensor studies have focussed on using sensor data for PD predictions based on statistical signal processing approaches, most notably, threshold-based algorithms which allow to establish movement and gait patterns, explored in the studies by Bachlin et al, 2010 [7], Ferrari et al, 2016 [8], Tunca et al, 2017 [9], and Mancini et al, 2021 [4].

More recently, successful application of machine learning methods to medical and healthcare problems have led to the growing number of studies dedicated to the application of supervised, unsupervised and deep learning to PD. High accuracy results of using supervised and deep learning methods to Parkinson’s were demonstrated in such studies as Camps et al, 2018, [5], Sama et al, 2018 [5], Zhao et al, 2018 [10], Hssayeni et al, 2020 [11].

To date, the wearable sensor has focused on three groups of use cases, which would benefit from the use of machine learning approaches, as presented in **Fig. 1**.



**Fig. 1 Wearables and Machine Learning for PD: Examples of Use Cases**

At the same time, the use of ML for medical wearables is still new, and the application of these systems to support medical decision support systems needs to be explored further. One of the questions that need to be answered is which machine learning approach is most suitable for different types of PD use cases in the remote monitoring home environment. The requirements for the data need to come from the end-users: PD patients themselves who will want to know how their data helps to support medical decisions; and PD clinicians who will be on the receiving end of the data collected in a remote, unmonitored environment.

Horst et al [12] noted that “In the context of personalised medicine, the determination of characteristics that are specific for gait patterns of a certain individual facilitates to support clinicians and researchers in the individualisation of their analyses, diagnoses and interventions.”

While deciding between different ML models, a choice is often presented as a trade-off between simpler but less accurate supervised models and more accurate “black box” deep learning models [13]. Both approaches have benefits and constraints discussed in this report. One of the constraints of the deep learning models, for example, is their high computational cost, although new microcontrollers might be able to address this. At the same time, the emergence of the new field of “Explainable AI” questions if this trade-off needs to be quite as strict. In their Harvard Data Science Review article [14], Rudin and Radin argue that “that interpretability might not hurt accuracy. Interpretability might even improve accuracy, as it permits an understanding of when the model … might be incorrect.”

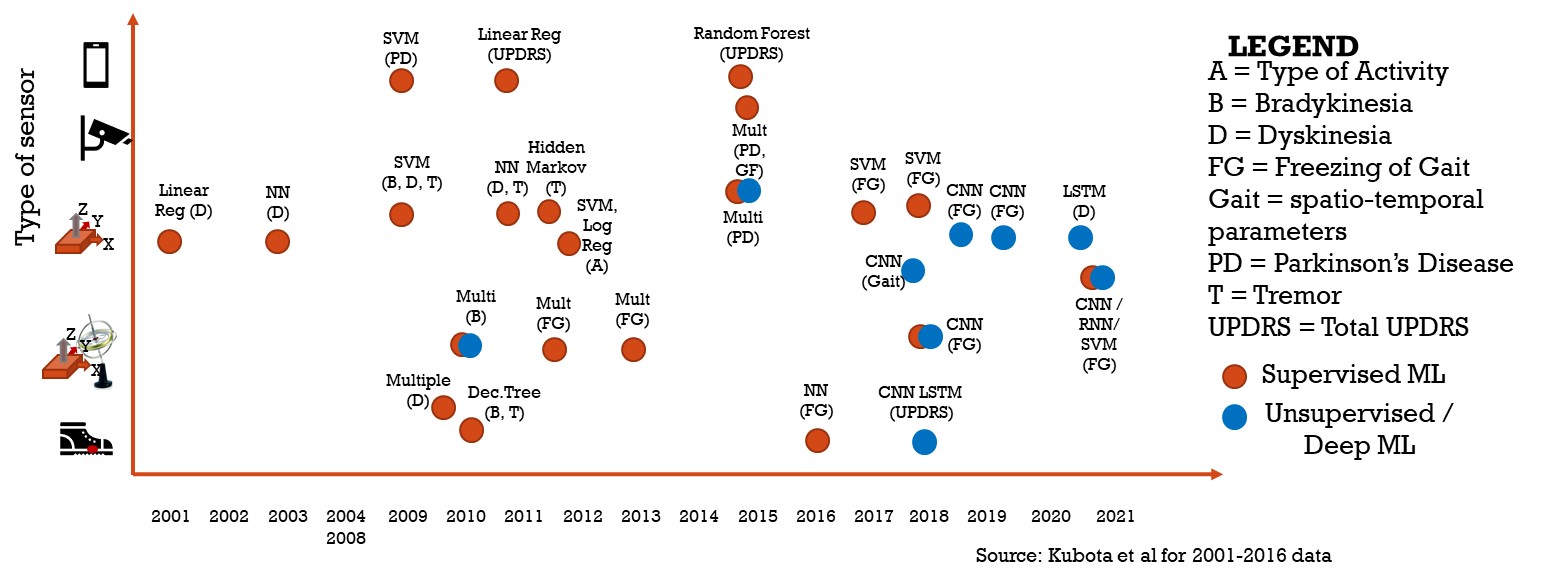
Thus, the problem statement for the development of wearable systems with machine learning algorithms needs to be reframed: these systems need to be robust, reliable and accurate, while proving the transparency inherent in supervised machine learning methods.

This study aims to evaluate the application of different types of machine learning approaches to main PD use cases, and explore how application of interpretable methods to deep learning models can improve their overall robustness and trustworthiness.

# Background Theory and Literature Review

A growing body of research has demonstrated that remote monitoring for Parkinson’s Disease (PD) in combination with machine learning can deliver significant value to both medical professionals and PD patients [15], [16], [17].

26 wearables studies used a variety of machine learning (ML) algorithms for PD research between 2001-2021, summarised in **Fig.2**. Of these, supervised ML methods have been applied to track specific daily activities to monitor and classify tremors, bradykinesia, dyskinesia, gait freezing and other gait disturbances. The research in more recent years has focussed primarily on using deep learning for freezing of gait, classification of Unified Parkinson’s Disease Rating Scale (UPDRS) stages, and classification of dyskinesia.



**Fig. 2 Summary of Machine Learning Studies for PD**

Use of wearables and machine learning for three different types of PD use cases have been explored over the past decade:

1. Machine Learning can use motor features and regression-based ML methods for Medical Prognosis, for example, PD early detection. Two studies published by Puhan et al in 2016 [18] and Del Sin et al in 2019 [19] explored motor features as potential prodromal markers of Parkinson's Disease. The presence of certain anomalies in these features could predict conversion to Parkinson's disease (PD) years before diagnosis. The 2016 study [18] concluded that “reduced arm swing is a well‐known clinical feature of Parkinson's disease (PD), often observed early in the course of the disease.” The 2019 study [19] confirmed that the variability and asymmetry of gait characteristics could also be used as prodromal markers for PD.
2. Machine Learning can be used for Medical Diagnosis, for example, PD Diagnosis and rating scale classification based on the patient’s motor signs and activity. Two most widely spread classification scales for PD are the Hoehn and Yahr (HY) scale and the Unified Parkinson’s Disease Rating Scale (UPDRS). Both list motor characteristics as defining elements in determining the phases of PD. However, the two scales are categorical rather than continuous, and very broad. A recent study from the International Parkinson and Movement Disorders Society Task Force on Technology into the Hoehn and Yahr (HY) scales noted that a more granular classification could be beneficial [20]. Use of wireless sensor data, which could be collected continuously and in significant volumes, could be used to improve the classification of PD.
3. Machine Learning can also be used for Medical Treatment and management of disease, for example, tracking PD medicine efficacy based on the patient’s motor signs, and Remote PD event monitoring, for example, detection of falls and freezing of gait (FoG), based on the pattern of body or gait movement. De Lima, 2017 [16] summarised different approaches to FoG, and Camps et al study published in 2018 [5] was the first paper to apply DL to the Freezing of Gait in PD patients, demonstrating clear gains in the accuracy of DL models. Such PD symptoms as Dyskinesia (involuntary movements), or Bradykinesia (slowness of movement) can be established based on the pattern of patients’ activity and measured by wearable sensors.

In 2008, a public Freezing of Gait (FoG) dataset Daphnet [7] was produced as part of the PD study via a collaboration between the Laboratory for Gait and Neurodynamics, Tel Aviv Sourasky Medical Center, Israel and the Wearable Computing Laboratory, ETH Zurich, Switzerland. Recordings were run at the Tel Aviv Sourasky Medical Center, producing time series accelerometer sensor data from wearable sensors placed on patients’ ankles, thighs and body trunks. 10 patients in total (7 males, 3 females) were monitored. The dataset is annotated, providing labels for Freezing of Gate and normal walking events. The dataset became the basis of a dozen of studies exploring the performance of different ML methods explored in the next section of the report.

## Approaches to the analysis of wearable sensor data

The analysis of wearable sensor PD data is broadly based on two types of approaches – statistical threshold-based algorithms and machine learning algorithms.

**Threshold – based algorithms** use signal processing methods, including algorithms tracking changes in signal energy, and to determine specific stages of the gait phases, such as Initial Contact (IC) and Toe-Off (TO).

These algorithms have been commonly applied to the recognition of specific events, such as detection of fall and Freezing of Gait (FoG), and a wide range of gait characteristics and parameters. Examples of this approach include the Freezing of Gait research performed by Baechelin et al in 2010 using the Daphnet FoG dataset [7], which used power spectral density threshold to differentiate between walking, standing and freezing episodes. More recently, the research by Mancini et al [4] derived an algorithm combining the power threshold calculated by Fast Fourier Transform (FFT) with the correlation between right and left angular velocity.

Other research demonstrated other uses of threshold – based algorithms to derive spatio-temporal gait characteristics. Ferrari et al, 2016 [8] used zero-velocity-update gait analysis system based on Kalman filter for real – time detection gait patters. Tunca et al, 2017 [9] extended the zero-velocity update and Kalman filtering methodology to non-hospital settings, to derive a rich set of standard gait metrics. Keloth et al, 2019 [21] established the variability of gait between PD and control subjects by means of measuring left and right foot angular velocity and angle differences.

The threshold- based approach remains popular due to its computational efficiency and transparency. At the same time, it presents a challenge when statistical model of the data is unknown from the start: Baechlin et al [7] observed that use of global thresholds led to lower specificity (true negatives, or proportion of actual predicted negative cases) and sensitivity (true positives, or proportion of actual predicted positive cases), which indicated that the user – independent model did not generalise as well as user-specific model. Threshold-based algorithms also struggle with high-dimensional data that contains a large number of features, making it difficult to identify average daily living (ADL) activities.

Use of **Machine Learning algorithms** for the analysis of wearable sensor data solves some of these challenges, and their application to PD monitoring has been widely studied in the last decade. Kubota et all [6] reviewed 17 wearables studies that used a variety of machine learning algorithms for PD research between 2001 and 2016. Since then, the research by Camps et al, 2018 [5] and Sigcha et al, 2020 [3] demonstrated that machine learning had better performance for FoG event detection, with sensitivity rates in excess of 90% compared to 85% rates achieved with the best threshold algorithms [4]. Sigcha et al [3] noted that “The proposed [Support Vector Machines] data representation presents advantages over previous handmade feature extraction methodologies and shows opportunities for the improvement of FOG detection systems to be applied in real time**.**”

In summary, machine learning extends traditional statistical methods, such as “parametric and nonparametric null hypothesis testing, linear and logistic regression, discriminant analysis, principal components, factor analysis, and cluster analysis”, “to cope with high dimensionality and nonlinearity, which is of particular importance in wearable sensor data.” [6]

## Supervised ML for PD wireless sensor data

As summarised in **Fig. 2**, the bulk of the ML research for PD conducted during 2009-2018 has used supervised learning algorithms for PD classification, such as SVM (Support Vector Machines), decision trees and random forest.

Supervised learning uses an algorithm that identifies a relationship between input data and output data, using the labels where “each training input must be associated with an output value” [6]. An output data represents either continuous set of values, which uses regression analysis, or a finite set of discrete data, which uses logistics regression analysis. In case of PD, both regression and classification can apply, depending on the problem that is being solved. While traditional statistical regression analysis is typically linear, supervised ML allows to establish non-linear, high-dimensional relationships.

However, labelling data collected from wearable sensors requires a complex set-up, typically performed in a hospital environment, when either a clinician or a patient validates an event. Thus, most data collected from wearable sensors is unlabelled, which has prompted the use of unsupervised machine learning and deep learning techniques.

## Unsupervised machine learning and deep learning for PD

Unsupervised learning is establishing input – output relationships for unlabelled raw sensor data. Unsupervised ML algorithms, such as K-means, cluster data into separate classes based on its characteristics. “K-means is fairly well established in PD studies that seek to identify subtypes of PD, such as those patients who are tremor dominant versus those with rapid motor function decline and cognitive impairment.” [6]

Of the 17 wireless sensor PD studies conducted during 2001-2016 and reviewed by Kubota [6], only 2 used unsupervised methods in combination with supervised, to classify bradykinesia (slowness of movement) and analyse the link between PD and mild cognitive impairment. Since then, 7 deep learning (DL) studies have been conducted in 2018-2021; two of these used a combination of supervised and deep learning approaches.

Deep learning learns representations of data with multiple levels of abstractions and is used for handling data without labels. Two different types of deep learning architectures have been explored for time – series sensor data in general and PD wearable data in particular. The bulk of research focussing on Convolutional Neural Network (CNN), which is now considered a state-of-the-art approach to modelling human activity, and on Long Short-Term Memory (LSTM), an architecture of artificial recurrent neural network (RNN), which is considered the best architecture for sequential data, such as time series.

The Convolutional Neural Networks (CNN) have been combined with the application of autoencoders (unsupervised learning techniques and data compression mechanisms that learn to map input data to output data automatically instead of being engineered by a human). For example, autoencoders have been used for denoising the sensor signal by Mohammadian et al., 2018 [22].

Sigcha (2020) [3] summarised that “recent studies propose the use of DL models for HAR (human activity recognition) and FOG (freezing of gait) detection. When working with sensor signals, the authors have successfully used deep networks with CNN and fully connected neural networks, while CNN work as an automatic feature extractor, the fully connected layers are used for classification.”

In the application of DL to detect Freezing of Gait events and several spatio- temporal gait parameters, such as stride length, DL models have performed at a better sensitivity and AUC (area under the curve – an indicator of the model performance) levels than supervised learning. San-Segundo et al., 2019 [23], has achieved AUC of 93.1% for FoG event detection based on the Daphnet dataset, compared to the AUC of 89% achieved by Mancini et al, 2021 [4] which used threshold based algorithm.

In 2018, Zhao, 2018 [10] used the Long Short-Term memory (LSTM, architecture that extends memory of neural networks), to rate the severity of PD disease from gait information using the sequential data of Vertical Ground Reaction Force (VGRF) recorded by foot sensors. The study developed a two-channel model combining LSTM and CNN to learn the spatio-temporal patterns behind the gait data. More recently, Hssayeni et al, 2021 [11] used deep learning LSTM model to estimate the severity of dyskinesia in PD patients on a dataset of 14 PD patients. The study addressed this as a regression problem, rather than a classification problem. The study achieved a high correlation (r= 0.86) with the scores assigned by a neurologist.

Despite higher accuracy, Camps et al [5] pointed out that deep learning models are difficult to train, with respect to time and computation power, especially in the real-time implementation scenarios. One of the suggestions to address this was through “generating adjustable models” that would allow a user a trade-off between performance against an overhead of retraining a model.

Mancini et al [4] summarised the challenge with deep learning as follows: “despite the higher sensitivity in detecting the occurrence of even shorter FOG episodes compared to the previous method (an accuracy above 90% was achieved), the deep learning approaches may require a higher computational cost, requiring up to several seconds from the occurrence of the episode to its detection, making those algorithms not suitable for real-time interventions, such as cueing [auditory signal that aims to interrupt a freezing of gait episode].” However, these constraints will not remain in place for long, as new types of floating – point powerful microcontrollers could potentially allow for the use of ML models in real-time and make them suitable for wearables.

## Transparency and interpretability of machine learning models for PD

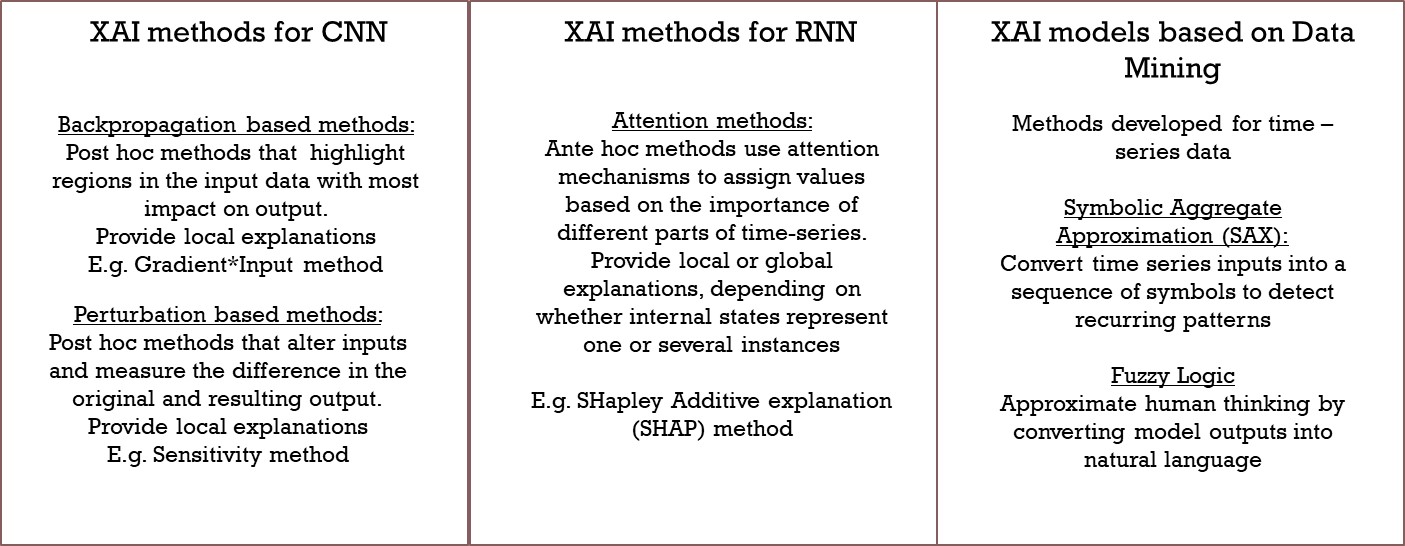
The growing use of non-linear machine leaning models raises an important question of transparency and interpretability that is needed for medical decision-making. Horst et al, 2019, [12] noted “For an implementation of machine learning in clinical diagnoses and therapeutic interventions, for example in terms of an automatic classification of gait disorders or (neurological) disease, the understanding about their decisions and decision-making seems to be inevitable. Since the lack of transparency has so far been a major drawback of preceding applications of machine learning, e.g. in medical applications (like gait analysis), further research on explaining, understanding and interpreting machine learning predictions should get attention.”

Very little discussion of model interpretability for wearables in PD has taken place to date. This is primarily because threshold – based models and linear machine learning models are transparent with regards to how the algorithms have derived the models’ output. At the same time, time series data collected from wearables is not always intuitively easy to interpret for an end-user of the data, such as a PD clinician, or a PD patient. Even when linear methods are used, visualisation techniques are a helpful tool to provide end-users with an easy way to understand the main factors contributing to the classification or prediction. Rojat [24] points out that “humans are not used to representing this temporal data in the form of a signal that varies as a function of time.” It can be argued that compared with such applications as computer vision and Natural Language Processing, it is more important to provide a clear explanation of time-series models, as these are not as easy to interpret for a human eye or ear as, for example, images or speech.

The subfield of “Explainable AI”, or XAI, has emerged in the last several years, to understand the workings of complex “black box” machine learning algorithms, and is now an active area of research. Dr. Wojciech Samek presented an overview of “Explainable AI” methods in his speeches during the Open Data Science Conference Europe (ODSC) 2018 [25], while Rojat et al [24] reviewed the application of explainable AI methods to time series data.

The XAI models add interpretability constraints, either as “ante-hoc”, which incorporate explainability into the structure of the model itself, or “post-hoc”, which explain relationships between features and predictions without changing the model [24].

The following approaches have been developed to add interpretability for a ML model: understanding the contribution of individual model features toward the output; decomposition of complex deep learning models into simpler models that explain how different features are combined to contribute to the outcome (for example, Deep Taylor decomposition); model deconvolution; and a series of methods to visualise the data [26]. Most of these have been developed for computer vision and NLP (Natural Language Processing) and later applied to time – series data (see Fig. 3). Other methods – for example, symbolic aggregate approximation and fuzzy logic based on data mining – have been developed specifically for time series data.



**Fig. 3 Explainable AI for Time Series Data (source: Rojat et al** [24]**)**

A new method Layer-wise Relevance Propagation (LRP), which combines backpropagation and perturbation approaches, was proposed by Bach et al, 2015 *[add citation]* and developed further by Dr. Samek [26]. The method decomposes a decision function fc, where each input x is assigned relevance Ri for its ith element, such that fc (x) = ∑i Ri . The contribution of each input neuron *i* is evaluated with respect to its contribution toward the model’s output and assigned an amount of relevance. The decision behind the model’s output is then explained in terms of input variables, similar to linear models.

LRP was tested in different applications, including gait analysis by Horst et al [12], who applied it in SVM and ANN models on a group of 57 subjects (29 female, 28 male) without any pathologies. The study applied LRP to identify individual human gait characteristics, decomposing the prediction of a DL model into input variables – ground reaction forces and full-body joint angles. The study could be extended to the gait analysis model based on IMU data, and be applied to the patients with gait pathology, such as those with Parkinson’s Disease.

## Deep Learning Models for Wearable PD Devices

While deep learning models deliver significantly improved accuracy, there are other considerations that determine their suitability for healthcare wearables. One of these is the challenge of using deep neural networks in real time deployments of wearables: the large size of the DL models (measured in MB, while the RAM capacity of wearable sensor devices is in KB), need for extensive compute capacity; and how power consumption required for devices that are always on while tracking PD patient activities.

The following wearable hardware constraints need to be addressed:

* Many wearables need to be very small to be worn on a patient’s body, which imposes further challenges on microcontrollers, sensors and battery inside them. This means that wearable sensors are subject to power / battery constraints, as many operate on coin cell batteries that use milliwatts of power, which means that they are not able to stream data in real time.
* The size of DL models on mobile Nets is currently measured in MBs, but would need to be in Kbps, in order to run on edge devices.

Yes, emergence of smart watches and other wearables that perform inference on the devices for speech and image recognition show that this can be done.

Tiny ML framework has emerged as the answer to running deep learning models on constrained edge devices. This is an end-to-end framework that optimises and reduces full-stack machine learning solutions – including hardware, system, software, and applications) to perform well on the edge devices (see **Fig.4**) [[27] [28]].

A picture containing diagram

Description automatically generated

**Fig. 4 TinyML WorkFlow [source: EdX TinyML Course]**

The benefits derived from the use of TinyML are as follows:

Real – time inference on edge devices:

* Wearables with ML models are capable of running inference models on the edge in real-time, which consume significantly less compute resource compared to sending data for processing into the cloud. This is because radio accounts for the largest portion of the power consumption on wearable devices [*source*].

Minimising data transfer to the cloud:

* Not all data from wearable devices needs to stream to the clinicians, but the storage capacity on wearable devices is limited. One example of streaming to the cloud: 2 bytes per measurement x 6 sensor streams (e.g. 6-axis IMU with accelerometer + gyroscope) x 20MHz = 240 KB/sec. For a Zigbee radio can stream ony at 30 KB/sec – 8 x lower; with BLE, which has data rates of 1MBps, the throughput is no longer a problem, but power consumption becomes the main limiting factor. The BLE power consumption is 1W, with peak current consumption of <30 mA. With a coin cell battery on a budget of milliwatts, it’ll be depleted within a day.

* Instead of streaming the time series signal readings from a wearable sensor, the data could be processed locally, and only event – driven meta-data (e.g. if a Freezing of Gait event took place after the intake of medicine). Similarly, a classification of the patient’s gait can be performed on the edge, with only a small amount of data transferred to the clinician.

# Summary

A growing body of research has demonstrated that remote monitoring for Parkinson’s Disease (PD) can deliver significant value to both medical professionals and PD patients [15], [16], [17]. Analysis of the wearable sensor data via machine learning algorithms can create a new stream of information for medical decision making, helping PD clinicians to make more informed, accurate, real-time evaluations of their patient’s status.

Both statistical and machine learning methods have been used in the studies of remote wireless sensor monitoring of PD patients. In the studies reviewed in this report, threshold algorithms have been applied to events – based PD use cases, such as freezing of gate, as well as for detecting walking activity and tracking a range of spatio-temporal gait parameters, such as step and stride length. These studies achieved accuracy levels between 66% (Bachlin et al, 2010) [7] and 85% (Mancini, 2021) [4]. Camps et al [5] acknowledged that “most of the biomedical time-series problems can be solved by simpler classifying or regression models.”

While linear threshold algorithms are efficient to deploy and transparent, several challenges exist: these algorithms require prior knowledge of gait models, and experimental adjustment of thresholds that does not always generalise well. Machine learning algorithms are needed for more complex classification use cases, especially those that involve large datasets with multiple features, and require non-linear models. ML approaches achieve better accuracy results for classification of events, e.g. in the range of 88% with supervised ML algorithms (Rodríguez-Martín et al, 2017) and 91% for DL models (Camps et al [5]).

Deep learning in particular has shown promise for such PD use cases as freezing of gait, while convolutional neural networks are now considered the state-of-the-art for activity recognition and movement monitoring using wearable sensors. However, further studies are needed to establish how well deep learning models would perform in real-time, home environment. Furthermore, more studies are needed into which PD monitoring indeed requires real-time monitoring that would be of most benefit to PD practitioners or patients.

Moreover, for PD remote monitoring to gain traction, machine learning models need to be generating accurate results and be trusted by developers (ML experts who create the model and would like to see quantitative evaluations); and by users (clinician- experts) who need to understand the rationale behind the model’s classification and trust the model’s decision. While computational efficiency might be addressed by more advanced microcontrollers, explainable AI methods need to be developed which provide users of the models with transparency and trust in the models’ decision – making. The use of Explainable AI (XAI) is emerging to address this need.

Proposed aim of this study:

The proposed focus of this study is to establish which ML models would be the best fit for the requirements of remote PD monitoring.

There are two proposed directions this study:

Study the optimisation of deep learning models for PD: Use TinyML framework to optimise DL models for PD. The aim of the study is to understand which DL models deliver the best inference for PD use cases (e.g. DNN or RNN, RNN with LSTM), and understand if these can be optimised for TinyML. The outcome will be to evaluate the performance of TinyML models, focussing on the accuracy in the first instance.

Study interpretability of deep learning models for PD: Following the trustworthiness framework, proposed by Rojat, 2021 [24], the aim of this study is to develop trust in the ML models, and make them acceptable for medical decision-making by proving their stability, interpretability, robustness and confidence levels. The research will aim to interpret which feature inputs most significantly contribute toward the output and decision of the model, using the LRP (Layer-Wise Relevance Propagation) method that was introduced by Bach et al, 2015, and Dr. Samek, 2019 [26], and applied to the analysis of individual human gait characteristics by Horst et al [12].

Proposed approach:

* Model classification/ prediction for a PD use cases, using different types of models:
  + Use case from MJFox Levodopa Study: Classifying tremors, dyskinesia (upper & lower limbs), bradykinesia (upper limbs only) (0-4 severity); OR Freezing of Gait data set
  + Supervised model(s) (e.g. SVM and decision tree);
  + Deep learning models: CNN and RNN with LSTM;
* Compare accuracy results for different ML algorithms

Direction for Tiny DL study for PD:

* Optimise the best performing DL models for PD using Tiny ML framework, which includes using quantization and pruning techniques.
* Compare accuracy of the DL models with and without Tiny ML, and with supervised ML models.
* Evaluate the size of the model; test its performance on Arduino with TinyML kit.

Direction for trustworthy DL study:

* For supervised ML, evaluate which features contribute toward the outcome (e.g. through visualisation?)
* Develop feature representation maps / heat maps explaining attribution toward the predictive outcome, using Layer-Wise Relevance Propagation as transparency method.
* Compare interpretability of different ML models – which inputs are highlighted for CNN vs LSTM models. Are they the same as for decision trees?
* TBC: simulate amount of traffic that will be sent if, for example, only metadata based on the events is sent to the cloud.
* what happens when noise and perturbation are introduced into the most salient inputs. How does each type of ML model behave? In addition to providing explainability, XAI can help develop new metrics and training practices, and improve robustness of the models.

Validation:

* What is the confidence behind the predictions for each of these?
* Validate results with a clinician:
  + Based on the accuracy of the TinyML models, will forecasts from wearable devices be acceptable to clinicians.
  + Based on the feature heat maps, does the model use the most relevant input for PD - specific predictions? Does it match what an expert would expect to see from this prediction?
  + How should results be delivered to a clinician, to ensure confidence in the model decisions. What is needed to make these models trustworthy to support medical decisions?

# List of References

[1] S. L. Kowal, T. M. Dall, R. Chakrabarti, M. V. Storm, and A. Jain, “The current and projected economic burden of Parkinson’s disease in the United States,” *Mov. Disord.*, vol. 28, no. 3, pp. 311–318, 2013, doi: 10.1002/mds.25292.

[2] A. T. Tzallas *et al.*, “Perform: A system for monitoring, Assessment and management of patients with Parkinson’s disease,” *Sensors (Switzerland)*, vol. 14, no. 11, pp. 21329–21357, 2014, doi: 10.3390/s141121329.

[3] L. Sigcha *et al.*, “Deep learning approaches for detecting freezing of gait in parkinson’s disease patients through on-body acceleration sensors,” *Sensors (Switzerland)*, vol. 20, no. 7, 2020, doi: 10.3390/s20071895.

[4] M. Mancini *et al.*, “Measuring freezing of gait during daily-life: an open-source, wearable sensors approach,” *J. Neuroeng. Rehabil.*, vol. 18, no. 1, pp. 1–13, 2021, doi: 10.1186/s12984-020-00774-3.

[5] J. Camps *et al.*, “Deep learning for freezing of gait detection in Parkinson’s disease patients in their homes using a waist-worn inertial measurement unit,” *Knowledge-Based Syst.*, vol. 139, pp. 119–131, 2018, doi: 10.1016/j.knosys.2017.10.017.

[6] K. J. Kubota, J. A. Chen, and M. A. Little, “Machine learning for large-scale wearable sensor data in Parkinson’s disease: Concepts, promises, pitfalls, and futures,” *Mov. Disord.*, vol. 31, no. 9, pp. 1314–1326, 2016, doi: 10.1002/mds.26693.

[7] M. Bächlin *et al.*, “Wearable assistant for Parkinsons disease patients with the freezing of gait symptom,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 436–446, 2010, doi: 10.1109/TITB.2009.2036165.

[8] A. Ferrari, P. Ginis, M. Hardegger, F. Casamassima, L. Rocchi, and L. Chiari, “A mobile Kalman-filter based solution for the real-time estimation of spatio-temporal gait parameters,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 7, pp. 764–773, 2016, doi: 10.1109/TNSRE.2015.2457511.

[9] C. Tunca, N. Pehlivan, N. Ak, B. Arnrich, G. Salur, and C. Ersoy, “Inertial sensor-based robust gait analysis in non-hospital settings for neurological disorders,” *Sensors (Switzerland)*, vol. 17, no. 4, pp. 1–29, 2017, doi: 10.3390/s17040825.

[10] A. Zhao, L. Qi, J. Li, J. Dong, and H. Yu, “A hybrid spatio-temporal model for detection and severity rating of Parkinson’s disease from gait data,” *Neurocomputing*, vol. 315, pp. 1–8, 2018, doi: 10.1016/j.neucom.2018.03.032.

[11] M. D. Hssayeni, J. Jimenez-Shahed, M. A. Burack, and B. Ghoraani, “Dyskinesia Severity Estimation in Patients with Parkinson’s Disease Using Wearable Sensors and A Deep LSTM Network,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2020-July, pp. 6001–6004, 2020, doi: 10.1109/EMBC44109.2020.9176847.

[12] F. Horst, S. Lapuschkin, W. Samek, K. R. Müller, and W. I. Schöllhorn, “Explaining the unique nature of individual gait patterns with deep learning,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–13, 2019, doi: 10.1038/s41598-019-38748-8.

[13] H. Yin and N. K. Jha, “A Health Decision Support System for Disease Diagnosis Based on Wearable Medical Sensors and Machine Learning Ensembles,” *IEEE Trans. Multi-Scale Comput. Syst.*, vol. 3, no. 4, pp. 228–241, 2017, doi: 10.1109/TMSCS.2017.2710194.

[14] C. Rudin and J. Radin, “Why Are We Using Black Box Models in AI When We Don’t Need To? A Lesson From An Explainable AI Competition,” *Harvard Data Sci. Rev.*, vol. 1, no. 2, pp. 1–9, 2019, doi: 10.1162/99608f92.5a8a3a3d.

[15] A. J. Espay *et al.*, “Technology in Parkinson’s disease: Challenges and opportunities,” *Mov. Disord.*, vol. 31, no. 9, pp. 1272–1282, 2016, doi: 10.1002/mds.26642.

[16] A. L. Silva de Lima *et al.*, “Home-based monitoring of falls using wearable sensors in Parkinson’s disease,” *Mov. Disord.*, vol. 35, no. 1, pp. 109–115, 2020, doi: 10.1002/mds.27830.

[17] A. L. S. De Lima *et al.*, “Feasibility of large-scale deployment of multiple wearable sensors in Parkinson’s disease,” *PLoS One*, vol. 12, no. 12, pp. 1–15, 2017, doi: 10.1371/journal.pone.0189161.

[18] M. A. Puhan *et al.*, “Arm Swing as a Potential New Prodromal Marker of Parkinson’s Disease,” *Mov. Disord.*, vol. 37, no. 4, pp. 784–790, 2017, doi: 10.1183/09031936.00063810.The.

[19] S. Del Din *et al.*, “Gait analysis with wearables predicts conversion to parkinson disease,” *Ann. Neurol.*, vol. 86, no. 3, pp. 357–367, 2019, doi: 10.1002/ana.25548.

[20] C. G. Goetz *et al.*, “Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: Status and recommendations,” *Mov. Disord.*, vol. 19, no. 9, pp. 1020–1028, 2004, doi: 10.1002/mds.20213.

[21] S. M. Keloth, R. Viswanathan, B. Jelfs, S. Arjunan, S. Raghav, and D. Kumar, “Which gait parameters and walking patterns show the significant differences between Parkinson’s disease and healthy participants?,” *Biosensors*, vol. 9, no. 2, 2019, doi: 10.3390/bios9020059.

[22] N. M. Rad, T. van Laarhoven, C. Furlanello, and E. Marchiori, “Novelty detection using deep normative modeling for imu-based abnormal movement monitoring in parkinson’s disease and autism spectrum disorders,” *Sensors (Switzerland)*, vol. 18, no. 10, 2018, doi: 10.3390/s18103533.

[23] R. San-Segundo, H. Navarro-Hellín, R. Torres-Sánchez, J. Hodgins, and F. de la Torre, “Increasing robustness in the detection of freezing of gait in Parkinson’s disease,” *Electron.*, vol. 8, no. 2, pp. 1–14, 2019, doi: 10.3390/electronics8020119.

[24] T. Rojat, R. Puget, D. Filliat, J. Del Ser, R. Gelin, and N. Díaz-Rodríguez, “Explainable Artificial Intelligence (XAI) on TimeSeries Data: A Survey,” 2021.

[25] D. W. Samek, “Towards Interpretable Deep Learning,” Europe 2019, 2018.

[26] D. W. Samek, “Explainable AI - Methods, Applications & Recent Developments,” ODSC Europe 2019, 2019.

[27] R. David *et al.*, “TensorFlow Lite Micro: Embedded Machine Learning on TinyML Systems,” 2020.

[28] H. Doyu, R. Morabito, and M. Brachmann, “A TinyMLaaS Ecosystem for Machine Learning in IoT: Overview and Research Challenges,” pp. 1–5, 2021, doi: 10.1109/vlsi-dat52063.2021.9427352.