**MSc Dissertation**

**Interim 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

2020-21

# Abstract

The 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 Parkinson’s Disease predictions based on statistical signal processing approaches, most notably, threshold-based algorithms which allow to establish movement and gait patterns. 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 approaches to PD. Both demonstrated high accuracy results for remotely tracking PD symptoms, such as tremor, dyskinesia, bradykinesia and freezing of gait.

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. The accuracy, however, may not be sufficient to justify then use of deep neural networks for PD wearables. Among the constraints of the deep learning models are their size, high computational cost and power consumption, as well as lack of transparency.

This study aims to evaluate the application of different types of machine learning approaches for monitoring such PD symptoms as tremor and dyskinesia. This will be based on accelerator sensor data from the MJFF Levodopa Wearable Sensors dataset from the study conducted in 2015.

The study will explore whether the application of TinyML methods, such as quantization (reformatting data from float32 format to int8 format) and pruning (reducing the number of inputs), to deep neural networks, combined with interpretable methods, can improve their overall robustness and trustworthiness and make their deployment onto embedded PD wearables possible. It will use prediction accuracy, model size, power consumption and computation latency, to evaluate the performance of the models.

# Table of Contents

[1) Abstract 2](#_Toc75379912)

[2) Table of Contents 3](#_Toc75379913)

[3) Introduction and Problem Statement 4](#_Toc75379914)

[4) Theory/Methodology 6](#_Toc75379915)

[4.1. Approaches to the analysis of wearable sensor data 7](#_Toc75379916)

[4.2. Supervised ML for PD wireless sensor data 8](#_Toc75379917)

[4.3. Unsupervised machine learning and deep learning for PD 8](#_Toc75379918)

[**4.4.** Goals of the Study 10](#_Toc75379919)

[**4.5.** Research Methodology and Implementation 10](#_Toc75379920)

[4.6. Expected Outcome and Impact 13](#_Toc75379921)

[5) Project Work Plan 15](#_Toc75379922)

[6) Results to date 16](#_Toc75379923)

[7) References 17](#_Toc75379924)

[8) Appendices 19](#_Toc75379925)

# Introduction and Problem Statement

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.

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 [5], Ferrari et al, 2016 [6], Tunca et al, 2017 [7], 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, [8], Sama et al, 2018 [8], Zhao et al, 2018 [9], Hssayeni et al, 2020 [10].

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.

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 [11]. Both approaches have benefits and constraints discussed in this report. Among the constraints of the deep learning models are their size, high computational cost and power consumption, as well as lack of transparency. The emergence of the TinyML framework that optimises the size and performance of deep neural networks for embedded devices, and the field of “Explainable AI” aim to change the parameters of this trade-off. In their Harvard Data Science Review article [12], 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.”

This study aims to evaluate the application of different types of machine learning approaches to main PD use cases and explore how application of TinyML methods, such as quantization and pruning, deep learning models, combined with interpretable methods can improve their overall robustness and trustworthiness, and make their deployment onto embedded PD wearables possible.

# Theory/Methodology

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 [13], [14], [15].

26 wearables studies used a variety of machine learning (ML) algorithms for PD research between 2001-2021, summarised in **Fig.1**. 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.

Chart, scatter chart

Description automatically generated

**Fig. 1 Summary of Machine Learning Studies for PD**

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. 2**.

Graphical user interface, text

Description automatically generated

**Fig. 2 Wearables and Machine Learning for PD: Examples of Use Cases**

## 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 [5], 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 [6] used zero-velocity-update gait analysis system based on Kalman filter for real – time detection gait patters. Tunca et al, 2017 [7] 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 [16] 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 [5] 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 [17] 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 [8] 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.” [17]

## Supervised ML for PD wireless sensor data

As summarised in **Fig. 1**, 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” [17]. 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. Prediction of an event (for example, FoG or dyskinesia) will apply binary classification, prediction of activity will use multi-label classification approach, while UPDRS score can use either multi-label classification or a regression to derive a numeric score. 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.” [17]

Of the 17 wireless sensor PD studies conducted during 2001-2016 and reviewed by Kubota [17], 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 [18].

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 [19], 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 [9] 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 [10] 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 [8] 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.

## Goals of the Study

The aim of the study is to evaluate which deep learning models can be used for wearable PD devices, and whether these can achieve better levels of accuracy than state of the art statistical and supervised machine learning models.

Deep neural nets are measured in MB and therefore too large to fit onto constrained wearable devices. This study will use the TinyML framework to optimise and compress the models. The ML framework uses quantization and pruning methods to convert pre-trained models from TensorFlow to TensorFlow Lite, and further into TensorFlow Micro. This study will apply the conversions and evaluate the size, latency and accuracy of the model.

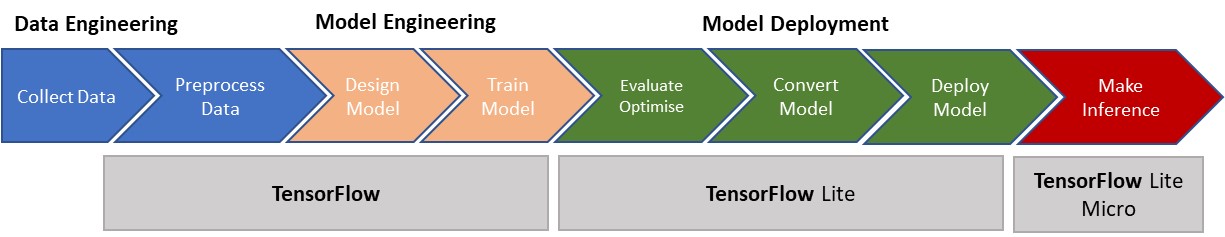
The objectives of the study:

* Perform supervised ML models and deep neural networks models for PD use cases, such as monitoring of symptoms (tremors, dyskinesia) and classification of activity.
* Compare accuracy of the models.
* Compress and optimise neural nets using TinyML framework, to minimise storage and memory footprint.
* Perform inference, and compare the accuracy of the compressed models, aiming to minimise the loss of model accuracy.
* Discuss results with a PD clinician to understand the acceptance of the proposed model performance.

## Research Methodology and Implementation

In this study, we will follow the Machine Learning workflow outlined in the Harvard TinyML Applications course, focussing on data engineering, model engineering and model deployment.

The product analytics step includes inference and visualisation of the results. The study plans to test the inference on the edge – device, which consists of Arduino Nano 33 BLE Sense microcontroller with Tiny Machine Learning Shield.



**Fig. 3 Machine Learning Workflow** [20]

**Data Engineering Considerations**

The MJFF (Michael J Fox Foundation) Levodopa Wearable Sensors Dataset [21] will be used for this study. The data described in **Fig.4** was collected during 4 days in 2015, using 5 Shimmer3 sensors that contained accelerometer on wrists, ankles and back. The data was collected from 16 patients, of which 4 female (aged 54-80), 12 male (aged 55-77). It includes 2 days of labelled data collected in the controlled Lab environment and 2 days of data collected in the home environment.

Text

Description automatically generated with low confidence

**Fig. 4 MJFF Levodopa Wearable Sensors Dataset (2015)** [21]

The data is used to classify tremors, dyskinesia (upper & lower limbs) and bradykinesia (upper limbs only). Clinicians classified the data according to 0-4 severity.

Potential limitations and biases:

* Limited number of PD patients was included in the sensor study. For this reason, analysis of some use case, for example, forecast of UPDRS scores, will be limited, as only a small number of cases will be available for each classification group.
* The patient dataset has only 4 female PD patients, skewing the results toward PD male patients.
* Only data from accelerometers was collected for the Levodopa Sensor dataset. This limits the application of the results of this study, as increasingly IMUs with at least 6-axis, including accelerometer and gyroscope sensors, are used for PD wearables studies.

**Model Engineering Considerations**

The study aims to compare accuracy and performance for supervised machine learning and deep learning models.

Supervised ML model will include Support Vector Machine for classification applications.

For deep learning model engineering, this study will focus on the two types of Deep Neural Networks – convolutional neural network (CNN) and Recurrent Neural Networks that is used for time series:

* Convolutional Neural Networks have been successfully deployed for activity recognition:
  + 2D convolutional layers – used for identifying features and patterns. Heim et al (2021) [22] concluded that latency optimisation is optimised through the choice of input connections. Faster latency is achieved if a number of inputs is an even number, or divisible by 4.
  + Dense layers – used for backpropagation and calculation of weights.
* Recurrent Neural Network with Long Short-Term Memory (LSTM) is used for time series data as this type of deep learning tracks temporal data patterns.

Heim et al [22] concluded that the best optimisation is achieved during the design process, choosing the best combination of layer types and dimensions.

**Model Deployment Considerations**

Once developed and validated, the model for embedded devices and microcontrollers will be optimised for performance, power consumption and latency. This will be done by using tools on the already-trained model: for example, a converter compresses a model developed in TensorFlow to the format suitable for running on an edge device.

* TensorFlow Lite is a runtime optimised for Android, iOS and embedded systems that run on a variant of Linux, e.g. RaspberryPi. During the conversion, many of the memory consuming elements of the neural network are removed, without compromising accuracy. Quantization can take place during this conversion, where data is reformatted from high-memory consuming float32 into a lower-memory format int8.
* TensorFlow Lite Micro is a core enabling technology for TinyML framework; it is built on top of TensorFlow Lite to downsize the model even further to work on microcontrollers.

Three main methods to compress the models have demonstrated inference efficiency, including quantization, weight sharing and network pruning.

A study by Han et al (2016) [23] implemented a three stage “deep compression” pipeline, combining pruning, quantization and Huffman coding. This approach targeted latency – sensitive applications running on mobiles, which required real – time inferences. The study concluded that pruning reduced the number of connections by 9x to 13x; Quantization then reduced the number of bits that represent each connection from 32 to 5, reaching compression rate between 27x and 31X. Finally, application of Huffman coding reduced overall storage requirements by 35X to 49X without reduction in accuracy. Combined, pruning and quantization compressed the network to 3% of the original size.

Heim et al (2021) [22] have conducted research that optimised NNs with a focus on “perceptible metrics” – the metrics that the end-user is exposed to, such as inference latency and energy consumption. This study experimented with 8-bit quantization, chosen so because it is supported by most MCU (compared with 4-bit or adaptive rate). It applied quantization and specialised kernel to two models to LeNet and ResNet neural nets, and concluded that operating on a floating point and fixed point made significant difference; achieving memory footprint reduction of 73% while losing only 0.05% of the accuracy when tested on the host (PC that was used to design and optimise the model). On the latency, while the floating point unit (FPU) was disabled, the unoptimized model achieved a 4x faster inference; applying software acceleration with the specialized neural network kernel (CMSIS-NN) to the quantized (optimised) model achieved additional 4x improvement in latency.

Potential limitations and biases:

* While optimised models will be deployed on Arduino Nano BLE with Machine Learning to test, it will not be possible to test the accuracy of inference, as the study will not include testing on PD patients. Thus “perceptible metrics” will need to be simulated.
* The dataset was based on the readings of 5 sensors worn on each limb and on the lower back. Real-life deployments will aim to minimise the number of sensors worn on a patient, and therefore, the ML models intended for use on a real wearable will need to be retrained for the right placement of sensors.

## Expected Outcome and Impact

This research will strengthen the understanding of which machine learning models provide the best fit for wearables used for Parkinson’s Disease applications in the home monitoring environment, which requires accurate and robust predictions, as well as ability to deploy the chosen model onto a constrained wearable device.

The expected outcome of this research is to evaluate the best optimisation methods for supervised and deep learning models, aiming at the smallest memory footprint and compute requirements, using TinyML framework.

The desired impact of this research is to improve the feasibility of using wearables for monitoring Parkinson’s Disease, and aid with the faster adoption of wearable systems for PD.

# Project Work Plan

The project will be following the work plan presented in **Fig.5**.



**Fig. 5 Project Work Plan**

# Results to date

No results are yet available for this study.

# 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] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] U. Harvard, “Fundamentals of TinyML.” [Online]. Available: https://online-learning.harvard.edu/course/fundamentals-tinyml?delta=0. [Accessed: 18-Jun-2021].

[21] “MJFF Levodopa Wearable Sensors Dataset.” 2015.

[22] L. Heim, A. Biri, Z. Qu, and L. Thiele, “Measuring what Really Matters: Optimizing Neural Networks for TinyML,” 2021.

[23] S. Han, H. Mao, and W. J. Dally, “Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding,” *4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.*, pp. 1–14, 2016.

# Appendices